Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A figure of merit for kinematic limitation of binary-pulsar tests of orbital decay

Muhammad Anas

anaszain260@gmail.com

independent researcher

Research Article

Keywords: gravitation, gravitational waves, neutron stars, pulsars, astrometry, relativistic binary stars

Posted Date: October 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10605350/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

A figure of merit for kinematic limitation of binary-pulsar tests of orbital decay

Muhammad Anas

Independent researcher

Correspondence: anaszain260@gmail.com

Abstract

The observed orbital period derivative $\dot{P}_b$ of a binary pulsar contains kinematic contributions — the Shklovskii term and the differential Galactic acceleration — that must be subtracted before comparison with the general-relativistic (GR) prediction for gravitational-wave emission. We ask whether the radiative test of GR is limited by the precision of the $\dot{P}_b$ measurement or by the uncertainty in the kinematic correction, and which measurement would most improve it. We define a dimensionless figure of merit $\mathcal{F}$, the ratio of the uncertainty in the kinematic correction (the distance error propagated jointly through the Shklovskii and Galactic terms, the proper-motion error, and the difference between two Galactic potential models, evaluated by Monte Carlo and split into components) to the measurement uncertainty in $\dot{P}_b^{\rm obs}$. We compute $\mathcal{F}$ for sixteen binary pulsars from the timing literature and for 28 field binaries in the ATNF Pulsar Catalogue (v2.8.1), and use a smoothed-bootstrap synthetic population of 2×10^5 systems to test the stability of the classification. Six of the sixteen literature systems are distance-limited ($\mathcal{F} \geq 1$), six borderline, and four measurement-limited; omitting the Galactic term understates the kinematic limitation for six systems, including B1913+16. Systems reach $\mathcal{F} \geq 1$ through large proper motion at small distance (the Shklovskii term) or large distance at low Galactic latitude (the Galactic term); for most the dominant uncertainty is the distance, but for J1909−3744 it is the Galactic potential model. Eleven of the 28 catalogue systems are distance-limited. $\mathcal{F}$ does not predict the achieved test precision. In the synthetic population, 7.9% of systems change class under realistic measurement noise (4.0% with parallax distances, 14.9% with DM distances), and the distribution of $\mathcal{F}$ is consistent with a single log-normal component, so the thresholds $\mathcal{F} = 0.3$ and 1 are conventions. $\mathcal{F}$ is a per-system diagnostic of which limitation dominates, not a predictor of test quality; where the Galactic-model term dominates, $\dot{P}_b$ is more informative as a probe of the Galactic potential than as a test of GR.

Keywords: astrometry — Galaxy: kinematics and dynamics — gravitation — gravitational waves — pulsars: general

1 Introduction

Binary pulsars have been used to test General Relativity (GR) since the discovery of PSR B1913+16 (Hulse & Taylor, 1975). A pulsar in a close orbit with a compact companion acts as a clock moving through a strong, time-varying gravitational field, and pulse arrival times constrain the orbital dynamics precisely. The orbital decay of B1913+16 provided the first evidence that gravitational waves (GWs) carry energy from such a system, at a rate consistent

with GR (Taylor & Weisberg, 1982; Weisberg & Huang, 2016). Five decades of subsequent work have turned binary pulsars into the most precise laboratories for strong-field and radiative gravity available before the era of direct GW detection (Freire & Wex, 2024; Wex, 2014; Will, 2014).

The theoretical prediction for the decay is the Peters–Mathews quadrupole formula (Peters & Mathews, 1963), which gives $\dot{P}_b$ for two point masses radiating gravitationally as a function of the component masses, orbital period, and eccentricity. For a system with all of these measured, GR predicts $\dot{P}_b$, and the comparison of prediction with observation is the radiative test. The component masses themselves are obtained from the timing solution, usually from two or more post-Keplerian (PK) parameters such as the periastron advance $\dot{\omega}$, the Einstein delay γ, and the Shapiro-delay parameters, so that the mass determination does not presuppose the validity of GR (Damour & Taylor, 1991; Wex, 2014).

Two broad classes of system contribute. Double neutron star (DNS) binaries such as B1913+16 and the Double Pulsar PSR J0737−3039A/B (Kramer et al., 2021) are the classical radiative-test systems, where the symmetry of the two components suppresses dipolar radiation and the test is essentially of the quadrupole formula. Neutron-star–white-dwarf (NS–WD) binaries such as PSR J1738+0333 and PSR J0348+0432 (Antoniadis et al., 2013; Freire et al., 2012) are highly asymmetric in gravitational binding energy, which makes them sensitive instead to the dipolar radiation predicted by many alternative theories of gravity; for these systems the observed agreement of $\dot{P}_b$ with the GR quadrupole prediction constrains that dipolar term (Freire & Wex, 2024; Wex, 2014). The figure of merit developed below is indifferent to this distinction, since it concerns the kinematic budget rather than the radiative physics, and we therefore include both classes in a single sample.

The complication, recognised early in the analysis of the Hulse–Taylor system (Damour & Taylor, 1991), is that the measured $\dot{P}_b$ is not the intrinsic decay due to gravitational radiation. The relative acceleration of the pulsar and Solar System in the Galactic potential, and the transverse motion of the pulsar across the sky (the Shklovskii effect; Shklovskii, 1970), both produce apparent changes in the orbital period. These kinematic contributions are subtracted before the residual is formed (Damour & Taylor, 1991; Lazaridis et al., 2009; Nice & Taylor, 1995). Both depend on the distance, and the Shklovskii term depends additionally on the proper motion. For B1913+16 the present limit on the test, about 0.2%, is set by the distance and Galactic-potential uncertainties rather than by the timing precision (Weisberg & Huang, 2016).

That distance limitation was anticipated before it became the dominant error. Bell & Bailes (1996) pointed out that the same kinematic $\dot{P}_b$ contribution that contaminates radiative tests can be turned around to measure distances, and noted in the same work that radiative tests of nearby relativistic binaries would be limited to accuracies of a few per cent by the distance, citing B1534+12 specifically. The method has since yielded model-independent "kinematic" distances for several millisecond pulsars (Bell & Bailes, 1996; Verbiest et al., 2008). The distance dependence is therefore well known, but it is usually discussed one system at a time, in the timing paper for each object.

Here we treat it across a population. The question we pose is narrow: for each binary pulsar with a measured $\dot{P}_b$, is the radiative test currently limited by how well $\dot{P}_b$ itself is measured, or by how well the kinematic correction is known? We do not introduce any new physical effect. The kinematic corrections are standard and the parameter values are from the published

timing literature and the ATNF Pulsar Catalogue (Manchester et al., 2005). The contribution is a single dimensionless quantity, computed uniformly across the sample, whose numerator is the full uncertainty of the kinematic correction: the distance error propagated jointly through the Shklovskii and Galactic terms, the proper-motion error, and the disagreement between two standard Galactic potential models, evaluated by Monte Carlo and split into its components so that the quantity also identifies which measurement would improve a given test. We test whether it predicts the precision a system actually achieves, extend it to the full catalogue, and use a synthetic population to assess how stable the resulting classification is against measurement error. All computations are performed with an open-source Python pipeline described in Sect. 2.4 and Appendix A.

2 Method

2.1 Radiative test

For component masses m_1, m_2 in an orbit of period P_b and eccentricity e, the GW orbital decay is (Peters & Mathews, 1963)

$$\dot{P}_\mathrm{b}^\mathrm{GR} = -\frac{192\pi G^{5/3}}{5c^5}\left(\frac{P_\mathrm{b}}{2\pi}\right)^{-5/3}\frac{m_1 m_2}{(m_1+m_2)^{1/3}}\,f(e), \tag{1}$$

with

$$f(e) = \left(1+\frac{73}{24}e^2+\frac{37}{96}e^4\right)(1-e^2)^{-7/2}. \tag{2}$$

The mass dependence enters only through the chirp mass $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1+m_2)^{1/5}$. The factor $f(e)$ rises steeply with eccentricity, so an eccentric system radiates much more strongly than a circular one at the same P_b (Peters & Mathews, 1963).

In practice the observed quantities entering Eqs. (1)–(2) are extracted from pulse times of arrival by a timing model fitted with standard software such as TEMPO2 (Edwards et al., 2006). The orbital decay enters either as an explicit $\dot{P}_\mathrm{b}$ in the Damour–Deruelle (DD) parametrisation, or, in the GR-assuming DDGR model, through a fitted excess (the XPBDOT parameter) that measures the difference between the observed decay and the GR prediction; we use whichever a given source reports and convert to a common convention.

2.2 Kinematic corrections

The observed decay decomposes as (Damour & Taylor, 1991; Nice & Taylor, 1995)

$$\dot{P}_\mathrm{b}^\mathrm{obs} = \dot{P}_\mathrm{b}^\mathrm{int} + \dot{P}_\mathrm{b}^\mathrm{Shk} + \dot{P}_\mathrm{b}^\mathrm{Gal}, \tag{3}$$

where $\dot{P}_\mathrm{b}^\mathrm{int}$ is the intrinsic decay and the other two terms are kinematic. The Shklovskii term (Shklovskii, 1970) is

$$\dot{P}_\mathrm{b}^\mathrm{Shk} = \frac{\mu^2 d}{c}\,P_\mathrm{b}, \tag{4}$$

with μ the total proper motion and d the distance. It is positive (an apparent period increase), so for a decaying orbit it makes the intrinsic decay larger in magnitude than the observed

one. The same expression underlies the use of $\dot{P}_{\rm b}$ as a kinematic distance estimator when the intrinsic decay is negligible or independently known (Bell & Bailes, 1996; Verbiest et al., 2008).

The Galactic term is the projection of the differential acceleration between the pulsar and the Solar System onto the line of sight. Following Damour & Taylor (1991), Nice & Taylor (1995), and Lazaridis et al. (2009), it is conventionally split into a component from the vertical acceleration toward the Galactic disc and a component from differential Galactic rotation,

$$\frac{\dot{P}_{\rm b}^{\rm Gal}}{P_{\rm b}} = -\frac{1}{c}\left(a_z|\sin b| + a_{\rm rot}\right), \tag{5}$$

where a_z is the vertical acceleration at the pulsar's height $z = d\sin b$ above the plane and $a_{\rm rot}$ is the planar differential-rotation term for a flat rotation curve,

$$a_{\rm rot} = \frac{\Theta_0^2}{R_0}\left(\cos l + \frac{\beta}{\beta^2 + \sin^2 l}\right)\cos b, \qquad \beta \equiv \frac{d}{R_0}\cos b - \cos l, \tag{6}$$

with R_0 and Θ_0 the Galactocentric distance and circular speed of the Sun (Lazaridis et al., 2009). We adopt $R_0 = 8.275$ kpc and $\Theta_0 = 240.5$ km s^{-1} (GRAVITY Collaboration, 2021). For a_z we use the parametrisation of Holmberg & Flynn (2004) in the form given by Lazaridis et al. (2009), $a_z/(10^{-9}\,{\rm cm\,s^{-2}}) = 2.27\,z_{\rm kpc} + 3.68\,(1 - e^{-4.31 z_{\rm kpc}})$. Eqs. (5)–(6) are our reference Galactic model. As an alternative we evaluate the line-of-sight acceleration directly from the Milky Way potential of McMillan (2017), as implemented in GALPY (Bovy, 2015); the difference between the two is used as a model uncertainty in Sect. 2.3. The two models agree to about ten per cent near the Sun and in the plane, but differ by factors of two or more along some lines of sight, because the flat-rotation-curve form of Eq. (6) is a poor approximation toward the inner Galaxy and the local a_z parametrisation and the global model differ by tens of per cent at heights of several hundred parsecs. Both kinematic terms scale with distance, so a distance error propagates into the correction and hence into the residual.

Distances themselves come from heterogeneous sources. Geometric distances from VLBI or timing parallax are model-independent. Distances from the Galactic electron-density models NE2001 (Cordes & Lazio, 2002) and YMW16 (Yao et al., 2017), inferred from the dispersion measure, carry large and not-well-quantified systematics, since the electron density along a given sightline is not tightly constrained by a smooth model; the two models can disagree by tens of per cent or, for some lines of sight, by a factor of two (Yao et al., 2017). A few NS–WD systems also have photometric distances from white-dwarf cooling models. We record the distance type for each system because it bears directly on the reliability of the kinematic correction.

2.3 Figure of merit

We quantify whether the kinematic subtraction or the $\dot{P}_{\rm b}$ measurement limits the test with

$$\mathcal{F} = \frac{\sigma\left(\dot{P}_{\rm b}^{\rm kin}\right)}{\sigma\left(\dot{P}_{\rm b}^{\rm obs}\right)}, \qquad \dot{P}_{\rm b}^{\rm kin} \equiv \dot{P}_{\rm b}^{\rm Shk} + \dot{P}_{\rm b}^{\rm Gal}, \tag{7}$$

where the denominator is the published measurement uncertainty. To first order, the numerator has three independent contributions:

$$\sigma^2\left(\dot{P}_{\rm b}^{\rm kin}\right) \simeq \left(\frac{\partial \dot{P}_{\rm b}^{\rm kin}}{\partial d}\,\sigma_d\right)^2 + \left(2\dot{P}_{\rm b}^{\rm Shk}\frac{\sigma_\mu}{\mu}\right)^2 + \sigma_{\rm mod}^2. \tag{8}$$

The first term propagates the distance error through the Shklovskii and Galactic terms together, since both depend on the same d; the factor of two in the second follows from the μ^2 dependence in Eq. (4); and $\sigma_{\rm mod} = |\dot{P}_{\rm b}^{\rm Gal}({\rm ref}) - \dot{P}_{\rm b}^{\rm Gal}({\rm McM17})|$ is the difference between the two Galactic models of Sect. 2.2 at the nominal distance. Dividing each term by $\sigma(\dot{P}_{\rm b}^{\rm obs})$ defines three components, $\mathcal{F}_d$, $\mathcal{F}_\mu$, and $\mathcal{F}_{\rm mod}$, whose quadrature sum is the first-order $\mathcal{F}$. The largest component identifies the "lever": the measurement — parallax, proper motion, or Galactic model — whose improvement would most reduce the kinematic uncertainty.

Omitting the Galactic term and $\sigma_{\rm mod}$ reduces Eq. (8) to the simplified form

$$\sigma\left(\dot{P}_{\rm b}^{\rm kin}\right) \simeq \left|\dot{P}_{\rm b}^{\rm Shk}\right| \sqrt{\left(\frac{\sigma_d}{d}\right)^2 + \left(\frac{2\sigma_\mu}{\mu}\right)^2}, \tag{9}$$

which we denote $\mathcal{F}_{\rm Shk}$ and report for comparison, since it is the form most easily evaluated from tabulated parameters.

First-order propagation is inaccurate when fractional errors are large, and the distances inferred from low-significance parallaxes have asymmetric distributions. We therefore evaluate the first two terms of Eq. (8) by Monte Carlo. We draw 2×10^4 realisations of d — by sampling the parallax from a Gaussian and inverting, for parallax and VLBI distances, or from a log-normal of width $\ln(1 + \sigma_d/d)$ for other distances — and of μ from a Gaussian. The kinematic uncertainty is then half the 16th–84th percentile width of the resulting $\dot{P}_{\rm b}^{\rm kin}$ samples, added in quadrature to $\sigma_{\rm mod}$. This Monte Carlo value is the adopted $\mathcal{F}$.

Because d and μ are themselves measurements, a system's class may depend on which values within their uncertainties are true. We quantify this by re-evaluating $\mathcal{F}$ at 2000 plausible true values of (d, μ) drawn as above. We report the 16th–84th percentile range of $\mathcal{F}$ and the probability that the system lies in its assigned class.

When $\mathcal{F} \ll 1$ the test is measurement-limited and the distance is adequate; when $\mathcal{F} \gtrsim 1$ it is limited by the kinematic correction. The quantity is the ratio of two numbers that timing papers already report or can compute; its purpose is to rank the population uniformly rather than to assess systems in isolation. The underlying judgement is not new: individual timing papers routinely note whether a given system's radiative test is limited by its $\dot{P}_{\rm b}$ measurement or by the kinematic correction, and some estimate the achievable precision over a plausible distance range for that purpose (Gautam et al., 2022; Stovall et al., 2018). What we add is a uniform, dimensionless version of that judgement applied across the sample, together with the component that dominates it. We treat $\mathcal{F} \sim 0.3$ to 1 as a band rather than a sharp boundary, and label systems as measurement-limited (M, $\mathcal{F} < 0.3$), borderline (B, $0.3 \leq \mathcal{F} < 1$), or distance-limited (D, $\mathcal{F} \geq 1$).

2.4 Implementation

All quantities in this paper are computed with a purpose-written Python pipeline built on NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020), pandas (McKinney, 2010), scikit-learn (Pedregosa et al., 2011), Matplotlib (Hunter, 2007), Astropy (Astropy Collaboration, 2022), and GALPY (Bovy, 2015); catalogue data are retrieved with PSRQPY (Pitkin, 2018). The pipeline evaluates Eqs. (1)–(9), performs the Monte Carlo propagation, queries and filters the catalogue, generates the synthetic population, and produces every figure and table. As a check of the implementation, it reproduces the Shklovskii term of J0737−3039A/B quoted in Sect. 3.3 to three significant figures. It is fully deterministic for a given random seed; the structure is described in Appendix A.

3 Sample and results

3.1 Sample

We collected from the literature, for each system, the orbital period, eccentricity, component masses, total proper motion, distance with uncertainty, and observed orbital decay with uncertainty; Galactic coordinates follow from the pulsar positions. The sample (Table 1; inputs in Table B1) contains sixteen ranked binary pulsars and mixes DNS and NS–WD systems, since $\mathcal{F}$ depends only on the kinematic budget and not on the companion type. Systems without a measured $\dot{P}_\mathrm{b}$ (for example wide-orbit DNS systems where the decay is undetectable on current baselines; Martinez et al., 2015) are outside the sample by construction. Two further systems with published $\dot{P}_\mathrm{b}$ are not ranked. PSR J1141−6545 has no proper motion in the timing solution we adopt (Bhat et al., 2008); the ATNF catalogue lists one, but with a fractional uncertainty of about 0.9, which gives $\mathcal{F} \approx 0.3$ (Sect. 3.5) with a class probability of only 0.5. PSR J2043+1711 is excluded because its observed $\dot{P}_\mathrm{b}$ implies a peculiar acceleration that departs from Galactic-potential models at the 4σ level (Fonseca et al., 2016), most plausibly from an unmodelled third body or stellar flyby rather than the standard kinematic terms, so the decomposition underlying $\mathcal{F}$ does not apply to it.

3.2 Ordering by the figure of merit

Six of the sixteen systems are distance-limited, six borderline, and four measurement-limited (Table 1, Fig. 1). Including the Galactic term and the model uncertainty matters. Compared with the simplified $\mathcal{F}_\mathrm{Shk}$, six systems change class, and every change is towards stronger kinematic limitation. J1756−2251, J1738+0333, J0737−3039A/B, and J1946+2052 move from M to B, and B1913+16 and J1757−1854 move from B to D. For B1913+16 this resolves a tension. The simplified $\mathcal{F}_\mathrm{Shk} = 0.31$ would call its test borderline, whereas Weisberg & Huang (2016) find that the test is limited by the distance and Galactic-potential uncertainties. With the Galactic term, $\mathcal{F} \approx 3.0$, in agreement with that assessment.

If distance-measurement type were the controlling variable, the DM-distance systems would dominate the distance-limited class. They do not: four of the six distance-limited systems (J1933−6211, J2145−0750, J1909−3744, and B1534+12) have non-DM distances — parallax distances or, for J1909−3744, the $\dot{P}_\mathrm{b}$-derived distance (the catalogue uses a parallax; Sect. 3.5) — and so does every system with $\mathcal{F} > 5$. But distance type is not irrelevant either,

and the two DM-distance systems in the D class reach it by a different route from the geometric ones (Sect. 4.2). B1534+12 ($\mathcal{F} \approx 15$) and J1909−3744 ($\mathcal{F} \approx 11$) have well-measured non-DM distances and very precise $\dot{P}_\mathrm{b}$, but proper motions of 25 and 37 mas yr^{-1}; because the Shklovskii term scales as $\mu^2 d$, its size, and hence its uncertainty, dwarfs the measurement error. B1913+16 and J1757−1854 have small proper motions but lie 9–12 kpc away close to the Galactic plane, where the Galactic term and its distance derivative are large.

The components show which measurement would help. For five of the six distance-limited systems, the dominant term is the distance (Table 1): even where the large proper motion makes the Shklovskii term large, the proper motion itself is measured far more precisely than the distance, so σ_d/d rather than $2\sigma_\mu/\mu$ sets the uncertainty. For B1534+12, $\mathcal{F}_d \approx 13$ against $\mathcal{F}_\mu \approx 0.8$. This is consistent with the dedicated VLBI parallax campaign that Ding et al. (2021) undertook for this system because its radiative test was distance-limited. The exception is J1909−3744, for which the distance, derived from $\dot{P}_\mathrm{b}$ itself, is precise to 0.3%. Its dominant term is the Galactic-model uncertainty ($\mathcal{F}_\mathrm{mod} \approx 8$), followed by the proper motion ($\mathcal{F}_\mu \approx 7$). The system is already used in the reverse sense, as a kinematic distance estimator rather than a radiative test (Liu et al., 2020).

Three class assignments are uncertain. J1756−2251, with $\sigma_d/d = 0.55$, has a probability of only 0.32 of lying in class B, and its $\mathcal{F}$ spans 0.16–0.56; J1946+2052 (0.61) and J0737−3039A/B (0.83) are the other two below 0.9. All other assignments have probabilities of at least 0.90.

3.3 Independent recomputation

As a check on the tabulated decompositions rather than as new determinations, we recomputed the Shklovskii term from Eq. (4) for the systems that report a separated kinematic budget. For J0737−3039A/B the recomputation gives $\dot{P}_\mathrm{b}^\mathrm{Shk} = +1.73 \times 10^{-16}$, and the decomposition $\dot{P}_\mathrm{b}^\mathrm{obs} = \dot{P}_\mathrm{b}^\mathrm{int} + \dot{P}_\mathrm{b}^\mathrm{Gal} + \dot{P}_\mathrm{b}^\mathrm{Shk}$ closes to the quoted precision (Kramer et al., 2021). For B1534+12 the recomputed term reproduces the value of Ding et al. (2021). For J1946+2052, where Meng et al. (2025) report only the combined residual, the recomputation splits it consistently with that residual. The Monte Carlo and first-order values of $\mathcal{F}$ agree to within 12% for fourteen of the sixteen systems. The exceptions are J1756−2251, whose Monte Carlo value is 15% higher because of its large distance uncertainty, and J1757−1854, where it is 28% lower, reflecting the non-linear dependence of the Galactic term on distance. These checks guard against transcription, unit, and linearisation errors only.

3.4 Test against achieved precision

A figure of merit of this kind invites a test: does $\mathcal{F}$ predict the radiative-test precision actually achieved? If it captured the limiting factor, large $\mathcal{F}$ should correspond to poorer fractional precision in the published $\dot{P}_\mathrm{b}^\mathrm{int}/\dot{P}_\mathrm{b}^\mathrm{GR}$ ratios. We compiled these ratios for the systems where a test of this form is reported (Antoniadis et al., 2013; Ding et al., 2021; Ferdman et al., 2014; Freire et al., 2012; Kramer et al., 2021; Meng et al., 2025; Weisberg & Huang, 2016) and compared the fractional precision with $\mathcal{F}$.

The prediction fails. The rank correlation between $\mathcal{F}$ and achieved fractional precision for the seven systems with comparable ratios is $\rho = -0.21$ (exact two-sided permutation $p = 0.66$, $n = 7$). The seven systems are B1913+16, B1534+12, J0737−3039A/B, J1946+2052, J1756−2251,

Table 1. Binary pulsars with published orbital-decay measurements and the figure of merit (Sect. 2.3), sorted by $\mathcal{F}$. Input parameters are in Table B1. Distances are geometric (parallax/VLBI), photometric (white-dwarf model), $\dot{P}_b$-derived, or from a Galactic electron-density model (DM); d is in kpc. $\mathcal{F}_{\rm Shk}$ is the simplified figure of merit of Eq. (9); $\mathcal{F}$ is the adopted Monte Carlo value including the Galactic term and the Galactic-model uncertainty; the range is the 16th–84th percentile of $\mathcal{F}$ over plausible true values of d and μ. $\mathcal{F}_d$, $\mathcal{F}_\mu$, and $\mathcal{F}_{\rm mod}$ are the first-order components of Eq. (8). "Lever" is the dominant component: distance (d), proper motion (μ), or Galactic model (Gal). The class is M ($\mathcal{F} < 0.3$), B ($0.3 \leq \mathcal{F} < 1$), or D ($\mathcal{F} \geq 1$), with the probability of that class in parentheses.

#	System	dist. type	d	$\mathcal{F}_{\rm Shk}$	$\mathcal{F}$	range	$\mathcal{F}_d$	$\mathcal{F}_\mu$	$\mathcal{F}_{\rm mod}$	lever	class
1	J0348+0432	photometric	2.1	0.003	0.0134	0.012–0.0149	0.0032	0.0012	0.013	Gal	M (1.00)
2	J2234+0611	parallax	0.95	0.079	0.0751	0.0732–0.0796	0.075	0.0034	0.015	d	M (1.00)
3	J1913+1102	DM	7.14	0.255	0.186	0.134–0.271	0.048	0.17	0.03	μ	M (0.91)
4	J2222−0137	VLBI	0.267	0.151	0.273	0.272–0.274	0.13	0.077	0.23	Gal	M (1.00)
5	J1946+2052	DM	3.5	0.233	0.302	0.215–0.473	0.33	0.069	0.066	d	B (0.61)
6	J1756−2251	parallax	0.73	0.032	0.307	0.159–0.558	0.27	0.006	0.031	d	B (0.32)
7	J0737−3039A/B	parallax	0.735	0.193	0.325	0.299–0.351	0.14	0.067	0.29	Gal	B (0.83)
8	J1829+2456	DM	1.2	0.56	0.365	0.328–0.489	0.39	0.029	0.099	d	B (0.91)
9	J1738+0333	parallax	1.47	0.181	0.492	0.446–0.545	0.21	0.012	0.44	Gal	B (1.00)
10	J1012+5307	parallax	0.83	0.839	0.836	0.793–0.9	0.81	0.11	0.2	d	B (0.99)
11	J1933−6211	parallax	1.6	1.25	1.63	1.31–2.06	1.3	0.11	0.9	d	D (1.00)
12	J2145−0750	parallax	0.5	1.86	1.66	1.08–2.26	1.4	0	0.48	d	D (0.90)
13	J1757−1854	unknown	11.5	0.421	2.27	1.8–3.53	3	0.2	1.1	d	D (1.00)
14	B1913+16	DM	9	0.306	3.05	2.05–5.11	3	0.074	1.7	d	D (1.00)
15	J1909−3744	$\dot{P}_b$	1.16	7.26	10.8	10.8–11	1.9	7	8.1	Gal	D (1.00)
16	B1534+12	parallax	0.94	13.2	14.5	13.4–15.8	13	0.78	6.3	d	D (1.00)

J1738+0333, and J0348+0432, with fractional precisions of about 0.2%, 2%, 0.06%, 0.09%, 3%, 14%, and 17%, respectively (the 3% for J1756−2251 is the formal uncertainty of the $\dot{P}_b^{\rm int}/\dot{P}_b^{\rm GR}$ ratio of 1.08 ± 0.03 of Ferdman et al., 2014, which is itself in tension with GR). The sign is opposite to that expected if $\mathcal{F}$ tracked precision, but with seven systems the correlation is statistically indistinguishable from zero. Systems with small $\mathcal{F}$ include some with the worst precision: J0348+0432 has $\mathcal{F} \approx 0.013$ but a test at the 17% level (Antoniadis et al., 2013), and J1738+0333 has $\mathcal{F} \approx 0.49$ but a test near 14% (Freire et al., 2012). J0737−3039A/B has $\mathcal{F} \approx 0.33$, higher than J0348+0432, but the best precision in the sample, near 0.06% (Kramer et al., 2021). B1913+16 ($\mathcal{F} \approx 3$) and B1534+12 ($\mathcal{F} \approx 15$) are both distance-limited, yet reach precisions of about 0.2% and 2%, respectively.

The explanation is direct. Achieved precision is set by the absolute size of the dominant uncertainty relative to $\dot{P}_b^{\rm GR}$, not by the ratio of the two uncertainties. For J0348+0432 the kinematic correction is negligible and well known — which is why $\mathcal{F}$ is small — but the orbital decay is only marginally measured, so the test is limited by the denominator, the measurement uncertainty. A small $\mathcal{F}$ therefore coexists with a poor test when the smallness comes from a poorly measured $\dot{P}_b$ rather than from clean kinematics. $\mathcal{F}$ identifies which limitation dominates; it does not state how good the test is. A third contribution, the uncertainty in $\dot{P}_b^{\rm GR}$ from the mass measurements, enters the achieved precision but not $\mathcal{F}$ (Sect. 4.4).

The sample for this comparison is small: most NS–WD systems in Table 1 report dipolar-radiation or $\dot{G}$ constraints rather than a $\dot{P}_b^{\rm int}/\dot{P}_b^{\rm GR}$ ratio, so the comparison rules out the strong predictive claim but does not characterise any relationship precisely.

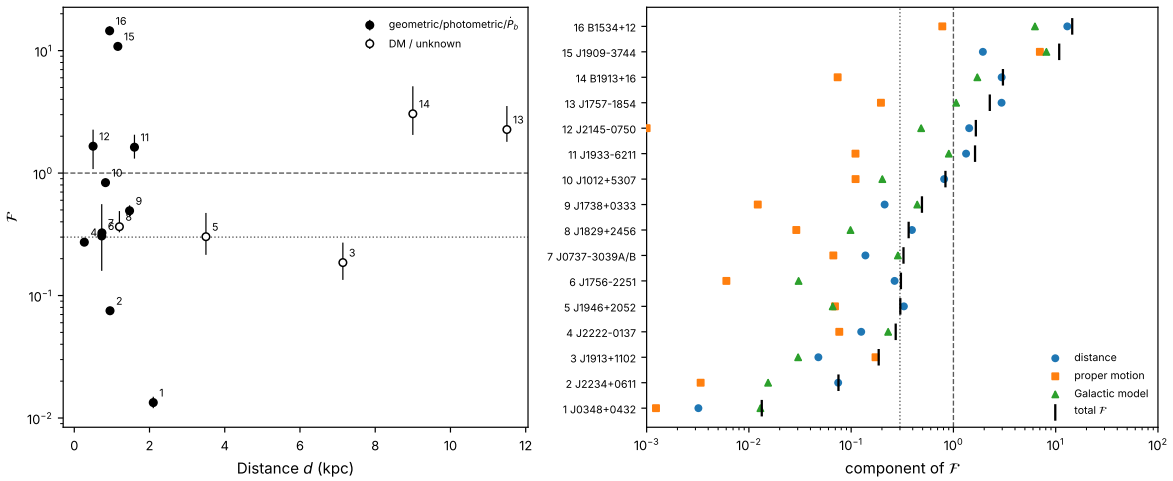


Figure 1. Left: figure of merit $\mathcal{F}$ versus distance for the sixteen ranked systems of Table 1, numbered by increasing $\mathcal{F}$. Error bars show the 16th–84th percentile range of $\mathcal{F}$ over plausible true values of d and μ. Filled markers: geometric, photometric, or $\dot{P}_b$-derived distances; open markers: DM or unknown. Dotted and dashed lines mark $\mathcal{F} = 0.3$ and 1. Right: the three components of $\mathcal{F}$ for each system — distance ($\mathcal{F}_d$), proper motion ($\mathcal{F}_\mu$), and Galactic model ($\mathcal{F}_{\rm mod}$) — with the total marked by a vertical bar. The largest component identifies the measurement whose improvement would most reduce the kinematic uncertainty.

3.5 Extension to the ATNF Pulsar Catalogue

To test whether the literature-sample results extend beyond hand-compiled systems, we computed $\mathcal{F}$ uniformly for every binary pulsar in the ATNF Pulsar Catalogue (v2.8.1; Manchester et al., 2005) that has the inputs it requires. The catalogue was queried with PSRQPY for binary pulsars with a measured orbital period, $\dot{P}_b$, and proper motion in both coordinates, which returned 331 systems. We retained the 60 with finite, positive values of P_b, μ, d, eccentricity, mass function, $\sigma(\dot{P}_b^{\rm obs})$, and σ_μ/μ. The fractional distance error was taken from the parallax where it is detected at more than 2σ, and was otherwise set to $\sigma_d/d = 0.25$, representative of the scatter between the NE2001 and YMW16 models (Yao et al., 2017).

We then removed systems for which the decomposition behind $\mathcal{F}$ does not apply, on the same grounds as the exclusion of J2043+1711. These were 20 globular-cluster pulsars, whose $\dot{P}_b$ is dominated by the cluster potential; eleven systems with ultra-light or non-degenerate (main-sequence) companions, whose orbital periods change through tidal and mass-loss effects rather than through GR and kinematics; and the B pulsar of the Double Pulsar, which shares its orbit with J0737−3039A. Three of the excluded companion systems also carry catalogue $\dot{P}_b$ uncertainties below 10^{-21}, orders of magnitude below any timing measurement, which we treat as catalogue artefacts. The final catalogue sample contains 28 field binaries (Table 2).

Of the 28 systems, ten are measurement-limited, seven borderline, and eleven distance-limited; the median $\mathcal{F}$ is ≈ 0.50. Nine of the eleven distance-limited systems have parallax distances; the other two, J1757−1854 and J1618−3921, have DM distances and lie at 19.6 and 5.5 kpc toward the inner Galaxy. The components divide the distance-limited systems by lever. The distance dominates for seven: J0437−4715, B1534+12, J1909−3744, B1913+16, J1012+5307, J1757−1854, and J1933−6211. The proper motion dominates for one, J0751+1807. The Galactic model dominates for three, J1713+0747, J1600−3053, and J1618−3921, which would not be distance-limited without the model term. J0437−4715, which is not in Table 1, has the largest

Table 2. Figure of merit for the 28 field binary pulsars in the ATNF Pulsar Catalogue v2.8.1 with measured $\dot{P}_b$, proper motion, and distance, after the exclusions described in Sect. 3.5, sorted by $\mathcal{F}$. All inputs are as listed in, or propagated from, the catalogue; no literature values are substituted. σ_d/d is from the parallax (plx) where detected at $> 2\sigma$, otherwise fixed at 0.25 (DM). $\sigma(\dot{P}_b^{\rm obs})$ is in 10^{-12} s s^{-1}. Other columns as in Table 1. Systems marked * also appear in Table 1.

#	PSR	P_b (d)	μ	d	σ_d/d	dist.	σ_μ/μ	$\sigma(\dot{P}_b^{\rm obs})$	$\mathcal{F}_{\rm Shk}$	$\mathcal{F}$	range	lever	class
1	J2045+3633	32.3	4.17	5.63	0.250	DM	0.060	5.3	0.0348	0.0244	0.0232–0.0314	Gal	M (1.00)
2	J1227−6208	6.721	6.06	9.47	0.250	DM	1.106	40	0.0273	0.0265	0.00305–0.114	μ	M (0.98)
3	J0737−3039A*	0.1023	4.34	1.1	0.161	plx	0.630	0.017	0.0332	0.0331	0.00621–0.0916	μ	M (1.00)
4	J0509+3801	0.3796	6.57	1.56	0.250	DM	0.273	0.07	0.0461	0.05	0.029–0.0839	μ	M (1.00)
5	J1518+4904	8.634	8.56	0.806	0.032	plx	0.040	0.22	0.0419	0.0532	0.0505–0.0559	μ	M (1.00)
6	J2019+2425	76.51	22.6	1.16	0.250	DM	0.145	60	0.0612	0.0571	0.0409–0.0823	μ	M (1.00)
7	J1455−3330	76.17	8.1	0.769	0.077	plx	0.014	2	0.0329	0.0771	0.0697–0.0853	Gal	M (1.00)
8	J1756−2251*	0.3196	2.42	0.73	0.250	DM	0.080	0.005	0.017	0.114	0.0971–0.166	d	M (1.00)
9	J2234+0611*	32	27.1	0.971	0.039	plx	0.026	2.5	0.124	0.119	0.115–0.131	μ	M (1.00)
10	J1141−6545	0.1977	3.41	10	0.250	DM	0.907	0.03	0.295	0.288	0.0611–1.07	μ	M (0.49)
11	J1952+2630	0.3919	5.98	10	0.250	DM	0.281	0.046	0.394	0.381	0.198–0.669	μ	B (0.62)
12	J2222−0137*	2.446	45	0.269	0.003	plx	0.005	0.008	0.381	0.442	0.441–0.448	μ	B (1.00)
13	J1918−0642	10.91	9.29	1.33	0.093	plx	0.013	0.07	0.365	0.483	0.436–0.54	d	B (1.00)
14	J1738+0333*	0.3548	8.67	1.69	0.085	plx	0.007	0.004	0.204	0.49	0.433–0.562	Gal	B (1.00)
15	J1640+2224	175.5	11.5	1.37	0.082	plx	0.007	2	0.279	0.505	0.459–0.56	Gal	B (1.00)
16	J1949+3106	1.95	5.85	7.47	0.250	DM	0.048	0.025	1.12	0.672	0.398–1.23	d	B (0.73)
17	J0613−0200	1.199	10.5	0.95	0.050	plx	0.006	0.002	0.68	0.765	0.725–0.807	d	B (1.00)
18	J1933−6211*	12.82	12.4	1	0.286	plx	0.027	0.1	1.21	1.43	1–1.9	d	D (0.84)
19	J1757−1854*	0.1835	4.51	19.6	0.250	DM	0.172	0.005	1.3	1.74	1.57–2.59	d	D (0.99)
20	J0751+1807	0.2631	13.5	0.6	0.047	plx	0.069	0.0005	1.76	1.77	1.54–2.04	μ	D (1.00)
21	J1600−3053	14.35	6.98	1.39	0.028	plx	0.010	0.03	0.232	1.9	1.82–1.99	Gal	D (1.00)
22	J1618−3921	22.75	2.8	5.51	0.250	DM	0.276	0.4	0.311	2	1.28–4.96	Gal	D (0.95)
23	J1012+5307*	0.6047	25.6	0.7	0.017	plx	0.004	0.0006	1.83	2.14	2.11–2.21	d	D (1.00)
24	J1713+0747	67.83	6.29	1.14	0.011	plx	0.001	0.07	0.107	3.46	3.41–3.52	Gal	D (1.00)
25	J1915+1606 (B1913+16)*	0.323	0.721	4.17	0.292	plx	0.110	0.001	0.0536	5.27	2.88–6.83	d	D (1.00)
26	J1537+1155 (B1534+12)*	0.4207	25.3	0.935	0.065	plx	0.011	0.0003	12.2	13.5	12.6–14.6	d	D (1.00)
27	J1909−3744*	1.533	37	1.14	0.012	plx	0.005	0.0007	10.9	13.6	13.4–13.8	d	D (1.00)
28	J0437−4715	5.741	141	0.157	0.006	plx	0.001	0.0015	15.8	15.5	15.6–15.8	d	D (1.00)

μ in mas yr^{-1}, d in kpc.

$\mathcal{F}$ in the catalogue sample (≈ 15.5); like J1909−3744, its $\dot{P}_b$ is dominated by the Shklovskii term and has been used to derive a kinematic distance (Verbiest et al., 2008). Compared with the simplified $\mathcal{F}_{\rm Shk}$, seven catalogue systems change class; six of the seven changes are towards stronger kinematic limitation. Four class assignments have probabilities below 0.9: J1141−6545 (0.49), J1952+2630 (0.62), J1949+3106 (0.73), and J1933−6211 (0.84).

Eleven of the sixteen ranked systems of Table 1 are in the catalogue sample; the other five lack one or more required catalogue entries. Seven of the eleven receive the same class from both sources, and for eight the two values of $\mathcal{F}$ agree within a factor of two. The larger differences reflect catalogue parameters that predate the literature values adopted here. The catalogue uncertainty on $\dot{P}_b$ for J0737−3039A is more than two orders of magnitude larger than that of Kramer et al. (2021), which lowers its catalogue $\mathcal{F}$ tenfold. For B1913+16 the catalogue lists a smaller proper motion and a parallax-based distance of 4.2 kpc, where Table 1 uses 9.0 kpc from Weisberg & Huang (2016), but both give class D. For J1909−3744 the catalogue uses a parallax distance ($\sigma_d/d = 0.012$) rather than the $\dot{P}_b$-derived one, so the distance rather than the Galactic model becomes the dominant term. For systems in both samples we regard Table 1 as authoritative, and the catalogue computation as a uniform extension to systems not compiled by hand.

3.6 Stability of the classification: a synthetic population

The catalogue values of d and μ are measurements, so a system near a threshold may be classified differently by an observer using slightly different values. To quantify this across a population, and to examine whether the distribution of $\mathcal{F}$ contains natural boundaries, we generated a synthetic population from the 28-system catalogue sample.

Generation. We drew 2×10^5 systems by smoothed bootstrap resampling (Efron & Tibshirani, 1993). The resampling was done jointly in the base-10 logarithms of six parameters: distance, proper motion, orbital period, eccentricity, mass function, and $\sigma(\dot{P}_{\rm b}^{\rm obs})$. Each draw is a randomly chosen catalogue system, perturbed by a multivariate Gaussian kernel whose covariance is the sample covariance scaled by Scott's rule (Scott, 1992), and shrunk towards the sample mean so that the variance of the population matches that of the catalogue sample. This preserves the observed correlations between parameters. Each synthetic system inherits the sky position, σ_d/d, σ_μ/μ, and distance method of the catalogue system it was drawn from. Eccentricities outside $10^{-8} \leq e < 0.95$ were redrawn; five iterations sufficed, and no draws were lost. Component masses were assigned by sampling a pulsar mass and inclination and solving the mass function for the companion mass within companion-type limits; this succeeded for 90.8% of draws. Masses enter only the GR prediction used to select a radiative-test subset, not $\mathcal{F}$. For speed, $\mathcal{F}$ is evaluated in the first-order form of Eq. (8), with the Galactic-model term tabulated along each sightline and interpolated in distance.

For each system we computed the "true" $\mathcal{F}$ from its drawn d and μ, and an "observed" $\mathcal{F}$ after perturbing d log-normally with width $\ln(1 + \sigma_d/d)$ and μ normally with width σ_μ. This mimics an observer who measures the system once and classifies it from those measurements. A system is misclassified if the two values of $\mathcal{F}$ fall in different classes.

Distribution. The distribution of $\log_{10} \mathcal{F}$ has a single peak, with standard deviation 0.92 dex (Fig. 2). A kernel density estimate has no local minimum. Gaussian mixture models (Pedregosa et al., 2011) with $k = 1$–6 components, fitted to a random subsample of 5×10^4 values and compared with the Bayesian information criterion (BIC; Schwarz, 1978), prefer a single component: $\Delta\mathrm{BIC} \approx +140$ for $k = 2$ relative to $k = 1$. The distribution of $\mathcal{F}$ is therefore consistent with a single log-normal, and contains no natural gap between regimes. The thresholds at $\mathcal{F} = 0.3$ and 1 are conventions, not features of the population.

Class fractions and misclassification. With the conventional thresholds, 36.9% of synthetic systems are measurement-limited, 21.9% borderline, and 41.2% distance-limited. Restricting to the 56 834 systems (28.4%) in which the GR decay would be detectable at $\geq 3\sigma$, and a radiative test is therefore possible in principle, the distance-limited fraction rises to 58.0% (M 21.6%, B 20.3%). The overall misclassification rate is 7.9% (Fig. 3). It is low for measurement-limited (5.9%) and distance-limited (4.4%) systems and highest for borderline systems (17.9%), which lie closest to both thresholds. It depends strongly on the distance method: 4.0% for systems with parallax distances (64% of the population) and 14.9% for those with DM distances. The classification of a system with a good geometric distance is therefore robust, whereas that of a DM-distance system near a threshold should be treated as provisional.

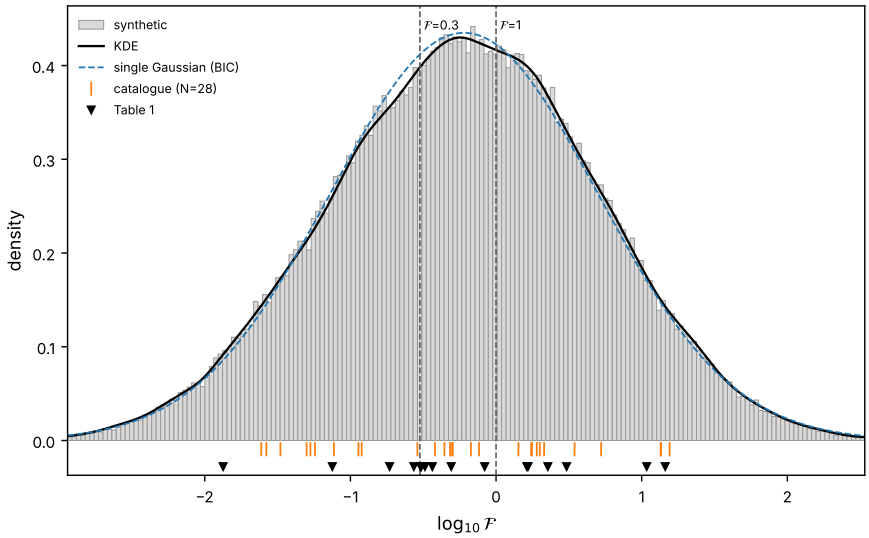


Figure 2. Distribution of $\log_{10}\mathcal{F}$ for 2×10^5 synthetic systems (histogram), with a kernel density estimate (black) and the BIC-preferred single-Gaussian model (dashed). Grey dashed lines mark the conventional thresholds $\mathcal{F}=0.3$ and 1. Orange ticks show the 28 catalogue systems of Table 2; black triangles show the sixteen literature systems of Table 1.

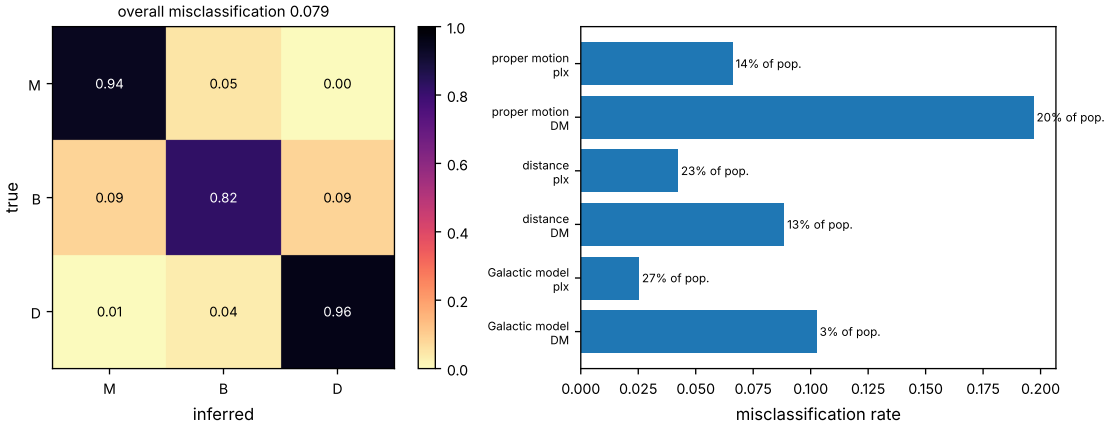


Figure 3. Left: confusion matrix between the class assigned from the true and from the noise-perturbed ("observed") $\mathcal{F}$ in the synthetic population, normalised by row (M, measurement-limited; B, borderline; D, distance-limited). Right: misclassification rate grouped by the dominant component of $\mathcal{F}$ and the distance method; labels give the fraction of the population in each group.

What drives $\mathcal{F}$. Among synthetic distance-limited systems, the dominant component is the distance for 49%, the Galactic model for 32%, and the proper motion for 19%; among measurement-limited systems, the proper-motion term is most often the largest of the (small) components (51%). The strongest rank correlation of $\mathcal{F}$ is with the measurement uncertainty $\sigma(\dot{P}_{\rm b}^{\rm obs})$ (Spearman $\rho=-0.64$), followed by the proper motion ($\rho=+0.34$) and the absolute Galactic latitude ($\rho=+0.19$); the correlations with distance ($\rho=-0.15$) and orbital period ($\rho=-0.13$) are weaker (Fig. 4). Whether a system is distance-limited is therefore governed first by how precisely its $\dot{P}_{\rm b}$ is measured, and then by the size of the kinematic term, not by the distance on its own.

We stress what the synthetic population can and cannot show. It is a smoothed resampling

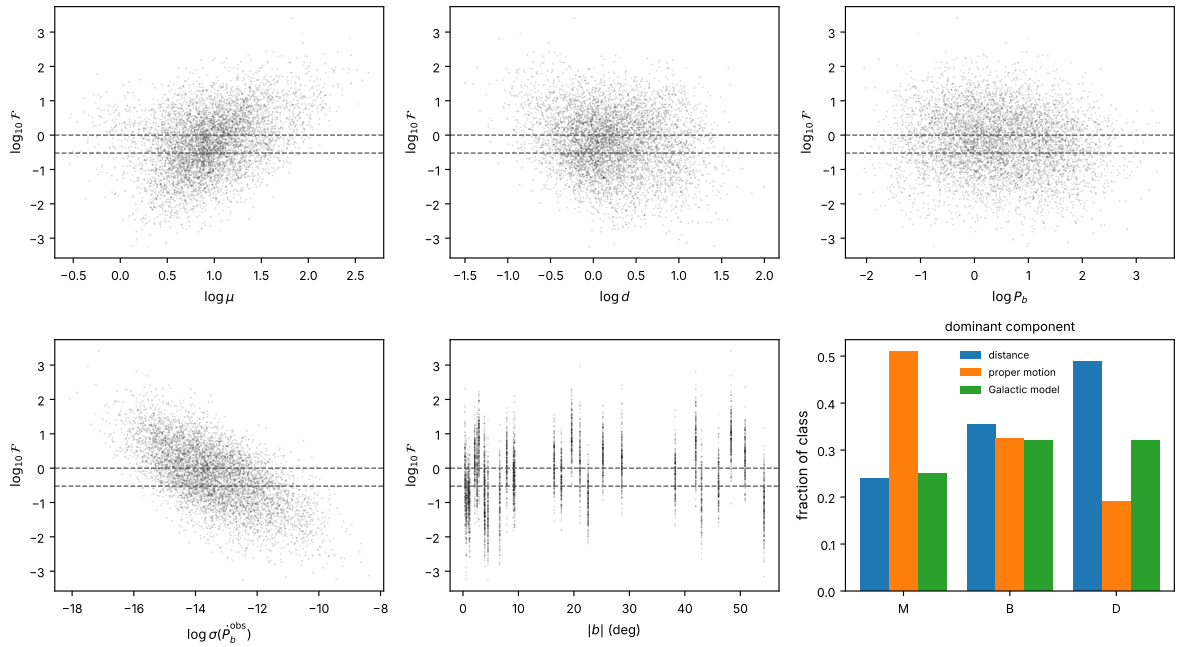


Figure 4. Dependence of $\log_{10}\mathcal{F}$ on its inputs in the synthetic population (6000 randomly selected systems): proper motion, distance, orbital period, measurement uncertainty $\sigma(\dot{P}_{\rm b}^{\rm obs})$, and absolute Galactic latitude. Dashed lines mark $\mathcal{F} = 0.3$ and 1. Bottom right: fraction of each class for which the distance, proper-motion, or Galactic-model component dominates.

of 28 real systems, not an independent population model, so it contains no information about binary pulsars unlike those already in the catalogue sample. Its value lies in the misclassification analysis, which requires many realisations of measurement noise, and in establishing that the distribution of $\mathcal{F}$ has no natural gap.

4 Discussion

4.1 Diagnostic use

Read as a diagnostic rather than a predictor, $\mathcal{F}$ and its components separate systems by the kind of improvement that would help. Systems with $\mathcal{F} \ll 1$ — J0348+0432 among them — gain nothing from astrometry or Galactic modelling and only from continued timing. For systems with $\mathcal{F} \gtrsim 1$, the lever identifies the investment.

- *Distance.* B1534+12, B1913+16, J1757−1854, J1933−6211, and J2145−0750 in the literature sample, and J0437−4715 and J1012+5307 in the catalogue, are limited mainly by σ_d/d. A more precise parallax — from VLBI or longer timing baselines — is the most productive measurement. For the DM-distance system J1757−1854, a first geometric distance would help most.

- *Proper motion.* Only J0751+1807 among the distance-limited systems is limited mainly by σ_μ/μ. Proper motions of the high-$\mathcal{F}$ systems are already measured to a small fraction of a per cent, so large μ makes the Shklovskii term large without making its uncertainty proper-motion-dominated.

- *Galactic model.* For J1909−3744, J1713+0747, J1600−3053, and J1618−3921, the dominant term is the disagreement between Galactic potential models; no improvement in astrometry or timing alone would move them below $\mathcal{F} = 1$ (Sect. 4.3).

This refines the view, common in individual timing papers, that high proper motion is what makes a test distance-limited. Large proper motion does make the kinematic term large, but the limiting uncertainty is usually the distance.

Several groups have noted that the radiative and dipolar-radiation tests from the leading NS–WD systems will soon be limited predominantly by distance uncertainty rather than by timing (Ding et al., 2020; Freire & Wex, 2024), and the kinematic distance method of Bell & Bailes (1996) continues to supply model-independent distances for the nearest systems (Verbiest et al., 2008). The components make that forecast system-specific. The synthetic population adds one qualification: for DM-distance systems near a threshold the class itself is uncertain at the 15% level, so the first investment for such a system may be a geometric distance, not because the test is known to be distance-limited but because without one it is unclear whether it is.

4.2 Two routes to distance limitation

B1534+12 and B1913+16 are both distance-limited, with $\mathcal{F} \approx 15$ and 3, respectively, but for different reasons. The two have comparable orbital periods of order hours, but proper motions of about 25 and 1.5 mas yr^{-1} (Ding et al., 2021; Weisberg & Huang, 2016). Through the μ^2 scaling of Eq. (4), that factor of seventeen becomes a factor of nearly three hundred in the Shklovskii term. For B1534+12, at 0.9 kpc and high Galactic latitude, the Shklovskii term dominates the kinematic budget. For B1913+16, 9 kpc away and 2° from the plane, the Shklovskii term is small but the Galactic term and its distance derivative are large. The first route requires a nearby, fast-moving system and occurs with good geometric distances; the second requires a distant system near the plane, where distances are usually from DM. This is why distance type does not control $\mathcal{F}$: both kinds of distance appear among the distance-limited systems, by different routes. The earlier observation that no DM-distance system is distance-limited holds only if the Galactic term is omitted.

4.3 When the Galactic model dominates

For systems whose largest component is $\mathcal{F}_{\rm mod}$, the measured $\dot{P}_{\rm b}$ constrains the Galactic potential along the line of sight more precisely than the two standard models agree with each other. For such systems the inference can be turned around, as in the kinematic-distance method of Bell & Bailes (1996): taking GR as given, $\dot{P}_{\rm b}$ measures the line-of-sight Galactic acceleration. Chakrabarti et al. (2021) used binary-pulsar accelerations in exactly this way to measure the vertical acceleration and mass density of the Galactic disc. The component split therefore identifies, from tabulated parameters alone, which systems are more informative about the Galaxy than about gravity. In our samples these are J1909−3744, J1713+0747, J1600−3053, and J1618−3921.

4.4 Caveats

Definition of $\mathcal{F}$. The model term $\sigma_{\rm mod}$ is the difference between two models, not a statistical uncertainty. It is conservative where one model is known to be inadequate, such as the flat-rotation-curve formula toward the inner Galaxy, and it is the most consequential assumption in the analysis. Omitting it changes the class of three literature systems (J1946+2052, J1738+0333, and J0737−3039A/B, all from B to M) and five catalogue systems (J1738+0333 and J1640+2224 from B to M, J1713+0747 from D to M, and J1600−3053 and J1618−3921 from D to B). $\mathcal{F}$ excludes the uncertainty in $\dot{P}_{\rm b}^{\rm GR}$ from the mass measurements, which contributes to the achieved test precision (Sect. 3.4) but not to the question of whether the kinematic correction or the measurement dominates. $\mathcal{F}$ is also a snapshot. For regularly sampled timing with white noise, $\sigma(\dot{P}_{\rm b}^{\rm obs})$ decreases as $T^{-5/2}$ with timing baseline T, whereas timing proper motions and parallaxes improve only as $T^{-3/2}$ and $T^{-1/2}$. $\mathcal{F}$ therefore grows with time, and systems drift towards distance limitation unless independent astrometry keeps pace.

Literature sample. The proper-motion uncertainties in Table B1 other than that of J1909−3744 are implied by the compiled parameters rather than transcribed from the original sources. Some distance uncertainties are symmetric approximations to asymmetric published intervals. Two distances are not independent of the timing: that of J1909−3744 is derived from its $\dot{P}_{\rm b}$, and the DM distance of B1534+12 is not independent of its timing distance because both electron-density models used that distance in their calibration (Ding et al., 2021). Distances from low-significance parallaxes are biased (Lutz & Kelker, 1973; Verbiest et al., 2012); among our systems this affects mainly B1913+16 and J1933−6211, whose catalogue parallaxes have signal-to-noise ratios below 5. $\mathcal{F}$ says nothing about systematic biases in $\dot{P}_{\rm b}^{\rm obs}$: J1756−2251 has a $\dot{P}_{\rm b}$ reported in tension with GR for reasons attributed to observational systematics (Ferdman et al., 2014), so it is not a clean test whatever its $\mathcal{F}$.

Catalogue sample. The catalogue inherits heterogeneous and sometimes outdated entries (Sect. 3.5). Its classification is robust to our two main choices. Setting σ_d/d for DM distances to 0.15 or 0.40 instead of 0.25 changes no class, and 0.50 changes one (J1949+3106, to D). Raising the parallax threshold from 2σ to 3σ changes none. The exclusion of cluster and non-degenerate-companion systems follows the same logic as the exclusion of J2043+1711, but it removes about half the systems with measured $\dot{P}_{\rm b}$.

Synthetic population. The population inherits all properties of its 28 parent systems, uses the first-order form of $\mathcal{F}$, and assumes Gaussian or log-normal measurement errors. Its class fractions reproduce those of the parent sample by construction.

With these qualifications, three results are robust because they hold in both samples and under all tested variations. B1534+12 is distance-limited through the Shklovskii term and B1913+16 through the Galactic term, and distance type does not by itself determine $\mathcal{F}$.

5 Conclusions

We ranked binary pulsars with published orbital-decay measurements by a figure of merit $\mathcal{F}$ comparing the full uncertainty of the kinematic correction with the measurement uncertainty

of the observed decay, and extended the ranking to the ATNF Pulsar Catalogue with an open-source pipeline. Our conclusions are as follows.

(i) Six of sixteen literature systems are distance-limited, six borderline, and four measurement-limited. Omitting the Galactic term understates the kinematic limitation for six systems; B1913+16 in particular becomes distance-limited, in agreement with its published error budget.

(ii) Systems become distance-limited by two routes: large proper motion at small distance, through the Shklovskii term, and large distance near the Galactic plane, through the Galactic term. Geometric and DM distances therefore both appear among distance-limited systems, and distance type alone does not determine $\mathcal{F}$.

(iii) For most distance-limited systems the dominant uncertainty is the distance, not the proper motion, so improved parallaxes are the most productive investment. For a few, including J1909−3744, it is the Galactic potential model; for these, $\dot{P}_b$ is more informative as a probe of the Galaxy than as a test of GR.

(iv) Among 28 field binaries in the ATNF catalogue, eleven are distance-limited, and J0437−4715 has the largest $\mathcal{F}$.

(v) $\mathcal{F}$ does not predict the achieved test precision, which is set by the absolute size of the dominant uncertainty rather than by the ratio.

(vi) Under realistic measurement noise, the class of 7.9% of systems changes — 4.0% with parallax distances but 14.9% with DM distances, and 18% of borderline systems. The distribution of $\mathcal{F}$ is consistent with a single log-normal, so the thresholds $\mathcal{F} = 0.3$ and 1 are conventions, not natural boundaries.

We therefore present $\mathcal{F}$ and its components as a per-system diagnostic. $\mathcal{F} \gg 1$ with a distance lever marks systems that need better parallaxes; a Galactic-model lever marks systems better used to probe the Galaxy; $\mathcal{F} \ll 1$ marks systems that benefit only from further timing; and DM-distance systems near a threshold need a geometric distance before the diagnosis is secure. No new physics is involved; the corrections are standard and the values are from the literature and the catalogue. Natural next steps are a statistical rather than two-model treatment of the Galactic-potential uncertainty, per-sightline DM-distance uncertainties, bias-corrected parallax distances, and forecasts of how $\mathcal{F}$ will evolve as timing baselines grow.

Acknowledgments

This work made use of the ATNF Pulsar Catalogue (Manchester et al., 2005), of PSRQPY (Pitkin, 2018), of GALPY (Bovy, 2015), and of NASA's Astrophysics Data System Bibliographic Services. The author is grateful to the pulsar-timing community whose published measurements made this compilation possible. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

The author declares no competing interests.

Data availability statement

The analysis pipeline, the parent sample, and all scripts used to produce the figures and tables are available at `https://github.com/muhammadanas-5/binary-pulsar-fom`. The catalogue query uses ATNF Pulsar Catalogue v2.8.1 (`https://www.atnf.csiro.au/research/pulsar/psrcat`). All results are reproducible with the random seeds stated in the code (0 for the per-system Monte Carlo, 42 for the synthetic population).

A Analysis pipeline

The pipeline consists of ten Python modules.

- `physics.py` implements Eqs. (1)–(5), (6) and the simplified figure of merit of Eq. (9), and inverts the mass function to assign companion masses.
- `galmodel.py` evaluates the line-of-sight Galactic acceleration in the McMillan (2017) potential with GALPY. It places the Sun at R_0 in the plane and projects the difference between the pulsar's and the Sun's acceleration onto the line of sight; its sign convention was checked against Eqs. (5)–(6), with which it agrees to about ten per cent for nearby, low-latitude systems.
- `budget.py` implements Eq. (8), its components, the Monte Carlo propagation, and the class probabilities. Because the full-potential evaluation is slow, the model term is tabulated on a logarithmic distance grid along each sightline and interpolated.
- `lit_inputs.py` holds the Table B1 inputs. It derives Galactic coordinates from the J2000 positions encoded in the pulsar names, which agree with catalogue coordinates to better than 0.1°.
- `catalogue.py` queries the catalogue, builds the parent sample (Sect. 3.5), and applies the exclusions (globular-cluster pulsars are identified by their cluster association or name), writing every excluded system and its reason to a file. Catalogue $\dot{P}_\mathrm{b}$ values are stored in heterogeneous units: entries with $|\dot{P}_\mathrm{b}| > 10^{-7}$ are interpreted as being in units of $10^{-12}\,\mathrm{s\,s^{-1}}$, following the TEMPO convention, and rescaled (three entries).
- `population.py` generates the synthetic population (Sect. 3.6).
- `analysis.py` contains the mixture-model scan, bootstrap, kernel density estimate, and misclassification matrix.
- `run_analysis.py` builds the catalogue sample and cross-checks it against Table 1; this cross-check is how the inconsistency in an earlier compilation of $\mathcal{F}$ for J1909−3744 was identified.

- `run_upgrade.py` evaluates the full figure of merit for both samples and the synthetic population, and writes all result tables, Figs. 1–4, and a machine-readable report.
- `make_tables.py` typesets Tables 1, 2, and B1 directly from the result files, so that the numbers in the paper are those of the code.

The full analysis runs in a few minutes on a laptop.

B Input parameters of the literature sample

Table B1. Input parameters of the sixteen literature systems of Table 1, in the same order. Numbers in parentheses are uncertainties in the last quoted digits. μ is in mas yr^{-1}, d in kpc, and $\dot{P}_{\rm b}^{\rm obs}$ in 10^{-12} s s^{-1}. WD, white-dwarf companion; NS, neutron-star companion. For J1909−3744, σ_μ/μ is from the ATNF catalogue v2.8.1. For the other systems it is the value implied by the proper-motion term of the simplified $\mathcal{F}$ computed from the compiled parameters; for J2145−0750 that term is negligible. *References.* (1) Weisberg & Huang (2016); (2) Kramer et al. (2021); (3) Ding et al. (2021); (4) Meng et al. (2025); (5) Ferdman et al. (2014); (6) Cameron et al. (2023); (7) Ferdman et al. (2020); (8) Freire et al. (2012); (9) Antoniadis et al. (2013); (10) Liu et al. (2020); (11) Ding et al. (2020); (12) Reardon et al. (2021); (13) Guo et al. (2021); (15) Stovall et al. (2019); (16) Haniewicz et al. (2021); (17) Geyer et al. (2023). Unranked: J1141−6545 (Bhat et al., 2008) and J2043+1711 (14; Fonseca et al., 2016), see Sect. 3.1.

#	System	comp	$P_{\rm b}$ (d)	μ	σ_μ/μ	d	$\dot{P}_{\rm b}^{\rm obs}$	dist. type	ref.
1	J0348+0432	WD	0.1024	5.35	0.0215	2.10(20)	−0.273(45)	photometric	9
2	J2234+0611	WD	32.001	27.1	0.0009	0.95(4)	+3.1(2.5)	parallax	15
3	J1913+1102	NS	0.2063	9.3	0.0971	7.14(1.5)	−0.48(3)	DM	7
4	J2222−0137	WD	2.4458	45.06	0.0011	0.267(1)	+0.251(8)	VLBI	13
5	J1946+2052	NS	0.0785	4.5	0.0446	3.5(1.0)	−1.8296(15)	DM	4
6	J1756−2251	NS	0.3196	2.42	0.0525	0.73(40)	−0.229(5)	parallax	5
7	J0737−3039A/B	NS	0.1023	3.31	0.0151	0.735(60)	−1.247920(78)	parallax	2
8	J1829+2456	NS	1.176	9.52	0.0065	1.2(3)	−0.029(12)	DM	16
9	J1738+0333	WD	0.3548	8.67	0.0023	1.47(10)	−0.0170(31)	parallax	8
10	J1012+5307	WD	0.6047	25.6	0.0040	0.83(5)	+0.061(5)	parallax	11
11	J1933−6211	WD	12.819	12.42	0.0083	1.6(3)	+0.7(1)	parallax	17
12	J2145−0750	WD	6.8389	13.15	—	0.50(15)	+0.13(2)	parallax	12
13	J1757−1854	NS	0.1835	4.4	0.0684	11.5(3.0)	−5.294(6)	unknown	6
14	B1913+16	NS	0.323	1.48	0.0276	9.0(2.0)	−2.423(1)	DM	1
15	J1909−3744	WD	1.5334	37.02	0.0048	1.158(3)	+0.5093(7)	$\dot{P}_{\rm b}$	10
16	B1534+12	NS	0.4207	25.29	0.0022	0.94(7)	−0.1366(3)	parallax	3

References

Antoniadis, J., Freire, P. C. C., Wex, N., et al. 2013, Science, 340, 1233232.

Astropy Collaboration. 2022, ApJ, 935, 167.

Bell, J. F., & Bailes, M. 1996, in Pulsars: Problems and Progress, ed. S. Johnston, M. A. Walker, & M. Bailes, ASP Conf. Ser. 105 (San Francisco: ASP), 209.

Bhat, N. D. R., Bailes, M., & Verbiest, J. P. W. 2008, Phys. Rev. D, 77, 124017.

Bovy, J. 2015, ApJS, 216, 29.

Cameron, A. D., Bailes, M., Champion, D. J., et al. 2023, MNRAS, 523, 5064.

Chakrabarti, S., Chang, P., Lam, M. T., Vigeland, S. J., & Quillen, A. C. 2021, ApJ, 907, L26.

Cordes, J. M., & Lazio, T. J. W. 2002, arXiv:astro-ph/0207156.

Damour, T., & Taylor, J. H. 1991, ApJ, 366, 501.
Ding, H., Deller, A. T., Fonseca, E., et al. 2021, ApJ, 921, L19.
Ding, H., Deller, A. T., Freire, P., et al. 2020, ApJ, 896, 85.
Edwards, R. T., Hobbs, G. B., & Manchester, R. N. 2006, MNRAS, 372, 1549.
Efron, B., & Tibshirani, R. J. 1993, An Introduction to the Bootstrap (New York: Chapman & Hall).
Ferdman, R. D., Freire, P. C. C., Perera, B. B. P., et al. 2020, Nature, 583, 211.
Ferdman, R. D., Stairs, I. H., Kramer, M., et al. 2014, MNRAS, 443, 2183.
Fonseca, E., Pennucci, T. T., Ellis, J. A., et al. 2016, ApJ, 832, 167.
Freire, P. C. C., & Wex, N. 2024, Living Rev. Relativ., 27, 5.
Freire, P. C. C., Wex, N., Esposito-Faresè, G., et al. 2012, MNRAS, 423, 3328.
Gautam, T., Freire, P. C. C., Batrakov, A., et al. 2022, A&A, 668, A187.
Geyer, M., Venkatraman Krishnan, V., Freire, P. C. C., et al. 2023, A&A, 674, A169.
GRAVITY Collaboration. 2021, A&A, 647, A59.
Guo, Y. J., Freire, P. C. C., Guillemot, L., et al. 2021, A&A, 654, A16.
Haniewicz, H. T., Ferdman, R. D., Freire, P. C. C., et al. 2021, MNRAS, 500, 4620.
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357.
Holmberg, J., & Flynn, C. 2004, MNRAS, 352, 440.
Hulse, R. A., & Taylor, J. H. 1975, ApJ, 195, L51.
Hunter, J. D. 2007, Comput. Sci. Eng., 9, 90.
Kramer, M., Stairs, I. H., Manchester, R. N., et al. 2021, Phys. Rev. X, 11, 041050.
Lazaridis, K., Wex, N., Jessner, A., et al. 2009, MNRAS, 400, 805.
Liu, K., Guillemot, L., Istrate, A. G., et al. 2020, MNRAS, 499, 2276.
Lutz, T. E., & Kelker, D. H. 1973, PASP, 85, 573.
Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993.
Martinez, J. G., Stovall, K., Freire, P. C. C., et al. 2015, ApJ, 812, 143.
McKinney, W. 2010, in Proc. 9th Python in Science Conf., ed. S. van der Walt & J. Millman, 56.
McMillan, P. J. 2017, MNRAS, 465, 76.
Meng, L., Freire, P. C. C., Stovall, K., et al. 2025, A&A, 704, A153.
Nice, D. J., & Taylor, J. H. 1995, ApJ, 441, 429.
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825.
Peters, P. C., & Mathews, J. 1963, Phys. Rev., 131, 435.
Pitkin, M. 2018, J. Open Source Softw., 3, 538.
Reardon, D. J., Shannon, R. M., Cameron, A. D., et al. 2021, MNRAS, 507, 2137.
Schwarz, G. 1978, Ann. Stat., 6, 461.
Scott, D. W. 1992, Multivariate Density Estimation: Theory, Practice, and Visualization (New York: Wiley).
Shklovskii, I. S. 1970, Soviet Ast., 13, 562.
Stovall, K., Freire, P. C. C., Antoniadis, J., et al. 2019, ApJ, 870, 74.
Stovall, K., Freire, P. C. C., Chatterjee, S., et al. 2018, ApJ, 854, L22.
Taylor, J. H., & Weisberg, J. M. 1982, ApJ, 253, 908.
Verbiest, J. P. W., Bailes, M., van Straten, W., et al. 2008, ApJ, 679, 675.
Verbiest, J. P. W., Weisberg, J. M., Chael, A. A., Lee, K. J., & Lorimer, D. R. 2012, ApJ, 755, 39.
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261.
Weisberg, J. M., & Huang, Y. 2016, ApJ, 829, 55.
Wex, N. 2014, in Frontiers in Relativistic Celestial Mechanics, Vol. 2, ed. S. M. Kopeikin (Berlin: de Gruyter), 39.
Will, C. M. 2014, Living Rev. Relativ., 17, 4.
Yao, J. M., Manchester, R. N., & Wang, N. 2017, ApJ, 835, 29.