How close dark matter haloes and MOND are to each other: three-dimensional tests based on *Gaia* DR2[*]

Yongda Zhu [ID],[1]† Hai-Xia Ma [ID],[2]† Xiao-Bo Dong [ID],[3]‡ Yang Huang [ID],[4,5] Tobias Mistele [ID],[6] Bo Peng,[7,8]
Qian Long,[3] Tianqi Wang,[8,9] Liang Chang [ID][3] and Xi Jin[8]

[1]*Department of Physics & Astronomy, University of California, Riverside, CA 92521, USA*
[2]*Division of Particle and Astrophysical Science, Nagoya University, Nagoya, Aichi 464-8601, Japan*
[3]*Yunnan Observatories, Chinese Academy of Sciences, Kunming, Yunnan 650011, China*
[4]*University of Chinese Academy of Sciences, Shijingshan District, Beijing 100049, China*
[5]*National Astronomical Observatories, Chinese Academy of Sciences, Chaoyang District, Beijing 100101, China*
[6]*Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, D-60438 Frankfurt am Main, Germany*
[7]*School of Information Engineering, Southwest University of Science and Technology, Mianyang, Sichuan 621010, China*
[8]*Lab of Solid-State Electronics, Department of Physics, University of Science and Technology of China, Hefei, Anhui 230026, China*
[9]*Linx Lab, Hisilicon, Shenzhen, Guangdong 518129, China*

ABSTRACT
Aiming at discriminating different gravitational potential models of the Milky Way, we perform tests based on the kinematic data powered by the *Gaia* DR2 astrometry over a large range of (R, z) locations. Invoking the complete form of Jeans equations that admit three integrals of motion, we use the independent R- and z-directional equations as two discriminators (T_R and T_z). We apply the formula for spatial distributions of radial and vertical velocity dispersions proposed by Binney et al., and successfully extend it to azimuthal components, $\sigma_\theta(R, z)$ and $V_\theta(R, z)$; the analytic form avoids the numerical artifacts caused by numerical differentiation in Jeans-equations calculation given the limited spatial resolutions of observations, and more importantly reduces the impact of kinematic substructures in the Galactic disc. It turns out that whereas the current kinematic data are able to reject Moffat's Modified Gravity (let alone the Newtonian baryon-only model), Milgrom's MOND is still not rejected. In fact, both the carefully calibrated fiducial model invoking a spherical dark matter (DM) halo and MOND are equally consistent with the data at almost all spatial locations (except that probably both have respective problems at low-$|z|$ locations), no matter which tracer population or which meaningful density profile is used. Since there is no free parameter at all in the quasi-linear MOND model we use, and the baryonic parameters are actually fine-tuned in the DM context, such an effective equivalence is surprising, and might be calling forth a transcending synthesis of the two paradigms.

Key words: gravitation – Galaxy: halo – Galaxy: kinematics and dynamics – dark matter.

1 INTRODUCTION

Is the 'missing mass problem' on (circum-)galactic scales due to the presence of dark matter (DM) or alternatively the delicate deviation of the underlying physical law from Newtonian gravity/dynamics? This is a fundamental and long outstanding question (see reviews, e.g. Feng 2010; Milgrom 2010b; Famaey & McGaugh 2012; Bullock & Boylan-Kolchin 2017; Banik & Zhao 2022). The dynamics of gas and stars in and around galaxies has been observed to be in excess of the Newtonian gravity of the total baryonic content of the galaxies; the observational evidence includes the rotational curves of disc galaxies, the stellar velocity dispersion fields of low-luminosity galaxies, and

the low-surface-brightness parts of luminous galaxies, and so on (e.g. Angus et al. 2015; Dabringhausen et al. 2016; McGaugh, Lelli & Schombert 2016).

Surprisingly and importantly, there are tight couplings (e.g. the Tully–Fisher relation; Tully & Fisher 1977; McGaugh et al. 2000) between the excess gravity and the baryonic content (see the above reviews). This fact inspired a gradually increasing number of researchers to interpret the 'excess' with modified gravity (or dynamics) theories such as the 'modified Newtonian dynamics' (MOND) proposed by Milgrom (1983) and the 'modified gravity' (MOG) by Moffat (2006), instead of the popular DM paradigm. So far, however, no observational test is conclusive for the two competing paradigms on (circum-)galactic scales.

Previously, almost all observational tests (see Famaey & McGaugh 2012; Banik & Zhao 2022) of DM versus MOND (or MOG)[1]

---

[*] This work is evolved from two undergraduate-training projects, which were initially supported by Department of Astronomy of USTC and Yunnan Observatories of CAS, respectively.
† These authors contributed equally to this work.
‡ E-mail: xbdong@ynao.ac.cn

[1] By 'DM', we mean Newtonian gravity with a DM component in addition to baryonic components for galaxies. In this paper, we focus on the 'extra

have employed either rotational velocity data commonly (in disc galaxies) or sometimes stellar velocity dispersion ($\sigma_\star$) data, with only a few exceptions (e.g. Angus et al. 2015, Lisanti et al. 2019, and Chrobáková et al. 2020; see also Nipoti et al. 2007 for methodological analysis), using both observed rotational curve (RC; in the galactic-disc plane) and observed $\sigma_\star$ information (particularly in the direction vertical to the disc) of a galaxy. By invoking Jeans equations, data of $\sigma_\star$ as well as streaming velocity $\bar{v}_\theta$ are linked to models of the galactic gravitational potential Φ (see section 4.8 of Binney & Tremaine 2008). The advantage of jointly using both RC and $\sigma_\star$ data is obvious with more constraints independently (Stubbs & Garg 2005).

Unfortunately, almost all the studies involving $\sigma_\star$ data in the literature adopted an unrealistic simplification of Jeans equations: they all assumed a two-integral distribution function, for instance, the popular $f = f(H, L_z)$, where H is the Hamiltonian of the system and L_z the z-direction angular momentum. Thus, the stellar velocity-dispersion tensor having $\sigma_R = \sigma_z$ and $\sigma_{Rz} = 0$ denoted in the cylindrical coordinate system (R, θ, z), i.e. the $\sigma_\star$ distribution in a meridional plane is isotropic, and the tilt angle of the velocity ellipsoid $\alpha = 0$. By doing so, the corresponding velocity-dispersion terms in Jeans equations are reduced or vanished accordingly, and the Jeans equations are closed (see section 2.1 of Nipoti et al. 2007, Section 2.2 of Angus et al. 2015, section III.B of Lisanti et al. 2019; cf. section 2.1 of Kipper et al. 2016). But the fact, well known for decades, is that $\sigma_R \neq \sigma_z$ and $\sigma_{Rz} \neq 0$ in the observed disc galaxies (e.g. the MW and M31; see Kipper et al. 2016 and references therein). Besides, there is more evidence supporting the viewpoint that the stellar orbits do respect, for which there is no analytic expression though, a third integral of motion (see Kipper et al. 2016; also section 3.2 and section 4.4 of Binney & Tremaine 2008). Specifically, concerning Jeans-equations modelling of the MW, the necessity of incorporating the cross-dispersion term σ_{Rz} (i.e. tilt angle) in Jeans equations has been thoroughly analysed (e.g. Hessman 2015, section 3 of Büdenbender, van de Ven & Watkins 2015 and more subsequent studies).

Besides the purpose to close Jeans equations, a practical reason of the above unrealistic simplification is to circumvent the calculating difficulty: the Jeans equations can only be solved *numerically* for all practical purposes with observational kinematic data used, and – to be worse – it usually requires algorithmic techniques to calculate the general form of Jeans equations (involving three distinct $\sigma_\star$ components and the cross term σ_{Rz} and their derivatives), given the *limited* observational data so far. Normally, it involves numerically calculating the partial derivatives of those $\sigma_\star$ components with respect to R and z (e.g. Chrobáková et al. 2020; cf. Section 4.1), which in principle demands dense sampling along the R and z directions (as well as careful numerical differentiation schemes or novel algorithms to minimize the notorious 'huge numerical errors'), and worse, is vulnerable to the impact of galactic substructures. The worrisome fact is that the stellar kinematics in galactic discs (e.g. the disc of the MW) is commonly affected by stellar substructures; or, in other words, galactic discs are full of kinematic substructures (Gaia Collaboration et al. 2018b).

In addition, in the aforementioned studies invoking Jeans equations, they not only simplified Jeans equations by assuming two-integral dynamics, but also usually approximated the solution of

Jeans equations with an algebraic formula between the averaged vertical σ_z^2 and the mass surface density locally at every radius R, i.e. $\sigma_z^2(R) \sim \Sigma(R)$ (commonly seen for external disc galaxies; see e.g. Angus et al. 2015). That simple formula was derived by neglecting the components of gravitational force in the $z = const$ planes (i.e. assuming that the gravitational force is in the z direction only, a so-called planar symmetry in the literature), which was actually wrongly assumed (or over-simplified) for the dynamics of stars (see section 6.1 of Piffl et al. 2014 and section 4.2 of McGaugh 2016 for the MW; footnote 2 of Nipoti et al. 2007 for external galaxies). This simplification is actually the simplest version of the old 'K_z method' so-called in the literature, and makes the system completely one-dimensional in the sense of both Jeans equations and Poisson equation. To be specific, following the notations of Read (2014) (see his Section 3.3), while $K_z(z)$ means vertical force, $-\frac{\partial \Phi}{\partial z}(z)$ literally, this simplest K_z method yet ignores both the 'tilt term' in z-directional Jeans equation and the 'rotation-curve term' in Poisson equation. Likewise, in some studies using MW data, the link between the vertical density profile ($\rho(z)$), and vertical distribution of the z-component velocity dispersion ($\sigma_z(z)$) of a tracer population was established by this simplest K_z method (see e.g. Section III.B of Lisanti et al. 2019).

Aiming at observationally discriminating between DM and alternative gravitational potential models, we employ the complete form of Jeans equations that admit three integrals of motion,[2] and perform tests on the latest kinematic data powered by the G*aia* DR2 astrometry. In the *Gaia* era, the measurement uncertainties (e.g. the effect onto kinematic quantities caused by systematic bias in distance estimation) are no longer the major concern (see Section 4.1). Because the general form (namely 3-integral) of Jeans equations are not closed, instead of solving it with the above-mentioned simplifications, we use the two independent Jeans equations, R- and z-directional, as two discriminators of the consistency between gravitational potential models and kinematic data. In order to (1) reduce the impact of various kinematic substructures in the Galactic disc, as well as (2) to avoid the numerical artifacts caused by numerical differentiation in Jeans-equations calculation given the limited spatial resolutions of the observational data, we apply the analytic form for $\sigma_R(R, z)$ and $\sigma_z(R, z)$ proposed by Binney et al. (2014), and successfully extend it to the azimuthal components $\sigma_\theta(R, z)$ and $V_\theta(R, z)$. Our comprehensive tests consistently point to the conclusion: Whereas the current kinematic data, with the precision and accuracy powered by *Gaia* DR2, is able to reject the MOG model (let alone the Newtonian baryon-only model, adopting the baryonic mass distribution priorly best-fitted in the DM paradigm), the MOND model is still not rejected, and behaves as good as the DM model. This is surprising, because while the fiducial DM model we adopt was carefully pre-fitted with all available Galactic kinematic data, and in fact has been kept improving elaborately by researchers during past decades (see Section 3.1 and the references therein), there is no free parameter at all in the MOND model (no bother of fitting), and the parameters of the baryonic mass model are actually fine-tuned in the DM context.

This paper is organized as follows. In Section 2, we describe the complete form of Jeans equations for axisymmetric systems, and propose the two measures T_R and T_z. In Section 3, we give the fiducial

---

mass/gravity' phenomena on circum-galactic and galactic scales only, i.e. within the gravitational binding of the so-called DM halo (in the DM language) hosting some galaxy. It is in this context that the statements like this sentence hereinafter should be understood.

[2]Regarding the methodology, one of our aspirations came from the critical analysis by M. Milgrom on the methodology of the DiskMass project, particularly on the analysis method of Angus et al. (2015); see Milgrom (2015) and Angus, Gentile & Famaey (2016) for the detail.

mass distribution model of the MW used in this work, with best-fit model parameters in the DM context (Section 3.1), and describe two alternative gravitational potential models, namely quasi-linear MOND (Section 3.2) and MOG (Section 3.3). In Section 4, we introduce the data we employ particularly in Section 4.1, we describe our further analysis of the 3D velocity data of Huang et al. (2020), and present our best-fit formulae for the spatial distributions, namely $\sigma_R(R, z)$, $\sigma_z(R, z)$, $\sigma_\theta(R, z)$, and $V_\theta(R, z)$. In Section 5, we present the results of comprehensive tests on the gravitational potential models, particularly the T_R and T_z tests using different tracer populations with various density profiles of tracers assumed (Sections 5.2 and 5.3); in Section 5.6, we discuss the physical implication as well as its practical application of our main result. In addition, in the Appendix, we present the results using a different parametrization of the Galactic mass model and corresponding kinematic data, which are consistent with the results in the main text. Section 6 summarizes the paper. Throughout the paper, we adopt a Galactocentric cylindrical system, with R being the projected Galactocentric distance increasing radially outwards, θ toward the Galactic rotation direction, and z in the direction of north Galactic pole.

2 TWO DISCRIMINATORS IN TERMS OF THE THREE-INTEGRAL JEANS EQUATIONS

Rotation curves, which involve rotational velocities (i.e. in the azimuthal direction) only, and are conventionally measured in the galactic mid-plane only, are one-dimensional: reflecting the azimuthally averaged R-directional acceleration; i.e. $V_c^2(R)/R = \partial\Phi/\partial R(R; z = 0)$.

Most previous applications of Jeans equations, as described in the Introduction, assumed two-integral dynamics and even additional simplifications, which are not consistent with the observed kinematic data of the WM, the subject of the present study.

Our own Galaxy provides three-dimensional data, i.e. the 3-directional components of velocity-related quantities (see Section 4). Moreover, it enable us to test gravitational potential models *at different spatial locations (R, z)*, or even at three-dimensional locations (R, θ, z) in the future. This is in stark contrast with external galaxies, where only vertically-averaged quantities are available, such as observed radial σ_z profiles (e.g. the DiskMass project; see Angus et al. 2015).[3] The point is, the set of kinematic data ($\sigma_\star$ and $\bar{v}_\theta$) at *every* (R, z) location can be regarded as an *independent* constraint to the gravitational models through Jeans equations, and thus the more data points – particularly those at relatively large z – the better the models get constrained.

To test gravitational models comprehensively, with three-dimensional kinematic quantities (namely their R-, z-, and θ-directional components) and at different (R, z) locations, we invoke the complete form of Jeans equations. For a steady-state collisionless gravitational system, Jeans equations relate the gravitational field of the system to the density and kinematic qualities of a certain tracer population (Section 4.8 of Binney & Tremaine 2008). We write the equations using the notations of Kipper et al. (2016). Because the mass models we use are axisymmetric (see Section 3.1), the two cross-term components of the velocity dispersion tensor are zero,

---

$\sigma_{R\theta} = \sigma_{\theta z} = 0$. Thus, the complete Jeans equations can be written as two independent equations in cylindrical coordinates:

$$\frac{\partial(\rho\sigma_R^2)}{\partial R} + \left(\frac{1 - k_\theta}{R} + \frac{\partial\kappa}{\partial z}\right)\rho\sigma_R^2 + \kappa\frac{\partial(\rho\sigma_R^2)}{\partial z} - \rho\frac{V_\theta^2}{R}$$
$$= -\rho\frac{\partial\Phi}{\partial R}, \tag{1}$$

$$\frac{\partial(\rho\sigma_z^2)}{\partial z} + \left(\frac{\xi}{R} + \frac{\partial\xi}{\partial R}\right)\rho\sigma_z^2 + \xi\frac{\partial(\rho\sigma_z^2)}{\partial R} = -\rho\frac{\partial\Phi}{\partial z}, \tag{2}$$

where $\kappa = \frac{1}{2}\tan(2\alpha)(1 - k_z)$, $\xi = \kappa/k_z$, and $V_\theta \equiv \bar{v}_\theta$, the averaged azimuthal velocity of tracers at every location. The parameter α is the tilt angle of the velocity ellipsoid, i.e. the angle by which the ellipsoid's longest axis at every position is tilted with respect to the galactic-disc plane. The other two parameters, k_z and k_θ, are the axial ratios of the ellipsoid: $k_z = \sigma_z^2/\sigma_R^2$ and $k_\theta = \sigma_\theta^2/\sigma_R^2$. Note that ρ in Jeans equations is tracer's density, while Φ is the total gravitational potential contributed by all components of the system.

Given observed kinematic data, a right gravitational model or theory should satisfy the two Jeans equations everywhere throughout the MW. As mentioned, because the equations are not closed, we define two measures as follows,

$$T_R = -\frac{1}{\rho}\left\{\frac{\partial(\rho\sigma_R^2)}{\partial R} + \left(\frac{1 - k_\theta}{R} + \frac{\partial\kappa}{\partial z}\right)\rho\sigma_R^2\right.$$
$$\left. + \kappa\frac{\partial(\rho\sigma_R^2)}{\partial z} - \rho\frac{V_\theta^2}{R}\right\}, \tag{3}$$

and

$$T_z = -\frac{1}{\rho}\left\{\frac{\partial(\rho\sigma_z^2)}{\partial z} + \left(\frac{\xi}{R} + \frac{\partial\xi}{\partial R}\right)\rho\sigma_z^2 + \xi\frac{\partial(\rho\sigma_z^2)}{\partial R}\right\}. \tag{4}$$

According to the Jeans equations (equations (1) and (2)), a correct gravitational model (Φ) should satisfy

$$T_R = \frac{\partial\Phi}{\partial R} \tag{5}$$

and

$$T_z = \frac{\partial\Phi}{\partial z}, \tag{6}$$

everywhere throughout the MW. We call the above two criteria 'T_R test' and 'T_z test', respectively. We will see (Section 5.3), the discriminating power of T_R test comes from the fact that it is fairly insensitive to the choice of tracer's density profile (namely the common prescriptions for galactic components), while the merit of T_z test is instead its sensitivity to tracer's density profile.

The measure T_R, in fact, is the observed R-directional acceleration, calculated from σ_R, σ_z (through κ), σ_θ (through k_θ), V_θ, and tilt angle α (through κ), as well as tracer's density profile $\rho(R, z)$. Thus T_R test means that the observed R-directional acceleration equals to the radial gradient of gravitational potential at any locations. It can be regarded as a generalized rotation-curve test, on and off the galactic mid-plane (Chrobáková et al. 2020).

Likewise, the measure T_z is the observed z-directional acceleration, calculated from σ_z and tilt angle α (through ξ), as well as tracer's density profile $\rho(R, z)$. T_z test means that the observed z-directional acceleration equals to the vertical potential gradient at any locations. In testing the vertical characteristics of gravitational models, T_z is more universal and accurate (i.e. without additional simplifications) than the old K_z ('vertical force') method as mentioned in the Introduction (see also Section 3.3 of Read 2014, and Section 5.2 below), that is generally either partially one-dimensional

---

[3] The σ_z profiles mean $\sigma_z(R)$ where σ_z is averaged or integrated over the z direction, similar to the form of RCs $V_c(R)$. Likewise, in radial profiles of line-of-sight (LOS) $\sigma_\star$ (namely $\sigma_{los}(R)$; e.g. Kipper et al. 2016), σ_{los} is averaged or integrated along the line of sight.

(e.g. neglecting the tilt term in vertical Jeans equation; e.g. McGaugh 2016), or even completely one-dimensional (in both vertical Jeans and Poisson equations; e.g. Lisanti et al. 2019).

Note that we deliberately use a different terminology 'T_z' (as well as 'T_R') rather than the old 'K_z', in order to avoid any possible prejudice resulting from the simplified use of the 'K_z' method prevailing in the literature, and to stress that our two measures by definition are *observed* vertical and radial *kinematic accelerations* calculated from tracers' density profile and kinematic data. By definition, K_z is vertical *field strength*, namely the negative of gradient, of (theoretical models of) gravitational potential. Because of the same consideration in this paper, we often use the words 'acceleration' versus 'field strength, force, or gradient' differently.

3 MASS DISTRIBUTION (POTENTIAL) MODELS OF THE MILKY WAY

In this work, we focus on the global potential field of the MW, particularly the outer part where the circular velocity and velocity dispersion are dominated by the supposed DM, we therefore choose to ignore kinematic substructures of stars, and non-axisymmetric structures (e.g. bars and spiral arms) that are dynamically important mainly in the inner part. Specifically, we use and compare axisymmetric mass models of the MW throughout the paper.

For all the models, we implement a light-weight C program to solve the axisymmetric potential on a 1280^3 grid using the direct sum method (Binney & Tremaine 2008). The grid is equally divided into cells, and every cell physically corresponds to a spatial size 60 pc on a side. We have checked the numerical convergence and verified our results, using FreeFem++ (Hecht 2012), a popular software solving partial differential equations with the finite-element method (FEM), which achieves both high-spatial resolution and high precision.

3.1 Fiducial model of the Galactic mass distribution

The fiducial mass model we use in the main text is the one prescribed by Wang, Hammer & Yang (2022). It adopts the mass distribution profile formulae and basic structural parameter values from the best-fit main model of McMillan (2017) for the bulge, stellar discs, and interstellar medium discs, and the Zhao's (Zhao 1996) profile for the DM halo. The density values (namely the normalization of the aforementioned profiles), as well as the scale lengths of thin and thick stellar discs and the other parameters of the DM halo, are constrained by Wang et al. (2022) with latest observations powered by *Gaia* DR2 (Gaia Collaboration et al. 2018a) and *Gaia* EDR3 (Gaia Collaboration et al. 2021). We briefly summarize the details of every components below.

The bulge's density profile is

$$\rho_{\rm b} = \frac{\rho_{0,\rm b}}{(1 + \frac{r'}{r_0})^\alpha} \exp\left[-\left(\frac{r'}{r_{\rm cut}}\right)^2\right], \qquad (7)$$

and, in cylindrical coordinates,

$$r' = \sqrt{R^2 + \left(\frac{z}{q}\right)^2}, \qquad (8)$$

with $\rho_{0,\rm b} = 9.5 \times 10^{10}\,{\rm M_\odot kpc^{-3}}$, $\alpha = 1.8$, $r_0 = 0.075\,{\rm kpc}$, $r_{\rm cut} = 2.1\,{\rm kpc}$, and axis ratio $q = 0.5$.

The stellar discs of the Milky Way are usually considered to be divided into two components: the thin disc and thick disc. Their mass

Table 1. Disk parameters for the Milky Way mass model we use.

	Thin	Thick	H I	H$_2$
$\Sigma_0[{\rm M_\odot\,pc^{-2}}]$	1003.12	167.93	53.1	2179.5
$R_{\rm d}[{\rm kpc}]$	2.42	3.17	7.0	1.5
$z_{\rm d}[{\rm kpc}]$	0.3	0.9	0.085	0.045
$R_{\rm m}[{\rm kpc}]$	–	–	4.0	12.0

distributions follow the following form

$$\rho_{\rm d}(R, z) = \frac{\Sigma_0}{2z_{\rm d}} \exp\left(-\frac{|z|}{z_{\rm d}} - \frac{R}{R_{\rm d}}\right), \qquad (9)$$

with corresponding scale height $z_{\rm d}$, scale length $R_{\rm d}$, and central surface density Σ_0.

The interstellar medium of the Milky Way includes two components: the H I and molecular H$_2$ gas discs. These discs follow the density law

$$\rho_{\rm g}(R, z) = \frac{\Sigma_0}{4z_{\rm d}} \exp\left(-\frac{R_{\rm m}}{R} - \frac{R}{R_{\rm d}}\right) {\rm sech}^2(z/2z_{\rm d}), \qquad (10)$$

with $R_{\rm m}$ being the associated scalelength of the central hole. The actual width of the hole is determined by both $R_{\rm d}$ and $R_{\rm m}$, with the maximum surface density (i.e. the rim of the hole) being at $R = \sqrt{R_{\rm d}R_{\rm m}}$. The parameters of the stellar and gas discs are listed in Table 1.

The DM halo is described by the Zhao's profile,

$$\rho_{\rm h}(r) = \rho_{0,\rm h}\left(\frac{r}{r_{\rm h}}\right)^{-\gamma}\left[1 + \left(\frac{r}{r_{\rm h}}\right)^\alpha\right]^{(\gamma-\beta)/\alpha}, \qquad (11)$$

where the full set of three free parameters (α, β, γ) can be calculated analytically. In this paper, we adopt the best derived value (α, β, γ) = (1.19, 2.95, 0.95) (see Table 2 of Wang et al. 2022). The Zhao's profile is more flexible than the widely used NFW (Navarro, Frenk & White 1996) profile, and will reduce to the normal NFW formula in the case of (α, β, γ) = (1, 3, 1). The remaining halo parameters are as follows: $\rho_{0,\rm h} = 1.55 \times 10^7\,{\rm M_\odot\,kpc^{-3}}$, $r_{\rm h} = 11.75\,{\rm kpc}$, and $q = 0.95$.

In this study (except in the Appendix), we use the distance from the Sun to the Galactic Centre $R_\odot = 8.122\,{\rm kpc}$ (Gravity Collaboration et al. 2018), and a nominal circular velocity $v_\odot \simeq 229.0\,{\rm km\,s^{-1}}$ at the radius of the Sun (e.g. Eilers et al. 2019). The fiducial model of the Galactic mass distribution was built under the same $R_\odot$ and $v_\odot$ constants, i.e. the same as Eilers et al. (2019) (Wang 2022, private communication).

We have explored other parametrizations of the Galactic mass distribution (as well as other kinematic data), including those under other sets of the solar position and velocity values ($R_\odot$ and $v_\odot$), and found that our conclusions remain intact. Such an examination, performed under the legacy $R_\odot$ and $v_\odot$ values, is presented in the Appendix.

Finally, because the model parameters of the above Galactic components were constrained in the DM context for fair comparison between DM and modified-gravity models, we need to make clear to what degree the data used in constraining the fiducial mass model (by McMillan 2017 and Wang et al. 2022) overlap the data we use here to discriminate gravitational models. Here, we summarize the data that were already used to fit the fiducial mass model, and list the overlapped parts with the data used in the present study. McMillan (2017) used various rotation-curve data, solar velocity (to constrain $v_\odot$), vertical-force data at $|z| = 1.1\,{\rm kpc}$ and $R = R_\odot$ of Kuijken & Gilmore (1991), and the upper limit of the total mass within the MW's

inner 50 kpc according to Wilkinson & Evans (1999). Wilkinson & Evans (1999) based their estimate on the distance and velocity data of 27 objects in the outer Galaxy (satellite galaxies and globular clusters at $R > 20$ kpc). Wang et al. (2022) used the rotation-curve data of Eilers et al. (2019), the vertical-force data at $|z| = 1.1$ kpc and $4 \lesssim R \lesssim 9$ kpc, $K_{z, 1.1 \, \mathrm{kpc}}(R)$, derived by Bovy & Rix (2013) based on G-type dwarf stars from SDSS/SEGUE survey (see also Section 5.4), and kinematic data of globular clusters. In relation to the data used in the present study (see Section 4), (1) the rotation-curve data, concerning radial accelerations in the Galactic plane, are essentially overlapped (particularly the best data obtained by Eilers et al. 2019); (2) the radial accelerations off the Galactic plane (namely rotation curves at $|z| > 0$) probed by our data are not available in either McMillan (2017) or Wang et al. (2022); (3) as to the data concerning vertical accelerations (e.g. the so-called 'vertical force' K_z data), in effect there is overlap to a certain degree, but the vertical accelerations probed by our data are not limited at $|z| = 1.1$ kpc; (4) the data of satellite galaxies and globular clusters used by McMillan (2017) and Wang et al. (2022) are completely irrelevant to our data.

3.2 Quasi-linear MOND

QUMOND is the quasi-linear realization (Milgrom 2010a) of the MOND theory (Milgrom 1983). MOND was initially proposed to explain the flat rotation curves of galaxies without DM. We refer the reader to recent reviews (such as Famaey & McGaugh 2012 and Banik & Zhao 2022) for detailed and lucid descriptions. In this work, essentially we treat QUMOND as a gravitational potential model rather than a 'modified gravity or dynamics' *theory*; i.e. we employ it in the fashion of $\rho_{\mathrm{b}} + \rho_{\mathrm{pdm}}$, with ρ_{pdm} as an alternative of popular DM haloes. Here, 'pdm' (or in capital letters) means 'phantom dark matter', a term coined to reflect that this MOND effect – such a virtual (phantom) stuff – would be interpreted by a Newtonist as a DM halo (see below). We calculate the QUMOND potential with the baryonic mass density profile prescribed in the fiducial mass model.

The MOND acceleration was originally written in the following way (the Milgrom 1983 formula):

$$\mu \left(\frac{g}{a_0} \right) \mathbf{g} = \mathbf{g}_{\mathrm{N}}, \tag{12}$$

where $\mu(x)$ is an interpolating function, and

$$\mu(x) \to 1 \text{ for } x \gg 1 \text{ and } \mu(x) \to x \text{ for } x \ll 1. \tag{13}$$

Here $\mathbf{g}_{\mathrm{N}}$ is the Newtonian gravitational acceleration $\mathbf{g}_{\mathrm{N}} = -\nabla \Phi_{\mathrm{N}}$, and Φ_{N} is the Newtonian potential:

$$\nabla^2 \Phi_{\mathrm{N}} = 4\pi G \rho_{\mathrm{b}}, \tag{14}$$

with ρ_{b} being the baryonic matter density. In terms of the simple Milgrom 1983 formula (equation 12), however, the acceleration field $\mathbf{g}$ is not derivable from a scalar potential, and consequently there is no conserved momentum.

QUMOND, just like its cousin AQUAL (aquadratic Lagrangian formulation of MOND, Bekenstein & Milgrom 1984), is a complete theory that is self-consistently derivable from an modified Newtonian gravitational action (see Famaey & McGaugh 2012 for the detail). QUMOND has the following Poisson equation:

$$\nabla^2 \Phi = \nabla \cdot \left[\nu \left(\frac{g_{\mathrm{N}}}{a_0} \right) \nabla \Phi_{\mathrm{N}} \right], \tag{15}$$

where scalar Φ is the QUMOND gravitational potential, and $\nu(y)$ is an interpolating function. The function $\nu(y)$ is related to the above

$\mu(x)$ by $\nu(y) = 1/\mu(x)$ and $y = x\mu(x)$. We can define $\tilde{\nu}(y) = \nu(y) - 1$, then equation (15) leads to

$$\nabla^2 \Phi = \nabla \cdot \left[\nabla \Phi_{\mathrm{N}} + \tilde{\nu} \left(\frac{g_{\mathrm{N}}}{a_0} \right) \nabla \Phi_{\mathrm{N}} \right], \tag{16}$$

or

$$\nabla^2 \Phi = 4\pi G (\rho_{\mathrm{b}} + \rho_{\mathrm{pdm}}). \tag{17}$$

Equation (17) reveals the merit of QUMOND: the gravitational potential can be ascribed formally to two matter sources in terms of normal Poisson equation, the baryonic matter and the afore-mentioned PDM. From a mathematical point of view, the PDM density ρ_{pdm} is conceptually equivalent to the density of 'DM haloes' (but with totally different physical content); see Section 5.5, also Milgrom 2010a and section 6.1.3 of Famaey & McGaugh 2012. Accordingly, there is a striking technical advantage (e.g. compared with AQUAL that involves a non-linear generalization of Poisson), which is obvious: QUMOND involves solving only *linear* differential equations (namely the normal Poisson equation). Thus, *all* the well-developed algorithms (e.g. Tree-PM) and codes (e.g. `Gadget` of Springel 2005) for Newtonian *N*-body numerical calculations and simulations are still usable in QUMOND.[4]

In practice, given baryonic ρ_{b} or Φ_{N}, ρ_{pdm} is calculated straightforward as follows,

$$\rho_{\mathrm{pdm}} = \frac{1}{4\pi G} \nabla \cdot \left(\tilde{\nu} \left(\frac{g_{\mathrm{N}}}{a_0} \right) \nabla \Phi_{\mathrm{N}} \right). \tag{18}$$

Correspondingly, we can trivially define a scalar ϕ_{pdm} as the PDM potential,

$$\nabla^2 \phi_{\mathrm{pdm}} = \nabla \cdot \left(\tilde{\nu} \left(\frac{g_{\mathrm{N}}}{a_0} \right) \nabla \Phi_{\mathrm{N}} \right) = 4\pi G \rho_{\mathrm{pdm}}, \tag{19}$$

then the QUMOND potential can be written as $\Phi = \Phi_{\mathrm{N}} + \phi_{\mathrm{pdm}}$.

In this work, the critical acceleration constant is held fixed to be the commonly used value $a_0 = 1.2 \times 10^{-10}$ m s^{-2} (Banik & Zhao 2022). The simple formula of $\nu(y)$ is adopted (Famaey & McGaugh 2012):

$$\nu(y) = \frac{1}{2} \sqrt{1 + \frac{4}{y}} + \frac{1}{2}. \tag{20}$$

That is, there is no free parameter at all in the QUMOND formula that we use in this study.

Note that in this work, we have not taken into account the so-called 'external field effect' (EFE) of MOND. EFE is a general characteristic of MOND (particularly its modified-gravity theories such as QUMOND), because MOND depends on the *total* acceleration with respect to some pre-defined (inertial) frames. But EFE does not necessarily exist in specific MOND theories (see Section 4.6 of Milgrom 2014), particularly in modified-inertia theories of MOND (see Milgrom 2011). Thus, in this work, we only practically use QUMOND as a practical (effective) formula to calculate the MOND potential of the MW baryons, and refrain from accounting for the subtlety of EFE. Anyway, practically, the gravitational strength of the external field around the MW is reasonably estimated to

---

[4]In the literature, there was a claim that due to the non-linearity of MOND, the Poisson solvers that are not based on grids/meshes, such as tree-codes, cannot be used (e.g. section 2.1 of Angus et al. 2013). This is not necessarily true for QUMOND, because one can build a temporary grid to implement equation (18), calculating PDM density from Newtonian potential, which is not difficult technically (section 6.1.3 of Famaey & McGaugh 2012).

be 0.01–0.03 a_0 (Wu et al. 2008), which is $\sim 10^2$ times smaller than the Newtonian gravitational strength at the (R, z) locations considered in this work; i.e. the EFE is negligible for our purpose. In addition, mention in passing that by defining $\mathbf{g} = -\nabla\Phi$, the complete MOND theories so far (such as QUMOND and AQUAL) assume the gravitational vector field is still curl-less, in contrast, the gravitational or acceleration field $\mathbf{g}$ defined in the pristine Milgrom (1983) formula (equation 12) is curled.

3.3 Moffat's MOG

We also test another alternative to DM, Modified Gravity (MOG; e.g. Moffat 2006; Moffat & Rahvar 2013), which is a covariant modification of Einstein gravity. Simply put, MOG adds two additional scalar fields and one vector field to explain the dynamics of astronomical systems based on the distribution of baryonic matter.

In the weak field approximation (e.g. in the MW), the effective potential for an extended distribution of baryonic matter (ρ) in MOG is as follows:

$$\Phi(\boldsymbol{x}) = -G_\infty \left[\int \frac{\rho(\boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|} \left(1 - \frac{G_\infty - G_N}{G_\infty} e^{-\mu|\boldsymbol{x} - \boldsymbol{x}'|} \right) \mathrm{d}^3\boldsymbol{x}' \right]. \tag{21}$$

with $G_\infty = (1 + \alpha)G_N$ being the modified gravitational constant. In this work, the two universal constants are held fixed to be $\alpha = 8.89$ and $\mu = 0.042\,\mathrm{kpc}^{-1}$, which are best fitted with the rotation-curve data of external galaxies by Moffat & Rahvar (2013).

4 DATA

We use recent kinematic observations, including the rotation curve and three-dimensional velocity dispersion of the MW to test the gravitational models. We only include data at $R > 4\,\mathrm{kpc}$ to avoid the complexity in the central region of the MW. Besides the data collected from the literature, we analyse and fit the spatial distributions along R and z directions of σ_R, σ_z, σ_θ, and V_θ (namely the mean azimuthal velocity, see Section 2). We basically follow the methodology of Binney et al. (2014), except for an additional innovation that we also give well-parametrized formulae for $\sigma_\theta(R, z)$ and $V_\theta(R, z)$, which are described below (Section 4.1).

Our own Galaxy is remarkable in testing gravitational models. There are already plenty of kinematic observations of both RC and $\sigma_\star$ (as well as $\bar{v}_\theta$). Moreover, although on the one hand our position inside the Galactic disc weakens the ability to measure the RC in the outer Galaxy, on the other hand it allows a three-dimensional measurement of the position and velocity of individual stars, particularly of those in the z direction far into the halo.

4.1 Spatial-distribution formulae for three-dimensional kinematics based on *Gaia* DR2

We analyse the three-dimensional velocity data of the `LAMOST` and `Gaia red clump sample` complied by Huang et al. (2020). This sample, consisting of $\approx 137\,000$ red clump stars (as the tracer population in this work), has a good coverage of the Galactic disc of $4 \leq R \leq 16\,\mathrm{kpc}$ and $|z| \leq 4\,\mathrm{kpc}$.

In order to reduce the impact of particular structures in the Galactic disc (e.g. stellar streams of various origins; Gaia Collaboration et al. 2018b), we fit the velocity dispersion, σ_R, σ_θ, and σ_z, to the smooth analytic forms with respect to R and z given by Binney et al. (2014, particularly cf. their tables 2 and 3), and thus acquire the 'macro' (namely spatially coarse-grained) kinematics. For the same reason,

Table 2. Best-fitting values of the parameters defined by equation (22) required to fit the dependence on (R, z) for σ_R, σ_θ, σ_z, V_θ, respectively.

	a_1	a_2	a_3	a_4
σ_R	1.177	0.688	32.196	0.105
σ_θ	0.698	0.661	9.437	0.509
σ_z	0.615	0.631	34.453	0.168
V_θ	6.914	0.008	2.418	−0.742

we make no efforts in distinguishing different stellar groups, although we appreciate the difference in the kinematics of stars with different age and metallicity (Huang et al. 2020).

In fact, we have tried using only the red clump stars in the thin disc (numbering $\approx 116\,000$; according to the [Fe/H]–[α/Fe] criterion by Huang et al. 2020) as the tracer population, which would enable us to have a better constraint on the density profile of the tracers (cf. Section 5.3), e.g. by simply adopting the geometrical thin-disc component in the fiducial mass model as the tracer's density profile. But, it turns out that if we do so, many spatial bins at $|z| > 1\,\mathrm{kpc}$ have not sufficient stars to fit the v_θ probability distribution (see below), and thus disable us to perform the T_R tests for those spatial locations. Because T_R is the important and robust measure to test the gravitational models (see Sections 5.2 and 5.3), we base our main results of this work on the entire red clump sample of Huang et al. (2020), and for safety we test our results by using three schemes of density profile for the tracers. Besides the results based on the thin-disc-only red clump stars are consistent with those based on the entire sample (see Section 5.3).

Binney et al. (2014) presented parametrized formulae for the spatial distributions of the two meridional-plane components of stellar velocity dispersions (i.e. velocity ellipsoid), namely $\sigma_1(R, z)$ and $\sigma_3(R, z)$, as follows (their equation 4):

$$\sigma(R, z) = \sigma_0 a_1 \exp[-a_2(R/R_\odot - 1)] \left[1 + (a_3 z/R)^2 \right]^{a_4}. \tag{22}$$

The above functional form comes out of physical intuition as well as their trial and error, and work well in practice. Here we adopt the same formulae for σ_R and σ_z, and set the parameters (see the constants in Table 2) to be free and constrained by our data. Formally there seems a difference in that the formulae of Binney et al. (2014) are for the two principal velocity dispersions, and here for those along the R and z directions. But in essence this is not a problem (considering the semi-empirical nature of the formulae), and has been verified by our experiment. Rather, this is partly the reason that we allow our best-fit constants can be different to some degree from those of Binney et al. (2014).

Binney et al. (2014) presented a novel fitting recipe (see their equations 7 & 8) to model the distributions of the azimuthal velocities (v_θ) of the tracers on every spatial location, i.e. for their *every (R, z) bin*; it is well-known, as the *asymmetric drift* phenomenon, that the v_θ distributions are highly non-Gaussian. The Binney et al. (2014) distribution function takes a form of *sigma-varying* Gaussian, i.e. with different σ (dispersion) for different v_θ, for the sample of tracer stars in a spatial bin, the fitting is extremely good when applied to observed data.

We make a further innovation on the shoulder of Binney et al. (2014) out of our exploration: The spatial distributions of the mean azimuthal velocity (V_θ) and its corresponding σ_θ (derived by the Binney et al. 2014 methodology, as described in the above paragraph), i.e. distributions over a range of spatial bins can be well

fitted by the formula of equation (22) also. The possibility of such an innovation was actually discussed by Binney et al. (2014, see their section 4.1), although the RAVE data they used only cover a small region within ∼2 kpc of the Sun. We now have a sample of V_θ and σ_θ data with larger coverage in the $R - z$ plane than Binney et al. (2014), which exhibit apparent trends of V_θ and σ_θ over large spatial scales enabling us to conduct such an exploration.

Following the methodology described in the above three paragraphs, we calculate the quantities $\sigma_R, \sigma_z, \sigma_\theta$, and V_θ for every spatial bins, and then fit their spatial distribution with equation (22). The σ_0 in the equation is fixed to be 30 km s^{-1}. Our best-fit parameters are listed in Table 2.

Besides the merit of the well-parametrized analytic formulae of $\sigma_R(R, z)$, $\sigma_z(R, z)$, $\sigma_\theta(R, z)$, and $V_\theta(R, z)$ *per se*, the analyticity of $\sigma_R(R, z)$ and $\sigma_z(R, z)$ leads to a great advantage in calculating T_R and T_z: derive the partial derivatives analytically (such as $\partial\sigma_R/\partial R$, $\partial\kappa/\partial z$, $\partial\sigma_z/\partial z$, etc.), free of the technical difficulties in calculating those partial derivatives numerically instead (e.g. 'huge' errors in such numerical implementation given the spatial resolutions so far; see e.g. Chrobáková et al. 2020). Again, we would like to stress that our major motivation of using these spatial-distribution formulae is to reduce the astrophysical 'impurities' such as kinematic substructures.

Regrading the spatial binning of the data, generally we divide the entire space (4 < R < 16 kpc and −4 < z < 4 kpc) into bins of $\Delta R = 0.2$ kpc and $\Delta z = 0.05$ kpc. We use the bins with more than 10 objects to fit equation (22) for $\sigma_R(R, z)$, $\sigma_z(R, z)$, and $V_\theta(R, z)$. As for σ_θ, in order to get relatively reliable fitting to its statistical distribution within a specific bin (Equation 7 of Binney et al. 2014), we only employ the bins with more than 50 objects, derive their σ_θ, and fit equation (22) for $\sigma_\theta(R, z)$. By and large, the spatial bins over the 'continuous' space of 6 < R < 12 kpc and −2.5 < z < 2.5 kpc have σ_θ measurements. Thus, the reliable R range for applying the best-fit spatial distributions (equation 22) is 6 < R < 12 kpc (without poor fitting on the boundaries because of abundant data at R < 6 and R > 12 kpc); the reliable |z| (the distance to the Galactic mid-plane) range is conservatively deemed to be from the resolution limit (see Section 5.4) to |z| = 2 kpc.

Regarding the measurement uncertainties of the velocity dispersion values in the spatial bins, the 1-σ statistical errors in σ_R are <2.8 km s^{-1}, those in σ_z are <2.5 km s^{-1}, and those in σ_θ, <3.4 km s^{-1}, the mean error in any one of the three quantities is 0.5 km s^{-1}. The above quoted errors already include the effect of systematic errors in distance estimation on the derived kinematic quantities, because the uncertainties of the 3D velocities given by Huang et al. (2020) have accounted for all kinds of error sources by Monte Carlo simulation. In fact, the total measurement uncertainty in distance is 5–10 per cent (see Section 5.2 of Huang et al. 2020) to which the contribution of systematic bias is minor by virtue of the power of *Gaia*. This is totally different from the situation prior to the *Gaia* era (cf. Section 5.3 of Binney et al. 2014).

The uncertainties in the fitted parameters of the spatial-distribution formulae (see equation (22)) are dominated by two parts: the statistical uncertainties of the kinematic quantities described in the above, and the physical fluctuations (i.e. deviations from the model owing to astrophysical reasons, on small spatial scales, say, with $R \lesssim$ 1 kpc and $z \lesssim 0.5$ kpc). In the analysis of this study (concerning data binning, etc.), the two parts are comparable to each other. And, when used in our Jeans-equations tests (Section 5.2), these uncertainties are relatively minor compared with the uncertainties in the density profile of tracers (see Section 5.3). We have checked that our results are not sensitive to binning schemes (including bin sizes and the aforementioned thresholds) or fitting methods. The details of the

data binning and analysis are beyond the scope of this study, and will be included in a future paper investigating the Galactic kinematics of *Gaia* stars.

The tilt angle information (required in the Jeans equations (1) and (2)) is taken from the measurement by Everall et al. (2019) for a sample of disc stars with *Gaia* DR2 astrometry, $\alpha = (0.952 \pm 0.007) \arctan(|z|/R)$.

4.2 Rotation curves and other data

The rotation curve data are from giant stars (Eilers et al. 2019), Classical Cepheids (Mróz et al. 2019), and the compilation by Chrobáková et al. (2020). They are all consistent with the Galactic constants we use, $R_\odot = 8.122$ kpc and $v_\odot \simeq 229$ km s^{-1}. We note that large scatters exist in the measured circular velocity between different works. Therefore, we compile the rotation curve by averaging V_c over bins of $\Delta R = 0.5$ kpc generally, and increase the bin size at large R to ensure sufficient S/N (see Fig. 1). The typical (mean) 1-σ error of the binned data is 12.1 km s^{-1}. The size of binning, based on our tests, does not impact our conclusions.

We also used the RAdial Velocity Experiment (RAVE) data (Binney et al. 2014) to perform the Jeans-equations tests (see below), and find a good consistence (within 1σ confidence) between the results based on the RAVE and *Gaia* data.

5 RESULTS AND DISCUSSION

5.1 Rotation-curve test

We compare the rotation curves predicted by models, $V_c(R) = \sqrt{R \, \partial\Phi/\partial R}$, to the observations. Fig. 1 shows the results. As expected, the Newtonian baryon-only model under-predicts $V_c(R)$ evidently, deviating from every binned data points by $\gtrsim 3\sigma$ generally. By adding a DM halo component, the fiducial MW model (see Section 3.1) appears to match the data well within the 1σ errors of almost all the bins. This is also the case for QUMOND. The MOG model appears broadly consistent with the data, albeit not as good as the fiducial DM model and QUMOND, and systematically smaller than most of the binned data points and the other two gravitational models.

In order to quantify how well the model predictions are consistent with the data, we calculate the reduced χ^2 with a degree of freedom d.o.f. = 27 regarding the 28 radial bins. Obviously, the Newtonian baryon-only model is rejected by the data with $\chi^2_{\nu,N} = 51.8 \gg 1$. The fiducial DM model agrees with the data with $\chi^2_{\nu,DM} = 0.5$; QUMOND is broadly consistent with the data with QesComm $\chi^2_{\nu,QUMOND} = 1.5$, and MOG is also acceptable with $\chi^2_{\nu,MOG} = 6.6 \sim \mathcal{O}(1)$, in contrast with the Newtonian baryon-only case. These χ^2 calculations are consistent with the above visual impression from Fig. 1.

As mentioned in Section 3.1, We have tested the four gravitational models with other prescriptions of the baryonic mass distributions, and with other RC data collected from the literature (see the Appendix), and found that all the tests give conclusions similar to the above.

5.2 Jeans-equations tests

We use Jeans-equations tests to examine how well the gravitational models agree with the data outside the Galactic plane. According to equations (3) & (4), we calculate T_R and T_z based on three-dimensional kinematic data. Then we compare them to the respective
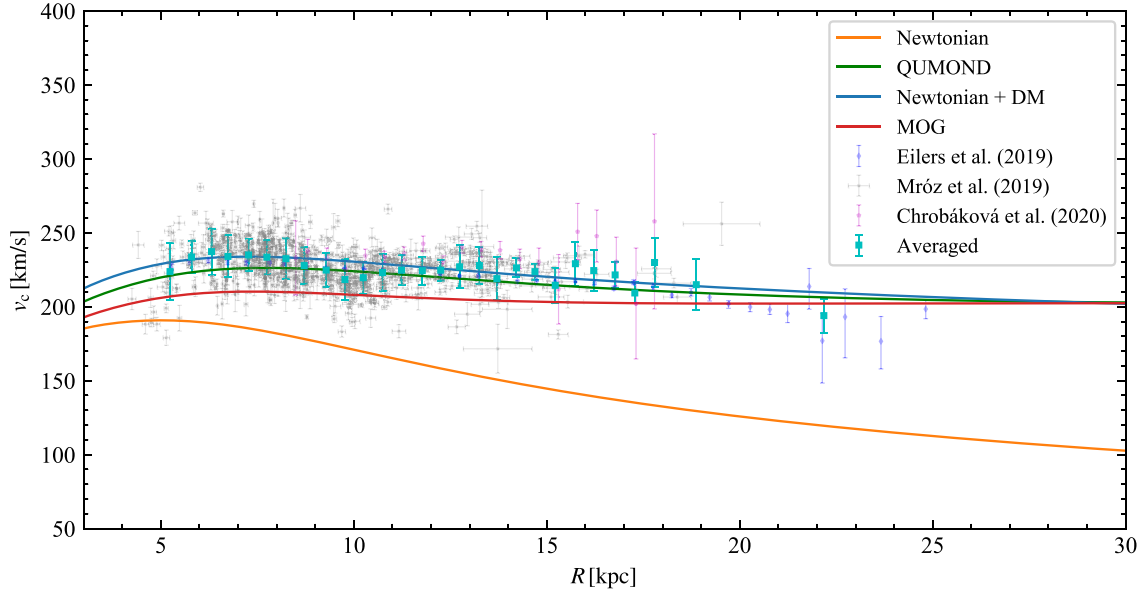
Figure 1. Rotation curves of the Milky Way, with the predicted ones of the gravitational models compared with the observations. The cyan data points with error bars ($\pm 1\sigma$) are our averaged rotation curve over spatial bins with $\Delta R = 0.5$ kpc (note that we increase the bin size at a few large-R bins), based on the data of Eilers et al. (2019), Mróz et al. (2019), and Chrobáková et al. (2020). The baryonic mass parameters is from Wang et al. (2022). The orange, blue, green, and red curves represent the Newtonian baryon-only, DM, QUMOND, and MOG models, respectively.

radial and vertical components of the potential gradients predicted by the gravitational models (namely, $\partial \Phi / \partial R$ and $\partial \Phi / \partial z$).

T_R and T_z, the observed radial and vertical accelerations, are derived from the tracer's density profile and kinematic data. Their uncertainties (1σ) are estimated in terms of standard error propagation, as follows:

$$\epsilon_{T_R} = \sqrt{\left(\frac{\partial T_R}{\partial \rho}\right)^2 \epsilon_\rho^2 + \sum_i \left(\frac{\partial T_R}{\partial X_i}\right)^2 \epsilon_{X_i}^2}, \quad (23)$$

and

$$\epsilon_{T_z} = \sqrt{\left(\frac{\partial T_z}{\partial \rho}\right)^2 \epsilon_\rho^2 + \sum_i \left(\frac{\partial T_z}{\partial X_i}\right)^2 \epsilon_{X_i}^2}. \quad (24)$$

Here, $\{X_i\}$ are the observed quantities, and $\{\epsilon_{X_i}\}$, their uncertainties. We also include a nominal uncertainty of 20 per cent for the tracer's density at each location (R, z).

As for the density profile of the tracer stars, we simply exploit the (weighted) whole Galactic disc (namely geometrical thin+thick discs; see their prescriptions in Section 3.1), but with appropriate proportion between the two disc components:

$$\rho(R, z) = 0.85 \times \rho_{d,thin}(R, z) + 0.15 \times \rho_{d,thick}(R, z). \quad (25)$$

The proportional factors (0.85 and 0.15) are the fractions in number of the two disc populations of the red clump stars according to their chemical classification (see Section 4.1). But we are not sure if, and how well, the red clump stars follow the spatial distribution of general stars (cf. Piffl et al. 2014); also not sure how well the chemically classified thin-disc clump stars are consistent with the dynamically best-fit thin disc of Wang et al. (2022). Thus in this study, we also use additional possible density profiles for the tracers, and the results are presented in next subsection.

In this subsection, we present the test results for the (R, z) locations in the way of illustrating T_R (or T_z) as a function of R, at different

altitudes (namely distance $|z|$) from the Galactic plane. This is because both the *Gaia*+LAMOST data and the RAVE data cover a limited range in $|z|$. The test results based on the RAVE data are consistent with those based on the *Gaia*+LAMOST data. Because the RAVE-derived T_R and T_z have large errors (see the Figures in the Appendix) and might mislead the reader's judgment, we do not plot the RAVE results in the figures of the main text, but plot them in the Appendix. Because the calculation of T_R requires σ_θ, the (R, z) space with sufficient data coverage for T_R test is $6 < R < 12$ kpc and $-2.5 < z < 2.5$ kpc (see Section 4.1).

Fig. 2 shows the results of T_R test. On the observational-data side, T_R monotonically decreases with R, which just reflects the trend of decreasing magnitude of radial acceleration along the radial direction. On the model side (the coloured lines in the figure), generally the radial gradients of the four gravitational models have considerably different magnitudes. The Newtonian baryon-only model is obviously far below the observed radial acceleration (T_R), for all (R, z) locations. Likewise, the MOG model is outside at least the 95 per cent confidence interval of the observed T_R, for all (R, z) locations. The fiducial DM model basically lies within the 95 per cent confidence interval of the data for all the R range at $|z| = 0.8$ kpc (middle panel) and $|z| = 1.2$ kpc (right-hand panel), except for the case of $|z| = 0.4$ kpc (left-hand panel), where the DM model goes outside the 95 per cent confidence interval for almost the entire R range. The QUMOND model behaves best: it lies within the 68 per cent (1σ) confidence interval for almost all (R, z) locations as displayed in the three panels.

According to Fig. 2, one may draw the conclusion that QUMOND fits the data best (within 68 per cent for almost all spatial locations); DM pass the T_R test basically, at least for all locations with $|z|$ greater than a certain height (we will see in Section 5.4 that in term of T_R test, the fiducial DM model is outside the 68 per cent confidence level for all locations at $z \lesssim 0.8$ kpc); Newtonian baryonic-only model and MOG obviously fail. Being conservative and for safety, yet we must
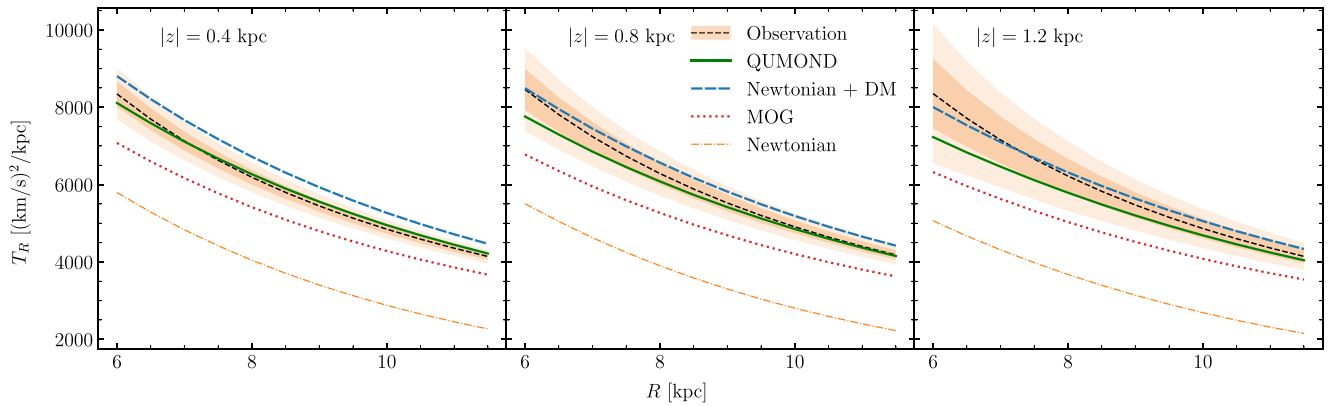
Figure 2. Radial Jeans-equation (T_R) tests of the gravitational models versus the data at various (R, z) locations, illustrated as a function of R at different altitudes ($|z|$). In every panel, the dashed black line represents the quantities calculated from the data of the entire *Gaia*+LAMOST sample of red clump stars (Huang et al. 2020); Dark and light shades show 68 and 95 per cent confidence intervals, respectively; The orange, blue, green, and red curves represent the Newtonian baryon-only, DM, QUMOND, and MOG models, respectively.
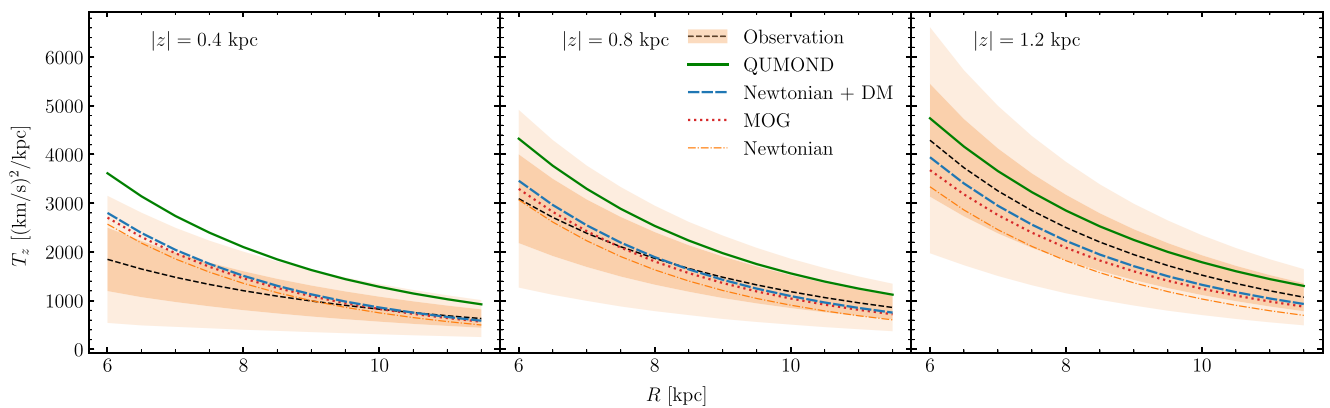
Figure 3. As Fig. 2, but showing the vertical Jeans-equation (T_z) test results.

note that the test depends on the tracer's density profile we adopt, and that at least DM and QUMOND cannot be discriminated for sure (see next subsection).

Independent of the T_R test, we now show the T_z test results in Fig. 3, which illustrate the distributions along R direction for the vertical gradients of the four potential models at different $|z|$ slices, with respect to the observed vertical accelerations (T_z). While the trend with R is similar to $T_R(R)$, the magnitude of T_z is in general much smaller than T_R of the same locations by at least a factor of 2. All the four models are broadly consistent with the observations within the 95 per cent confidence interval. While the Newtonian baryon-only, the fiducial DM, and MOG models lie close to each other, and are all within the 68 per cent confidence interval for almost all locations, yet QUMOND lies with the 68 per cent only at $|z| = 1.2$ kpc (right-hand panel). To be worse, for the locations at $R < 8.5$ kpc and $|z| = 0.4$ kpc, QUMOND is outside the 95 per cent confidence (left-hand panel); we will see in Section 5.4 that in term of T_z test QUMOND is outside the 68 per cent confidence level for almost all locations at $z \lesssim 0.8$ kpc probably, and within that confidence for all locations at $z \gtrsim 0.8$ kpc.

The discriminating power of T_z here is not so strong as T_R, as seen from the above test results. The theoretical reason is that, as mentioned above, in disc galaxies generally the vertical component of potential gradient ($\partial \Phi / \partial z$, namely the so-called 'vertical force' in

the literature) is much smaller than the radial gradient. Thus the differences of vertical field strength between those best-fit gravitational models are squeezed together compared with the differences in their radial strength (comparing Figs 2 and 3). The observational reason is that the relative errors (namely the ratios of the aforementioned ϵ_{X_i} to X_i) of σ_z is larger than that of σ_R by a factor of $\sim$2 (cf. figs 11–14 of Binney et al. 2014), which are the dominating error terms of the observed vertical and radial accelerations T_z and T_R, respectively. Thus, as displayed in Figs 2 and 3, the error bars of T_z (and importantly the relative errors) are much larger than those of T_R at the same spatial locations.

5.3 The common results from using different tracers' density profiles

As stated above, the major caveat of this work is that we lack the knowledge of the shape of the density profile of the tracer population (see $\rho(R, z)$ in Section 2), and have to represent it by using the profiles of general populations of the disc stars, such as the weighted thin+thick geometrical disc model of Section 3.1 as we adopt in the preceding subsection. Using different density profiles to represent the tracers' spatial distribution may give different values of T_R and T_z, and thus change the relationship between T_R and $\partial \Phi / \partial R$ (and between T_z and $\partial \Phi / \partial z$).

In this subsection, we assess the impact of the uncertainty in tracer's density profile to our Jeans-equations tests, with the following strategy. We believe that the real shape of the tracer's density profile should be embraced by the two main populations of disc stars, namely the geometrical thin and thick discs. Thus we also use the thin-disc and thick-disc profiles (prescribed in Section 3.1) to calculate T_R and T_z, and then safely base our conclusions about Jeans-equations tests on the common results shared by the schemes of using the three kinds of density profiles.

Compared with the above weighted thin+thick disc profile scheme, the thin-disc profile scheme results in larger values of both T_R and T_z (particularly for large-$|z|$ locations), whereas the thick-disc profile scheme leads to smaller values. Interestingly, while T_z changes dramatically in the two schemes (increased by factors of 1.8–3.1 in the thin-disc scheme, and decreased by a factor of 0.4–0.8 in the thick scheme), T_R changes mildly in the two schemes (increased by factors of 1.0–1.3 in the thin-disc scheme, and decreased by a factor of 0.9–1.0 in the thick scheme). That is, while the T_z test is sensitive to the tracer's density profile, the T_R test is relatively insensitive and thus robust.

The most important T_R and T_z result in common among the three profile schemes (excluding the thick-disc profile scheme for T_z test; see the next paragraph), in a sentence, is the following: both the fiducial DM model and MOND always lie in 95 per cent confidence intervals with respect to T_R and T_z for almost all locations with $|z|$ greater than a certain altitude (probably $\gtrsim 0.5$ kpc, see next subsection), while the MOG model lie farther away from the T_R data at many locations (let alone the baryon-only Newtonian model). In addition, there is a second notable point: On the side of gravitational models, the DM model is always larger than MOND in the radial field strength, yet always smaller than MOND in the vertical; what is more, relative to the observed accelerations at low-$|z|$ locations, the radial field strength of the DM model may even systematically larger than T_R (outside the 95 per cent confidence) while the vertical field strength of MOND may even systematically larger than T_z (outside the 95 per cent confidence), which will analysed in detail in next subsection.

The thick-disc profile scheme of the T_z tests yields that all the four gravitational models lie beyond the 95 per cent confidence intervals of the data for almost all spatial locations (see Fig. 5, middle panel). This fact indicates that the real density profile of the tracers, i.e. the red clump stars of Huang et al. (2020), is closer to the thin-disc profile than the thick-disc one. This inference is definitely correct because, as we recall, the sample of Huang et al. (2020) is dominated by thin-disc red clump stars (116 000 of 137 000; see Section 4.1). Thus, the T_z tests for the total red clump sample equipped with the thick-disc profile does not means that this scheme rules out all the four gravitational models, but means that T_z test is sensitive to tracer's density profile. This inspires us to consider the merit of this sensitive dependence in the end of this subsection.

We demonstrate the test results of the two additional schemes (thin-disc profile and thick-disc profile) in Fig. 4 (T_R tests) and Fig. 5 (T_z tests), together with the weighted thin+thick profile scheme as the reference. To present more new information, besides the $T_R(R)$ and $T_z(R)$ results of the additional two schemes for $z = 1.2$ kpc, we plot the results of the three schemes for higher altitudes ($z = 1.5$ and 2.0 kpc), where our data reach and the three density profiles for the tracers differ from each other significantly. From the figures we can easily see the above-stated features of the test results of the three schemes, particularly the most important result in common.

Besides as already mentioned in Section 4.1, we have tried to use only the thin-disc red clump stars chemically selected from the

Huang et al. (2020) sample to perform the T_R and T_z tests. In this trial, the number of the data points for T_R test (i.e. the spatial locations with valid σ_θ and V_θ) considerably decreases comparing with the above analysis, and thus the power of T_R test is impaired. The test results of the available data points, with the thin-disc density profile correspondingly, are consistent with those presented in the right-hand panels of Fig. 4. The number of the data points for T_z test decreases not so significantly, and thus we can perform all the tests, as did in Fig. 5. First, of course the T_z tests of the trial case sensitively reject the schemes adopting the density profiles of the weighted total disc and the thick disc (see left and middle panels of Fig. 6). Second, indeed, the trial tests equipped with the thin-disc density profile get somehow improved than the corresponding ones of the entire-sample case: MOND and the other three models (the three clustering closely in the T_z plots) all together lies within the 95 per cent confidence interval for almost all locations and even within the 68 per cent interval for a large fraction of the locations (please compare the respective right-hand panels of Figs 5 and 6). Anyway, no matter whether in terms of T_R or T_z tests, the conclusion remains the same as we conservatively state in the above (namely, the result in common).

Concerning the dependence of T_R on tracer's density profile, we have seen from the above analysis that the dependence is not negligible, at least for the commonly assumed density profiles in the literature (namely the prevailing prescriptions of the Galactic disc components). Thus we would like to caution that if one use R-directional Jeans equation to calculate certain quantities (e.g. the rotation curves on and off the Galactic plane, Chrobáková et al. 2020), the uncertainty caused by tracer's density profile has to be accounted for.

More importantly, concerning the sensitive dependence of T_z on tracer's density profile, actually there is a potential application. It is generally difficult to directly derive the density profile of the tracer population (e.g. red clump stars) with high completeness (cf. Piffl et al. 2014). Instead, if we can constrain the other quantities, i.e. gravitational potential and velocity dispersion, then we will be able to place tight constraints on the spatial distribution of a specific population of stars (i.e. tracers), by taking advantage of the sensitive T_z measure.

In practice, one can even use the two measures in turn as follows. First, the T_z measure is employed to pick up plausible models for the tracer's density profile (based on a grossly correct gravitational model). Then T_R is used to discriminate various gravitational models with subtle discrepancies. The two steps can be iterated to get both the best parameterized tracer's density profile and gravitational model.

5.4 Combining radial and vertical dynamics at low altitudes: vexing for both MOND and spherical DM haloes?

From the above two subsections, all the meaningful tests come to a convergent result that the Newtonian baryon-only model and MOG are rejected, and the fiducial DM model and MOND are consistent with the T_R and T_z data generally.[5] However, there appear systematical trends at low-$|z|$ locations discomforting for both DM and MOND. For a deeper investigation of the possible low-$|z|$ problem, in this subsection we plot the T_R and T_z tests as functions of $|z|$, at three radial positions $R = 7$, 8, and 9 kpc. Since the thin-disc stars in our sample are not capable to give T_R tests over large $|z|$

---

[5]Concerning MOG and MOND, certainly there is a caveat: they are tested by assuming the baryonic matter distribution is the prior one (Section 3.1) that was best-fitted in the DM paradigm.
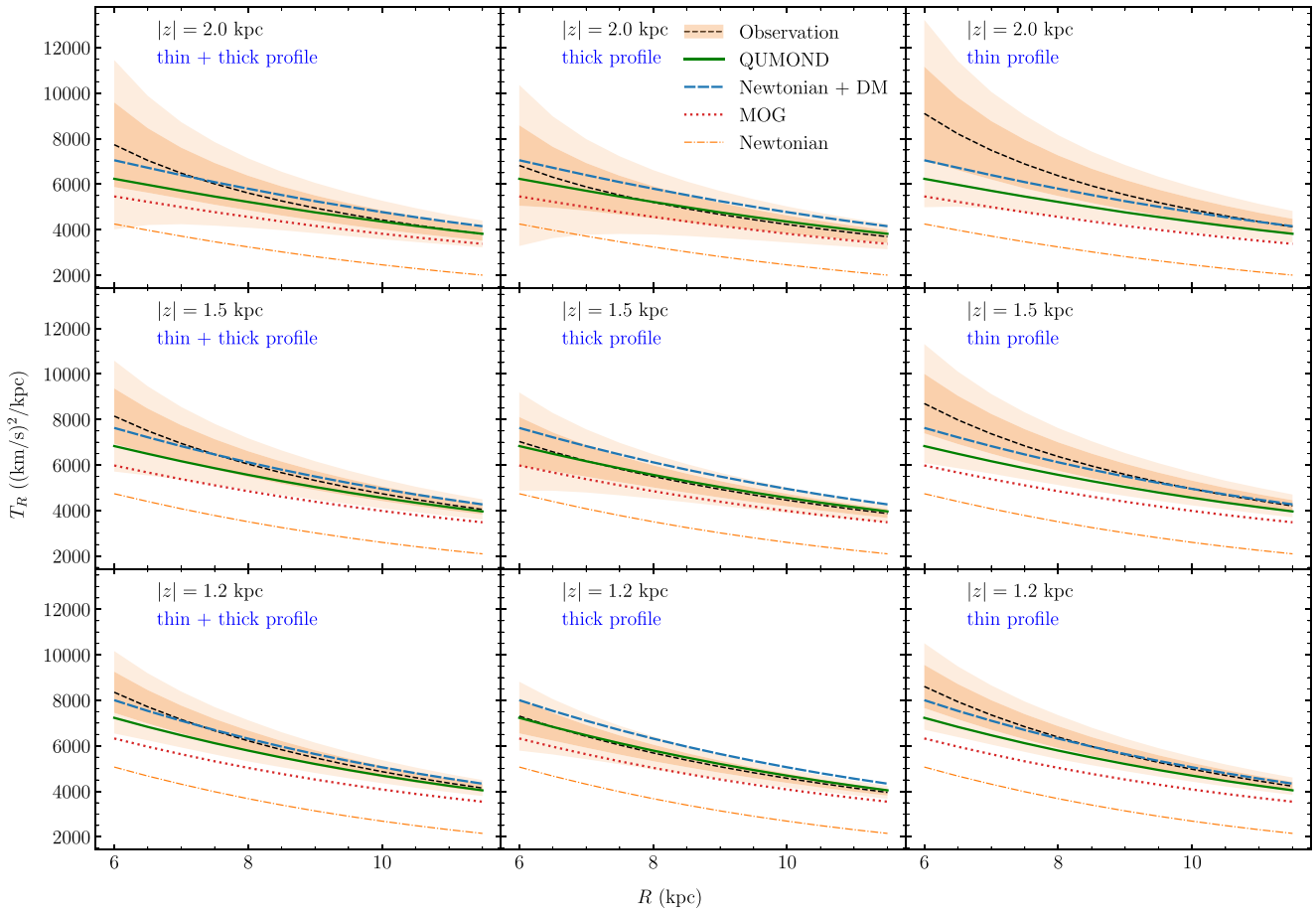
Figure 4. T_R test results assuming different tracer's density profiles based on the entire *Gaia*+LAMOST sample of red clump stars (Huang et al. 2020). Three schemes are shown: the total-disc profile (weighted thin+thick discs, left-hand panel), thick-disc profile (middle panel), and thin-disc profile (right-hand panel). Denotations are the same as in Fig. 2.

range, (see also Section 5.3), here we only exploit the test results based on the total sample equipped with the weighted thin+thick disc profile.

Regarding the resolution limits in the z direction to T_R, T_z and the corresponding radial and vertical components of field strengths of the four gravitational models ($\mathbf{g}_N$, $\mathbf{g}_{DM}$, $\mathbf{g}_{MOND}$, and $\mathbf{g}_{MOG}$), we estimate as follows. The spatial binning size in z is 50 pc for the kinematic data (see Section 4.1), then according to Nyquist's sampling theorem, the resolution limit to the kinematic quantities (e.g. σ_z) is twice. T_R and T_z involves the first derivative of those kinematic quantities with respect to z, so their spatial resolution limit requires at least two adjacent resolved units, i.e. four times the binning size namely 0.2 kpc. On the gravitational models' side, likewise, the spatial resolution limit to the above field strengths is four times the size of a grid cell, namely 0.24 kpc. Thus, in the figures we only plot the range from $|z| = 0.25$ kpc (the resolution) to 2 kpc that our data reliably cover.

In the T_R–z plots (Fig. 7) the radial field strength of the fiducial DM model lies outside the 95 per cent confidence interval at the locations with $|z| \lesssim 0.5$ kpc, and does not enter the 68 per cent confidence until $|z| \gtrsim 0.8$ kpc; this trend of inconsistency with the T_R data gets somehow worse with R moving outwards. On the contrary, the radial field strength of MOND always lies in the 68 per cent confidence interval at every location.

In the T_z–z plots (Fig. 8) the vertical field strength of the fiducial DM model always lies in the 68 per cent confidence interval of every locations. On the contrary, the vertical field strength of MOND lies outside the 95 per cent confidence interval at the locations with $|z| \lesssim 0.5$ kpc, and does not enter the 68 per cent confidence until $|z| \gtrsim 0.8$ kpc; this trend of inconsistency with the T_z data gets somehow alleviated with R moving outwards.

In summary, when $|z| \gtrsim 0.8$ kpc, both the fiducial DM model and MOND lies within the 68 per cent confidence of T_R and T_z for all locations. But, at low altitudes (say $|z| \lesssim 0.5$ kpc), there may be problematic: DM with respect to T_R, and MOND with respect to T_z. The exact $|z|$ values have something to do with the tracer population, which is subtle to handle as we demonstrated in Section 5.3; we defer this issue to future work. There is a possibility that the real Galactic gravitational potential, particularly its inner part, is in between the fiducial DM model with a spherical DM halo and the MOND, i.e. in the DM language, the halo may be oblate (cf. Figs 9 and 10 in next subsection).

Note that in Figs 7 and 8, there are qualitative differences in the shape as a function of $|z|$ between the kinematic accelerations (namely $T_R(z)$ and $T_z(z)$), and the field strengths of the four models ($\mathbf{g}_R(z)$ and $\mathbf{g}_z(z)$). The $T_R(z)$ (or $T_z(z)$) shapes, in the range shown in the figures, are convex, while the shapes of the four $\mathbf{g}_R(z)$ (or $\mathbf{g}_z(z)$) lines look similar and are not so curved. The reason is that
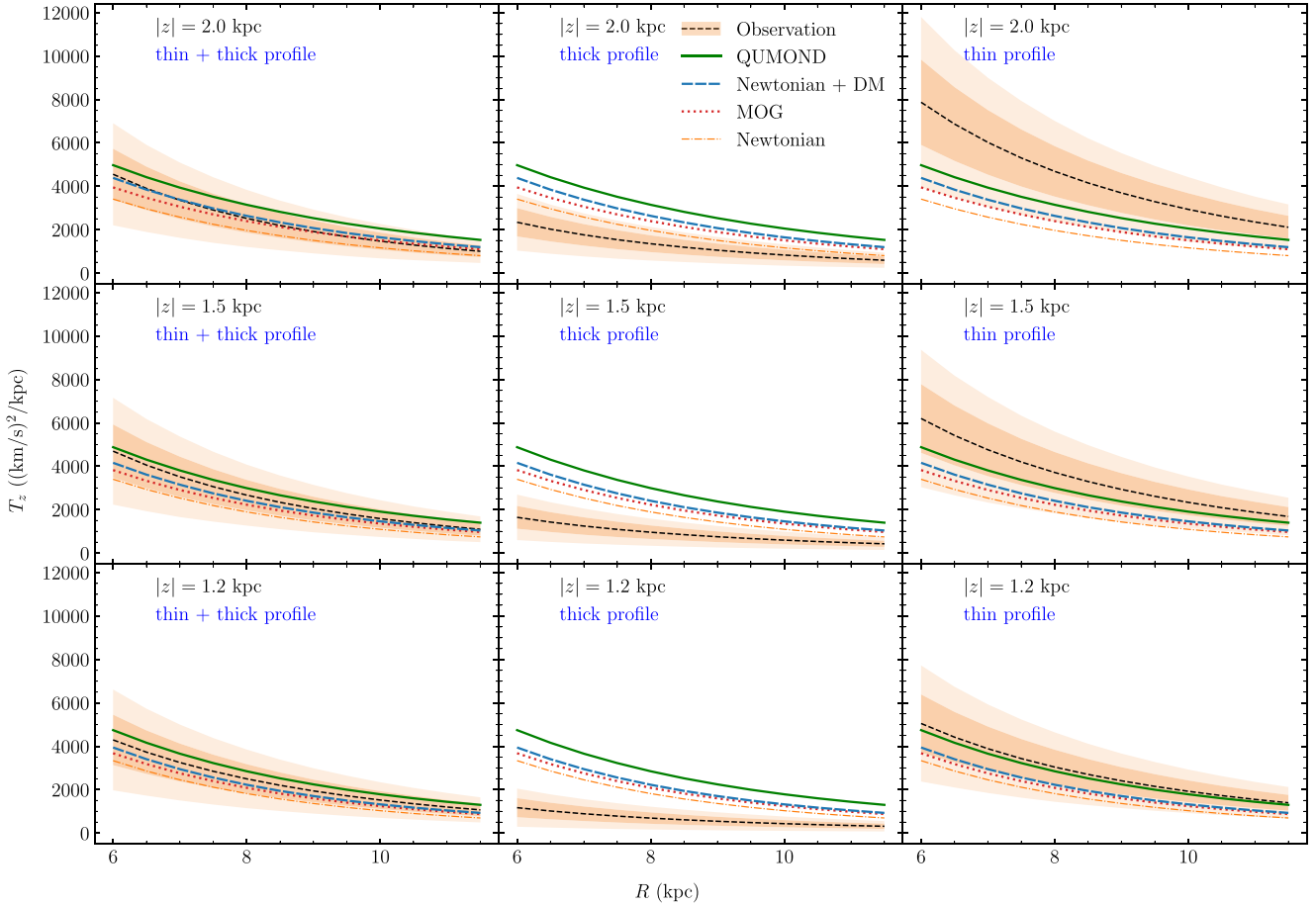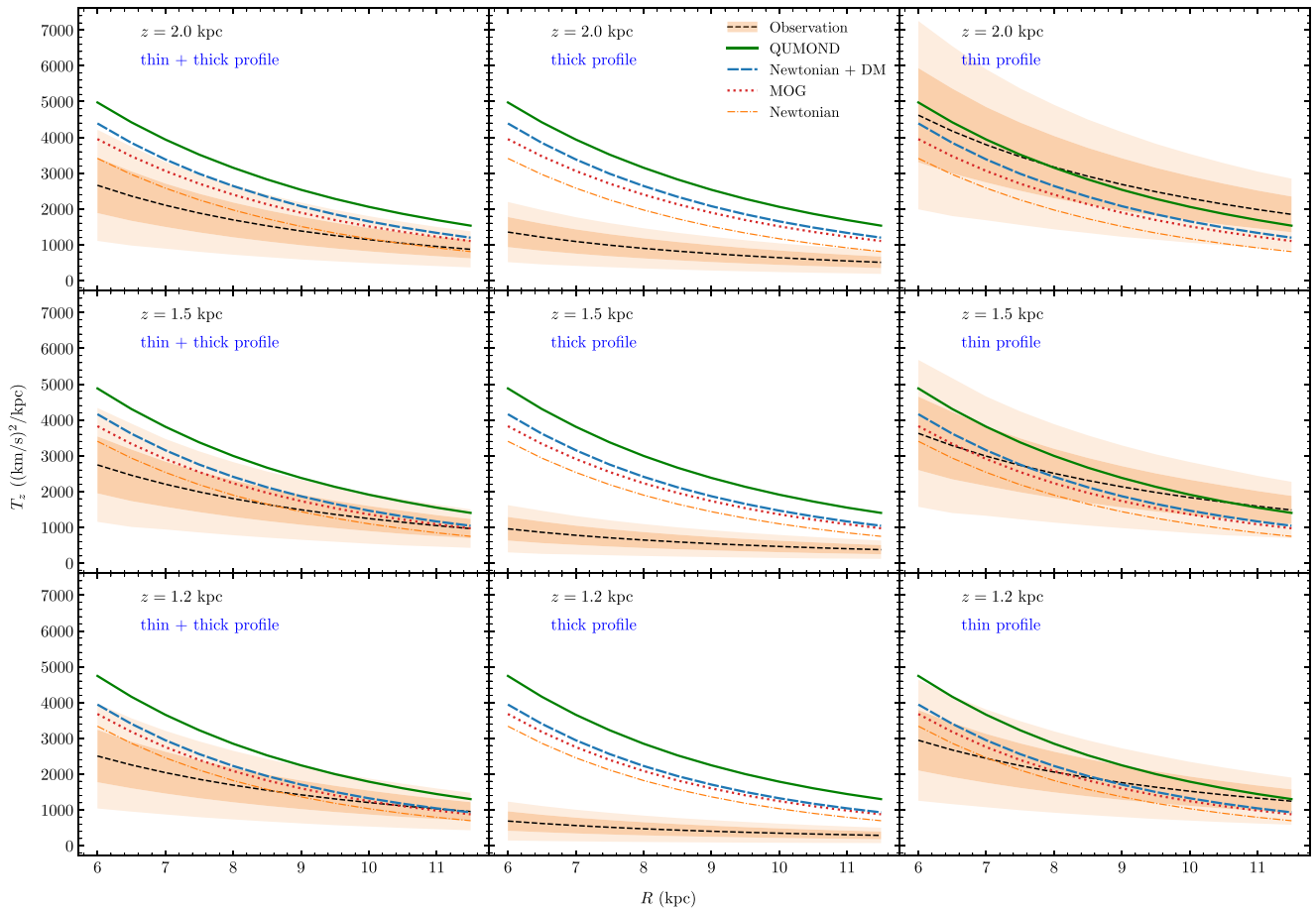
Figure 5. As Fig. 4, but for T_z test results.

the functions underlying the kinematic and dynamical quantities are different. The dynamical $\mathbf{g}(z)$ lines are basically determined by either the DM halo function (in the case of the DM model) or the baryonic matter distribution (the other three models), and both the DM halo and baryonic distribution functions decay monotonically farther out (see Section 3.1). The kinematic $T_R(z)$ and $T_z(z)$, on the other hand, are determined by the functions, $\rho(R, z)$ (the tracer's density distribution we adopt) and equation (22) (spatial-distribution function of velocity quantities), and their derivatives, the T_R and T_z shapes with respect to $|z|$ are thus complex. We can imagine, both $T_R(z)$ and $T_z(z)$ would increase rapidly with $|z|$ when $|z|$ larger than a certain value because of the exponential decay of $\rho(z)$; this just means that the assumed tracer's density profile, likely as well as the extrapolation of the spatial-distribution function of velocity quantities, breaks down in that $|z|$ range. It is right because of the above reason that in Figs 7 and 8, we only plot the range $|z| \leq 2$ kpc (see Section 4.1), and compare the models (**g**) with the data in terms of confidence intervals only.

In the literature, it is being hotly debated as to the shape of the Galactic DM halo is oblate, spherical, or prolate, with observational evidence both for and against an oblate shape of the inner Galactic gravitational potential (see Hattori, Valluri & Vasiliev 2021 and the references therein). In our above analysis of the possible small-altitude problem, as the exact $|z|$ range and the degree of DM and MOND deviating from the data depend somehow on the tracer population and its density profile we use, thus at this point we leave this problem open.

In the history of MOND research, it is a vexing issue about MOND's possible over-prediction of vertical acceleration; see section 3.1.2 of Banik & Zhao (2022) for a detailed account. We have noticed that Lisanti et al. (2019) came to the strong conclusion that gravitational models of MOND type failed to simultaneously explain both the rotational velocity and vertical motion of stars in the solar neighborhood. In our opinion, there are technical reasons that explain the tension between their conclusion and our not-so-discriminating one. There are several problems in their data and modelling method. The most serious is the key data set they used: the observed number density ($n(z)$) and vertical velocity dispersion ($\sigma_z(z)$) of three mono-abundance stellar populations at $R = R_\odot$. The same data set has been thoroughly analysed by Büdenbender et al. (2015), which turned out that the DM densities estimated by the different stellar populations are inconsistent with each other (see particularly their Fig. 3 and Section 3), owing to a major reason that the data set did not measure the cross-dispersion component σ_{Rz} of the velocity ellipsoid. Hessman (2015) also analysed that data set, and achieved the same diagnostic as Büdenbender et al. (2015), along with his other caveats on vertical Jeans-equation modelling; in fact, as stated in the Introduction, importance of the cross term σ_{Rz} has been well proved in past decade. By the way, the rotation-curve information Lisanti et al. (2019) used was limited to a single location, the Solar radius (cf. McGaugh 2016). Concerning their modelling method linking $n(z)$ and $\sigma_z(z)$, which is the completely one-dimensional Jeans modelling (namely the simplest K_z method),

Figure 6. As Fig. 5, but for T_z test results using only the data of the thin-disc red clump stars chemically selected from the Huang et al. (2020) sample.

now it is clear that neither the 'tilt term' in vertical Jeans equation nor the 'rotation-curve term' in Poisson equation can be neglected (see the sixth paragraph of the Introduction and the references therein).

McGaugh (2016) performed K_z analysis with the 'rotation-curve term' considered and yet without accounting for the 'tilt term' involving σ_{Rz} (see his equation 12), based on the K_z (vertical force) data measured by Bovy & Rix (2013). The K_z data were derived by complicated action-based distribution function modelling, with one assumption being that both σ_R and σ_z are not dependent on z, i.e. their vertical profiles being constant (see Section 3.1 of Bovy & Rix 2013). A caution mentioned in passing: the 'rotation-curve term' in Poisson equation (jargon used in this paper; also Read 2014) was called 'tilt term' in section 4.7 of McGaugh (2016). Hessman (2015) also critically analysed the $K_z(z)$ problem of the Bovy & Rix data set, along with his comments on the 'accuracy versus precision' issue of the advanced yet complicated (and thus over-simplified practically) method of distribution function modelling (particularly cf. his Section 3). Recently Binney & Vasiliev (2022) described in detail the problems of the (unrealistic) quasi-isothermal distribution function model adopted in Bovy & Rix (2013) for Galactic-disc populations.

Besides the above-inspected studies based on the Galactic data, there are studies based on stellar velocity dispersion ($\sigma_\star$) and other properties of galactic-disc stars of external galaxies (listed in section 3.1.2 of Banik & Zhao (2022); see also the Introduction). Just like the *status quo* of those Galactic studies, the external-

galaxies ones are also inconclusive; one reason lies in the difficulty of measuring both $\sigma_\star$ and requisite other properties (e.g. scale height, or stellar mass-to-light ratio or alike) consistently from the same stellar population (see e.g. Milgrom 2015; Angus et al. 2016; Aniyan et al. 2021).

5.5 Exploring the 'extra mass/gravity'

Echoing the early names of the DM problem, such as missing, hidden, excess or extra mass, and excess or extra gravity with interest, we explore the extra mass or extra gravity in excess of the Newtonian baryonic one for the DM, QUMOND, and MOG models.

We first explore the differences in gravitational potential predicted by the three models (denoted as $\Phi_{\rm model}$) compared with the Newtonian baryon-only case ($\Phi_{\rm N}$), namely $\Delta\Phi = \Phi_{\rm N} - \Phi_{\rm model}$, also we explore the corresponding gradients of the potential difference, namely the vector difference in field strength, $\mathbf{g}_{\rm model} - \mathbf{g}_{\rm N} \equiv \nabla(\Phi_{\rm N} - \Phi_{\rm model})$. Hereafter we call them extra potential and extra gravity, respectively; yet by definition the two are interchangeable essentially. Fig. 9 plots the distributions of the three $\Delta\Phi$ and the corresponding extra gravity in the meridian plane. The extra potential of the fiducial model (namely the DM halo; see the middle panel) is spherically symmetric as prescribed by the Zhao's profile. QUMOND (left-hand panel) gives a comparable extra potential in magnitude to the DM case, but the shape of the extra potential is fairly flatten in the z-direction (i.e. an *oblate* gravitational potential). MOG yields a slightly oblate extra potential (right-hand panel); this
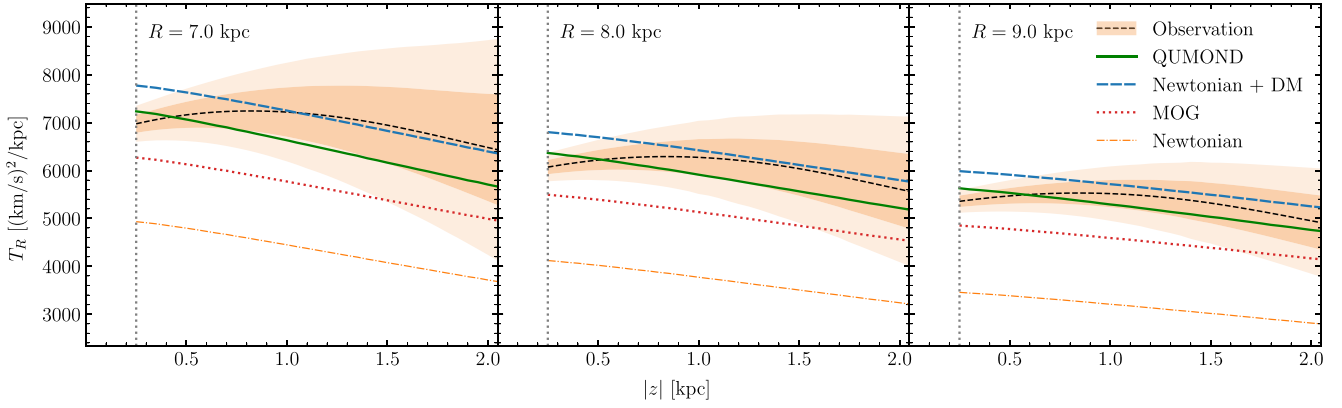
Figure 7. Radial Jeans-equation (T_R) tests of the gravitational models versus the data at various (R, z) locations, illustrated as a function of z at three radial positions. In every panel, the dashed black line represents the quantities calculated from the data of the entire *Gaia*+LAMOST sample of red clump stars (Huang et al. 2020); Dark and light shades show 68 and 95 per cent confidence intervals, respectively; The orange, blue, green, and red curves represent the Newtonian baryon-only, DM, QUMOND, and MOG models, respectively. The vertical grey dotted line denotes the spatial resolution limit in the z direction to the field strengths and observed acceleration (T_R).
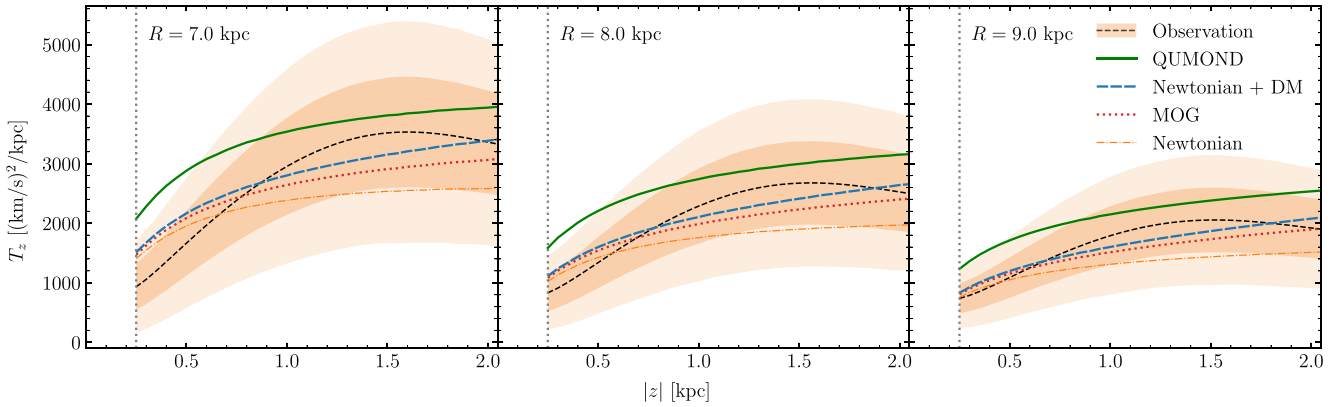
Figure 8. As Fig. 7, but showing the vertical Jeans-equation (T_z) test results.

is clearer in Fig. 10, which can be interpreted as the divergence of the 'extra gravity' field. In addition, the magnitude of the MOG extra potential is $\approx$1.5 times of the QUMOND or DM one on average. The magnitude of the extra gravity in MOG is instead fairly smaller than the other two gravitational models (see the bottom row), which is in fact consistent with the systematic smallness of MOG in the rotation-curve test (i.e. the gravitational acceleration at $z = 0$). From Sections 5.2 and 5.3, we have seen that our Jeans-equations tests (mainly the T_R) disfavour the MOG model, yet presently cannot judge for sure which one of the fiducial DM model (namely spherical halo) and MOND, or some one in between, matches the data to a better degree.

Next, we translate the 'extra potential' (the above $\Delta\Phi$) into the effective 'extra mass' in the Newtonian sense, simply using normal Poisson equation. In the case of the DM model, this translation is physical and exact, and the extra mass is just the DM halo. We must caution, however, that such a translation is merely mathematical for any modified-gravity models, and the concept of 'extra mass' is even misleading (for the case of MOG; see below).

In the case of QUMOND, interestingly, this translation is meaningful (albeit without any physical content), and the 'extra mass' is the very concept of 'phantom dark matter' described in Section 3.2.

This is because the QUMOND formulation has a great merit that its gravitational potential can be naturally decomposed, and ascribed in the Newtonian sense to two matter components: the baryonic matter (the real) and the effective DM (the phantom). The effective PDM density distribution on the $R - z$ plane is shown in Fig. 10 (left-hand panel). Compared with the density distribution of the DM halo of the fiducial mass model (the middle panel), the QUMOND PDM is morphologically closer to a traditional (quasi-)spherical DM halo plus a disc-shaped component, which is consistent with that presented in, e.g. Wu et al. (2008).

In the case of MOG, just as generic modified-gravity theories (e.g. the AQUAL realization of MOND proposed by Bekenstein & Milgrom 1984), such an 'extra mass' translation is merely effective; i.e. the extra mass distribution (plus the baryonic one) is used in the DM paradigm to mimic the MOG gravitational potential. We plot the MOG's 'extra mass' in the right-hand panel of Fig. 10, just for an intellectual curiosity. We stress again that our tests are based on MOG's gravitational potential, not on the density distribution of the 'extra mass' described in this subsection. Certainly, it is correct and useful to view the 'extra mass' (Fig. 10) as the divergence of the 'extra gravity' plotted in the bottom row of Fig. 9.
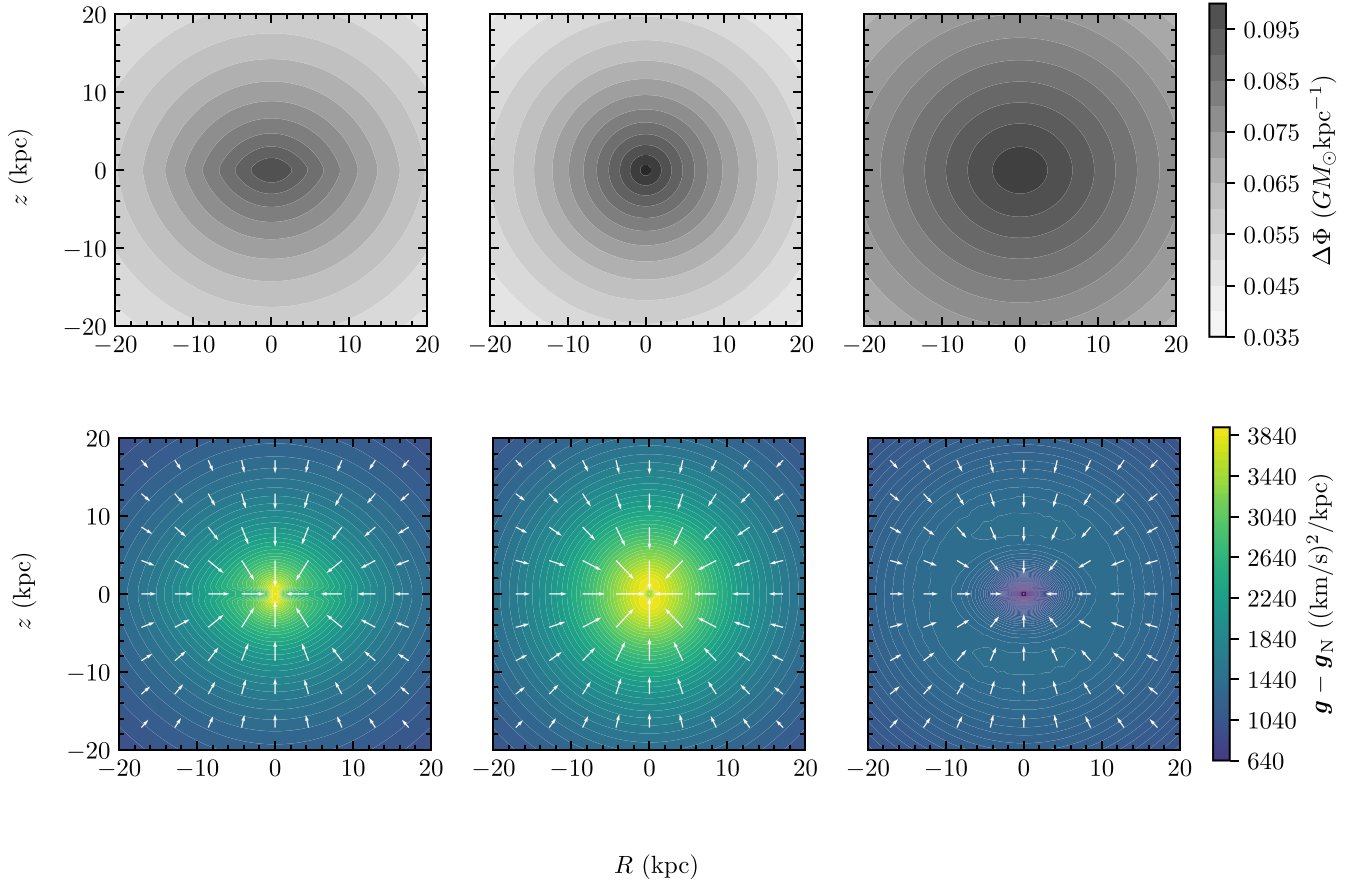
Figure 9. Top row: Gravitational potential difference between the Newtonian baryon-only model and other models ($\Delta\Phi = \Phi_{\mathrm{N}} - \Phi_{\mathrm{model}}$). The left-hand, middle, and right-hand panels are for the QUMOND, DM, and MOG cases, respectively. Generally the MOG model is not favoured by our Jeans-equations tests. At present, it is yet an open question to what degrees DM and MOND, respectively, represent the real gravitational field of the MW. **Bottom row:** The corresponding field-strength difference between the Newtonian baryon-only model and other models, $\mathbf{g}_{\mathrm{model}} - \mathbf{g}_{\mathrm{N}}$. The direction of the vectors is denoted by arrows, and their magnitude is colour coded. Note that around the centre, the darker the MOG's extra gravity is (in blue and even purple) the weaker is the field strength.
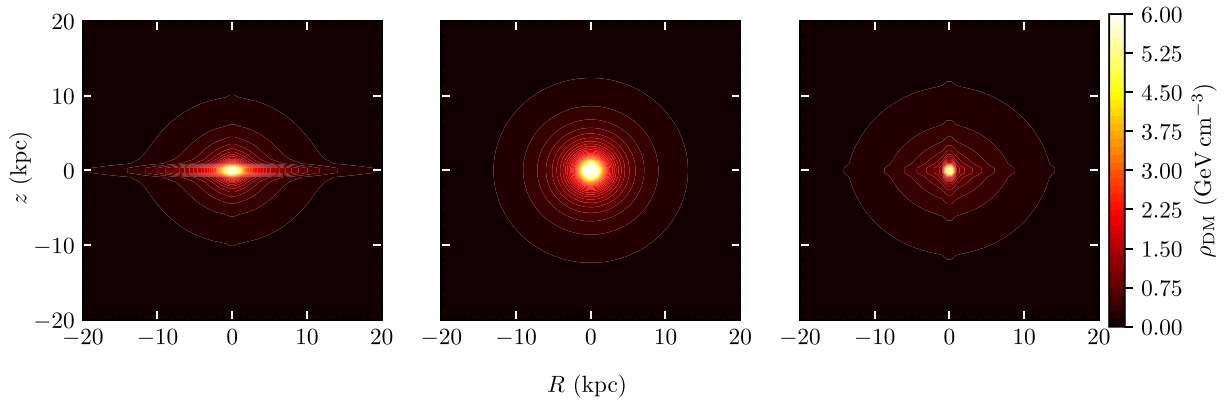
Figure 10. The 'extra mass' distribution translated directly from the 'extra potential' (Fig. 9) in terms of the normal Poisson equation. **Left:** The density of the effective DM (namely 'phantom dark matter') predicted by QUMOND. **Middle:** The density of the DM halo in the fiducial mass model. **Right:** The 'extra-mass' density of the MOG case. Caution that the MOG's 'extra mass' is just a mimic in the DM paradigm, and basically useless if not viewed as merely the divergence of the 'extra gravity' field but interpreted as 'mass' (see the text for the detail).

5.6 On the effective equivalence between MOND and DM

After decades of search, DM particles have not been found still (Feng 2010). Particularly, from the observational standpoint, the tight correlations between DM and baryonic matter cannot be explained satisfactorily within the DM framework (Bullock & Boylan-Kolchin 2017). On the other hand, MOND (or its generally covariant descendants), taken in its present form, has not been proved to be a mature fundamental theory. It seems that we still have a long way to go discovering the nature of the 'dark matter problem'. Just in the above context, we are excited by the present study, tightening the effective equivalence between MOND and DM on circumgalactic and galactic scales, or called 'CDM–MOND degeneracy' (Banik & Zhao 2022); to be precise, it is the effective equivalence between the PDM of MOND and (possibly oblate) DM haloes, in the sense of acting as gravitational-potential models. A possibility that the effective equivalence is hinting[6]: a new synthesis may arise, reconciling, and transcending both MOND and DM paradigms. The thinking behind is as follows. First of all, all the observed correlations between 'DM' and baryonic matter can be explained easily and elegantly by the simple Milgrom (1983) law (namely the essence of MOND), basically without any a free parameter. This surprising fact suggests the delicate mechanism of the interaction between baryons and 'DM' (particles, or fields, or effective ones) for the future theory, either in the form of a new gravity (say, an effect of quantum gravity, or even a new dynamics/law of nature?), or in the form of a new ingredient within the established quantum field theory, or in a third way. Furthermore, if we take a broader vision, which sees dark energy and DM as two facets of a single origin as some researchers have pursued (e.g. Zhao & Li 2010), then the effective equivalence would point to quantum vacuum, as Milgrom's critical acceleration constant suggests (by $2\pi a_0 \approx c H_0 \approx c^2 \sqrt{\Lambda/3}$). Finally, we would like to remark that, if there is any minimum value in the above vision, MOND might be better interpreted as an effect of modified inertia (e.g. Milgrom 1999), and even hints at nonlocality (nonlocal inertia of Milgrom 1999, albeit being non-quantumlike for now), and reminds us of the role of quantum vacuum as 'fluid of virtual particles'. Although being exciting, this kind of thinking is speculative so far, and here we refrain from brain-storming farther.[7]

---

[6]We must admit that the effectiveness of the equivalence between MOND and DM as gravitational-potential models is only within the best observational constraints available so far, and further tightened by comparing with the MOG case (see Section 5.2), i.e. effective to some degree only. Of course, their equivalence is not absolute: as illustrated by Figs 9 and 10, the two are different *per se*. Besides, plausibly they both deserve to be transcended, as discussed in this subsection.

[7] We want to add a final remark: In all covariant modified-gravity theories so far, which are different from the modified-inertia interpretation as Milgrom stressed, one or more additional fields are required; those fields have energy and thus are additional sources of gravity, but their stress-energy distribution does not follow that of normal matter (although with other kinds of delicate coupling mechanisms between the additional fields and normal matter; see e.g. Hossenfelder 2017). Thus, virtually it is interchangeable to call them *additional fields*, *modification of gravity*, *additional (non-normal) stuff* or directly *non-baryonic DM*; this is in fact one broader theoretical background inspiring us to think about the effective equivalence between MOND and DM on galactic scales. That is, in the direction of modified-gravity interpretation of MOND, the theoretical developments also point to, and have already suggested, a transcending synthesis of the two paradigms (particularly cf. Section V of Hossenfelder 2017). After all, from a modern viewpoint of quantum field theory, the two paradigms can be conceptually viewed as effective theories for 'collective excitations' of quantum vacuum (Wen 2003).

On the other hand, thinking practically, we can exploit the effective equivalence. As demonstrated in the present study for any practical purposes when researchers want to study the kinematics on galactic scales, they can safely use the QUMOND formula (i.e. the gravitational field of the 'phantom dark matter') as an alternative of DM halo models. This approach will save the researchers from handling various prerequisites and fine tuning the cumbersome parameters of DM haloes.

6 SUMMARY

In terms of the complete form of Jeans equations that admit three integrals of motion, we perform tests on gravitational models for the Milky Way, based on the latest three-dimensional (i.e. R-, z-, and θ-directional) kinematic data over a large range of (R, z) locations. Our primary aim is to discriminate between MOND and DM halo models with MOG (as well as Newtonian baryon-only model) as comparison. The kinematic data we use here are mainly based on the sample of red clump stars compiled by Huang et al. (2020), which are powered by the *Gaia* DR2 astrometry.

In the *Gaia* era (from the DR2 onward), previous long-standing problems concerning observational data (e.g. systematic bias in distance estimation) are gone. The major factors that affect dynamical modelling of the Milky Way now are of astrophysical origin (the complexity of real galaxies), e.g. kinematic substructures, still rich discrepancies inside a certain tracer population, and so on (see Sections 4.1 and 5.3). As far as the data we use are concerned, the typical 1-σ error in the rotation–curve data to fit is 12 km s^{-1}, and in the velocity-dispersion data fitting the spatial-distribution formulae (see below), 0.5 km s^{-1}.

Regarding the stellar kinematics that we derive based on the data of Huang et al. (2020), aside from the analytic form proposed by Binney et al. (2014) for the spatial distributions of σ_R and σ_z, we find that the spatial distributions of σ_θ and V_θ also can be well fitted by the same functional expression, namely in the form of $\sigma_\theta(R, z)$ and $V_\theta(R, z)$. We fit the function to the four sets of data, respectively, and obtain best-fitting parameters for the spatial distributions of the four kinematic quantities (see Table 2). We then use the kinematic data calculated in terms of the formulae to perform the T_R and T_z tests on every spatial locations. The advantage is at least two-fold: (1) free of the numerical artifacts caused by numerical differentiation given the limited spatial resolutions of the observational data, and more importantly (2) reducing the impact of various kinematic substructures in the Galactic disc.

The main results of our comprehensive tests (Sections 5.1, 5.2, 5.3, and 5.4) are summarized as follows:

(i) The Newtonian baryon-only model, as expected, is rejected not only by the rotation-curve test (namely dynamics in the Galactic-disc plane), but also by the R-directional Jeans-equation test (T_R) for all spatial (R, z) locations.

(ii) Concerning the three models with 'extra mass or gravity' (fiducial DM model with a spherical halo, MOND and MOG), rotation-curve data alone (with $z = 0$) cannot reject any one of them for sure (see Figs 1 and A1).

(iii) The most important result in common among the Jeans-equation tests with meaningful tracers' density-profile schemes is the following: both the fiducial DM model and MOND always lie in 95 per cent confidence intervals in terms of both T_R and T_z (the observed radial kinematic accelerations) for almost all locations with $|z|$ greater than a certain altitude ($|z| \gtrsim 0.5$ kpc probably), while the MOG model lie farther away from the T_R data at many locations

(assuming the prior baryonic matter distribution best-fitted in the DM paradigm). In particular, both DM and MOND models are equally consistent with the T_R and T_z data within 68 per cent confidence of every locations at $|z| \gtrsim 0.8$ kpc.

(iv) At low-$|z|$ locations, there may be problematic trends for MOND and the fiducial DM model with a spherical halo, respectively: the radial field strength of the DM model seems systematically larger than T_R while the vertical field strength of MOND seems systematically larger than T_z. To be specific, in the Jeans tests based on the entire red-clump star sample equipped with the weighted total disc density profile, at locations with $0.5 \lesssim |z| \lesssim 0.8$ kpc, DM is in the 68–95 per cent confidence while MOND within 68 per cent in terms of T_R, and MOND is in the 68–95 per cent confidence while DM within 68 per cent in terms of T_z; at $|z| \lesssim 0.5$ kpc, DM is outside the 95 per cent T_R confidence of every locations, and MOND is outside the 95 per cent T_z confidence of every locations. The exact $|z|$ range and the degree of DM and MOND deviating from the data depend somehow on the tracer population and its density profile, and thus are uncertain at this point. There is a possibility that the real Galactic gravitational potential, particularly its inner part, is in between the fiducial DM model with a spherical DM halo and MOND; that is, in the DM language, the inner halo may be oblate.

First of all, the above test results consistently point to an observational conclusion: Even in the condition of current kinematic data with the precision and accuracy powered by *Gaia* DR2 (and the measurement uncertainties are no longer the major concern from now on), which is able to reject the MOG model (let alone the Newtonian baryon-only model; and see the caveat in Footnote 5), the MOND model is still not rejected, and behaves as good as the fiducial DM model through Jeans-equations tests on all spatial locations over $5 < R < 12$ kpc and $-2.5 < z < 2.5$ kpc (namely the (R, z) space with sufficient data coverage). This is surprising, because (1) there is no free parameter at all in the QUMOND model, i.e. without any fitting (let alone pre-fitting), and (2) the parameters of the baryonic mass model are actually fine-tuned in the DM context; on the contrary, the fiducial DM model we adopt was fitted already with all available Galactic kinematic data (even the same as part of the rotation-curve data set we use), and has been kept improving elaborately for decades. Secondly, both the fiducial DM model with a spherical halo and MOND may have their respective vexing facet at low-attitude location (see the forth item above), which awaits further investigations.

The physical implication of the above test results, what excites us the most, is the concept that we are tempted to put forward in this paper: the effective equivalence of DM and MOND on circumgalactic and galactic scales (see Section 5.6 and Footnote 6). There may be a value in this concept (as this kind of equivalence is effective and hints at both paradigms being effective): A new synthesis may arise, reconciling, and transcending both MOND and DM. On the other hand, from a pragmatic standpoint (the two being equivalent or degenerate gravitational-potential models for now), we can exploit the effective equivalence in this way: when researchers want to study the kinematics on galactic scales, they can use the QUMOND formula (i.e. the gravitational field of the 'phantom dark matter') as an alternative of DM halo models. This is safe at least on the precision and accuracy level of kinematic data derived from *Gaia* DR2. This approach will save the researchers from handling various prerequisites and fine tuning the cumbersome parameters of DM haloes.

Besides the above astrophysical outputs, the present work discovers the instrumental advantages of the two measures, T_R and T_z. The two measures defined kinematically in terms of the complete form of Jeans equations (in axisymmetry at this point; Section 2) represent the observed radial and vertical accelerations (fed with kinematic data). They both exploit three-dimensional kinematics, admit three-integral dynamics, and respect three-dimensional Poisson equation. Thus, first of all, as stated in the Introduction (also in Section 2), they surpass previous commonly used methods, such as rotation-curve test that is essentially a one-dimensional method (namely concerning the R-directional dynamics only), the simple K_z method that is completely one-dimensional also (simplifying a galaxy as z-directional slabs), and most Jeans-equation applications in the literature (assuming dynamics of two integrals of motion only). More importantly, out of the present work (Section 5.3), we find that T_R test is fairly insensitive to the choice of tracer's density profile, and thus is robust in discriminating gravitational models, while the merit of T_z test is instead its sensitivity to tracer's density profile.

Looking forward to the near future, we expect to use more (as well as better) kinematic data from, e.g. *Gaia* DR3 astrometry and ongoing large-scale spectroscopic surveys; importantly, to perform more realistic treatments in galactic modelling (e.g. handling substructures and refining tracer populations), and update the Jeans-equations tests of the present study. Immediately, we would like to make full use of the two measures, T_R and T_z, in the iterative way as described in Section 5.3: first, employ T_z to constrain the model parameters of tracer's density profile (based on a grossly correct gravitational model); second, employ T_R to discriminate gravitational models with subtle discrepancies; then the two steps are iterated to consistently obtain the best realistic tracer's density profile and best gravitational model. By doing so, we hope to achieve the final goal: what a gravitational potential can represent the real Milky Way.

DATA AVAILABILITY

The raw data underlying this study are available from the *Gaia* archive at https://gea.esac.esa.int/archive/, and the LAMOST data releases at http://www.lamost.org/public/?locale = en . The distance and 3D velocity data, as well as other information, of the red

clump stars of Huang et al. (2020) we use in this work are publicly available by following the link provided by Y. Huang: https://zenodo.org/record/3875974 . All the data-analysis code of this work, as well as all data and results at intermediate levels, are publicly available at https://github.com/ydzhuastro/JeansTest-MW .

A NOTE IN PROOF

While this article was in press, we were made aware of Milgrom (2022) which shows that it is possible to construct modified-inertia models of MOND where vertical accelerations are less enhanced than in modified-gravity models of MOND. Such modified-inertia models may occupy the middle ground between spherical DM halos and QUMOND; i.e., may fit the real, probably oblate gravitational potential of the Milky Way suggested in subsection 5.4 of this paper.

REFERENCES

Angus G. W., Diaferio A., Famaey B., van der Heyden K. J., 2013, MNRAS, 436, 202
Angus G. W., Gentile G., Famaey B., 2016, A&A, 585, A17
Angus G. W., Gentile G., Swaters R., Famaey B., Diaferio A., McGaugh S. S., van der Heyden K. J., 2015, MNRAS, 451, 3551
Aniyan S., Ponomareva A. A., Freeman K. C., Arnaboldi M., Gerhard O. E., Coccato L., Kuijken K., Merrifield M., 2021, MNRAS, 500, 3579
Banik I., Zhao H., 2022, Symmetry, 14, 1331
Bekenstein J., Milgrom M., 1984, ApJ, 286, 7
Binney J. et al., 2014, MNRAS, 439, 1231
Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton University Press, Princeton, NJ
Binney J., Vasiliev E., 2022, MNRAS, preprint (arXiv:2206.03523)
Bissantz N., Gerhard O., 2002, MNRAS, 330, 591
Bovy J., Rix H.-W., 2013, ApJ, 779, 115
Büdenbender A., van de Ven G., Watkins L. L., 2015, MNRAS, 452, 956
Bullock J. S., Boylan-Kolchin M., 2017, ARA&A, 55, 343
Chrobáková Ž., López-Corredoira M., Sylos Labini F., Wang H.-F., Nagy R., 2020, A&A, 642, A95
Dabringhausen J., Kroupa P., Famaey B., Fellhauer M., 2016, MNRAS, 463, 1865
Dehnen W., Binney J., 1998, MNRAS, 294, 429
Eilers A.-C., Hogg D. W., Rix H.-W., Ness M. K., 2019, ApJ, 871, 120
Everall A., Evans N. W., Belokurov V., Schönrich R., 2019, MNRAS, 489, 910
Famaey B., McGaugh S. S., 2012, Living Rev. Relativ., 15, 10
Feng J. L., 2010, ARA&A, 48, 495
Gaia Collaboration et al., 2018a, A&A, 616, A1
Gaia Collaboration et al., 2018b, A&A, 616, A11
Gaia Collaboration et al., 2021, A&A, 649, A1
Gravity Collaboration et al., 2018, A&A, 615, L15
Hattori K., Valluri M., Vasiliev E., 2021, MNRAS, 508, 5468
Hecht F., 2012, J. Numer. Math., 20, 251
Hessman F. V., 2015, A&A, 579, A123
Hossenfelder S., 2017, Phys. Rev. D, 95, 124018
Huang Y. et al., 2020, ApJS, 249, 29
Kipper R., Tenjes P., Tihhonova O., Tamm A., Tempel E., 2016, MNRAS, 460, 2720
Kuijken K., Gilmore G., 1991, ApJ, 367, L9
Lisanti M., Moschella M., Outmezguine N. J., Slone O., 2019, Phys. Rev. D, 100, 083009
McGaugh S. S., 2016, ApJ, 816, 42
McGaugh S. S., Lelli F., Schombert J. M., 2016, Phys. Rev. Lett., 117, 201101
McGaugh S. S., Schombert J. M., Bothun G. D., de Blok W. J. G., 2000, ApJ, 533, L99
McMillan P. J., 2017, MNRAS, 465, 76
Milgrom M., 1983, ApJ, 270, 365
Milgrom M., 1999, Physics Letters A, 253, 273
Milgrom M., 2010a, MNRAS, 403, 886
Milgrom M., 2010b, in Alimi J.-M., Fuözfa A., eds, AIP Conf. Proc. Vol. 1241, Invisible Universe. Am. Inst. Phys., New York, p. 139
Milgrom M., 2011, A Phys. Pol. B, 42, 2175
Milgrom M., 2014, MNRAS, 437, 2531
Milgrom M., 2015, preprint (arXiv:1511.08087)
Milgrom M., 2022, Phys. Rev. D, 106, 064060
Moffat J. W., 2006, J. Cosmology Astropart. Phys., 2006, 004
Moffat J. W., Rahvar S., 2013, MNRAS, 436, 1439
Mróz P. et al., 2019, ApJ, 870, L10
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Nipoti C., Londrillo P., Zhao H., Ciotti L., 2007, MNRAS, 379, 597
Pato M., Iocco F., 2017, SoftwareX, 6, 54
Piffl T. et al., 2014, MNRAS, 445, 3133
Read J. I., 2014, J. Phys. G Nucl. Phys., 41, 063101
Springel V., 2005, MNRAS, 364, 1105
Stubbs C. W., Garg A., 2005, preprint (arXiv:astro-ph/0512067)
Tully R. B., Fisher J. R., 1977, A&A, 54, 661
Wang J., Hammer F., Yang Y., 2022, MNRAS, 510, 2242
Wen X.-G., 2003, Phys. Rev. D, 68, 065003
Wilkinson M. I., Evans N. W., 1999, MNRAS, 310, 645
Wu X., Famaey B., Gentile G., Perets H., Zhao H., 2008, MNRAS, 386, 2199
Zhao H., 1996, MNRAS, 278, 488
Zhao H., Li B., 2010, ApJ, 712, 130

## APPENDIX: TESTS BASED ON THE LEGACY GALACTIC CONSTANTS

As we stated in Section 3.1 and Section 5.1, We have exploited other parameterizations of the Galactic mass distribution and other kinematic data in the literature, including those under other sets of the solar position and velocity constants ($R_\odot$ and $v_\odot$), and found that our conclusions presented in the main text remain intact. In the Appendix, we present such an examination: the results based on the mass models and kinematic data adopting the legacy Galactic constants $R_\odot = 8$ kpc and $v_\odot = 220$ km s^{-1}.

In order to use all kinds of rotation curve data in the literature, we employ the software galkin (Pato & Iocco 2017). galkin is a powerful tool that contains the largest compilation of rotation curve data of the MW from the literature, and can bring the data uniformly to a specific Galactic coordinate system the user sets (in the Appendeix here, we use the usual Galactocentric cylindrical system with the above legacy Galactic constants).

To be consistent with the legacy Galactic constants in this appendix, we adopt the Galactic mass model that was best fit with the 'Weaker R_0 prior' by McMillan (2017), where the best-fit $R_\odot$ and v_0 are consistent with the legacy values within 1σ uncertainty. McMillan (2017) use the Bissantz & Gerhard (2002) parametric formula for the bulge (with an axisymmetric approximation), exponential profiles for the stellar discs, the Dehnen & Binney (1998) model for the interstellar medium discs, and the NFW profile for the DM halo. The

Table A1. Parameters of the 'Weaker R_0 prior' mass model of McMillan (2017).

disc	Thin	Thick	H I	H₂
Σ_0[M$_\odot$ kpc^{-2}]	9.52e8	1.20e8	5.31e7	2.18e9
R_d[kpc]	2.40	3.47	7.0	1.5
z_d[kpc]	0.3	0.9	0.085	0.045
R_m[kpc]	–	–	4.0	12.0
Spheroid	Halo	Bulge	–	–
ρ_0[M$_\odot$ kpc^{-3}]	6.98e6	1.02e11	–	–
r_h[kpc]	21.21	–	–	–
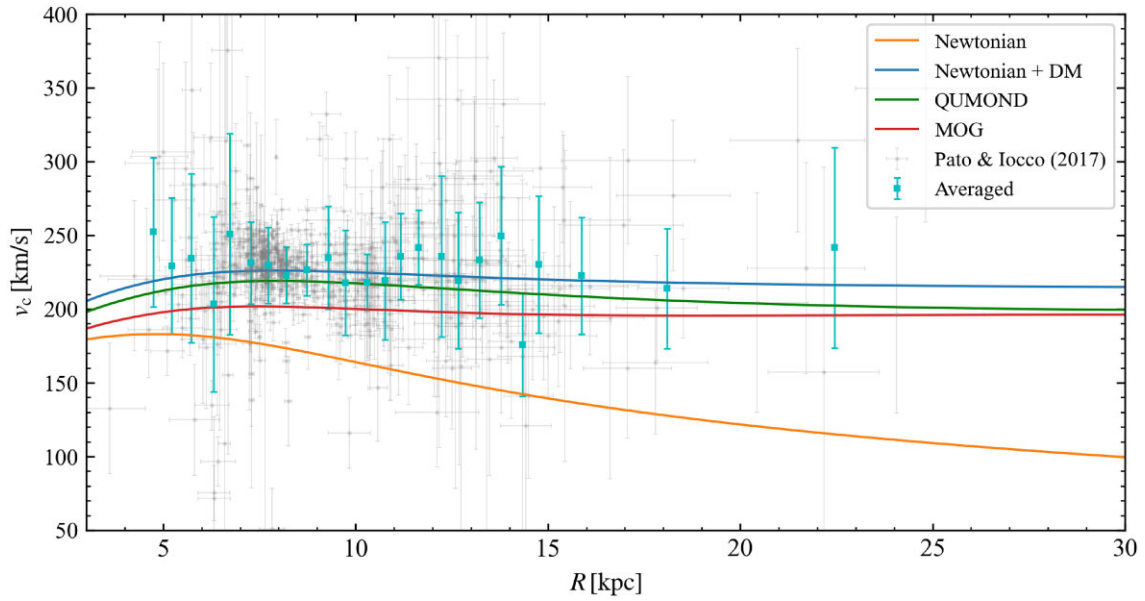
Figure A1. As Fig. 1, but for the mass model and observational data based on $R_\odot = 8$ kpc and $v_0 = 220$ km s^{-1}. The cyan data points with $\pm 1\sigma$ error bars are our averaged rotation curve over spatial bins with $\Delta R = 0.5$ kpc (we increase the bin size at large R), based on the data compiled by galkin (Pato & Iocco 2017). The baryonic parameters are from the 'Weaker R_0 prior' mass model of McMillan (2017).
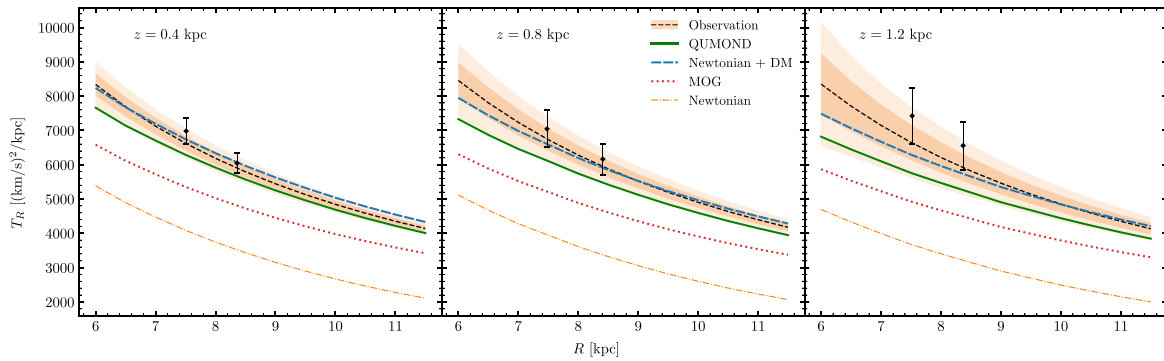
Figure A2. As Fig. 2, but showing the T_R test results for the 'Weaker R_0 prior' mass model of McMillan (2017). Also plotted is the results based on the RAVE sample (black dots with $\pm 1\sigma$ error bars).
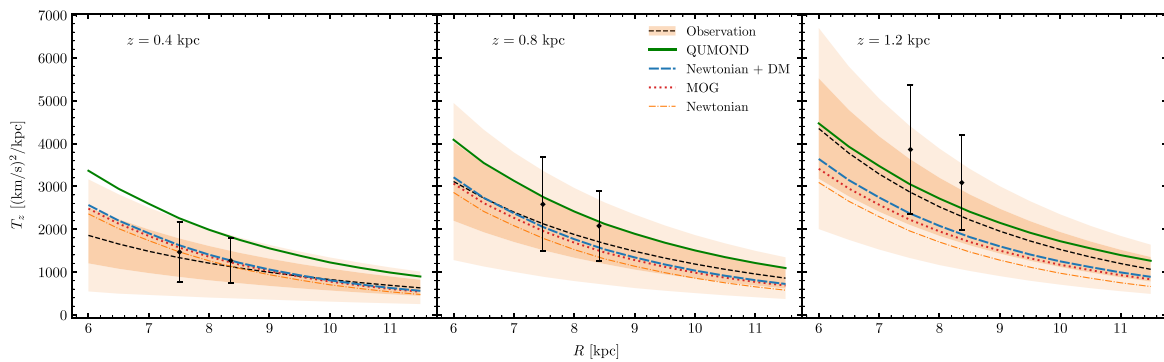
Figure A3. As Fig. A2, but showing the T_z test results.

NFW halo profile is as follows:

$$\rho_h = \frac{\rho_{0,h}}{x^\gamma (1+x)^{3-\gamma}},$$ (A1)

where $x = r/r_h$, with r_h being the scale radius and $\gamma = 1$. The parametrizations of the other profiles are detailed in Section 3.1. The specific parameters are listed in Table A1. In fact, all mainstream Galactic mass models that prevailed in the literature have no substantial difference from each other; this is at least true for the goal of our present study.

Fig. A1 shows the rotation curve for the 'Weaker R_0 prior' model. We also calculate the reduced χ^2 with a degree of freedom d.o.f. = 23 regarding to 24 radial bins of the data. The DM, QUMOND, and MOG models have reduced $\chi^2_{\nu,\mathrm{DM}} = 0.2$, $\chi^2_{\nu,\mathrm{QUMOND}} = 0.2$, and $\chi^2_{\nu,\mathrm{MOG}} = 0.5$. This is consistent with the visual inspection from Fig. A1, i.e. all the three gravitational models match most of the binned data points within their $\pm 1\sigma$ measurement uncertainties. In contrast, not surprisingly at all, the Newtonian baryon-only model is not favoured with $\chi^2_{\nu,\mathrm{N}} = 3.4$, larger than the other three models by an order of magnitude. We note that the χ^2_ν values here are smaller than those in Section 5.1 because the measurement uncertainties of the binned data points are large; this is reasonable considering that the galkin compilation is heterogeneous.

We also present the Jeans-equations tests based on the 'Weaker R_0 prior' mass model of McMillan (2017), and the three-dimensional velocity data as described in Section 4. The results are presented in Figs A2 and A3. As mentioned in Section 5.2, we also show the RAVE-based results whose T_R and T_z have large errors. It is evident that these new T_R and T_z tests are also consistent with our results in Section 5 based on the W22 mass model.

This paper has been typeset from a TeX/LaTeX file prepared by the author.