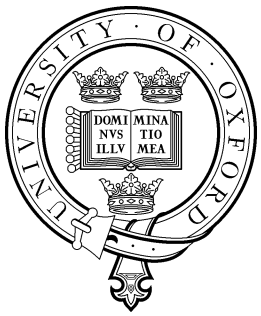
Fundamental Physics from Galaxies



Deaglan John Bartlett

Oriel College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2022

Acknowledgements

I am indebted to my supervisors Pedro Ferreira and Harry Desmond for their guidance and constant support and from whom I have learnt so much. I am extremely grateful for my family and friends, who do so much for me and have been a continued source of encouragement and support throughout my DPhil. I thank Jens Jasche for his collaboration for Chapters 2 and 5, Julien Devriendt and Adrianne Slyz for their invaluable input for Chapter 3, Andrija Kostić and Guilhem Lavaux for their work on Chapter 5, Dexter Bergsdal for his initial studies for Chapter 2, and Eliza Dickie for early contributions to Chapter 6. I would like to thank Harry Desmond for allowing me to use the `CSiBORG` simulations, which were created with the assistance of Julien Devriendt, Jens Jasche, Guilhem Lavaux and Adrianne Slyz. I am very grateful to Scott Barrows, Gabor Orosz and Sandor Frey for providing data necessary for Chapters 3 and 4. I would also like to thank David Alonso, Emilio Bellini, Céline Combet, Seth Digel, Scott Dodelson, Lam Hui, Moritz Hütten, Harley Katz, Eiichiro Komatsu, Kazuya Koyama, David Maurin, Samuel McDermott, Martin Rey, Jeremy Sakstein, Fabian Schmidt, Hans Winther and Miguel Zumalacarregui for useful input and discussions. I thank Jonathan Patterson for smoothly running the Glamdring Cluster hosted by the University of Oxford, where most of the data processing was performed.

I would like to thank the Horizon-AGN collaboration for allowing me to use the simulation data, and particularly Stephane Rouberol for smoothly running the Horizon Cluster hosted by the Institut d'Astrophysique de Paris where most of the processing of the raw simulation data was performed. Some of the numerical work also made use of the DiRAC Data Intensive service at Leicester, operated by the University of Leicester IT Services, which forms part of the STFC DiRAC HPC Facility (`www.dirac.ac.uk`). The equipment was funded by BEIS capital funding via STFC capital grants ST/K000373/1 and ST/R002363/1 and STFC DiRAC Operations grant ST/R001014/1. DiRAC is part of the National e-Infrastructure.

I am supported by STFC and Oriel College, Oxford. Projects in this thesis have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 693024). The work in Chapters 2 and 5 was done within the Aquila Consortium[1].

[1] https://www.aquila-consortium.org/

Abstract

Galactic-scale tests have proven to be powerful tools in constraining fundamental physics in previously under-explored regions of parameter space. In this thesis we use astrophysical systems to test some of the fundamental principles governing our current theories of the Universe, through the development of source-by-source, Monte Carlo-based forward models.

We consider modifications to the propagation of light by one of three effects: quantum gravity (QG), a non-zero photon mass and a violation of the Weak Equivalence Principle (WEP). We use spectral lag data of Gamma Ray Bursts from the BATSE satellite to constrain the photon mass to be $m_\gamma < 4.0 \times 10^{-5}\, h\, \mathrm{eV}/c^2$ and the QG length scale to be $\ell_{\mathrm{QG}} < 5.3 \times 10^{-18}\, h\ \mathrm{GeV}^{-1}$ at 95% confidence, WEP to $\Delta\gamma < 2.1 \times 10^{-15}$ at 1σ confidence between photon energies of 25 keV and 325 keV, and we demonstrate that these constraints are robust to how one models other contributions to the signal.

We investigate Galileon modified gravity theories by studying the offsets between the centre of a galaxy and its host supermassive black hole (BH). We constrain the Galileon coupling to be $\Delta G/G_{\mathrm{N}} < 0.16$ at 1σ confidence for Galileons with crossover scale $r_{\mathrm{C}} \gtrsim H_0^{-1}$. Inspired by the aforementioned test of modified gravity, we study spatially offset BHs in the Horizon-AGN simulation and compare these to observations, finding i) the fraction of spatially offset BHs increases with cosmic time, ii) BHs live on prograde orbits in the plane of the galaxy with an orbital radius that decays with time but stalls near $z = 0$, and iii) the magnitudes of offsets from the galaxy centres are substantially larger in the simulation than in observations.

By cross-correlating dark matter density fields inferred from the spatial distribution of galaxies with gamma ray data from the *Fermi* Large Area Telescope, marginalising over uncertainties in this reconstruction, small-scale structure and parameters describing astrophysical contributions to

the observed gamma ray sky, we place constraints on the dark matter annihilation cross-sections and decay rates. We rule out the thermal relic cross-section or s-wave annihilation for all $m_\chi \lesssim 7\,\mathrm{GeV}/c^2$ at 95% confidence if the annihilation produces Z bosons, gluons or quarks less massive than the bottom quark. We infer a contribution to the gamma ray sky with the same spatial distribution as dark matter decay at 3.3σ. Although this could be due to dark matter decay via these channels with a decay rate $\Gamma \approx 3 \times 10^{-28}\,\mathrm{s}^{-1}$, we find that a power-law spectrum of index $p = -2.75^{+0.71}_{-0.46}$ is preferred by the data.

Finally, we outline a framework for assessing the reliability of the methods used in this thesis by constructing and testing more advanced models using cosmological hydrodynamical simulations. As a case study, we use the Horizon-AGN simulation to investigate warping of stellar disks and offsets between gas and stars within galaxies, which are powerful probes of screened fifth-forces.

Contents

List of Figures

Chapter 1

Introduction

1.1 The Standard Models

The standard models of particle physics and cosmology have been remarkably successful in reproducing decades' worth of observations across a wide range of environments. The former is described by the language of quantum field theory; particles are excitations of these fields and can be classified into bosons and fermions based on their symmetry properties. Interactions between particles under three of the four known fundamental forces of nature (electromagnetism and the strong and weak nuclear forces) can be explained as exchanges of gauge bosons. The fourth "force" is currently understood as being distinct from the other three insofar as it is not really a force at all. Whereas particle physics can be understood as the evolution of particles which inhabit a spacetime, we understand gravity to be the curvature of spacetime itself.

For over a century we have described the interplay between the matter content of the Universe and the curvature of spacetime by General Relativity (GR). Ostensibly it is a very simple theory: matter travels through spacetime on geodesics and spacetime is curved due to the presence of matter. The motivation behind these statements is the universality of gravity, as captured by the principle of equivalence. There exist several formulations of the Equivalence Principle, of varying strengths. The Weak Equivalence Principle (WEP) states that all freely-falling test objects follow the same trajectories given the same initial conditions, irrespective of their composition or structure, i.e. the inertial mass and gravitational masses are equal for all objects. The Einstein Equivalence Principle (EEP) captures our inability to detect the existence of a gravitational field if restricted to a sufficiently small volume of spacetime; in such a volume the non-gravitational laws of physics are those of special relativity (SR). The Strong Equivalence Principle (SEP) takes this further and asserts that all the laws of

physics reduce to SR in a sufficiently small region of spacetime. These principles are highly suggestive of a geometric theory; if all free-falling particles move on geodesics then we can obey the WEP and the local SR-like behaviour can be attributed to local flatness of a manifold.

Indeed, by viewing spacetime as a manifold with a metric $g_{\mu\nu}$, one can construct the action for GR

$$S = \frac{1}{16\pi G}\int \mathrm{d}^4x\sqrt{-g}\,(R - 2\Lambda) + S_\mathrm{m}, \tag{1.1}$$

under the assumption that the action is comprised of scalars which depend only on the metric and its first and second derivatives. Here we have introduced Newton's gravitational constant, G, the cosmological constant, Λ, an action for matter, S_m, and the determinant of the metric, $g \equiv \det g_{\mu\nu}$. We also introduced the Ricci scalar, $R \equiv R^\mu{}_\mu$, where the Ricci tensor, $R_{\mu\nu}$, is defined in terms of the Riemann tensor, $R^\mu{}_{\nu\sigma\rho}$, as $R_{\mu\nu} \equiv R^\lambda{}_{\mu\lambda\nu}$, where

$$R^\mu{}_{\nu\sigma\rho} = \partial_\sigma\Gamma^\mu_{\rho\nu} - \partial_\rho\Gamma^\mu_{\sigma\nu} + \Gamma^\mu_{\sigma\lambda}\Gamma^\lambda_{\rho\nu} - \Gamma^\mu_{\rho\lambda}\Gamma^\lambda_{\sigma\nu}, \tag{1.2}$$

and we use the metric connection

$$\Gamma^\sigma_{\mu\nu} = \frac{1}{2}g^{\sigma\rho}\left(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu}\right). \tag{1.3}$$

Since our dynamical degree of freedom is the metric, we can extremise Equation 1.1 with respect to small variations of the metric to obtain the Einstein Field Equation

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.4}$$

where the Energy-Momentum Tensor is

$$T_{\mu\nu} = -2\frac{1}{\sqrt{-g}}\frac{\delta S_\mathrm{m}}{\delta g^{\mu\nu}}. \tag{1.5}$$

If one knows the matter content of the Universe (i.e. $T_{\mu\nu}$), then by solving Equation 1.4, one also knows the geometry of spacetime (or vice versa). This equation has passed every experimental test it has thus far encountered: the perihelion precession of Mercury, the bending of light rays in a gravitation field, gravitational redshift, the Shapiro time delay, the generation of gravitational waves. One can broadly classify these phenomena into one of three regimes: small scales (laboratory or Solar System, e.g. lunar laser ranging [10, 11]), cosmological scales (e.g. cosmic microwave background [12]), or the strong-field regime (e.g. gravitational waves from binary black holes [13]). Let us now consider the cosmological regime to see what form $T_{\mu\nu}$ can take in order to explain the evolution of and the formation of structure in the Universe.

1.1.1 The concordance cosmological model

The concordance cosmological model (ΛCDM) has been the best-fitting model for over 20 years. As with many descriptions of the Universe on cosmological scales, it assumes the Copernican Principle; on sufficiently large scales the Universe is spatially homogeneous and isotropic. Informally, this is the notion that there is no special place in the Universe. We can therefore construct spacelike slices of our four-dimensional spacetime such that our metric is the Friedmann–Lemaître–Robertson–Walker metric [14–20]

$$ds^2 = -\mathrm{d}t^2 + a^2(t)\,\gamma_{ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{1.6}$$

where γ_{ij} is a maximally symmetric three-dimensional metric, t is our time coordinate and $a(t)$ is the scale factor. Throughout this thesis we will only consider flat universes such that

$$ds^2 = -\mathrm{d}t^2 + a^2(t)\left(\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2\right), \tag{1.7}$$

where r is a radial coordinate and the metric of the two-sphere is $\mathrm{d}\Omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2$ in usual polar coordinates. One can also introduce the conformal time, η, such that $\mathrm{d}t^2 = a^2(\eta)\,\mathrm{d}\eta^2$.

For simplicity, let us begin by considering a universe comprised of a single perfect fluid. A perfect fluid's Energy-Momentum Tensor can be expressed in terms of its four-velocity, u_μ, as

$$T_{\mu\nu} = (\rho + p)\,u_\mu u_\nu + p g_{\mu\nu}, \tag{1.8}$$

where ρ and p are the rest-frame energy density and isotropic pressure, respectively. Assuming a simple equation of state $p = w\rho$ for constant w, conservation of $T_{\mu\nu}$ leads to

$$\nabla_\mu T^{\mu\nu} = 0 \implies \rho \propto a^{-3(1+w)}. \tag{1.9}$$

Substituting this into Equation 1.4, we obtain the Friedmann equations [14]

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho, \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3p), \tag{1.10}$$

where an overdot refers to a derivative with respect to t. For universes with multiple components we sum over the right hand side of these equations. Defining the Hubble parameter, H, to be H_0 (the Hubble constant) today, where we set $a = 1$, the first of these equations can be written as

$$\left(\frac{H}{H_0}\right)^2 = \Omega_\mathrm{r} a^{-4} + \Omega_\mathrm{m} a^{-3} + \Omega_\Lambda, \tag{1.11}$$

where we have two fluids - radiation ($w_{\rm r} = 1/3$) and matter ($w_{\rm m} = 0$) - for which we define

$$\Omega_i \equiv \frac{8\pi G}{3H_0^2}\rho_i, \tag{1.12}$$

and a non-zero cosmological constant, which can be considered as an additional component of $T_{\mu\nu}$ with $w_\Lambda = -1$ and density

$$\rho_\Lambda = \frac{\Lambda}{8\pi G}. \tag{1.13}$$

The addition of curvature would give another term on the right hand side of Equation 1.11 of the form $\Omega_\kappa a^{-2}$.

The end of the 20$^{\rm th}$ century saw the startling discovery that the Universe's expansion is accelerating [21, 22]. Due to the success of GR, one potential explanation is that this acceleration is due to a negative pressure component of stress-energy: dark energy [23]. Our current cosmological model (ΛCDM) assumes this takes the form of a cosmological constant, hence the introduction of the Ω_Λ component above. For large scale factors, one sees that Equation 1.11 is solved by an era of exponential expansion, and hence our Universe is accelerating and becoming asymptotically de Sitter.

Of course, the real Universe is not perfectly homogenous or isotropic and thus we should add perturbations [24] to Equation 1.7 in order to determine how structure forms. We can write our perturbed metric as [25]

$$ds^2 = a^2(\eta)\left(-(1+2A)\,{\rm d}\eta^2 + 2B_i {\rm d}x^i {\rm d}\eta + (\delta_{ij} + h_{ij})\,{\rm d}x^i {\rm d}x^j\right). \tag{1.14}$$

Considering scalar perturbations only and working in the Newtonian gauge, this becomes [26]

$$ds^2 = a^2(\eta)\left(-(1+2\Psi)\,{\rm d}\eta^2 + (1-2\Phi)\,\delta_{ij}{\rm d}x^i {\rm d}x^j\right). \tag{1.15}$$

Given scalar perturbations $\delta\rho_i$ to the background densities ρ_i, we define $\delta_i \equiv \delta\rho_i/\bar{\rho}_i$, and the peculiar velocities of these fluid perturbations are defined to be $\vec{v}_i \equiv \vec{\nabla} v_i$. To linear order, and assuming that all components are perfect fluids, the scalar perturbations evolve as [26]

$$\Phi' = -\frac{a'}{a}\Phi - \frac{1}{2}a^2\sum_i (1+w_i)\rho_i v_i, \tag{1.16}$$

$$\delta_i' = (1+w_i)\left(3\Phi' + v_i k^2\right), \tag{1.17}$$

$$v_i' = 3\frac{a'}{a}\left(w_i - \frac{1}{3}\right)v_i - \left(\Phi + \frac{w_i}{1+w_i}\delta_i\right), \tag{1.18}$$

where a prime denotes a derivative with respect to conformal time and k is the wavenumber for the perturbations. One can also express the potentials algebraically in terms of the other components via

$$\Phi = \Psi = \frac{1}{2k^2} \sum_i \left(3\frac{\dot{a}}{a} (1 + w_i) v_i - \delta_i \right) a^2 \rho_i, \tag{1.19}$$

which looks like Poisson's equation, with an additional velocity-dependent term. We take the fluid index i to range over $i \in \{\mathrm{r}, \mathrm{m}, \Lambda\}$ for radiation $w_\mathrm{r} = \frac{1}{3}$, matter $w_\mathrm{m} = 0$, and dark energy $w_\Lambda = -1$ respectively, but in general only perturb the first two of these. In reality one should describe photons via a distribution function in photon momentum and not as a perfect fluid, and consider interactions between photons and baryons, however for the remainder of the thesis we will neglect the contribution from radiation and thus ignore this detail.

Now that we have the equations of motion which govern the evolution of species in the Universe, one needs to provide the initial conditions. From observations of the Cosmic Microwave Background (CMB) [12], it is known that the Universe closely resembled a Gaussian Random Field in the distant past (such a density distribution can be achieved through a short period of accelerated expansion; inflation). Therefore, one can say that the density fluctuations at high redshift (δ^I, where I stands for initial conditions (IC)) is (approximately) Gaussian. This is then fully described by the matter power spectrum, $P(k)$, which defines the statistics of density fluctuations

$$\langle \Delta(\boldsymbol{k}) \Delta^*(\boldsymbol{k}') \rangle \equiv (2\pi)^3 P(k) \delta^\mathrm{D}(\boldsymbol{k} - \boldsymbol{k}'), \tag{1.20}$$

where δ^D is the Dirac delta function. Note that the process of structure formation will make the present day density field non-Gaussian (e.g. filaments are typically associated with higher order statistics). Given some initial form of $P(k)$, one can solve the perturbation equations in order to predict its form today. From inflationary theory, one expects the initial power spectrum, $P_\mathcal{R}(k)$, of the comoving curvature perturbation, $\mathcal{R}$, to be nearly scale invariant

$$P_\mathcal{R}(k) = A_\mathrm{s} \left(\frac{k}{k_\star} \right)^{n_\mathrm{s}-1}, \tag{1.21}$$

where $k_\star$ is a reference wavenumber, and, in the Newtonian gauge,

$$\mathcal{R} = -\Phi + \mathcal{H} v, \tag{1.22}$$

for velocity perturbation v and $\mathcal{H} \equiv a'/a$.

Until now we have not attempted to distinguish between different types of matter. Measurements such as the rotation curves of galaxies [27], the CMB [12], gravitational lensing and large scale structure [28] demonstrate that the total matter budget of our Universe must contain other particles than those in the Standard Model (SM) (referred to as baryonic matter). In fact, we find that baryonic matter is sub-dominant: its cosmological mean density is approximately five times smaller than that of dark matter (DM) [12]. Furthermore, one must be careful in how one treats baryons since neutrinos behave very differently to other SM particles.

The current cosmological model is described by only six parameters: the baryonic, $\Omega_{\rm b}$, and DM, $\Omega_{\rm c}$, density parameters, the Hubble constant, H_0, the scalar spectral index, $n_{\rm s}$, the curvature fluctuation amplitude, $A_{\rm s}$, and the reionisation optical depth, τ. All other parameters can be derived from these six, and thus sometimes a different set of parameters is chosen. For example, instead of $A_{\rm s}$, one often quotes σ_8 which is the root-mean-square density fluctuation when the field is smoothed with a top-hat filter of radius $8\,h^{-1}$Mpc (and $H_0 = 100h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$).

1.1.2 Extensions to the Standard Models

Despite the phenomenal success of ΛCDM in describing individual observations independently, when one compares different cosmological probes we find that the inferred parameters do not always agree. For example, the inferred value of H_0 from local observations does not match that from the CMB (the Hubble tension) [29] and the value of σ_8 obtained from weak lensing does not agree with that from the CMB [30]. Moreover, there exists small-scale observational inconsistencies such as the cusp-core or missing satellites problems [31, 32]. Of course, these problems could be due to unknown systematics or simply a lack of understanding of the complicated baryonic processes which also affect our measurements and analyses. However, there are also theoretical challenges to the current model, most notably that almost the entire energy density of the Universe is comprised of two components which do not appear in the SM (dark matter and dark energy) and the realisation that dark energy has issues with radiative instability and UV-sensitivity (the cosmological constant problem [33]). These considerations suggest that altering some of the assumptions of ΛCDM or the SM could produce a model which more faithfully reproduces the observed Universe. In this section we will discuss some of these possibilities.

1.1.2.1 Theory of Gravity

Although introduced in Section 1.1.1 as a geometric theory, GR can equivalently be described as the unique Lorentz-invariant, low-energy theory of a massless spin-2 particle [34]. Given our inability to successfully quantise this theory, it is not unreasonable to assume that GR is merely an effective field theory of some UV-complete description of gravity, and thus, although it performs exceptionally in its range of validity, this range is limited. Moreover, given the observational and theoretical challenges of GR, perhaps GR is also not the correct description of gravity on large scales. In either case, it is interesting to see what the observational impact of breaking the assumptions of GR is.

Modified gravity theories generically violate the SEP, which refers to the equivalence of free fall independent of an object's gravitational binding energy or composition [35]. Despite little success so far in explaining cosmic acceleration [36], studies of modified gravity are essential as they provide consistent, plausible alternatives to GR and could relieve some of the current tensions (e.g. [37–40]).

On astrophysical scales, modified gravity theories lead to a fifth fundamental force [41]. This force is not detected in laboratory or Solar System experiments, so must be 'screened' to remain hidden in these environments. To understand such screening mechanisms, we will follow the approach of [42] and consider a scalar field, ϕ, which has a background value ϕ_0 and a perturbation $\varphi = \phi - \phi_0$. Assuming a locally flat spacetime, one would consider a Lagrangian

$$\mathcal{L} = \frac{Z(\phi_0)}{2}\partial_\mu\varphi\partial^\mu\varphi + \frac{m_\phi^2(\phi_0)}{2}\varphi^2 - \delta g_{\mu\nu}\delta T^{\mu\nu}, \tag{1.23}$$

to second order in φ. The metric perturbation is

$$\delta g_{\mu\nu} = \frac{\beta(\phi_0)}{m_{\rm pl}}\varphi\eta_{\mu\nu} - \gamma(\phi_0)\,\partial_\mu\partial_\nu\varphi + \delta(\phi_0)\,\partial_\mu\varphi\partial_\nu\varphi + \ldots \tag{1.24}$$

For generality, in our discussion we let Z, m_ϕ, β, γ and δ be arbitrary functions of the background field, but for a given theory one can obtain these from the unperturbed Lagrangian by expanding about a background value, ϕ_0. Since we interpret m_ϕ as the effective mass of the field, we see that φ has a mass dimension equal to one. Limiting ourselves to the case of a static background and static matter overdensity $\delta T^{00} = \rho$, we have

$$K^{\mu\nu}(\phi_0)\,\partial_\mu\partial_\nu\varphi - m^2\varphi = -\frac{\beta(\phi_0)}{m_{\rm pl}}\delta T, \tag{1.25}$$

where

$$K^{\mu\nu}(\phi_0) = Z(\phi_0)\,\eta^{\mu\nu} - 2\delta(\phi_0)\,\delta T^{\mu\nu}. \tag{1.26}$$

This looks like a generalised Poisson equation,

$$Z^{ij}\partial_i\partial_j\varphi + \tilde{m}^2\varphi = 8\pi G_{\rm N}\alpha\rho, \tag{1.27}$$

where ρ is the matter density, $G_{\rm N}$ is Newton's constant, Z^{ij} is a generalised kinetic coefficient, $\tilde{m}$ is the mass of the scalar and α the strength of its coupling to matter. This scalar then produces an additional acceleration (the fifth force), which leads to an effective enhancement of the gravitational force,

$$G_{\rm N} \to G_{\rm N}\left(1 + \frac{\Delta G}{G_{\rm N}}\right) = G_{\rm N}\left(1 + 2\alpha^2 e^{-mr}\right). \tag{1.28}$$

In the absence of a screening mechanism, the parameters α and m must be tuned to very small values in order to obey Solar System and laboratory tests of gravity [43] (see [36, 44–46] for reviews of screened modified gravity theories). Screening dynamically suppresses the kinetic, mass or coupling terms in Equation 1.27 by allowing them to depend on the background scalar field, resulting in either 'kinetic' (e.g. K-mouflage [47] and Vainshtein [48]) or 'thin-shell' (e.g. chameleon [49, 50], symmetron [51] and dilaton [52]) screening. An alternative mechanism utilises DM-baryon interactions to produce a DM density-dependent gravitational constant [53].

In 'thin-shell' screening, if the mass of the scalar field, $m\left(\phi_0\right)$, increases sharply within a massive object, then the scalar field sourced by this mass is Yukawa suppressed. Instead, only a thin shell of matter near the surface of an object can source the scalar field in an appreciable manner. If the thin shell has mass ΔM and M is the total mass of the object, this is equivalent to multiplying $\beta\left(\phi_0\right)$ by $\Delta M/M$ in the equation of motion. Provided $\Delta M/M \ll 1$, the scalar field is suppressed. For the case of a modified Poisson equation, φ can be related to the Newtonian potential, Φ, as

$$\frac{\varphi}{m_{\rm pl}} = \frac{2\beta\left(\phi_0\right)}{Z\left(\phi_0\right)}\Phi, \tag{1.29}$$

so the scalar field is suppressed compared the Newtonian potential (and therefore screened) if

$$|\Phi| \geq \frac{\phi_0}{2m_{\rm pl}\beta\left(\phi_{\rm out}\right)}, \tag{1.30}$$

since in a sufficiently dense object $\varphi \approx -\phi_0$, and we have normalised our field such that $Z = 1$. Therefore, an object is thin-shell screened if its Newtonian potential is sufficiently large. Note that the mass of the scalar field is not just a function of density, but of the background scalar field. In this way, the field can be screened in regions of large Newtonian potential, even if the matter density is zero (e.g. throughout the Solar System).

The 'kinetic' mechanisms utilise the observation that if Z is sufficiently large, then the coupling α will be very small. Since Z is merely a constant at leading order, for this mechanism we consider higher order terms

$$Z = 1 + a\left(\phi_0\right) r_{\rm c}^2 \frac{\Box\varphi}{m_{\rm pl}} + b\left(\phi_0\right) \frac{\left(\partial\varphi\right)^2}{\tilde{\Lambda}^4} + c\left(\phi_0\right) \frac{\Box^2\varphi}{\tilde{\Lambda}^5} + \ldots, \tag{1.31}$$

where $r_{\rm c}$ is the cross-over scale and $\tilde{\Lambda}$ is the strong coupling scale. If the term proportional to $\Box\varphi$ dominates, then we have a field that is Vainshtein screened, such that the field is screened when

$$\left|\nabla^2\Phi\right| \geq \frac{1}{2\beta\left(\phi_0\right) r_{\rm c}^2}, \tag{1.32}$$

i.e. when the spatial curvature is sufficiently large. The K-mouflage mechanism is similar but is relevant when the $\left(\partial\varphi\right)^2$ term dominates, with the screening criterion

$$\left|\nabla\Phi\right| \geq \frac{\Lambda^2}{2\beta\left(\phi_0\right) m_{\rm pl}}. \tag{1.33}$$

It is clear that one has a similar screening effect when the $\Box^2\varphi$ term dominates, but this is governed by $\nabla^4\Phi$. In all these cases the screening mechanism is due to the important effects of higher order operators, and therefore in some cases this may be beyond the validity of an effective field theory approach.

In this thesis we will test one example of both of these screening mechanisms. In Chapter 4 we will consider Galileon theories: scalar-tensor theories which obey the Galileon symmetry $\varphi \to \varphi + b + c_\mu x^\mu$ and are Vainshtein screened. When discussing how to calibrate galactic-scale tests of fundamental physics in Chapter 6, we will use $f(R)$ gravity as a case study. In these theories one replaces R in Equation 1.1 by a generic function of the Ricci scalar, which is equivalent to a thin-shell screened scalar-tensor theory.

1.1.2.2 The Nature of Dark Matter

The actual particle nature of DM is not yet known, despite it having five times the average density of baryonic matter. Theoretically favourable candidates are weakly interacting massive particles (WIMPs) [54–56], including, but not limited to, the lightest supersymmetric particle in supersymmetric theories. In particular, in Chapter 5 we will consider particles of mass m_χ which are able to annihilate, with a thermally averaged cross-section $\langle\sigma v\rangle$. If these WIMPs are thermal relics of the early Universe,

then one can compute $\langle \sigma v \rangle$ from the current abundance of DM. We outline the approximate calculation below, similar to as found in [57]; see [6] for a more careful treatment.

As previously discussed, if DM was a perfect fluid with $w = 0$, then the number density of DM, n, would depend on the scale factor, a, as $n \propto a^{-3}$. Let us now suppose that DM can self-annihilate to produce SM particles. If this can occur, then the reverse reaction is also possible, such that the rate of DM depletion can be expressed as

$$\frac{\mathrm{d}n}{\mathrm{d}t} + 3Hn = \langle \sigma v \rangle \left(n_{\mathrm{eq}}^2 - n^2 \right), \tag{1.34}$$

where n_{eq} is the number density of DM at chemical equilibrium. For high annihilation rates, n must approach n_{eq}, whereas we have the $w = 0$ behaviour if the annihilation rate is low. The transition between these regimes will occur when $\langle \sigma v \rangle n^2 \sim Hn$. This crossover is named 'freezeout' and happens when the characteristic time scale for a DM annihilation to occur is of order the Hubble time.

When in equilibrium, one can approximately express the temperature, T, dependence of n_{eq} as

$$n_{\mathrm{eq}} \sim \begin{cases} (m_\chi T)^{\frac{3}{2}} e^{-m_\chi / T} & , \quad T \ll m_\chi \\ T^3 & , \quad T \gg m_\chi. \end{cases} \tag{1.35}$$

One therefore sees that freezeout for a non-relativistic DM particle will occur when

$$(m_\chi T)^{\frac{3}{2}} e^{-m_\chi / T} \langle \sigma v \rangle \sim H. \tag{1.36}$$

Assuming that this occurs during radiation domination, then one can approximate $H \approx H(x = 1) x^{-2}$, where $x \equiv m_\chi / T$. One therefore sees that freezout must occur at $x = x_{\mathrm{f}}$, where

$$x_{\mathrm{f}}^{1/2} \exp(-x_{\mathrm{f}}) \sim \frac{H(x = 1)}{m_\chi^3 \langle \sigma v \rangle}, \tag{1.37}$$

which one can solve to obtain the freezeout temperature, T_{f}, at a given m_χ and $\langle \sigma v \rangle$.

One can also estimate x_{f} from the current DM density. If n_γ denotes the photon number density, then one can write

$$\frac{m_\chi n_{\mathrm{eq}}}{n_\gamma (x_{\mathrm{f}})} \sim \frac{m_\chi H(x = 1) x_{\mathrm{f}}^{-2}}{(m_\chi / x_{\mathrm{f}})^3 \langle \sigma v \rangle} = \frac{H(x = 1) x_{\mathrm{f}}}{m_\chi^2 \langle \sigma v \rangle}, \tag{1.38}$$

where we have used the relativistic equation for the photon number density. This is simply the ratio of the mass density of DM to the number density of photons. Since both are proportional to a^{-3} after freezeout, this should match the current value, and thus we have a second expression for x_{f} at given m_χ and $\langle \sigma v \rangle$.

For both of the above relations to hold, we require $x_\mathrm{f} \sim 25$, and hence $\langle\sigma v\rangle \sim 10^{-26}\,\mathrm{cm^3 s^{-1}}$ if m_χ is a few hundred GeV/c^2. The canonically quoted value is $\langle\sigma v\rangle_\mathrm{th} = 3\times10^{-26}\,\mathrm{cm^3 s^{-1}}$ [6] (the thermal relic cross-section), and more careful calculation finds that this is relatively independent of mass.

This is an interesting result, since if the annihilation occurred due to an interaction with coupling α_χ, then one would expect $\langle\sigma v\rangle \sim \alpha_\chi^2/m_\chi^2$. The thermal relic cross-section at these masses suggests that $\alpha_\chi \sim 0.01$, which is suspiciously similar to the electroweak coupling of the SM.

1.2 Statistical Methods

1.2.1 Bayesian statistics

There are broadly two goals of cosmological data analysis. Firstly, if one has a physical model, $\mathcal{M}_1$, which is described by some parameters, $\boldsymbol{\theta}$, then one wishes to compare such a physical model to data, $\mathcal{D}$, to determine which parameters best describe the data. In essence, one wishes to determine the posterior distribution $\mathcal{P}(\boldsymbol{\theta}|\mathcal{D})$, i.e. the probability distribution for $\boldsymbol{\theta}$ given the observed data. However, our physical model $\mathcal{M}_1$ does not directly give us this distribution. Instead, we are often in a slightly different situation where one has a model to predict how likely a set of observations are given $\mathcal{M}_1$ and $\boldsymbol{\theta}$: $\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})$ (the likelihood model). To determine the posterior from the likelihood, one must apply Bayes' theorem

$$\mathcal{P}(\boldsymbol{\theta}|\mathcal{D}) = \frac{\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})\,P(\boldsymbol{\theta})}{\mathcal{Z}(\mathcal{D})}, \tag{1.39}$$

where we have introduced the prior $P(\boldsymbol{\theta})$ to describe our *a priori* knowledge of the probability of $\boldsymbol{\theta}$. The remaining term in Equation 1.39, $\mathcal{Z}$, provides the normalisation for the posterior

$$\mathcal{Z}(\mathcal{D}) = \int \mathcal{L}(\mathcal{D}|\boldsymbol{\theta})\,P(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}, \tag{1.40}$$

and is known as the Bayesian evidence.

The second purpose of an analysis is to compare two or more different physical models, e.g. $\mathcal{M}_1$ and $\mathcal{M}_2$, to ascertain which is a better description of the data. In the notation of Equations 1.39 and 1.40 we have suppressed the dependence on our physical model, e.g. $\mathcal{Z}(\mathcal{D})$ is actually $\mathcal{Z}(\mathcal{D}|\mathcal{M}_1)$. Again, this is not what we are interested in, but rather the probability of a physical model given data, $\mathcal{P}(\mathcal{M}_1|\mathcal{D})$. We should therefore apply Bayes' theorem again to obtain

$$\mathcal{P}(\mathcal{M}_1|\mathcal{D}) = \frac{\mathcal{Z}(\mathcal{D}|\mathcal{M}_1)\,\varpi(\mathcal{M}_1)}{Z(\mathcal{D})}, \tag{1.41}$$

where now the denominator is independent of the physical model and we have introduced a prior for this model, $\varpi(\mathcal{M}_1)$. If one assumes that two physical models are equally likely *a priori* ($\varpi(\mathcal{M}_1) = \varpi(\mathcal{M}_2)$), then by computing the Bayes factor

$$\log K = \log \mathcal{Z}(\mathcal{D}|\mathcal{M}_1) - \log \mathcal{Z}(\mathcal{D}|\mathcal{M}_2), \tag{1.42}$$

one can determine which physical model is preferred by the data ($\mathcal{M}_1$ is preferred if $\log K > 0$).

Although this is the correct way to compare physical models within a Bayesian framework, there is a clear problem with this method if we do not have a well motivated $P(\boldsymbol{\theta})$. By design, to calculate the Bayesian evidence and thus the Bayes factor, one must integrate over all parameter space (Equation 1.40), such that we do not have to choose a set of parameters at which to calculate the posterior. Suppose that we set a deliberately wide, uniform prior on our parameters, $\boldsymbol{\theta}$. If this is much wider than the posterior, then if we make our prior twice as wide for each element of $\boldsymbol{\theta}$, $\mathcal{P}(\boldsymbol{\theta}|\mathcal{D})$ and its summaries will remain approximately unchanged, however $\log \mathcal{Z}(D)$ will be smaller by $\mathcal{K} \log 2$, if $\boldsymbol{\theta}$ has length $\mathcal{K}$. Now, if we only made this transformation for one of our physical models, we can artificially change the sign of $\log K$ and thus change which physical model is preferred. In all analyses in this thesis we are in this situation, and we will thus rely on 'less Bayesian' model comparison techniques.

1.2.2 Model comparison

To avoid producing arbitrary complex models, in any form of model comparison there should be penalty for models with a large number of parameters. Throughout this thesis, we often use the Bayesian Information Criterion (BIC) [58]

$$\mathrm{BIC} = \mathcal{K} \log \mathcal{N} - 2 \log \hat{\mathcal{L}}, \tag{1.43}$$

where $\mathcal{K}$ is the number of model parameters, we have $\mathcal{N}$ data points, and $\hat{\mathcal{L}}$ is the maximum likelihood estimate, at which $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. To demonstrate why this is appropriate, let us perform the integral in Equation 1.40 using Laplace's method

$$\begin{aligned} \mathcal{Z}(\mathcal{D}) &= \int \exp\left(\log\left(\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})\, P(\boldsymbol{\theta})\right)\right) \mathrm{d}\boldsymbol{\theta} \\ &\approx \int \exp\left(\log \hat{\mathcal{L}} + \frac{1}{2}\left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)^{\mathrm{T}} \mathcal{H} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right)\right) \mathrm{d}\boldsymbol{\theta} = \hat{\mathcal{L}} \sqrt{\frac{(2\pi)^{\mathcal{K}}}{|\det(\mathcal{H})|}}, \end{aligned} \tag{1.44}$$

where $\mathcal{H}_{ij} = \partial_i \partial_j \left[\log\left(\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})\, P(\boldsymbol{\theta})\right)\right]|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$. If we now assume a uniform prior and $\mathcal{N} \gg 1$, we see that $\mathcal{H}$ is the second derivative of $\mathrm{E}[\mathcal{N} \log p]$, where p is the individual

likelihood for each of our $\mathcal{N}$ independent data points. Hence, $|\det(\mathcal{H})| = \mathcal{N}^{\mathcal{K}} |\det(\mathcal{I})|$, where $\mathcal{I}$ is the Fisher information for a single data point, which, if the likelihood for a given data point is f, is defined to be

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta_i}\log f\right)\left(\frac{\partial}{\partial\theta_j}\log f\right)\middle|\,\boldsymbol{\theta}\right], \tag{1.45}$$

where E denotes the expectation value. $\mathcal{I}$ is a measure of the information content an observable random variable has about the parameters describing the distribution from which the observable is drawn.

For $\mathcal{N} \gg 1$, we therefore find

$$\log \mathcal{Z}(\mathcal{D}) \approx \log \hat{\mathcal{L}} - \frac{\mathcal{K}}{2}\log N = -2 \times \mathrm{BIC}, \tag{1.46}$$

to leading order in $\mathcal{N}$. Therefore, (minus half) the BIC approximates the logarithm of the Bayesian evidence for large $\mathcal{N}$, but is independent of our prior width. The model which maximises the BIC is preferred under this procedure.

In cases where we have access to the full posterior, as suggested by Handley and Lemos [59] we set $\mathcal{K} = \tilde{d}$, where the Bayesian model dimensionality (BMD) is defined to be

$$\frac{\tilde{d}}{2} \equiv \int \mathcal{P}(\boldsymbol{\theta}|\mathcal{D})\left(\log\frac{\mathcal{P}(\boldsymbol{\theta}|\mathcal{D})}{P(\boldsymbol{\theta})}\right)^2 \mathrm{d}\boldsymbol{\theta}, \tag{1.47}$$

and can be computed using e.g. the ANESTHETIC software package [60]. Although not as 'Bayesian' as comparing the evidence, we prefer this statistic due to its insensitivity to the prior.

In Chapter 3 we will also wish to compare two observational distributions to determine whether they are equivalent, and will do this using the two-sided Kolmogorov–Smirnov (KS) test [61]. This test is designed to determine the probability that two samples have been drawn from the same probability distribution. To perform the test, we first define the empirical distribution function for a set of observations $\{X_i\}$ to be

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[-\infty,x_i]}(X_i), \tag{1.48}$$

where $\mathbb{1}_{[-\infty,x_i]}(X_i)$ is 1 for $X_i \leq x$ and 0 otherwise. For the two-sided KS test we then compute

$$D_{nm} = \sup_x |F_{1,n}(x) - F_{2,n}(x)|, \tag{1.49}$$

where $F_{1,n}$, $F_{2,n}$ are the empirical distribution functions for datasets 1 and 2 respectively. Intuitively, one can view this as the maximum difference in the cumulative

distribution functions of the two sets of observations at fixed x. One then can reject the null hypothesis that the two samples are drawn from the same population distribution at level α if

$$D_{nm} > \sqrt{-\frac{1}{2}\log\left(\frac{\alpha}{2}\right)\frac{n+m}{nm}}, \tag{1.50}$$

where there are n (m) observations in dataset 1 (2).

1.2.3 Sampling techniques

It is often not possible to express the posterior distribution in an analytic form, and thus we wish to find a numerical representation for $\mathcal{P}(\boldsymbol{\theta}|\mathcal{D})$. This will enable us to marginalise over a subset of our parameters or find statistical summaries of the distribution. This can be achieved by sampling from such a distribution to determine $N \gg 1$ samples, $\{\boldsymbol{\theta}_i : i = 1, \ldots, N\}$. One can then approximate the expectation value of any function of the parameters, $f(\boldsymbol{\theta})$, as

$$\langle f(\boldsymbol{\theta})\rangle = \int \mathrm{d}\boldsymbol{\theta}\, f(\boldsymbol{\theta})\,\mathcal{P}(\boldsymbol{\theta}|\mathcal{D}) \approx \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{\theta}_i)\,. \tag{1.51}$$

The natural question now arises as to how one should generate such samples. In this section we will summarise two different approaches to this problem.

1.2.3.1 Markov Chain Monte Carlo

The first solution requires one to construct a Markov Chain - a stochastic sequence of events where the probability of the outcome of next step is determined solely by the preceding step - such that in the limit of an infinite number of samples, the probability distribution of the samples stored from each step of the chain converges on the target distribution, $\pi(\boldsymbol{x})$. The aim of a Markov Chain Monte Carlo (MCMC) algorithm is to obtain the posterior distribution from $\pi(\boldsymbol{x})$. We will now outline various choices for such an algorithm.

1.2.3.2 Metropolis-Hastings Algorithm

Suppose we are at some point in our Markov Chain with parameters $\boldsymbol{x}$. We now wish to determine where we should move to in parameter space for the next step in the chain, $\boldsymbol{x}'$. The Metropolis-Hastings algorithm begins by assuming the principle of

detailed balance, i.e. that the probability of being at $\boldsymbol{x}$ and moving to $\boldsymbol{x}'$ must be equal to the probability of the reverse process of being at $\boldsymbol{x}'$ and moving to $\boldsymbol{x}$:

$$P\left(\boldsymbol{x}'|\boldsymbol{x}\right)P\left(\boldsymbol{x}\right) = P\left(\boldsymbol{x}|\boldsymbol{x}'\right)P\left(\boldsymbol{x}'\right). \tag{1.52}$$

There are two stages to determining where to move in parameter space. First, one must propose moving from $\boldsymbol{x}$ to $\boldsymbol{x}'$ with probability $g\left(\boldsymbol{x}'|\boldsymbol{x}\right)$. Once we have proposed a move, we will then accept our new sample with a probability given by the acceptance distribution, $A\left(\boldsymbol{x}'|\boldsymbol{x}\right)$, such that

$$P\left(\boldsymbol{x}'|\boldsymbol{x}\right) = g\left(\boldsymbol{x}'|\boldsymbol{x}\right)A\left(\boldsymbol{x}'|\boldsymbol{x}\right). \tag{1.53}$$

Substituting into our requirement for detailed balance, we have

$$\frac{A\left(\boldsymbol{x}'|\boldsymbol{x}\right)}{A\left(\boldsymbol{x}|\boldsymbol{x}'\right)} = \frac{g\left(\boldsymbol{x}|\boldsymbol{x}'\right)}{g\left(\boldsymbol{x}'|\boldsymbol{x}\right)}\frac{P\left(\boldsymbol{x}'\right)}{P\left(\boldsymbol{x}\right)}. \tag{1.54}$$

A simple choice now exists for our acceptance distribution

$$A\left(\boldsymbol{x}'|\boldsymbol{x}\right) = \min\left(1, \frac{g\left(\boldsymbol{x}|\boldsymbol{x}'\right)}{g\left(\boldsymbol{x}'|\boldsymbol{x}\right)}\frac{P\left(\boldsymbol{x}'\right)}{P\left(\boldsymbol{x}\right)}\right), \tag{1.55}$$

which clearly allows detailed balance. The samples we store will have a distribution approaching $\pi\left(\boldsymbol{x}\right) = P\left(\boldsymbol{x}\right)$ as we increase the length of our chain. Note that since we are only interested in the ratio of our posteriors, for this algorithm we only need to compute the likelihood and prior, and do not require a further calculation of the evidence, which may be non-trivial.

We are yet to specify the proposal distribution $g\left(\boldsymbol{x}'|\boldsymbol{x}\right)$. One can see that the efficiency of our sampling (the fraction of samples which we accept) is going to depend strongly on this choice. However, it does not affect the accuracy of our MCMC chain due to the Metropolis-Hastings acceptance criterion. The appropriate choice is problem specific. In this thesis we will often use an affine invariant ensemble sampler, as implemented in the EMCEE package [62].

Not only do we want a high acceptance rate for efficient computation, but the aim of the MCMC algorithm is to draw *independent* samples, otherwise we cannot compute Equation 1.51. Obviously, the next sample in a chain depends on the previous sample (by the definition of a Markov Chain), however we want the future steps in the chain to 'forget' about the current state as quickly as possible. To quantify the independence of the samples, we compute the autocorrelation length; samples separated by an autocorrelation length are approximately independent.

1.2.3.3 Hamiltonian Monte Carlo

The trajectory through parameter space for a conventional MCMC algorithm resembles a random walk. As one increases the number of parameters in the physical model, the increased dimensionality of the space to explore renders such an approach impractical due to the decreasing efficiency of the sampling. The number of samples required increases exponentially with the number of parameters; the so-called curse of dimensionality.

The Hamiltonian Monte Carlo (HMC) algorithm [63] is designed to resolve this issue by (up to numerical precision) guaranteeing that the next step we propose in our Markov chain will be accepted. For each parameter in $\boldsymbol{x} = (x_1, x_2, \ldots)$, we now introduce a conjugate momentum $\boldsymbol{p} = (p_1, p_2, \ldots)$ and define a Hamiltonian for the system

$$H = \frac{1}{2} p_i M_{ij}^{-1} p_j - \log P(\boldsymbol{x}), \tag{1.56}$$

for some mass matrix M_{ij}, where we sum over repeated indices. This is the Hamiltonian for a massive particle moving in a high-dimensional parameter space with a potential $\Psi(\boldsymbol{x}) = -\log P(\boldsymbol{x})$. We now describe our trajectory in parameter space by some pseudo-time, τ, such that we obey Hamilton's equations

$$\frac{\mathrm{d}x_i}{\mathrm{d}\tau} = \frac{\partial H}{\partial p_i}, \quad \frac{\mathrm{d}p_i}{\mathrm{d}\tau} = -\frac{\partial H}{\partial x_i} = -\frac{\partial \Psi(\boldsymbol{x})}{\partial x_i}. \tag{1.57}$$

Note that the particle is going to preferentially move to the high likelihood regions, since this is where the potential is smallest. After allowing our target distribution to be

$$\pi(\boldsymbol{x}) \propto e^{-H} = P(\boldsymbol{x}) \exp\left(-\frac{1}{2} p_i M_{ij}^{-1} p_j\right), \tag{1.58}$$

and integrating the equations of motion for the particle from some initial pseudo-time, τ_i, to some final pseudo-time, τ_f, we use the standard Metropolis-Hastings acceptance criterion (Equation 1.55) to find our acceptance probability

$$A(\boldsymbol{x}', \boldsymbol{p}' | \boldsymbol{x}, \boldsymbol{p}) = \min\left[1, \exp\left(-H(\boldsymbol{x}', \boldsymbol{p}') + H(\boldsymbol{x}, \boldsymbol{p})\right)\right]. \tag{1.59}$$

By conservation of the Hamiltonian, we are guaranteed to accept our new state, provided we have no numerical errors. By marginalising over the conjugate momenta, we obtain our desired distribution, $P(\boldsymbol{x})$. In practice, this means we only store $\boldsymbol{x}$ after each step in the chain and can discard the conjugate momenta. To obtain the next step, one simply needs to draw the initial momenta from the appropriate Gaussian distribution before integrating the equations of motion.

The HMC algorithm is therefore particularly efficient for exploring high-dimensions probability distributions for two reasons: it uses conservation of the Hamiltonian to ensure a high acceptance rate, and it utilises gradient information of $P(\boldsymbol{x})$ so that it is attracted to high-likelihood regions. One needs to tune the mass matrix, M_{ij}, and integration time, $\tau_f - \tau_i$, for the problem at hand (in fact $\tau_f - \tau_i$ should be varied to avoid the impact of resonances), however the method will prove essential in Section 1.3 when we wish to infer $\mathcal{O}(10^6 - 10^7)$ parameters.

1.2.3.4 Nested Sampling

One of the main advantages of the MCMC methods discussed in the preceding sections is that they rely on the relative probability between two points, and thus one does not need to compute the evidence. In some situations (e.g. if one needs to compute the Bayes factor) it is desirable to obtain the evidence, and thus other techniques should be employed. The aim of nested sampling algorithms is to compute $\mathcal{Z}(\mathcal{D})$ itself and, as a by-product, obtain samples for the posterior.

The basic principle can be seen as follows [64]. Let us first define the prior mass, $X(\lambda)$, to be

$$X(\lambda) = \int_{\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})>\lambda} P(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}, \tag{1.60}$$

which is the prior volume for which the likelihood exceeds λ. This must be a monotonically decreasing function of λ and $0 \leq X(\lambda) \leq 1$. If one can produce this function's inverse, $f(X(\lambda)) = \lambda$, then the evidence will become trivial to compute

$$\mathcal{Z}(\mathcal{D}) = \int_0^1 f(X)\,\mathrm{d}X. \tag{1.61}$$

One should therefore find the contours of equal likelihood (to obtain $X(\lambda)$) and then compute this integral. This differs from standard MCMC techniques, since here we wish to always move in the direction of increasing likelihood as we continue our exploration of parameter space, whereas we move in both directions with standard methods. The exact method by which this occurs is sampler-specific; in this thesis we use both the PYMULTINEST [65–67] and ULTRANEST[1] [68–70] packages.

1.2.4 The likelihood model

As introduced in Section 1.2.1, the likelihood model, $\mathcal{L}(\mathcal{D}|\boldsymbol{\theta})$, is a model which determines how similar the observed data, $\mathcal{D}$, are to the predicted data given a physical

[1] `https://johannesbuchner.github.io/UltraNest/`

model, $\mathcal{M}$, and its parameters, $\boldsymbol{\theta}$. The form of the likelihood can be informed from the physical model. For example, if one has a physical model which predicts the mean and covariance of a Gaussian Random Field, then the appropriate likelihood is a Gaussian with that mean and covariance. Alternatively, if one is studying the number of observed photons in pixels across the sky, then one would choose a Poisson likelihood with a mean predicted by the physical model. The addition of noise can complicate this scenario, however if one has a model for the form of this noise, e.g. Gaussian noise, then one can convolve the likelihood for the uncontaminated observable with the noise likelihood to obtain the likelihood of the observed data. In the astrophysical tests of fundamental physics considered in this thesis, we will often be in a different situation where one cannot write down a form for $\mathcal{L}$ and conclusively state that it must be the true likelihood. In this case, one should view $\mathcal{L}$ just like $\mathcal{M}$: its form is simply an analysis choice and one should compare different likelihoods to determine which is most favoured by the data. Not doing so could lead to biased results, in much the same way as using a sub-optimal $\mathcal{M}$.

This subtlety in the analysis arises when one tries to explicitly write a form for the likelihood. One can circumvent this challenge through the use of Implicit Likelihood Inference (ILI), where one avoids this step entirely and the likelihood appears implicitly in the analysis (this is also commonly referred to as Likelihood Free Inference (LFI) or Simulation Based Inference (SBI)). The simplest example of ILI is Approximate Bayesian Computation (ABC) [71], where one randomly draws parameters from the prior distribution and computes the predicted data using $\mathcal{M}$. Utilising some definition of the distance between the predicted and observed data, one simply keeps all samples which lie within a distance ϵ of the observations. The distribution of these samples gives the posterior. This likelihood-free rejection sampling approach can become unfeasible if one wants to minimise the number of model evaluations or has a large number of parameters since the rejection rate will often be high. Instead, one can introduce a more efficient proposal step to reduce the number of simulations (e.g. Bayesian Optimisation for Likelihood-Free Inference (BOLFI) [72, 73]) or linearise the physical model (e.g. Simulator Expansion for Likelihood-Free Inference (SELFI) [74]) to reduce the computational cost of small moves in parameter space. Since one is often only interested in summaries of the posterior, a more radical approach is to not compute the posterior at all, but only its mean and covariance through the use of e.g. moment networks [75].

In this thesis we will commonly use ILI techniques to obtain the posteriors of intermediate parameters which we will later compare to the data. The ILI technique

which we will often use is based on Density-Estimation Likelihood-Free Inference (DELFI) [76–78]. This method aims to fit an estimator of the likelihood or posterior from data-parameter pairs. Specifically, we will draw parameters $\boldsymbol{\theta}$ from their priors, evaluate a physical model $\mathcal{M}$, and then fit the distribution of these predictions to a flexible estimator. As in [78], we will typically fit these distributions to a Gaussian Mixture Model (GMM), although we will always choose a parametrisation whereby one does not need to apply neural density estimators (unlike [78]). By fitting a GMM we obtain a continuous representation of the Monte Carlo data, which we will often convolve with analytic distributions describing parts of the model where a Monte Carlo procedure is unnecessary.

1.2.5 Forward modelling and field level inference

Cosmology has historically focused on the study of two-point statistics; with limited observations and computational resources, one can place constraints on cosmological parameters and models without needing to know the phases of the initial conditions of the local Universe. For example, in studies of large scale structure, one could attempt to measure the power spectrum of matter perturbations (Equation 1.20) and compare the observed $P(k)$ to that expected from theory. Since this will depend on cosmological parameters, one can place constraints on these values using such methods.

As we move into an era of increased observational and computing power, one should question whether we can go beyond this traditional approach. One could try to measure higher order statistics such as the bispectrum or trispectrum, although this still discards potentially useful information available when one studies each object individually. In general, one does not observe the whole sky, and thus one must include selection effects, which will couple different Fourier modes. Moreover, once we have obtained a power spectrum or correlation function, it is not obvious what form the likelihood should take.

Therefore, in this thesis we will always compare data at the 'field level' or on an 'object-by-object' basis, rather than first computing their summary statistics and will therefore implicitly capture not just the two-point statistics, but all higher orders too. It has been shown that tighter constraints on cosmological parameters are possible using this method than through two-point analyses [79–81], since one does not discard the information contained in these higher order statistics.

Not only does one need to consider whether to perform an analysis on an object-by-object basis or at the level of a summary, but there is a second analysis choice

to make. The observed data will usually be subject to contamination from masks, selection effects and noise, whereas the simplest physical model will assume 'perfect' data, e.g. one can measure the full sky with arbitrary accuracy, with an isotropic exposure and without noise. One can either attempt to undo the selection effects of the data to extract the underlying data, or one can augment the physical model to try to predict what the observed data looks like. This second procedure, dubbed 'forward modelling' is a mapping from theory to data space, such that we do our analysis through direct comparison to observations. Although this is a conceptually cleaner methodology and allows a more straightforward implementation of the noise model, it is typically difficult to perform since one needs to increase the dimensionality of parameter space to incorporate models for the imperfections in the data. The treatment of noise is significantly more challenging in the reverse modelling scenario (mapping from data to theory space) so one typically would rely on frequentist statistics, for example by subtracting the maximum likelihood template for some contamination to the underlying signal. In this thesis we will be attempting to identify correlations in extremely noisy data, where there is often not an obvious way to 'subtract' the noise. We will therefore exclusively use forward modelling techniques in this thesis due to the importance of the noise models in our analyses and to allow us to marginalise over these models in a Bayesian manner.

1.3 Bayesian Origin Reconstruction from Galaxies

In this section we describe a method for processing large scale structure data which goes beyond two-point statistics, where one can measure the true (i.e. not just the statistical) DM distribution of the local Universe using the `BORG` (Bayesian Origin Reconstruction in Galaxies) algorithm [4, 5, 82–86].

1.3.1 The BORG algorithm

Despite having five times the average density of baryonic matter, since DM is only known to interact gravitationally, we must infer its presence indirectly through its effect on other components of the Universe. In particular, we expect that galaxies reside with quasi-spherical 'halos' of DM, such that the distribution of galaxies throughout the Universe can be used to trace the distribution of DM. Hence, if one knows what the connection between the galaxy and DM density is (the so-called bias), one can infer the large scale distribution of DM. From a Bayesian perspective, this enables us to derive and evaluate a likelihood for the observed galaxy distribution

given a DM distribution. Of course, it is not the likelihood which concerns us, but the posterior. By Bayes' theorem (Equation 1.39) we therefore need to know what the prior distribution of DM is. This is in fact a highly non-trivial task and, due to the non-linear equations of structure formation, an analytic expression for the prior of the present-day matter distribution is not known. One could introduce maximally agnostic priors (either a Gaussian or log-normal distribution depending on whether you know the scale of the density fluctuations) however this is far from ideal.

We are saved from a theoretical and observational view point, since the density field is (approximately) Gaussian at high redshifts. We can therefore place a Gaussian prior on the initial density fluctuations,

$$P\left(\delta^{\mathrm{I}}\right) = \frac{1}{\sqrt{\det\left(2\pi S\right)}} \exp\left(-\frac{1}{2}\delta_i^{\mathrm{I}} S_{ij}^{-1} \delta_j^{\mathrm{I}}\right), \tag{1.62}$$

where we have discretised our density field such that i labels the voxel of interest. Given a current matter density field, δ^{F}, it is tempting to integrate the equations of motion backwards to high redshift and then evalate this prior. There a number of issues with this approach. First, the equations of motion are second order and thus contain two solutions; one of these grows with time and the other decays. By integrating backwards, one can match onto one of these decaying modes, which will grow to spuriously large values when integrating backwards, making such a procedure numerically infeasible. Secondly, if one has a masked region of the sky, it is unclear how one should fill in the unobserved region in order to have a full density field to evolve backwards in time. The procedure for reconstructing the initial conditions from present day observations in such a manner is ill-posed [87, 88].

The `BORG` algorithm approaches this differently. Instead of trying to integrate the equations backwards in time to unambiguously find the initial conditions (which is impossible), it instead works in reverse. The algorithm aims to reconstruct the initial conditions (ICs) of the local DM density field by applying a Bayesian forward model for the number densities of observed galaxies in voxels and marginalising over galaxy bias parameters. A key assumption is that gravitational structure formation is a deterministic process, such that the likelihood of a final density field given an initial one is

$$P\left(\{\delta_i^{\mathrm{F}}\}|\{\delta_i^{\mathrm{I}}\}\right) = \prod_i \delta^{\mathrm{D}}\left(\delta_i^{\mathrm{F}} - G_i\left(\{\delta_i^{\mathrm{I}}\}\right)\right), \tag{1.63}$$

where the set of functions $G_i\left(\{\delta_i^{\mathrm{I}}\}\right)$ is the structure formation model. In this way, one can constrain the ICs of the local Universe by first evolving these forward to the present day under some structure formation model and then by comparing these to

the observed galaxy number counts by applying a bias model. Evaluating $G_i\left(\{\delta_i^{\rm I}\}\right)$ involves simulating the evolution of the Universe from the ICs to a redshift of zero. Structure formation models implemented to date include first and second order Lagrangian Perturbation Theory, Particle-Mesh simulations and the COmoving Lagrangian Acceleration (CMB) method [89].

This procedure requires one to infer a large number of parameters. If one wishes to infer the ICs on a grid with 256 grid points per side, this results in $\mathcal{O}\left(10^7\right)$ (highly degenerate) parameters, and thus a HMC procedure (Section 1.2.3.3) is necessary. Not only this, but one also needs to infer and marginalise over nuisance parameters, e.g. those describing galaxy bias or foreground templates. The `BORG` algorithm is therefore designed as a highly flexible block sampling procedure to allow for arbitrary complexity in the model.

The result of this sampling is a Markov Chain of feasible realisations of the ICs of the local Universe. A particular advantage of the `BORG` algorithm is that it interpolates the density field into the unobserved regions of the sky in a physically consistent manner, such that for a given MCMC sample, one cannot distinguish between the observed and unobserved regions. This is not true across the whole chain; the ensemble mean reflects the observational strategy such that the mean $\delta^{\rm F}$ will be zero in the unobserved regions, corresponding to the mean cosmological density. In the observed regions, however, the data are highly constraining such that one can identify structures which exists in all samples, and these regions have a smaller (fractional) variance in $\delta^{\rm F}$ across the chain.

1.3.2 Constrained simulations

Throughout this thesis, we use the `CSiBORG` (Constrained Simulations in `BORG`) suite of constrained N-body simulations of the local universe [90–93]. The full `CSiBORG` suite takes $\sim$100 sets of $z = 69$ ICs from the posterior of the particle-mesh `BORG` reconstruction of the 2M++ volume [5, 94] separated by several autocorrelation lengths. These cover a box length of $677.77\,h^{-1}$Mpc with 256^3 voxels, yielding a resolution of $2.65\,h^{-1}$Mpc. Within a smaller sphere of radius $155\,h^{-1}$Mpc centred on the Milky Way, the ICs are augmented with white noise to a resolution of 2048^3, giving a particle mass of $4.38 \times 10^9 M_\odot$. Each set of ICs is then used to run a DM-only `RAMSES` [95] simulation to $z = 0$, refining only in the higher-resolution central sphere (although keeping the larger cube to include longer-wavelength modes). This produces $\sim$100 N-body realisations of the local DM structure. By sampling the realisations one can marginalise over both the uncertainties in the constraints on the ICs derived from

2M++ and the unconstrained smaller-scale modes, and hence over the local DM density field itself. `BORG` and `CSiBORG` use the cosmology $T_{\rm CMB} = 2.728$ K, $\Omega_{\rm m} = 0.307$, $\Omega_\Lambda = 0.693$, $\Omega_{\rm b} = 0.04825$, $H_0 = 70.5$ km s^{-1} Mpc^{-1}, $\sigma_8 = 0.8288$, $n_{\rm s} = 0.9611$.

1.4 Thesis Overview

The program of work outlined in thesis aims to utilise astrophysical systems as precision probes of fundamental physics through simulation-based, Bayesian statistical forward-modelling techniques. These methods differ from conventional analyses in two main ways. First, we utilise an under-explored region of parameter space, since relatively few tests of ΛCDM and the SM have been conducted on galactic scales [96], although the field of astrophysical tests of gravity is nonetheless emerging as powerful and complementary to traditional methods [36]. There are clear advantages to studying galaxies. We are no longer restricted to the linear regime, so remove the associated limits on our results introduced by cosmic variance. However, the same non-linearities that make galaxies such rich laboratories for investigating fundamental physics also complicate any such analysis, because the complex astrophysical effects that shape galaxies act as critical systematics. This will require the development of a sophisticated treatment of astrophysical and observational noise. The second difference compared to conventional analyses is that we will always focus on field-level inference; we will make predictions on an object-by-object (or pixel-by-pixel) basis and not on their statistical summaries. This will enable us to implicitly capture all n-point statistics.

In Chapter 2 we will develop the inference techniques required to test fundamental physics on astrophysical scales by constraining Quantum Gravity (QG), the photon mass and the WEP with the spectral lags of Gamma Ray Bursts (GRBs). Before transferring these techniques to tests of scalar-tensor theories in Chapter 4 using the locations of the supermassive black hole (SMBH) in galaxies, we will investigate this phenomenon in ΛCDM through the use of hydrodynamical cosmological simulations in Chapter 3. In Chapter 5 we will constrain the self-annihilation cross-section and decay rate of DM by cross-correlating DM density fields inferred from the distribution of galaxies with the gamma ray background. These chapters will require us to develop empirical noise models describing alternative contributions to the observed signals, so in Chapter 6 we introduce a framework to assess the validity of these models and we carry out such an assessment for a galactic-scale test of $f(R)$ gravity. We conclude in Chapter 7. The work in Chapter 2 was first presented in [97, 98], Chapter 3 is based

on [99], Chapter 4 utilises work from [90], the analysis of Chapter 5 was presented in [100] and Chapter 6 was discussed in [101].

Chapter 2

Fundamental Physics and Gamma Ray Bursts

2.1 Introduction

High energy astrophysical transients at cosmological distances allow us to test the fundamental assumptions of the standard models of cosmology and particles physics, such as Lorentz invariance (LI), the massless nature of the photon or the WEP, (for a review, see [102]). If any of these assumptions are incorrect, photons of different energies propagate differently through spacetime, which could be observable in the spectral lags of GRBs.

If the photon velocity is energy-dependent, then photons of different energies from a distant source would arrive at different times, even if they were emitted simultaneously. For example, the photon's dispersion relation could be modified if it has a nonzero mass, m_γ, such that the photon velocity, v, depends on the energy, E, as

$$v = \sqrt{1 - \frac{m_\gamma^2}{E^2}} \approx 1 - \frac{1}{2}\frac{m_\gamma^2}{E^2}. \tag{2.1}$$

Alternatively, the hitherto elusive unification of quantum mechanics and GR is expected to exhibit so-called QG effects at energies of order the Planck scale, $E_\mathrm{P} \approx 1.2 \times 10^{19}\,\mathrm{GeV}$. By extension of the uncertainty principle, one may expect spacetime no longer to appear smooth on distance scales $\Delta x \sim 1/E_\mathrm{P}$ [103], and thus have a non-trivial refractive index for particles propagating through it. Hence, at low energies in QG theories, the photon velocity becomes [104]

$$v \approx 1 - \xi\frac{E}{E_\mathrm{QG}}, \tag{2.2}$$

where $\xi = \pm 1$ and E_QG is the QG energy scale, constituting LI violation. Such a linear modification to v is expected in a range of QG models (see Section 1 of [105]

and references therein). We dub the $\xi = +1$ and $\xi = -1$ models "subluminal QG" and "superluminal QG" respectively, based on the value of v for nonzero E. One may expect that $\xi = +1$ otherwise photons would quickly lose energy due to gravitational Čerenkov radiation [106], however we consider both signs in this chapter since current constraints [107] from Čerenkov radiation only consider models which would lead to even powers of E in Equation 2.2. Superluminal photons would also decay into electron–positron pairs above a threshold energy, providing an independent test of such theories [108].

Constraints on $E_{\rm QG}$ from GRBs have previously been obtained using a variety of methods [105, 109–130], with the most stringent lower bounds of $E_{\rm QG} > 9.3\times10^{19}\,{\rm GeV}$ for the subluminal ($\xi = +1$) and $E_{\rm QG} > 1.3\times10^{20}\,{\rm GeV}$ for the superluminal ($\xi = -1$) models arising from GRB 090510 [115]. In many cases these constraints are obtained using a handful of GRBs, do not propagate uncertainties in the redshifts of sources, or suffer from uncertain systematics in the model for other contributions to the spectral lag.

It is clear from Equation 2.1 that tighter constraints on m_γ can be obtained using lower frequency photons, and thus fast radio bursts (FRBs), pulsars and magnetars provide useful probes of m_γ [102, 131–140]. However, the scaling with frequency ($v \propto E^{-2}$) is identical to the time delay due to dispersion by electrons [141], which is negligible for gamma rays, but leads to degeneracies at radio frequencies. The majority of constraints from radio frequencies neglect this important contribution [102], although recently [133–135] the plasma effect has been incorporated into a Bayesian analysis of FRBs, leading to constraints of $m_\gamma < 4\times10^{-15}\,{\rm eV}/c^2$. Slightly tighter constraints of $2.9\times10^{-15}\,{\rm eV}/c^2$ have also been obtained [136], but these do not include the plasma effect in the analysis. GRBs have previously been used to constrain the photon mass, but these either compare to radio frequencies [137] or do not consider alternative causes for the time delay [138].

One does not require an energy-dependent photon velocity to produce a time delay between photons of different energy. A basic feature of metric theories of gravity is that objects on either timelike or null geodesics experience apparent time delays due to motion through regions of varying spacetime curvature, caused by the difference between proper and observer time induced by a gravitational field. In general metric theories the time delay is proportional to the integral of the fluctuation in Newtonian potential ϕ along the line of sight to the source; within the PPN framework it has a prefactor $1+\gamma$ where γ is the Parametrised Post-Newtonian (PPN) light-bending parameter [142], and $\gamma = 1$ in GR. This effect, first derived by Shapiro in 1964

[143], has now been accurately measured in the Solar System: the best constraint on the fractional deviation of the time delay from the prediction of General Relativity (GR) is currently at the 10^{-5} level, using data from the radio link with the Cassini spacecraft [144]. For sources at greater distance, however, the lack of knowledge of the time of emission of a signal precludes direct measurement of the delay.

Nevertheless, the time delay effect has found use in testing various aspects of the theory of gravity and the standard model of cosmology. This is done by *comparing* time delays, either between nearby geodesics or between different types of object following the same geodesic. The former may be achieved by comparing the time of reception of photons originating from a common source but traversing different paths due to gravitational lensing. When the source is a time-varying quasar and the lens a single massive elliptical galaxy or cluster, this method has been used to constrain the Hubble parameter, H_0, on which time delay distances depend [145, 146]. Conversely, comparing time delays between different objects travelling along the same geodesic allows the WEP to be tested by investigating whether they experience time delays identically. This can be achieved either when the source emits light at varying frequencies which can be independently measured [147], or when it emits other types of energy in addition to light, such as gravitational waves [148]. Within the PPN framework, this constrains the difference between their γ factors.

Constraining γ through time delays is therefore equivalent to determining $\delta\phi$. In typical analyses, $\delta\phi \simeq \phi$ is modelled as arising from one or a few isolated sources near the line of sight that are believed to be predominantly responsible for sourcing the potential. However, the long range of the gravitational potential ($\phi \sim 1/r$) casts doubt on the multiple point masses approximation, since ϕ is sensitive to the distribution of distant sources and thus should be considered in a cosmological context [149–151].

There are thus three aims of this chapter. First, we address the issues highlighted when constraining $E_{\rm QG}$ by constructing probabilistic source-by-source forward models of the time delays of GRBs from the BATSE satellite [152, 153] and marginalising over empirical models describing astrophysical and observational contributions to the measured time delay. Second, we provide constraints on the photon mass that are independent of radio observations and are thus insensitive to potential systematics in modelling the propagation of photons at radio frequency. Finally, we aim to account fully for the contributions to the time delay from all mass in the non-linear cosmological density field when constraining the energy variation of γ.

In Section 2.2 we describe the observations of GRBs used in this chapter. We then constrain the photon's dispersion relation in Section 2.3 by developing forward models for the time delay, introducing models for competing astrophysical and observational effects, and deriving the likelihood function. These techniques are then utilised in Section 2.4 to constrain EP violation, where instead of the geometric time delay we must compute the Shapiro time delay due to mass in the local Universe. Section 2.5 concludes.

2.2 Observational Data

The spectral lag data we use are a sample of the BATSE detections of GRBs catalogued in [152]. The temporal evolution of a GRB's luminosity, $L(t)$, typically contains many spikes, making a simple definition of the time delay based on a fit to $L(t)$ challenging. Instead, spectral lags are typically computed to be the peak in the cross-correlation function of $L(t)$ between two energy channels [154], and is the definition used in this chapter. We make use of a set of sources compiled in [153], where the four energy bins considered are sensitive to the ranges Ch1: 25-60 keV, Ch2: 60-110 keV, Ch3: 110-325 keV and Ch4: >325 keV. These four bins result in up to 6 time delay pairs per source, Δt_{ij}, where i, j label the channels used. Without loss of generality we define $i > j$. In the cases where no Δt_{ij} is recorded for a source due to low signal-to-noise, we ignore that particular pair but still consider the others. These sources also have pseudo-redshifts calculated using the spectral peak energy-peak luminosity relation [155].

The physical mechanisms that result in the spectral lags of GRBs are unknown; the search for the nature of spectral lag is an ongoing research topic where most of the focus lies on investigating possible effects at the source [see e.g. 156–162]. It has, for example, been shown that the effects of spectral lag can be recreated from simple source models utilising rapid bulk acceleration on relativistic jet shells [162]. Therefore, when modelling the intrinsic contribution to these time delays in Section 2.3.2 we will have to rely on empirical models, as opposed to the ideal case where we can calibrate our noise model with simulations [101].

2.3 Quantum Gravity and the Photon Mass

2.3.1 Forward modelling the time delay

As noted by [163], for energy-dependent photon speeds, one cannot simply multiply the difference in photon velocity in the observer's frame by the distance travelled by a photon, but one must consider the cosmological redshift of the photon. This results in a time delay between photons of observed energy E_i and E_j, $\Delta t_{ij} \equiv t_j - t_i$, of

$$\Delta t_{ij}^{(\mathrm{QG})} = \xi \frac{E_i - E_j}{E_{\mathrm{QG}}} \int_0^z \frac{1+z'}{H\left(z'\right)} \mathrm{d}z' \equiv \xi \frac{\Delta E_{ij}}{E_{\mathrm{QG}}} I_{\mathrm{QG}}, \tag{2.3}$$

for the QG scenario, and

$$\Delta t_{ij}^{(\mathrm{MP})} = \frac{m_\gamma^2}{2} \left(\frac{1}{E_i^2} - \frac{1}{E_j^2} \right) \int_0^z \frac{1}{H\left(z'\right)\left(1+z'\right)^2} \mathrm{d}z' \equiv \frac{m_\gamma^2}{2} \left(\frac{1}{E_j^2} - \frac{1}{E_i^2} \right) I_{\mathrm{MP}} \tag{2.4}$$

for the massive photon. We will assume a ΛCDM cosmology to determine the Hubble parameter $H(z)$; in general one should consider a variety of cosmological models [164–167], although this is beyond the scope of this work. In both cases we see that the parameter of interest (E_{QG} or m_γ) and the observed energy bands appear as scaling factors. We therefore simply need to compute the predicted theoretical time delay, $\Delta t_{ij}^{(\mathrm{th})} \in \{\Delta t_{ij}^{(\mathrm{QG})}, \Delta t_{ij}^{(\mathrm{MP})}\}$, using Equations 2.3 and 2.4 for some fiducial E_{QG} or m_γ, and then rescale these parameters linearly for E_{QG} and quadratically for m_γ (Equations 2.3 and 2.4) according to the E_{QG} or m_γ being sampled.

Since the redshift values of our sources are uncertain, we draw $N_{\mathrm{MC}} = 10^4$ redshifts per source from a two-sided Gaussian with upper and lower uncertainties equal to the uncertainties calculated in [155]. Doubling N_{MC} yields identical results, indicating that the number of samples is adequate. For each sample we evaluate the integrals I_{QG} and I_{MP} using the `scipy.integrate` subpackage [168]. To determine the appropriate E_i to use in Equations 2.3 and 2.4, for each sample we draw an energy randomly from a distribution proportional to the best-fit spectral model for that GRB as given in the BATSE 5B Gamma-Ray Burst Spectral Catalog [169]. Furthermore, at each iteration we draw the parameters for the model from Gaussian distributions with means and widths as given in the catalogue.

The resulting N_{MC} samples are assumed to follow a GMM [170], such that the likelihood function for $\Delta t_{ij}^{(\mathrm{th})}$ for some source s is

$$\mathcal{L}_s\left(\Delta t_{ij}^{(\mathrm{th})}\right) = \sum_\alpha \frac{w_{sij}^{(\alpha)}}{\sqrt{2\pi}\tau_{sij}^{(\alpha)}} \exp\left[-\frac{\left(\Delta t_{ij}^{(\mathrm{th})} - \lambda_{sij}^{(\alpha)}\right)^2}{2{\tau_{sij}^{(\alpha)}}^2} \right], \tag{2.5}$$

where

$$\sum_{\alpha} w_{sij}^{(\alpha)} = 1, \quad w_{sij}^{(\alpha)} \geq 0, \tag{2.6}$$

and the sum runs over the number of Gaussian components. We compute an independent GMM for each source, and choose the number of Gaussians which minimise the BIC (Equation 1.43).

2.3.2 Modelling the noise

Quantum gravity or a photon mass are not the only types of physics that can lead to spectral lags: these may also be generated through intrinsic differences in the emission of photons of different wavelength at the source or their propagation through the medium surrounding the GRB, or through instrumental effects at the observer. Without a robust physical model for the time delays these lead to, we model them using a generic functional form (a sum of Gaussians) with free parameters that we marginalise over in constraining m_γ and E_{QG}. We refer to any contribution to the time delays other than QG or a photon mass as "noise."

We suppose that the time delay between frequencies ν_i and ν_j can be written as the sum of the following independent terms

$$\Delta t_{ij}^{(\mathrm{obs})}(r) = \Delta t_{ij}^{(\mathrm{th})}(r) + A(r)\left(\nu_i^{-2} - \nu_j^{-2}\right) + B_{ij}. \tag{2.7}$$

Besides the term arising from an energy-dependent propagation speed, our time delay contains two other contributions. The first of these (containing $A(r)$) describes the dispersion due to electrons. The second (B_{ij}) represents the combination of an intrinsic time delay at the source and the instrument response, which we assume is independent of observed angle and redshift.

The mean of A depends on the temporal evolution of the comoving electron density, $\bar{n}_{e,c}$, as

$$\bar{A}(r) = \frac{e^2}{2\pi m_e}\frac{1}{4\pi\epsilon_0}\int_0^{z(r)} \frac{\mathrm{d}z'}{H(z')}(1+z')\,\bar{n}_{e,c}(z'), \tag{2.8}$$

and fluctuations about this depend on fluctuations in the electron density and other relativistic effects [141]. The comoving electron density can be modelled as

$$\bar{n}_{e,c}(z) = \frac{3H_0^2\Omega_b}{8\pi G m_p}\frac{x_e(z)\,(1+x_H(z))}{2}, \tag{2.9}$$

where x_e is the free electron fraction, x_H is the hydrogen mass fraction, Ω_b is the baryon density fraction, and m_p is the proton mass. x_e is proportional to the fraction of electrons in the intergalactic medium, $f_{\mathrm{IGM}}(z)$, which slightly increases with

redshift, from 0.8 for $z \lesssim 0.4$ to 0.9 at $z \gtrsim 1.5$ [171–173]. For simplicity, assuming $x_e = 1$ and $x_H = 0.75$, we find the contribution from the electron plasma to be

$$A(r)\left(\nu_i^{-2} - \nu_j^{-2}\right) = \left(\int_0^{z(r)} \frac{H_0}{H(z')}(1+z')\,\mathrm{d}z'\right)\left(\left(\frac{E_i}{\mathrm{keV}}\right)^{-2} - \left(\frac{E_j}{\mathrm{keV}}\right)^{-2}\right) \times 5.6\times10^{-17}\,\mathrm{s}. \tag{2.10}$$

Since in this chapter we will be considering GRBs, we will find that this contribution is negligible for the probed frequencies, and thus we will neglect it in our analysis. If we were to consider radio bursts we would need to consider this term, as we would expect contributions of $\mathcal{O}(1\,\mathrm{s})$.

Given that we are neglecting the contribution from the electron plasma in Equation 2.7, to determine the likelihood of an observed time delay, we must finally convolve Equation 2.5 with the likelihood for B_{ij}. We assume that for each pair of frequencies the distribution of B_{ij} can be written as the sum of N_G Gaussians,

$$\mathcal{L}(B_{ij}) = \sum_\beta \frac{\omega_{ij}^{(\beta)}}{\sqrt{2\pi}\sigma_{ij}^{(\beta)}} \exp\left[-\frac{\left(B_{ij} - \mu_{ij}^{(\beta)}\right)^2}{2{\sigma_{ij}^{(\beta)}}^2}\right], \tag{2.11}$$

where

$$\sum_\beta \omega_{ij}^{(\beta)} = 1, \quad \omega_{ij}^{(\beta)} \geq 0, \tag{2.12}$$

and $\beta \in \{0, 1, \ldots, N_\mathrm{G} - 1\}$. Without loss of generality, we define the Gaussians such that $\omega_{ij}^{(\beta)} \geq \omega_{ij}^{(\beta+1)}$.

In Equation 2.11 we have assumed that the noise only depends on the observed photon energies. Inspired by Ellis *et al.* [110], we also consider noise models in which the means and widths of one or more of the Gaussians are redshift dependent,

$$\mu_{ij}^{(\beta)} \to \mu_{ij}^{(\beta)}(1+z_s), \quad \sigma_{ij}^{(\beta)} \to \sigma_{ij}^{(\beta)}(1+z_s), \tag{2.13}$$

where z_s is the quoted pseudoredshift of source s. These models capture an intrinsic contribution to the time delay from the source such as a "magnetic-jet" model for GRB emission [122, 174], whereas one would expect the redshift-independent models to describe observational effects. By including these, we now have a wider range of noise models to choose from, increasing our confidence that the optimum model lies within this set.

2.3.2.1 Likelihood model

The likelihood of a given observed time delay, $\Delta t_{ij}^{(\mathrm{obs})}$, for source s is given by the convolution of Equation 2.5 (once we have appropriately scaled $\Delta t_{ij}^{(\mathrm{th})}$ and the GMM parameters) and Equation 2.11,

$$\mathcal{L}_s\left(\Delta t_{ij}^{(\mathrm{obs})}\right) = \sum_{\alpha\beta} \frac{w_s^{(\alpha)}\omega_{ij}^{(\beta)}}{\sqrt{2\pi\left(\tau_{sij}^{(\alpha)2} + \sigma_{ij}^{(\beta)2} + \varepsilon_{sij}^2\right)}} \exp\left[-\frac{\left(\Delta t_{ij}^{(\mathrm{obs})} - \lambda_{sij}^{(\alpha)} - \mu_{ij}^{(\beta)}\right)^2}{2\left(\tau_{sij}^{(\alpha)2} + \sigma_{ij}^{(\beta)2} + \varepsilon_{sij}^2\right)}\right], \tag{2.14}$$

where we have also included the quoted measurement uncertainty in the spectral lag, ε_{sij}.

Assuming that all frequency pairs and sources are independent, the total likelihood for our dataset $\mathcal{D}$ is

$$\mathcal{L}\left(\mathcal{D}|\boldsymbol{\theta}\right) = \prod_{sij} \mathcal{L}_s\left(\Delta t_{ij}^{(\mathrm{obs})}\right), \tag{2.15}$$

where $\boldsymbol{\theta} \equiv \{\mathcal{A}, \mu_{ij}^{(\beta)}, \sigma_{ij}^{(\beta)}, \omega_{ij}^{(\beta)}\}$, and $\mathcal{A} = m_\gamma$ or $\ell_{\mathrm{QG}} \equiv \xi E_{\mathrm{QG}}^{-1}$ depending on the theory considered. We choose to fit for the QG length scale, ℓ_{QG}, instead of E_{QG} since the infinite upper limit of the prior on E_{QG} becomes a zero lower limit on the prior for ℓ_{QG}. A separate set of noise parameters is fitted to each pair of frequencies, but we consider the target of interest, $\mathcal{A}$, to be universal.

The deliberately wide priors, $P(\boldsymbol{\theta})$, in Table 2.1 lead to difficulties in interpreting the Bayes ratio (Equation 1.42). To determine the appropriate N_{G}, we instead compare the models by calculating the BIC (Equation 1.43); the best-fitting model minimises this statistic. To find the maximum likelihood and thus the BIC, we first optimise using the Nelder-Mead algorithm [175] with a simplex consisting of parameters drawn randomly from the prior. We repeat this ten times then compute the Hessian at the maximum likelihood point (MLP), $\hat{\mathcal{L}}$. Drawing 256 walkers from a Gaussian centred on the MLP with this Hessian, we run the EMCEE sampler [62] for 10,000 steps to find a new estimate of the MLP using the 2.56×10^6 samples. If our estimate of the Hessian is not positive definite, we draw the walkers from log-normal distributions of unit width, centred on the MLP. We find that $\hat{\mathcal{L}}$ changes by less than 2 per cent for any N_{G} and for both theories considered if we only use the first 5,000 steps, which is much smaller than the change in BIC between different N_{G}.

For computational convenience, we now use these posterior samples to restrict the size of the prior: we find the samples for which the change in χ^2 ($\Delta\chi^2 = -2\Delta\log\mathcal{L}$) from the MLP is 25 times the number of observed frequency pairs (5σ for a Gaussian

Table 2.1: Priors on photon mass, QG length scale and parameters describing the empirical noise model (Equation 2.11). All priors are uniform in the range given.

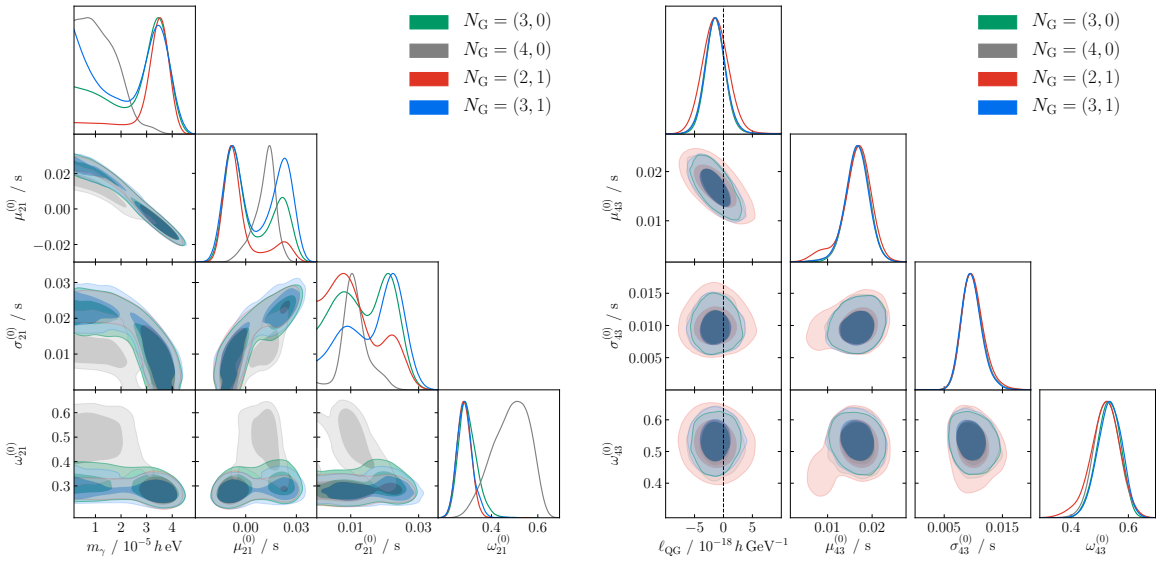
Parameter	Prior
m_γ / meV/c^2	$[0, 1]$
ℓ_{QG} / GeV^{-1}	$[-10^{-14}, 10^{-14}]$
$\mu_{ij}^{(\beta)}$ / s	$[-15, 15]$
$\sigma_{ij}^{(\beta)}$ / s	$[0, 15]$
$\omega_{ij}^{(\beta)}$	$[0,1], \quad \sum_\beta \omega_{ij}^{(\beta)} = 1, \quad \omega_{ij}^{(\beta)} \geq \omega_{ij}^{(\beta+1)}$

likelihood) and set the new prior such that it (just) encompasses these points. For some parameters we keep the prior wider than this to ensure that our results are not dominated by the choice of the prior. We now use Bayes' theorem (Equation 1.39) to obtain the posterior distribution of θ, and evidence with the nested sampling Monte Carlo algorithm MLFriends [68, 69] using the UltraNest package [70]. Since the prior is still treated as uniform and we do not use the Bayes ratio, reducing the size of the prior does not affect our results since it simply changes $\mathcal{P}(\boldsymbol{\theta}|\mathcal{D})$ by a multiplicative constant except in regions where it is already negligible.

2.3.3 Results

For both theories considered, we find that $N_{\mathrm{G}} = 3$ and 4 Gaussians have comparable BIC values, and that the best noise models contain either one redshift-dependent Gaussian or are completely independent of redshift, suggesting that observational effects dominate the noise. Using the `GetDist` package [176], in Figure 2.1 we show the corner plots for m_γ and ℓ_{QG} and the parameters for the component of the noise model which is most correlated with the signal. We note that, from the energy dependence of Equations 2.1 and 2.2, it is unsurprising that m_γ is most sensitive to the noise parameters from the frequency pair with the lowest energies, but for ℓ_{QG} this is the pair with the largest range of energies.

We find that, for the QG theories, the results are relatively independent of the noise model, and the best-fit model gives a constraint of $|\ell_{\mathrm{QG}}| < 5.3 \times 10^{-18}\, h\, \mathrm{GeV}^{-1}$ at 95% confidence, where $h \equiv H_0/(100\,\mathrm{km\,s^{-1}\,Mpc^{-1}})$ for Hubble constant H_0. We quote our results in terms of h to remain agnostic as to the true value of H_0, given the "Hubble tension." The maximum of the marginalised one-dimensional posterior is, perhaps coincidentally, in the same direction as [111] (accounting for the different sign in the definition), indicating a slight preference for a superluminal QG theory. For the photon mass inference we see that m_γ is correlated with the highest weighted



(a) (b)

Figure 2.1: Constraints on (a) the photon mass, m_γ, and (b) the QG length scale, ℓ_{QG}, and the noise parameters of the Gaussian which is most correlated with the signal. The legend gives the number of redshift independent and dependent Gaussians used to describe the noise, respectively. A QG model with superluminal photon speed at high energies ($\xi = -1$) is defined to have a negative ℓ_{QG}. For the QG theories there is little degeneracy with the noise, whereas m_γ is highly correlated with the mean of the highest-weighted Gaussian, making the constraint on m_γ more sensitive to the noise model employed.

Gaussian for frequency pair $(i, j) = (2, 1)$ and that the marginalised one-dimensional posterior is more sensitive to the noise model. In all cases we find $m_\gamma < 4.0 \times 10^{-5}\, h\,\mathrm{eV}$ at 95% confidence.

Besides the clear proportionality with h, we find that our constraints are relatively independent of cosmological parameters; varying Ω_{m0} in the range $0.25 - 0.35$ with fixed h changes the constraints by $\lesssim 10\%$.

We previously assumed that the uncertainty on the pseudoredshift of a source can be described by a two-tailed Gaussian. To test the impact of this assumption, we run the analysis assuming zero redshift error, and find the constraints on m_γ and ℓ_{QG} tighten by 18% and 3% respectively. Due to the pseudoredshift calibration, we removed all GRBs with pseudoredshifts above $z_{\mathrm{max}} = 4.5$. Increasing this to $z_{\mathrm{max}} = 6$ slightly tightens the constraints by 3% for the massive photon case and 5% for the QG theories. If we include all GRBs from [153] then the constraint on m_γ is again virtually unchanged, whereas we find a nonzero (at 3σ confidence) value of $\ell_{\mathrm{QG}} = -1.6 \times 10^{-18}\, h\,\mathrm{GeV}^{-1}$. We find that this "detection" is driven by two GRBs (4B 910619 and 4B 921112-) at $z \sim 7.5$ which have negative time delays. Upon excising these potential outliers, the constraint is again consistent with zero at 1σ confidence.

2.4 Equivalence Principle Violation

In the previous section we developed a framework of source-by-source, probabilistic forward modelling which we used to constrain two theories which give rise to photon propagation speeds, and we demonstrated how one can incorporate and compare empirical noise models. In those theories, the theoretical time delay was dominated by the geometric contribution, i.e. the time delay depended only on the distance of the source from the observer and not the direction. In this section we utilise the tools we have developed to consider a more complicated problem, where we constrain EP violation through direction-dependent time delays caused by the mass distribution of the local Universe. We describe these time delays in Section 2.4.1 and how we model them in Section 2.4.2. The results are presented in Section 2.4.3 and discussed in Section 2.4.4.

2.4.1 Time delay differences in metric theories of gravity

2.4.1.1 Equivalence principle violation

Consider a perturbed FRW-like metric in the Newtonian gauge,

$$\mathrm{d}s^2 = -\left(1+2\delta\phi\right)\mathrm{d}t^2 + a^2\left(t\right)\left(1-2\gamma\delta\phi\right)\mathrm{d}\boldsymbol{r}^2, \tag{2.16}$$

where we have introduced the PPN parameter γ to allow for deviations from General Relativity (where $\gamma = 1$).

For a massless particle, the gravitational time delay to a source at distance r_s is given by $(1+\gamma)\, t_{\mathrm{grav}}$, where

$$t_{\mathrm{grav}} = -\int_0^{r_s} \mathrm{d}r\, \delta\phi_0(\boldsymbol{r}) D(r), \tag{2.17}$$

where $D(r)$ is the linear growth factor, $\delta\phi_0$ is the potential fluctuation evaluated using the present-day matter field and we have used the weak field limit $\delta\phi \ll 1$. For a set of point masses $\phi = \sum_i GM_i/r_i$, but in a general density field $\delta\phi$ must be found by solving the Poisson equation

$$\nabla^2\delta\phi(\boldsymbol{r}) = 4\pi G\delta\rho(\boldsymbol{r}), \tag{2.18}$$

where $\delta\rho$ is the total matter density fluctuation. This is most readily solved in Fourier space. Thus determining time delays amounts to mapping out the three dimensional density field to at least the redshift of the source, as described in Section 1.3.

There exist several formulations of the Equivalence Principle, of varying strengths. The WEP, our focus here, states that all freely-falling test objects follow the same trajectories given the same initial conditions, irrespective of their composition or structure. This requires that all objects from a given source experience the *same* time delay, regardless of their composition or energy.

A potential contribution to the difference in the time of reception of photons from GRBs is a difference in the gravitational time delay, which can be parametrised as a difference in the PPN parameter γ between two wavelengths. Thus by comparing the measured time delays with the line-of-sight integral in Equation 2.17, one can put constraints on how γ changes with photon frequency, and therefore on any violation of the WEP.

2.4.1.2 Time delay angular power spectrum

The angular power spectrum of the gravitational time delay is [149, 177]

$$C_\ell = \frac{2}{\pi}\int \mathrm{d}k\, k^2 P_\phi\left(k\right)\left|\int_0^{r_\mathrm{s}} \mathrm{d}r\, D\left(r\right) j_\ell\left(kr\right)\right|^2, \tag{2.19}$$

such that the mean-squared time delay is

$$\left\langle t^2_\mathrm{grav}\right\rangle = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell, \tag{2.20}$$

where P_ϕ is the power spectrum of the potential. For $k \ll 1/r_\mathrm{s}$, the spherical Bessel function can be expanded as

$$j_\ell\left(x\right) = \frac{\sqrt{\pi}}{2^{\ell+1}\Gamma\left(\ell+\frac{3}{2}\right)} x^\ell + \mathcal{O}\left(x^{2+\ell}\right), \tag{2.21}$$

so that

$$C_\ell = \frac{4}{\pi^3}\frac{\sqrt{\pi}}{2^{\ell+1}\Gamma\left(\ell+\frac{3}{2}\right)}\int \frac{\mathrm{d}k}{k}\left[T_\phi\left(k\right)\right]^2 A_\mathrm{s} k^{n_\mathrm{s}-1} k^{2\ell}\left(\left|\int \mathrm{d}r\, D\left(r\right) r^\ell\right|^2 + \mathcal{O}\left(k^{2+2\ell}\right)\right), \tag{2.22}$$

where T_ϕ is the potential transfer function and we assume a nearly scale-invariant primordial power spectrum

$$\mathcal{P}_\mathcal{R}\left(k\right) = A_\mathrm{s} k^{n_\mathrm{s}-1}. \tag{2.23}$$

On super-Hubble scales, $T_\phi\left(k\right) \sim$ constant, so the smallest power of k in the integral has an exponent $n_\mathrm{s} - 2 + 2\ell$. For $\ell \geq 1$, this is $\geq n_\mathrm{s}$, which, given that $n_\mathrm{s} \sim 0.97$ [178], means that, for $\ell \geq 1$, the integral does not diverge at small k. However for $\ell = 0$, the exponent is $n_\mathrm{s} - 2 < -1$, and so the integral diverges. We therefore see that the monopole diverges due to the contribution from large scales. This is not to say that there is an infinite time delay for a given universe, but that its variance across all possible universes is infinite.

The problematic diverging monopole was first noted by Reischke *et al.* [179], who showed how to circumvent this issue by computing the angular power spectrum and forecasted constraints of $\Delta\gamma < 10^{-15}$ using FRBs. Similarly, for our forward modelling approach in Section 2.4.2, we will only be able to predict angular fluctuations about the mean Shapiro time delay at a given redshift, and not its absolute value.

2.4.2 Methods

In this section we use the large-scale structure information described in Section 1.3 to produce a source-by-source probabilistic forward model for the expected Shapiro time delay difference for a given $\Delta\gamma_{ij}$. Combining this with an empirical model describing other contributions to the measured time delays (noise) outlined in Section 2.3.2, we calculate the likelihood function and constrain $\Delta\gamma_{ij}$ and the parameters describing the noise via a MCMC algorithm. The parameters which are fixed in this section are summarised in Table 2.2.

2.4.2.1 Calculating the Shapiro time delay

As discussed in Section 2.4.1.2, the variance of the mean Shapiro time delay at a given redshift across all possible universes diverges. We can rephrase our problem so that this is not an issue. Let us decompose the gravitational time delay into two parts

$$t_{\text{grav}}(\boldsymbol{r}) = t_0(r) + \delta t_{\text{grav}}(\boldsymbol{r}), \tag{2.24}$$

where $t_0(r)$ is the mean Shapiro time delay across all angles at a given r (i.e. the monopole), and δt_{grav} gives the fluctuation about this. These fluctuations can be decomposed into the sum of three terms

$$\delta t_{\text{grav}}(\boldsymbol{r}) = \delta t^{(L)}_{\text{grav}}(\boldsymbol{r}) + \delta t^{(R)}_{\text{grav}}(\boldsymbol{r}) + \delta t^{(S)}_{\text{grav}}(\boldsymbol{r}) \tag{2.25}$$

corresponding to the long wavelength (L; $k < k_{\text{min}}$), resolved (R; $k_{\text{min}} \leq k < k_{\text{max}}$) and short wavelength (S; $k \geq k_{\text{max}}$) contributions respectively. The 'resolved' contribution can be determined using the inferred matter fields from `BORG` .

To determine the resolution and box size required to accurately reconstruct the time delay fluctuation, we compare the results of calculating the angular power spectrum with a finite resolution and box size to the continuous, infinite volume case. Obviously the latter is impossible in practice, but we approximate this limit by choosing a sufficiently small minimum ($10^{-5}\,\text{Mpc}^{-1}$) and large maximum ($100\,\text{Mpc}^{-1}$) k. To approximate the finite volume result, we consider a box length L_{box} with N_{box} grid cells along each side, so we have minimum and maximum (non-zero) k of

$$k_{\text{min}} = \frac{2\pi}{L_{\text{box}}}, \quad k_{\text{max}} = \frac{\pi\,(N_{\text{box}} - 1)}{L_{\text{box}}}. \tag{2.26}$$

Note that we do not include a factor of $\sqrt{3}$ in k_{max} as the sphere in k-space of radius k_{max} does not fully fit inside the first Brillouin zone if this is included. To mimic using

Table 2.2: Parameters used to constrain the Equivalence Principle-violating contribution to the time delay. Above the horizontal line are the parameters used to forward model the time delay and below are the parameters passed to MULTINEST in the MCMC analysis. In the final column we give the value chosen for each parameter, although we show in Section 2.4.4 that our results are unchanged for reasonable alternative values.

Parameter	**Description**	**Value**
$L_{\rm box}$ / $h^{-1}\,$Mpc	Side length of box used to reconstruct local density field.	4000
$N_{\rm box}$	Number of grid points per side of box used to reconstruct local density field.	256
k_0 / Mpc^{-1}	Minimum wavenumber used to calculate long wavelength contribution.	10^{-5}
$\ell_{\rm max}$	Maximum multipole used to compute time delay from angular power spectrum.	2000
$N_{\rm side}$	Resolution of HEALPix map used to calculate the monopole.	64
$N_{\rm mon}$	Number of redshifts between $z = 0.1$ and 2 used to calculate the monopole.	20
$N_{\rm z,R}$	Number of redshift points used to calculate the second term in Equation 2.29.	512
$N_{\rm z,L}$	Number of redshift points used to calculate long-wavelength time delay contributions.	1024
$N_{\rm k,L}$	Number of wavenumber points used to calculate unconstrained time delay contributions.	512
$N_{\rm MC}$	Number of Monte Carlo runs to get the distribution of time delays for the template signal for a given density field.	10^3
$N_{\rm bin}$	Number of redshift bins to determine redshift evolution of noise model.	5
$N_{\rm B}$	Number of density field reconstructions sampled from the `BORG` chain.	18
`n_live_points`	Number of live points used in MULTINEST sampling.	800
`importance_nested_sampling`	Whether to use importance nested sampling with MULTINEST.	True
`multimodal`	Whether to allow mode separation in MULTINEST.	True
`evidence_tolerance`	Evidence tolerance for MULTINEST.	0.5
`sampling_efficiency`	Sampling efficiency for MULTINEST.	0.8
`const_efficiency_mode`	Whether to use constant efficiency mode in MULTINEST.	False

this box, we top-hat filter the potential power spectrum, allowing modes between $k_{\rm min}$ and $k_{\rm max}$. We calculate two quantities: (i) the C_ℓ's themselves, and (ii) the RMS fluctuation in the time delay using Equation 2.20, where we find $\ell_{\rm max} = 2000$ to be more than sufficient (halving this to $\ell_{\rm max} = 1000$ does not change the final result).

In Figure 2.2 we compute the C_ℓ's for a source at redshift $z = 0.10$ using the Core Cosmology Library [180]. We compare the results from the continuous case with the SDSS-III/BOSS ($L_{\rm box} = 4000h^{-1}\,{\rm Mpc}$, $N_{\rm box} = 256$) and 2M++ ($L_{\rm box} = 677.7h^{-1}\,{\rm Mpc}$, $N_{\rm box} = 256$) reconstructions. We note that the maximum distance a source can be from the observer in the latter two cases is $L_{\rm box}\sqrt{3}/2$, which is 5.17 Gpc for the SDSS-III/BOSS and 876 Mpc for the 2M++ reconstruction, corresponding to redshifts of 1.9 and 0.21 respectively. Thus our test source at $z = 0.10$ is within both boxes.

The continuous case gives the same result as SDSS-III/BOSS to ~ 4 parts in 10^5 as the C_ℓ's only disagree at $\ell \gg 1$ while the result is dominated by low ℓ. The lack of power at low k results in the 2M++ box underestimating the low ℓ contribution, and thus the RMS time delay. The points of disagreement are as expected, since for this redshift, using the Limber approximation [181] $\ell \sim kr$, we find that $k_{\rm min}$ and $k_{\rm max}$ correspond to multipoles

$$\ell_{\rm min}^{\rm SDSS-III/BOSS} \sim 0.46, \quad \ell_{\rm max}^{\rm SDSS-III/BOSS} \sim 59 \tag{2.27}$$

$$\ell_{\rm min}^{\rm 2M++} \sim 2.7, \quad \ell_{\rm max}^{\rm 2M++} \sim 346, \tag{2.28}$$

which is approximately where we see the results diverging from the continuous case. Since the total time delay is dominated by low ℓ, we conclude that a large box size is far more important than a fine resolution, justifying our choice of the SDSS-III/BOSS over 2M++ reconstruction. For simplicity, we will henceforth refer to the SDSS-III/BOSS `BORG` reconstruction simply as '`BORG` '.

Nonetheless, for sources at higher redshift, the density field not contained within the inferred volume contributes significantly to the gravitational time delay. For sources outside of the `BORG` volume, Equation 2.17 can be split into a contribution from the observer to the edge of the box, $r_{\rm b}$, and then from $r_{\rm b}$ to the source:

$$t_{\rm grav} = -\int_0^{r_{\rm b}} {\rm d}r\, \delta\phi_0(\boldsymbol{r}) D(r) - \int_{r_{\rm b}}^{r_s} {\rm d}r\, \delta\phi_0(\boldsymbol{r}) D(r). \tag{2.29}$$

When we compute the first term (or Equation 2.17 for sources inside the box) we do not have a fluctuation, since the mean time delay at the upper limit of the integral

is not zero. We therefore must find and subtract this mean. We do this by computing the time delays on HEALPix[1] [182, 183] maps with $N_{\rm side} = 64$ at $N_{\rm mon} = 20$ logarithmically-spaced redshifts between $z = 0.1$ and $z = 2$. We then compute the monopoles of these maps with HEALPY. To determine the monopole at some intermediate redshift we linearly interpolate between the sampled redshifts.

The ensemble mean of the resulting time delay fluctuation map is plotted in Figure 2.3. One can see that the typical fluctuation from the resolved contribution is larger than $\sim 10^{11}\,\mathrm{s}$ and thus, given that the measured time delays are typically $\mathcal{O}(0.1\,\mathrm{s})$, one would expect $|\Delta\gamma_{ij}|$ to be smaller than $\sim 10^{-12}$.

If the source is outside the volume, we now add an unconstrained contribution to the fluctuation (but not the monopole) drawn from a Gaussian with a width given by Equation 2.20, but only using k modes accessible to BORG (Equation 2.26) and integrating between $r_{\rm b}$ and $r_{\rm s}$ using $N_{\rm z,R} = 512$ intermediate points. This corresponds to fluctuations due to the second term in Equation 2.29. We find $\ell_{\rm max} = 2000$ to be sufficient for this calculation.

To marginalise over the uncertainties in the pseudo-redshifts, angular position and unconstrained regions, we use simulation-based Bayesian forward modelling to create predictions from the statistical models. In particular, for a given density field from the converged part of the BORG MCMC chain, we take $N_{\rm MC} = 10^3$ Monte Carlo draws from the input distributions to build the likelihood. For each iteration, we draw an angular location from Gaussian distributions characterised by the positional uncertainty given in [153] and a redshift from a two-sided Gaussian using the upper and lower uncertainties determined by [155]. The time delay for a source at this position is then calculated for this density field. From Equation 2.7, we see that the gravitational contribution to the time delay between any two frequencies is proportional to $\Delta\gamma_{ij}$. We therefore only need to run this Monte Carlo procedure once per source to construct a template signal with $\Delta\gamma_{ij} = 1$ since this can be trivially reintroduced as a scaling factor later.

2.4.2.2 Monte Carlo modelling

Now that we have $N_{\rm MC}$ samples per source and per BORG density field of our probabilistic model for $\delta t^{(R)}_{\rm grav}$, we must convert these samples into a distribution to use in our likelihood analysis.

[1] http://healpix.sf.net

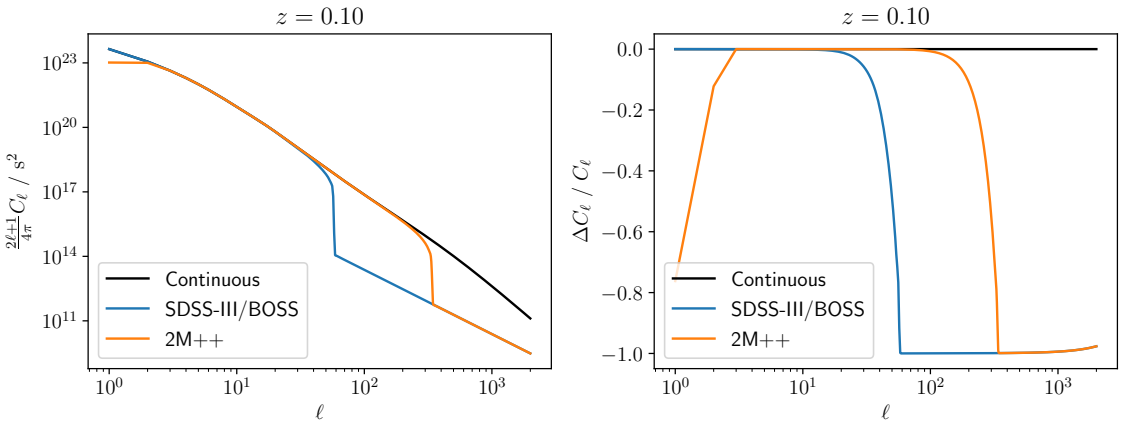


Figure 2.2: The predicted Shapiro time delay fluctuation angular power spectrum for a source at $z = 0.10$ calculated with the Core Cosmology Library. For the SDSS-III/BOSS and 2M++ calculations, we only include k modes within the range sampled in the simulated volume due to the finite box length and resolution. The right hand panel shows the fractional difference between these and the continuous case. Using SDSS-III/BOSS gives practically the same result for the time delay as the continuous case, whereas 2M++ underestimates this by ~ 32 per cent. The result is driven by the smallest ℓ, making large box size far more important than high spatial resolution.

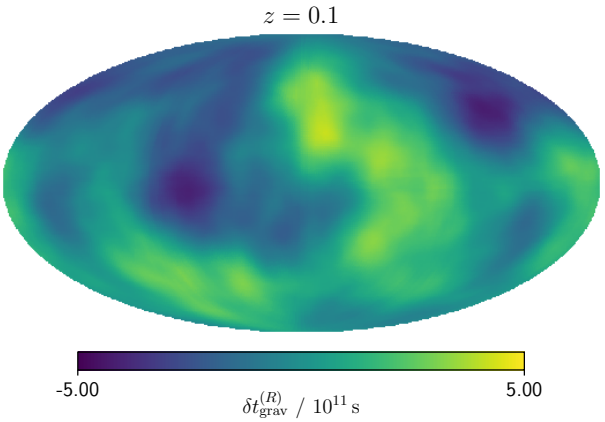


Figure 2.3: Mollweide projection in equatorial coordinates of the ensemble mean of the time delay fluctuations at $z = 0.1$ from resolved wavelengths. The typical scale is $\sim 10^{11}\,\mathrm{s}$ which, since the majority of observed time delays are $\mathcal{O}(0.1\,\mathrm{s})$, indicates that constraints at least as tight as $|\Delta\gamma_{ij}| \lesssim 10^{-12}$ should be possible.

We model the samples as a GMM [170], where the likelihood function for some source s and `BORG` density field b is

$$\mathcal{L}_{sb}\left(\delta t_{\rm grav}^{(R)}|\Delta\gamma_{ij}=1\right)=\sum_{\alpha}\frac{w_{sb}^{(\alpha)}}{\sqrt{2\pi}\tau_{sb}^{(\alpha)}}\exp\left[-\frac{\left(\delta t_{\rm grav}^{(R)}-\lambda_{sb}^{(\alpha)}\right)^2}{2{\tau_{sb}^{(\alpha)}}^2}\right], \tag{2.30}$$

where

$$\sum_{\alpha}w_{\rm sb}^{(\alpha)}=1,\quad w_{\rm sb}^{(\alpha)}\geq 0, \tag{2.31}$$

and the sum runs over the Gaussians. The number of Gaussians is chosen to minimise the BIC (Equation 1.43). Independent Gaussians are obtained for each source and `BORG` density field. To account for a different $\Delta\gamma_{ij}$, we must transform the means and widths of the Gaussians in the GMM as

$$\tilde{\lambda}_{sbij}^{(\alpha)}=\Delta\gamma_{ij}\lambda_{sb}^{(\alpha)},\quad \tilde{\tau}_{sbij}^{(\alpha)}=|\Delta\gamma_{ij}|\,\tau_{sb}^{(\alpha)}. \tag{2.32}$$

2.4.2.3 Adding larger scale modes

Thus far we have only calculated the 'resolved' contribution to the Shapiro time delay fluctuations. Since the large-k contributions are negligible, we do not consider these further. To incorporate the long wavelength modes, we evaluate Equation 2.20 for each source, where we integrate between $k_0=10^{-5}\,{\rm Mpc}^{-1}$ and $k_{\rm min}$ (Equation 2.26) using $N_{\rm k,L}=512$ intermediate points. We integrate up to the three-sigma redshift in the pseduo-redshift distribution with $N_{\rm z,L}=1024$ points and again use $\ell_{\rm max}=2000$. For source s this gives the width of the distribution of long-wavelength contributions, ξ_s, which we assume to be Gaussian distributed since the density should also be Gaussian distributed on large scales. We must convolve this long-wavelength distribution with our GMM to get the total likelihood of $\delta t_{\rm grav}$,

$$\mathcal{L}_{sb}\left(\delta t_{\rm grav}\right)=\sum_{\alpha}\frac{w_{sb}^{(\alpha)}}{\sqrt{2\pi\left({\tilde{\tau}_{sbij}^{(\alpha)}}^2+\tilde{\xi}_{sij}^2\right)}}\exp\left[-\frac{\left(\delta t_{\rm grav}-\tilde{\lambda}_{sb}^{(\alpha)}\right)^2}{2\left({\tilde{\tau}_{sbij}^{(\alpha)}}^2+\tilde{\xi}_{sij}^2\right)}\right], \tag{2.33}$$

where

$$\tilde{\xi}_{sij}\equiv|\Delta\gamma_{ij}|\,\xi_s. \tag{2.34}$$

2.4.3 Results

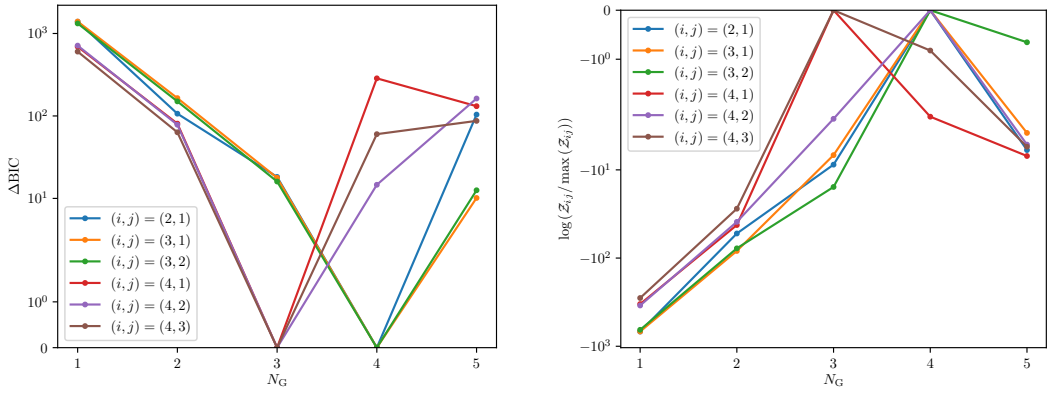
In Figure 2.4 we plot the BIC as a function of the number of Gaussian components in the noise model, $N_{\rm G}$, where we set the number of model parameters equal to the BMD (Equation 1.47). We find that the change in BIC is large when we use non-optimal $N_{\rm G}$, indicating that the best $N_{\rm G}$ is unambiguous. The same trend is found when we fix $\Delta\gamma_{ij} = 0$ (i.e. fit the observations with just the noise model). For $N_{\rm G} = 5$, we find that the weight of the lowest-weighted Gaussian, $\omega_{ij}^{(4)}$, is extremely small: $\omega_{ij}^{(4)} < 0.03$ at 68% confidence for all frequency pairs. It is unsurprising then that the BIC increases for this case, since we are adding three new but negligible parameters to the noise model. Similar results are found if we compare models using the Bayes factor; the only discrepancy is for $(i, j) = (4, 2)$ where the BIC prefers $N_{\rm G} = 3$ whereas the Bayes factor suggests that $N_{\rm G} = 4$ is optimal. The constraints are similar for both models.

For the optimal $N_{\rm G}$ we plot the marginalised one-dimensional posteriors of $\Delta\gamma_{ij}$ in Figure 2.5, and in Figure 2.6 we show the corner plot for the time delay pair with the weakest constraints on $\Delta\gamma_{ij}$. Since the unconstrained large scale contribution is modelled as a Gaussian of zero mean and finite width, if such scales alone were important then the one-dimensional posteriors would be symmetric about $\Delta\gamma_{ij} = 0$. That this is not true for all frequency pairs shows that our constraints on the density field are relevant.

For all time delay pairs we find $\Delta\gamma_{ij}$ to be consistent with zero. The 1σ constraints (defined to be half the difference between the $84^{\rm th}$ and $16^{\rm th}$ percentile) are tabulated in Table 2.3. The tabulated results use the best-fit $N_{\rm G}$; however, we find that these change by no more than 25% across the range $N_{\rm G} = 3 - 5$.

We find that our weakest constraint is for the time delay between the highest and lowest energy channels ($(i, j) = (4, 1)$), where we find $\Delta\gamma_{41} < 2.1 \times 10^{-15}$ at 1σ confidence. This is unsurprising because each pulse of a GRB's lightcurve is known to peak first at higher energy due to spectral evolution [184–187], although the reasons for this are not fully understood [156–162]. This results in the majority of observed time delays obeying $\Delta t_{ij} > 0$ for $i > j$, so one expects the largest time delay for this pair and hence the weakest constraint.

Given the form of our noise model, if $\Delta\gamma_{ij} = 0$ then one expects $\{\tilde{\mu}_{ij}^{(0)}\}$ to be redshift independent. We check this in Figure 2.7 and indeed find that for all pairs any variation in $\{\tilde{\mu}_{ij}^{(0)}\}$ is comparable to its uncertainty, indicating that our assumption is reasonable. Furthermore, the time delay pair giving the weakest constraint



(a) (b)

Figure 2.4: Noise model comparison, including the Equivalence Principle-violating term, using (a) the BIC (Equation 1.43) with the number of parameters of our model equal to the BMD (Equation 1.47), and (b) the Bayes factor (Equation 1.42). The best-fitting model should minimise the BIC and maximise the Bayes factor. For $i < 4$ we see that $N_G = 4$ is the optimal noise model using either statistic. Although the Bayes factor prefers $N_G = 4$ for $(i, j) = (4, 2)$, we adopt $N_G = 3$ for $i = 4$ since this is favoured by the BIC, but find similar constraints on $\Delta\gamma_{42}$ for both $N_G = 3$ and $N_G = 4$.

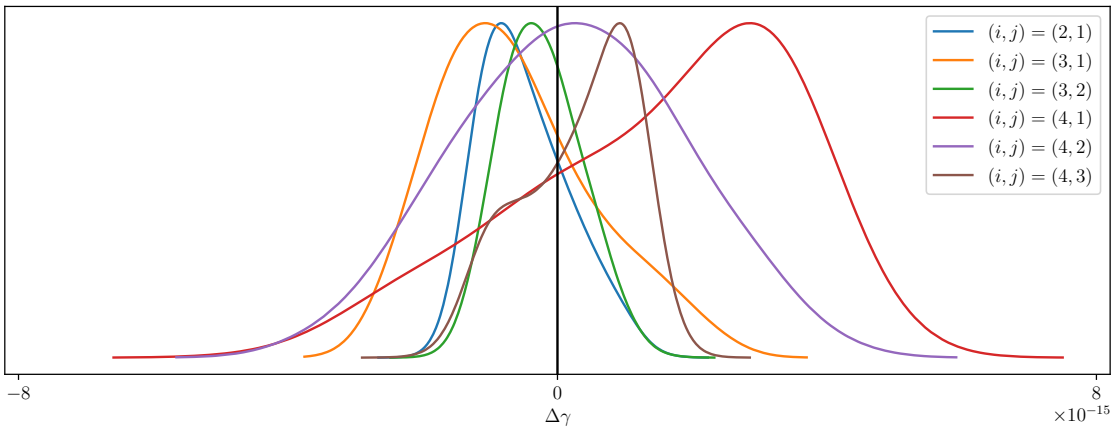


Figure 2.5: Posteriors on $\Delta\gamma_{ij}$ for the different time delay pairs, marginalised over the noise parameters. These constraints use the noise models which minimise the BIC (Equation 1.43).

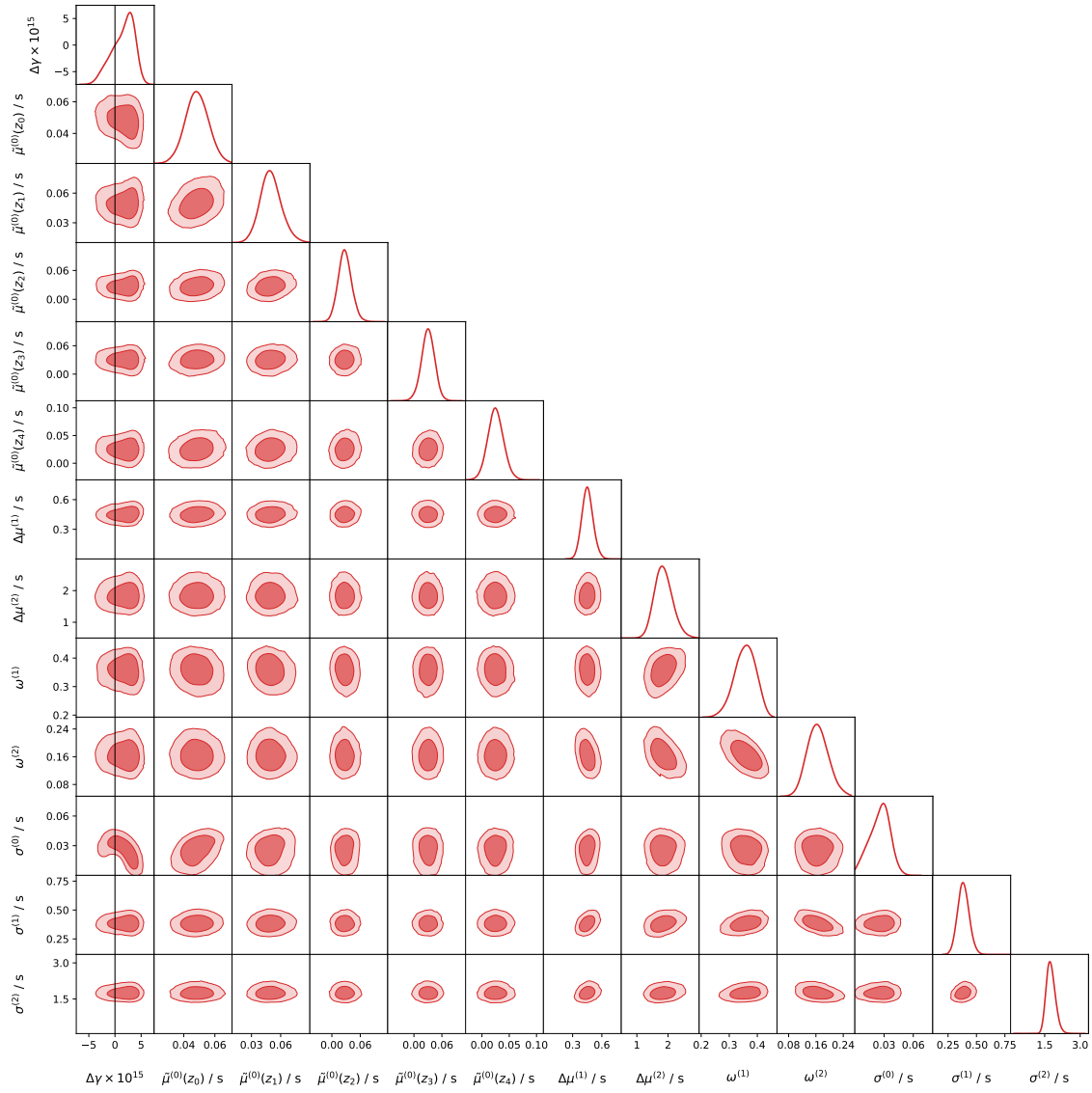


Figure 2.6: Constraints on $\Delta\gamma_{ij}$ and the parameters describing other contributions to the time delay for $(i, j) = (4, 1)$. The ij indices have been suppressed for clarity since all parameters are only for this pair. The contours show the 1 and 2σ confidence intervals.

Table 2.3: The 1σ constraints on $\Delta\gamma_{ij}$ for the different time delay pairs.

(i, j)	Constraint on $\Delta\gamma_{ij} \times 10^{15}$
$(2, 1)$	0.68
$(3, 1)$	1.19
$(3, 2)$	0.62
$(4, 1)$	2.13
$(4, 2)$	1.72
$(4, 3)$	0.92

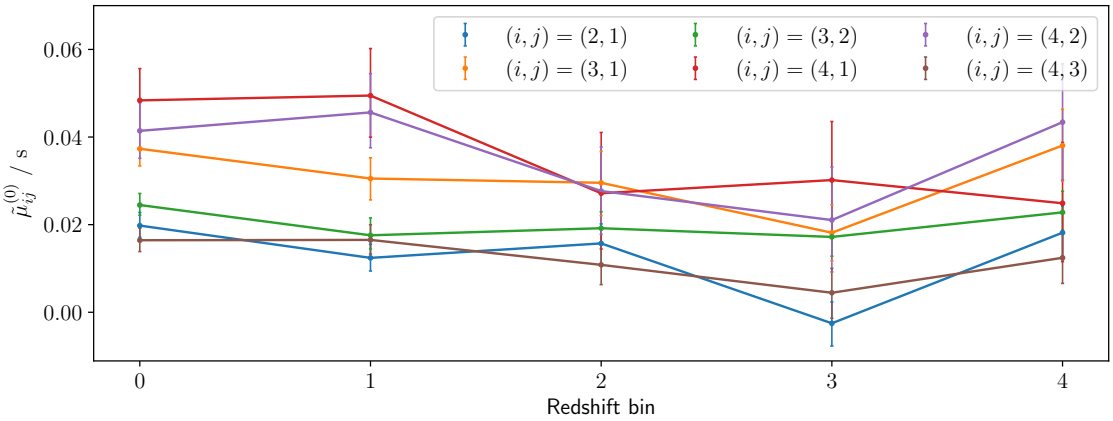


Figure 2.7: Evolution of the means of the highest-weighted Gaussian in the noise model, $\tilde{\mu}_{ij}^{(0)}$, with redshift, where it is assumed that all sources in the same redshift bin have the same $\tilde{\mu}_{ij}^{(0)}$. Any variation with redshift is comparable to the uncertainty on $\tilde{\mu}_{ij}^{(0)}$, as one would expect for a redshift-independent intrinsic time delay contribution, since $\Delta\gamma_{ij}$ is consistent with zero.

$((i,j) = (4,1))$ has noise parameters most suggestive of redshift evolution, as would be expected.

2.4.4 Discussion

2.4.4.1 Systematic uncertainties

Our probabilistic forward model is designed to propagate uncertainties in the source localisation and density field reconstruction, which we marginalise over via a MCMC algorithm. We use $N_{\rm MC} = 10^3$ Monte Carlo samples per source and per BORG density field to estimate the likelihood, but we also check that this is sufficiently large to fully sample the distributions we wish to marginalise over. Running the inference using $N_{\rm MC} = 500$ yields identical constraints on $\Delta\gamma_{ij}$ and the noise parameters as the fiducial case of $N_{\rm MC} = 10^3$, indicating that the number of samples is adequate.

In order to measure the fluctuations in the Shapiro time delay using BORG , we had to subtract the monopole at a given redshift. This involved sampling $N_{\rm mon} = 20$ logarithmically spaced redshifts, where for each redshift we computed the monopole using a $N_{\rm side} = 64$ HEALPix map. In Figure 2.8 we plot the ensemble mean of the inferred monopole, t_0, to subtract from the BORG contribution as a function of redshift, z, and see that t_0 is a smoothly varying function of z, indicating that $N_{\rm mon} = 20$ should be sufficient. We find that if we were to use small values of $N_{\rm side}$ we would calculate the wrong monopole at high redshift, but that the values quickly converge

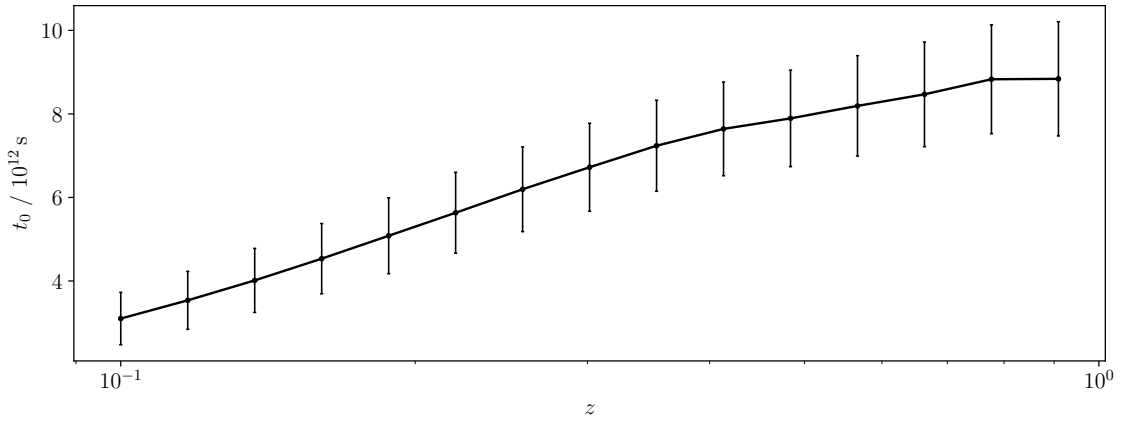


Figure 2.8: Evolution of the ensemble mean of the monopole of Shapiro time delays, t_0, with redshift, z, calculated using wavelengths resolved within the BORG box. Note that the true monopole from all wavelengths cannot be predicted statistically (Section 2.4.1.2). The monopole is found to vary smoothly with redshift, suggesting that our interpolation procedure is reasonable. The ensemble mean is positive due to local massive structures.

with increasing N_{side}. The maximum fractional difference between the calculated t_0 at $N_{\text{side}} = 32$ and $N_{\text{side}} = 64$ is 3×10^{-3}, indicating that our map's resolution is sufficient. Moreover, we run the inference again with $N_{\text{side}} = 32$ and $N_{\text{mon}} = 10$ and find 1σ constraints on $\Delta\gamma_{41}$ of 2.1×10^{-15} and 2.0×10^{-15} respectively, suggesting that our constraints are robust to these choices.

In Section 2.4.2.1, at each Monte Carlo iteration we drew a redshift from a two-sided Gaussian. One may be concerned that this is not the correct distribution. To test the effect of the redshift uncertainty, we repeat the analysis but fix the redshift to the mean value; i.e., we assume zero redshift error. We find the constraints change by $\sim$10% compared to our fiducial method, indicating that our constraints are not dominated by redshift uncertainty, and thus that the exact redshift distribution is not important.

Finally, to account for the monopole of the Shapiro time delay (which we are unable to predict or constrain by itself), we introduce a redshift dependence in our noise model, binning our sources into $N_{\text{bin}} = 5$ linearly spaced redshift bins. Repeating the inference with $N_{\text{bin}} = 1$ or $N_{\text{bin}} = 10$ changes our constraint to $\gamma_{41} < 1.8 \times 10^{-15}$ and $\gamma_{41} < 2.5 \times 10^{-15}$ respectively. This is as one would expect; the constraint tightens as we decrease the number of parameters in the noise model and weakens as this increases. However this change is only 0.1 dex, indicating that our constraint is relatively insensitive to the choice of N_{bin}.

2.4.4.2 Comparison with the literature

The majority of previous attempts to constrain violations of the WEP via $\Delta\gamma_{ij}$ with GRBs [153, 188–191], FRBs [136, 147, 192–194], Supernovae [195, 196], Gravitational Waves [132, 197–204], Blazar Flares [205–209], or Pulsars [210–213] have assumed that the gravitational potential is dominated by the contribution from the Milky Way and/or other massive objects such as the Laniakea supercluster. This has two major shortcomings for distant sources: first, the gravitational potential in Equation 2.17 should be a fluctuation about the cosmological mean (and thus can take either sign, unlike in the multiple-source approximation where it is strictly additive), and second, the long range behaviour of the gravitational potential means that we cannot neglect the large-scale distribution of mass [149–151]. These studies claim that only including a few sources underestimates the Shapiro time delay, making the constraints obtained conservative [102]. However, since the potential can take either sign, this reasoning is incorrect. Furthermore, the modelling of other contributions to the time delay in these studies are relatively simplistic compared to our analysis; for example, Yu *et al.* [153] assume all noise can be modelled as a single Gaussian, which we have shown to be a poor approximation. Not only have we considered more sophisticated noise models but also we have demonstrated that our constraints are robust to the choice of model.

As well as being more robust than previous work, our constraints are also stronger, and are comparable to the forecasts of [179] for FRBs. The first attempts to include the cosmological contribution to the time delays yielded constraints weaker than $\Delta\gamma_{ij} < 10^{-13}$ [149, 214]. Using the same sources as us and the point-mass methodology, Yu *et al.* [153] find $\Delta\gamma_{41} < 1.3 \times 10^{-13}$, which is a factor of $\sim$40 weaker than our constraints. As a consistency check, we re-run our inference for $(i, j) = (4, 1)$ but (incorrectly) assume that the gravitational potential is dominated by the Laniakea supercluster (of mass $M = 10^{17} M_{\odot}$, at a distance $d = 79\,\mathrm{Mpc}$, with RA=$10^{\mathrm{h}}32^{\mathrm{m}}$ and Dec=-46°00′) as is done by Yu *et al.* [153]. Also mimicking Yu *et al.* [153], we now assume that the redshifts and angular positions of the sources have no uncertainty. We find constraints of $\Delta\gamma_{41} < 1.1 \times 10^{-14}$, which are tighter than [153]. This is due to our use of the optimal noise model with $N_{\mathrm{G}} = 3$; if we use $N_{\mathrm{G}} = 1$ then our constraint is $\Delta\gamma_{41} < 1.1 \times 10^{-13}$, similar to that of [153]. One might have expected that increasing the number of Gaussian components describing the noise would weaken the constraint, not strengthen them. We find that for $N_{\mathrm{G}} = 1$ the constraint is dominated by the few sources with the largest time delays. This is because the WEP-violating term must account for the wide tails in the measured time delays, upweighting larger

values of $|\Delta\gamma|$. When allowing $N_{\rm G} > 1$, the broader Gaussians capture the tails instead, favouring smaller $|\Delta\gamma|$. This is preferable behaviour because the constraint on $\Delta\gamma$ should come from the angular correlation of the measured time delays with those predicted by `BORG` , rather than from the width of the measured time delay distribution itself.

The long Shapiro time delays for extragalactic sources result in tight constraints on $\Delta\gamma_{ij}$; however, one can only measure the time delay *differences* and not the absolute time delays; we constrain $\Delta\gamma_{ij}$ and not γ_i. Therefore, although our constraints may appear tighter than Solar System measurements from the *Cassini* spacecraft [144] or Very Long Baseline Interferometry [215, 216] of $\gamma - 1 = (2.1 \pm 2.3) \times 10^{-5}$ and $(-0.8 \pm 1.2) \times 10^{-4}$, respectively, these have the advantage of constraining the PPN parameter itself and thus can differentiate between different theories that obey the Equivalence Principle. Nonetheless, our constraints indicate that the Equivalence Principle should be obeyed to within 2.1×10^{-15} for photons in the energy range $25 - 325\,\mathrm{keV}$.

2.4.4.3 Further applications

In this section we have used the `BORG` reconstruction of the SDSS-III/BOSS galaxy compilation to predict the Shapiro time delay to GRBs. The mean pseudo-redshifts of the sources compiled by Yu *et al.* [153] is ~ 2, and thus we had to include unconstrained contributions to the calculation between the edge of the constrained volume and the sources (Equation 2.29). This, coupled with the importance of long wavelength modes (Section 2.4.2.1), suggests that our constraints could be improved by the next generation of galaxy surveys, such as with *Euclid* [217] or the Rubin Observatory [218], which will be sensitive up to $z \sim 2$. To estimate the potential improvement, we run the end-to-end inference assuming that we can accurately reconstruct the density field up to $z = 1.26$ and to twice this redshift by generating Gaussian Random Fields. We find that the constraint improves by a factor of $\sim 30\%$ with this extra information. By combining the SDSS-III/BOSS and 2M++ reconstructions one could also reduce the uncertainty of the low-redshift part of the calculation, since 2M++ provides better constraints on the local density field.

We note that in this section we have used $\Delta\gamma_{ij}$ as a phenomenological parameter with which to quantify WEP violation of photons through Equation 2.17. This is an arbitrary choice, and one could equally phrase equivalence principle violation as a difference in e.g. the gravitational constant [219]. Theories that predict a non-null signal in our test (e.g. massive photons) may have $\gamma = 1$ in the usual PPN sense.

In these cases, and neglecting cosmological redshift for illustrative purposes, one would equate $\Delta\gamma_{ij} = \Delta v_{ij}$ and thus for the above models our bound on $\Delta\gamma_{41}$ would correspond to $m_\gamma \lesssim 10^{-6}\,\mathrm{eV}$ and $E_{\mathrm{QG}} \gtrsim 10^{11}\,\mathrm{GeV}$, if $E_1 = 25\,\mathrm{keV}$ and $E_4 = 325\,\mathrm{keV}$. For such models the geometric term dominates the time delay, so it is unsurprising that these bounds are not stronger than those obtained in Section 2.3.

Strong lensing of distant objects smears their light into an Einstein ring. The paths of photons observed across the ring trace out two cones that intersect at the lens plane. The relative time delays of these photons therefore contain information not only on the distribution of mass in the lens that sources a large part of the potential, but also on the path length itself which the time delay is proportional to. This length is a function of H_0, enabling this fundamental cosmological parameter to be constrained by measuring the time delays across the ring [145, 146]. As in this work, the modelling is potentially sensitive to mass further away from the geodesics that the photons follow than either the lens itself or the few additional massive objects modelled in [145]. Provided one has sufficient spatial resolution to determine time delay differences across the Einstein ring, by using the time delay maps from constrained density fields one can test the validity of the external source assumptions.

Similarly, in GR photons and gravitational waves are predicted to follow the same geodesics. Recently, by considering the effect of four massive halos along the line of sight, it was shown [220] at $> 5\sigma$ confidence that GW170817 [148] underwent gravitational lensing. As with the case of measuring H_0, one could use constrained density fields to determine the impact of mass away from the line of sight when computing the time delay.

In this chapter we have only considered the time delay *differences* between different frequencies from a given source. The only cosmological regime in which the time delay itself may be directly measurable is the CMB, where different delays between different regions of the sky imply varying times of recombination, and correspondingly varying temperature of the blackbody radiation [221]. Standard autocorrelation techniques cannot currently detect such a delay; however, cross-correlation with other fields appears to be a promising avenue [222]. We leave it to future work to determine whether the large scale structure information from the `BORG` algorithm could afford a detection of this phenomenon.

2.5 Conclusions

In this chapter we have considered two theories in which the photon propagation speed depends on energy: a quadratic correction due to nonzero photon mass, m_γ, and a QG scenario in which the photon speed depends linearly on energy, as is expected in a wide range of models. By forward-modelling the expected time delays of photons of different frequencies for a large sample of GRBs, we find constraints on the photon mass of $m_\gamma < 4.0 \times 10^{-5}\, h\,\mathrm{eV}$ and on the QG length scale of $|\ell_{\mathrm{QG}}| < 5.3 \times 10^{-18}\, h\ \mathrm{GeV}^{-1}$ at 95% confidence. Our constraints on m_γ are significantly less stringent than those from radio observations, however they are much less sensitive to the effects of dispersion by electrons, which has the same frequency dependence in the dispersion relation as a massive photon.

We then constructed a source-by-source, Monte Carlo-based forward model for the Shapiro time delay from GRBs detected by the BATSE satellite. Most previous attempts to calculate the Shapiro delay in order to constrain EP violation through $\Delta\gamma_{ij}$ have assumed that the gravitational potential is dominated by only a few local sources and incorrectly argue that this produces conservative constraints. We worked in a cosmological context by combining the constrained local density field determined using the `BORG` algorithm with unconstrained, long-wavelength modes. Propagating uncertainties in the density field reconstruction via Monte Carlo sampling and marginalising over an empirical model characterising other contributions to the time delay, we derive constraints on $\Delta\gamma_{ij}$ between the four energy channels, and for all pairs find constraints at least as tight as $\Delta\gamma_{ij} < 2.1 \times 10^{-15}$ at 1σ confidence. These constraints are a factor $\sim$30 times tighter than previous results that use a cosmological model and $\sim$40 times tighter than if one neglects the cosmological contribution.

A large number of previous attempts to constrain QG with the spectral lag of GRBs assume a simple noise model in which the non-QG contribution to the time delay is proportional to $(1 + z)$ and is constant for all sources, even though Ellis *et al.* [110] demonstrated that ignoring stochasticity dramatically changes the results. Moreover, these studies often only use a small sample of GRBs (sometime only one), but one requires a statistical sample to provide trustworthy constraints. Our constraints are among the tightest astrophysical constraints on QG which use multiple sources (see Table 1 of [102]) and we have demonstrated that these are robust to how one models other astrophysical and observational contributions to the spectral lag. Our constraints are comparable to Ellis *et al.* [105], who use the irregularity, kurtosis and skewness of GRBs instead of spectral lag to find $E_{\mathrm{QG}} \equiv |\ell_{\mathrm{QG}}|^{-1} \gtrsim \mathrm{few} \times 10^{17}\,\mathrm{GeV}$.

It is expected that detecting GRBs at >100 GeV should be routine in the future [223]; with more, higher energy measurements one should begin to probe $E_{\rm QG}$ near the Planck energy, $E_{\rm P}$. Since one expects $E_{\rm QG} \sim E_{\rm P}$, either a nonzero or null detection of LI violation at these scales will significantly constrain which QG theories are allowed. With very few other known tests of QG, it is therefore important that future work should develop more theoretically motivated noise models for GRB spectral lag than we have used here to ensure that any detection or rejection of new physics is not due to incorrect modelling of the astrophysical processes governing GRB emission.

Our modelling is applicable to alternative multi-messenger probes of the WEP, although these may require different models for the other contributions to the time delays. For example, for FRBs one would need accurate maps of the electron density of the universe to forward model the contribution of scattering from the electron plasma, which depends on the integrated electron density (and is thus direction dependent) and scales as ν^{-2}. Furthermore, cosmological calculations of the Shapiro time delay can determine the accuracy of lens modelling when considering time delays across the Einstein ring of a strongly lensed source or find use in the analysis of the CMB. Identifying potential systematics in these probes of the Universe is vitally important, especially in the context of the Hubble tension [224].

Chapter 3

Spatially offset black holes in ΛCDM

3.1 Introduction

Most galaxies are now known to harbour SMBHs near their centres. The strong correlations between SMBH mass and galactic properties such as velocity dispersion [225–227] and stellar and bulge mass [228, 229] show that they are far from passive onlookers in their hosts' evolution, but rather play a key role in shaping the galaxy population.

The high bolometric luminosities of active galactic nuclei (AGN) are due to accretion of matter onto SMBHs. Tens of thousands of AGN have been discovered in the central regions of galaxies, although the precise location of the black hole (BH) need not be coincident with the galactic centre. Upon the merger of two galaxies a BH binary may form near the centre of the merged system, which can coalesce due to stellar and gaseous interactions [230]. Gravitational wave emission upon coalescence can, by linear momentum conservation, cause the centre of mass to recoil [231, 232], thus resulting in a BH offset from the galactic centre. It should be possible to observe the coalescence of these binaries in the early Universe with the *Laser Interferometer Space Antenna (LISA)* [233]. Other processes can result in this phenomenon such as three-body interactions between BHs if there are two successive mergers [234, 235], or subhalo accretion, which transfers energy to the BH by dynamical friction, resulting in offsets of tens of parsecs [236]. A population of offset and wandering BHs [237] is therefore expected.

The same three-body interactions that can physically eject BHs from galaxies can result in numerical artefacts in cosmological simulations through numerical heating of BH particles [238]. This is especially prevalent when the BH and DM particles

have similar masses. Many simulations implement expedient but unphysical schemes to alleviate this problem, such as 'teleporting' the BH back to the local potential minimum at each time step. This is not the case however in the Horizon-AGN (H-AGN) simulation, which uses a more physically motivated model – dynamical friction between the BH and surrounding gas – to impose a drag force on recoiling BHs [239, 240]. By studying the properties of offset BHs in this simulation, we develop both a physical picture of the evolution of individual systems and a statistical sense of the behaviour of the entire population. This will enable us to assess the efficacy of the dynamical friction model.

Knowledge of the location of central BHs is essential for models of galaxy formation. The number of wandering BHs depends on the degree of dissipation in galaxy mergers, which in turn determines the BHs' evolution through, and hence scatter around, the $M_{\rm BH} - \sigma$ relation [241]. The lower gas densities around an off-centre BH restricts BH accretion, which could result in a lower mass BH than one constrained to reside at the centre [242]. The lower accretion rate would quench BH feedback, and hence reduce the impact on the surrounding gas, stars and DM. This is particularly important in attempts to resolve small-scale problems in ΛCDM such as the core-cusp problem through BH feedback [236, 243]. The location of SMBHs is also important to the study of DM microphysics: self-interacting dark matter (SIDM) [244, 245], for example, lowers central DM densities in halos and hence lengthens dynamical friction timescales and increases the fraction of off-centre BHs [246]. SIDM also suppresses BH growth and feedback, allowing higher star formation rates in ΛSIDM galaxies than their ΛCDM counterparts [247].

The structure of this chapter is as follows. In section 3.2 we introduce the H-AGN simulation, and in section 3.3 we summarise the observational datasets against which we will compare the simulation results. We outline the methods used to make this comparison in section 3.4 and present the results in section 3.5. These are discussed in section 3.6 and our conclusions are given in section 3.7.

3.2 Horizon-AGN

H-AGN[1] [239] is a large-volume cosmological hydrodynamical simulation, run with the Adaptive Mesh Refinement code, RAMSES [95]. The adopted standard ΛCDM cosmology is compatible with a WMAP-7 cosmology [248] and thus has total matter density $\Omega_{\rm m} = 0.272$, dark energy density $\Omega_\Lambda = 0.728$, amplitude of the matter power spectrum

[1] `http://www.horizon-simulation.org/about.html`

$\sigma_8 = 0.81$, baryon density $\Omega_\mathrm{b} = 0.045$, Hubble constant $H_0 = 70.4\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, and power-spectrum slope $n_\mathrm{s} = 0.967$. Throughout this chapter we use WMAP-7 parameters, to remain consistent with H-AGN. The size of the box is $L_\mathrm{box} = 100h^{-1}\,\mathrm{Mpc}$ and contains 1024^3 DM particles, resulting in a DM mass resolution of $M_\mathrm{DM} = 8\times10^7\mathrm{M}_\odot$.

The simulation incorporates prescriptions for background UV heating, gas cooling (including the contribution from metals released by stellar feedback) and feedback from stellar winds and type Ia and type II supernovae assuming a Salpeter initial mass function (IMF) [249, 250]. Star formation follows a Schmidt law with a 1 per cent efficiency [251] and a star formation density threshold of $n_0 = 0.1\,\mathrm{H\,cm^{-3}}$.

A cell is refined up to an effective physical resolution of $\Delta x = 1\,\mathrm{kpc}$, with a new refinement level added if the mass in a cell is more than 8 times that of the initial mass resolution. The force softening scale is $\sim 2\,\mathrm{kpc}$.

3.2.1 Black hole formation, growth and feedback on ambient gas

If the combined gas and stellar density exceeds the threshold for star formation in a cell and if the stellar velocity dispersion within that cell surpasses $100\,\mathrm{km\,s^{-1}}$, a BH is created with an initial seed mass of $10^5\,\mathrm{M}_\odot$. A BH cannot form if this occurs within 50 comoving kpc of another BH, preventing many BHs from forming in the same galaxy. BHs grow through mergers and accretion, with the accretion rate given by the Bondi-Hoyle-Littleton rate multiplied by a dimensionless boost factor [252]

$$\alpha = \begin{cases} (\rho_\mathrm{gas}/\rho_0)^2 & \text{if} \quad \rho_\mathrm{gas} > \rho_0 \\ 1 & \text{otherwise,} \end{cases} \tag{3.1}$$

for gas density ρ_gas, and is capped at the Eddington rate with an assumed radiative efficiency of $\epsilon_\mathrm{r} = 0.1$ for the Shakura and Sunyaev [253] accretion onto a Schwarzschild BH. ρ_0 is the mass density of hydrogen if its number density is $n_0 = 0.1\mathrm{H\,cm^{-3}}$. This boost factor accounts for the inability to model the colder and higher density regions of the interstellar medium.

The AGN feedback is a combination of two different modes: the 'radio mode' for $\chi < 0.01$ and the 'quasar mode' otherwise, where

$$\chi = \frac{\dot{M}_\mathrm{BH}}{\dot{M}_\mathrm{Edd}}, \tag{3.2}$$

for BH and Eddington accretion rates $\dot{M}_\mathrm{BH}$ and $\dot{M}_\mathrm{Edd}$ respectively. The quasar mode isotropically ejects thermal energy with deposition rate $\dot{E}_\mathrm{AGN} = \epsilon_\mathrm{f}\epsilon_\mathrm{r}\dot{M}_\mathrm{BH}c^2$ into the

gas within a sphere of radius Δx. The efficiency $\epsilon_{\rm f}$ is taken to be 0.15 as this reproduces the correlations between BHs and galaxies and the BH density in our local Universe [see 254]. In contrast, the radio mode releases the feedback energy into a bipolar, cylindrical outflow with height $2\Delta x$ and radius Δx as in Omma *et al.* [255]. The jet velocity is $10^4\,\mathrm{km\,s^{-1}}$ and the radio mode has an increased efficiency, with $\epsilon_{\rm f} = 1.3$. Note that 'radio mode' does not necessarily mean 'radio loud'. For a detailed analysis of how the H-AGN simulation compares to observational radio data see Slyz *et al.* [256].

Due to the finite resolution of the simulation, for the rest of the chapter we ignore BHs with masses $M_{\rm B} < 2 \times 10^7\,\mathrm{M_\odot}$ [see 240].

3.2.2 Dynamical friction

Simulating the dynamical friction, which ensures the BH's trajectory decays towards the centre of the galaxy [257, 258], is notoriously challenging for cosmological simulations because the gas cannot be tracked all the way down to the BH [259]. Some simulations [260–262] therefore anchor the BH to the centre of their DM halos. Offset BHs might not efficiently accrete since high density gas tends to be centrally located [263], so the feedback of such BHs is quenched [236]. Consequently, pinning the BH to the halo centre can result in unrealistic BH and galactic evolution.

Artificial advection schemes exist to overcome this, but these are not without their drawbacks [see 242]. It is necessary to introduce a sub-grid model for the dynamical friction induced by the gas surrounding the BH. In H-AGN dynamical friction is modelled as

$$F_{\rm DF} = f_{\rm gas} 4\pi\alpha\rho_{\rm gas}\left(\frac{GM_{\rm BH}}{\bar{c}_s}\right)^2, \tag{3.3}$$

where $\bar{c}_s$ is the average sound speed and the coefficient $f_{\rm gas} \in [0, 2]$ is a function of the Mach number $\mathcal{M} = \bar{u}/\bar{c}_s < 1$ [264, 265], where $\bar{u}$ is the average velocity of the gas relative to the BH. The average density and sound speed around the BH are computed using kernel weighting of neighbouring cell values, whereas the average relative velocity is set to a constant value typical of gas velocity dispersion in the ISM, $\bar{u} = u_{\rm max} = 10\,\mathrm{km\,s^{-1}}$, as described in Dubois *et al.* [254].

3.2.3 Assigning black holes to galaxies

Since the BHs are not labelled by their host halo number in the simulation, we must do this in post-processing. In this section we outline how this is achieved. Halos and galaxies are identified using the ADAPTAHOP structure finder [266, 267] applied

to DM and star particles, respectively. In both cases, we need a minimum of 50 particles and the smoothed density field, calculated using the 20 nearest neighbours, must exceed 178 times the mean total matter density [268]. The centre is taken as the position of the densest particle, after a shrinking sphere approach is used [269]. Therefore the halo centre is defined to be the position of the densest DM particle, and the galactic centre is at the location of the densest star particle.

We look for galaxies within 10 per cent of the virial radius, $r_{\rm vir}$, of a DM halo and match the most massive unassigned galaxy to that halo [270] to produce a galaxy+halo structure. We consider each halo in turn, moving from the most to least massive. A BH is assigned to a galaxy+halo structure if it is within twice the effective radius, $r_{\rm eff}$, of the galaxy and 10 per cent of $r_{\rm vir}$ of the halo. At this point, some of the galaxy+halo structures contain multiple BHs, since the galaxies contain up to 59 BHs within $2r_{\rm eff}$ of their centres. To decide which is the central BH, we can make one of several choices

1. Select the most massive BH [as in 240].

2. Ignore any BH with $L < L_{\rm cut}$, where $L = \epsilon_{\rm r}\dot{M}_{\rm BH}c^2$ and $L_{\rm cut}$ is some cut-off luminosity to be decided. Select the BH from the remaining candidates that is closest to the centre of the galaxy. This is inspired by Volonteri *et al.* [240], who find that higher luminosity BHs tend to reside closer to galaxies' centres, so we expect this cut to preferentially choose BHs near the centre.

3. Simply select the BH closest to the centre of the galaxy.

In each case we work hierarchically through the halos, going from the most to least massive. Once a BH has been assigned to a galaxy+halo structure, it is removed from the list of available BHs.

We can now calculate the distance between the centre of the galaxy and its central BH, $r_{\rm GB}$, using the three dimensional information. We can also project the offsets onto the plane of the sky according to an observer at the centre of the simulation volume, to determine the two-dimensional offset that would be observed.

We start by investigating the effects of the various selection procedures since conclusions made about the offset population are sensitive to the way we assign BHs to galaxies. For the remainder of this section, we use the $z = 0.1$ output for H-AGN and consider the three-dimensional offsets.

To decide what to use as $L_{\rm cut}$, in Figure 3.1 we plot the BH luminosity distribution. Since we are interested in the high-luminosity region, as we expect high-luminosity

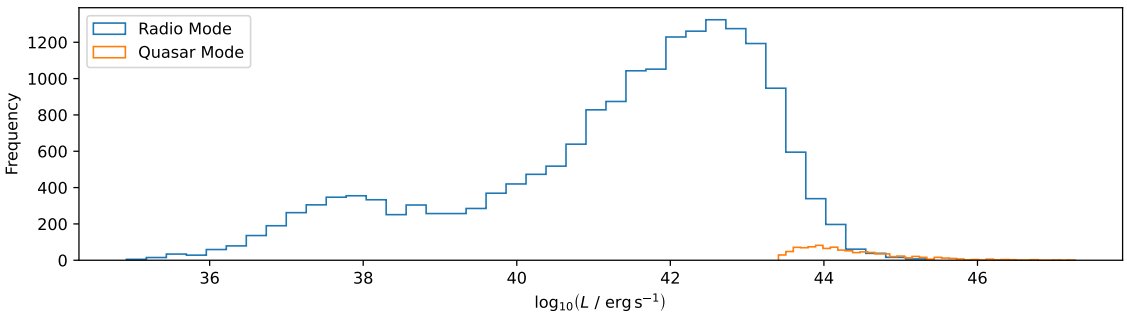


Figure 3.1: Distribution of the luminosity, L, of BHs in H-AGN at $z = 0.1$ with masses $M > 2 \times 10^7\,\mathrm{M}_\odot$. The distribution is bimodal with peaks at $L \sim 10^{38}$ and $10^{43}\,\mathrm{erg\,s^{-1}}$. BHs only operate in the quasar mode at very high luminosity and are greatly outnumbered by those in the radio mode.

BHs to reside near the centre [240], we perform two cuts near the peaks of the bimodal distribution, at $L_{\mathrm{cut}} = 10^{38}$ and $10^{43}\,\mathrm{erg\,s^{-1}}$.

In Figure 3.2 we plot the halo-galaxy, r_{HG}, and BH-galaxy, r_{GB}, distances in terms of r_{eff} with and without a luminosity cut. It is clear that for $L_{\mathrm{cut}} = 10^{43}\,\mathrm{erg\,s^{-1}}$, we preferentially select BHs closer to the centre of the galaxy than without this cut. In particular, $r_{\mathrm{GB}} < r_{\mathrm{eff}}$ for most BHs and the sharp cut-off observed at $r_{\mathrm{GB}} = 2r_{\mathrm{eff}}$ is less visible with this luminosity cut. We also see in Figure 3.2 that if we assign the most massive BH as the central one, the distribution is almost identical to if we selected the closest BH without a luminosity cut. 99.3 per cent of galaxies which have a central BH using the 'closest' method have the same central BH in the 'most massive' method. If use $L_{\mathrm{cut}} = 10^{38}\,\mathrm{erg\,s^{-1}}$, the distribution is almost identical to those without a luminosity cut.

It may seem paradoxical that the 'closest' method without a luminosity cut is more likely to select BHs that are further out than when we impose a non-zero L_{cut} (which does the same thing, however only after removing low-luminosity BHs) and is comparable to the 'mass' method, which only uses BH mass. The reason is to do with which galaxies are retained in the sample. With the 'closest' method all galaxies with BHs are retained, while the 'L_{cut}' method only keeps galaxies with particularly luminous BHs. Since more luminous BHs tend to lie closer to their host galaxies' centres, the L_{cut} method preferentially selects galaxies with small offsets. While the 'closest' method will select the same BH in those galaxies (or one even closer to the centre), for the remaining galaxies with less luminous BHs the offsets are larger.

We therefore see that simply using the selection criteria of [240] produces a large tail of low luminosity BHs in the galaxy-BH offset distribution and that a cut is

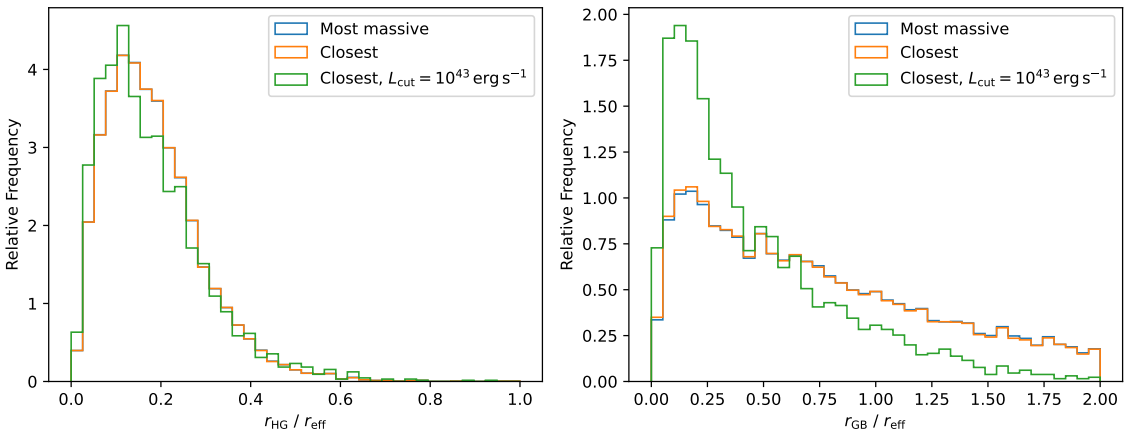


Figure 3.2: Distribution of halo-galaxy, $r_{\rm HG}$, and galaxy-BH, $r_{\rm GB}$, offsets in H-AGN at $z = 0.1$ as a multiple of the galaxy's effective radius, $r_{\rm eff}$, for systems obeying one of three selection cuts. In one case we select the most massive BH and in the other cases we select the BH closest to the galactic centre. For the case with a luminosity cut, $L_{\rm cut}$, we reject all BHs with luminosity $L < L_{\rm cut}$ before making the assignment. The galactic and halo centres are always close, whereas we can have very different offset BHs if we do not introduce a sufficiently stringent luminosity cut. The values of the halo-BH offsets, $r_{\rm HB}$, are very similar to $r_{\rm GB}$ and are thus not plotted. There is little difference between choosing the most massive or closest BH as the central BH.

necessary to preferentially select the BHs near the centre of the galaxy. We use a range of selection criteria in this work. Our quasar and radio mode samples (see subsection 3.5.1) select the closest BH to the centre of the galaxy and impose cuts on accretion efficiency and luminosity. We detail the selection cuts used to compare to observational data in section 3.3.

3.3 Observational data & simulation selection cuts

Offset BHs have been observed through an array of techniques up to redshift ~ 1.5. In Figure 3.3 we plot the BH displacement with respect to the centre of the host galaxy (e.g. the brightest part of the optical emission) projected onto the plane of the sky as a function of redshift for a number of observational samples. In this section we describe the various datasets and the region of the $r_{\rm GB} - z$ plane they are sensitive too, and these are summarised in Table 3.2. Where appropriate, we also detail the selection cuts made to the H-AGN sample in order to mimic those observations, as summarised in Table 3.1. We find that our AGN cut, where we only select BHs with $\chi > 0.01$, makes the samples almost identical, with each containing

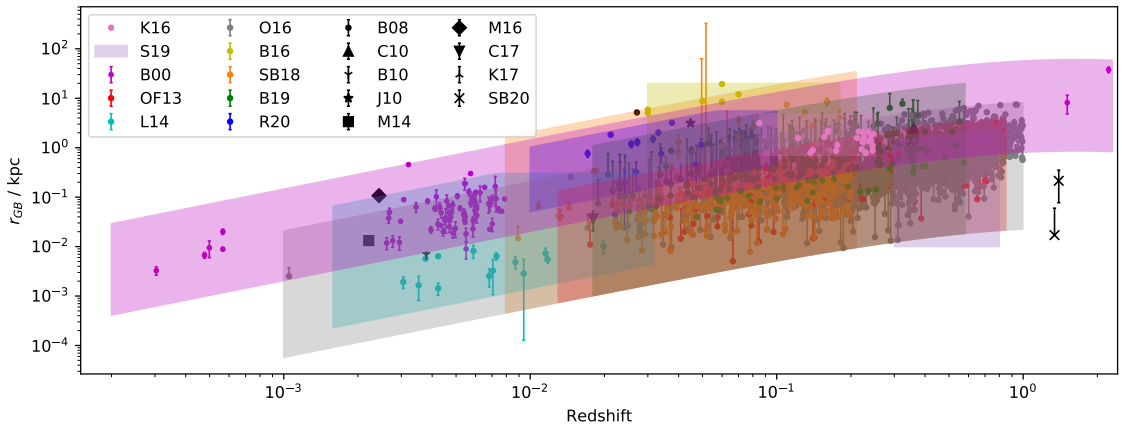


Figure 3.3: Offset as a function of redshift for various observational datasets. The shaded regions indicate the approximate the regions each dataset is sensitive to, based on resolution and cuts made to the data. The OF13, SB18 and B19 data only show upper errorbars since the lower errorbars are of the same magnitude, but the logarithmic y axis scale makes the plot confusing if the lower errors are also included. The lower error bar for one of the SB20 points has also been omitted as the offset is consistent with zero and this cannot be displayed with a logarithmic y scale. The S19 region does not contain any points here, although in reality there are 8210 measurements (not publicly available). Black points correspond to studies of only one or two systems.

≤ 533 systems. The remaining cuts for specific samples remove up to an additional few tens of systems.

We note that offsets between the point identified as the BH and the galactic centre could be 'true', where systems with large true offsets are likely to be mergers [271]. However, misassociation, extended sources, double or lensed quasars, statistical outliers due to an extended tail [272], or the presence of a jet [273, 274] could also be responsible. Even smooth elliptical dust lanes could produce an apparent offset. Any values of the observed fraction of offset BHs should therefore be considered as upper limits for the intrinsic offset fraction, and this necessitates statistical methods to disentangle true vs merely apparent offsets.

3.3.1 Binggeli et al. 2000 (B00)

Binggeli *et al.* [275] investigated a sample of 78 'nucleated' dwarf galaxies (dE,N) in the Virgo cluster using a previous photometric study [276, 277]. They compare the position of the nucleus (brightest pixel) to the optical centre of the galaxy. The observed relation between the nuclear magnitude and ellipticity is predicted to be due

Table 3.1: Selection criteria of the observational datasets that we use. We apply these also to H-AGN to allow a fairer comparison. $L_{\rm R}$ and $L_{\rm X}$ are the minimum allowed radio and X-ray luminosities, respectively.

		Minimum luminosity / $\mathrm{erg\,s^{-1}}$		
	AGN cut	$L_{\rm R}\,(5\,\mathrm{GHz})$	$L_{\rm R}\,(8.46\,\mathrm{GHz})$	$L_{\rm X}$
OF13	✓			
L14	✓	1.0×10^{36}		
O16	✓			
B16	✓			1.0×10^{42}
SB18	✓		1.4×10^{38}	
B19	✓			1.0×10^{42}
R20	✓			

Table 3.2: Summary of observational datasets used in this work. The number of objects may be smaller than given in the reference due to additional selection cuts made here. The resolutions are approximate, and are a combination of quoted, mean and fitted values. We provide the astrometric precision in the source plane for the SB20 sample, hence the small value. We do not give a resolution for S19, since these data are upper limits in the range $\sim 5 - 20\,\mathrm{kpc}$, nor for K16, since these are velocity offsets that are only approximately converted to physical offsets. The final columns indicates the wavelength used to determine the position of the BH.

Name	Reference	Objects	Resolution ($''$)	Redshift	BH Position
B00	Binggeli *et al.* [275]	78	0.42	$z \sim 5 \times 10^{-3}$	Optical
OF13	Orosz and Frey [278]	233	0.05	$0.01 < z < 0.85$	Radio
L14	Lena *et al.* [279]	14	0.01	$z < 0.02$	Optical
K16	Kim *et al.* [280]	26	-	$z < 0.25$	Hα (Broad)
O16	Orosz *et al.* [281]	1327	0.07	$z < 1$	Radio
B16	Barrows *et al.* [271]	48	0.06	$z < 0.2$	X-ray
SB18	Skipper and Browne [282]	345	0.06	$z < 0.2$	Radio
B19	Barrows *et al.* [283]	254	0.18	$z < 0.58$	X-ray
S19	Shen *et al.* [284]	8210	-	$0.3 < z < 0.8$	Optical
R20	Reines *et al.* [285]	13	0.25	$z < 0.055$	Radio
SB20	Spingola and Barnacka [286]	2	5×10^{-6}	$z \sim 1.4$	Radio

to a central massive compact object (i.e. a BH). Objects in their sample are required to have an apparent B-band magnitude brighter than 18 mag and those which have bright stars near the centre are removed. The lower limit in Figure 3.3 is set by the typical standard deviation of mean off-centre distance of $0\farcs1$ and the upper limit is the maximum offset in the sample.

Since Binggeli *et al.* do not directly determine whether a BH resides in their nucleus, and there are no obvious selection cuts on BH luminosity, mass or accretion rate based on these observations, we do not produce a H-AGN sample designed to mimic these data.

3.3.2 Orosz & Frey 2013 (OF13)

Orosz and Frey [278] find the optical counterparts of 1297 radio sources from the International Celestial Reference Frame (ICRF2) in SDSS DR9 [287], of which 233 are classified by SDSS DR9 as extended (i.e. galaxy-AGN offsets) and the remainder are spatially unresolved and therefore appear as point-like quasars. The AGN position is determined from ICRF2 due to its superior astrometric precision as these are Very Long Baseline Interferometry (VLBI) observations, with the optical counterpart from SDSS required to lie within $0\farcs5$, which is the upper limit in Figure 3.3. The sample covers the redshift range $0.01 < z < 0.85$. The cut-off in the published data of $3\sigma \sim 0\farcs17$ sets the lower limit in Figure 3.3.

As we are only interested in galaxy-AGN offsets so we exclude the quasar-like SDSS sources in this dataset. We also cut the H-AGN systems so that $\chi > 0.01$.

3.3.3 Lena et al. 2014 (L14)

Lena *et al.* [279] analyse archival *Hubble Space Telescope* (*HST*) images of 14 nearby ($d < 100$ Mpc) bright elliptical galaxies containing low luminosity AGN. The selected galaxies were required to have an optically bright central point-like source and be free of heavy nuclear obscuration or other photometric irregularities. The offset is measured as the displacement between the photocentre (flux-weighted average of the centres of elliptical isophotes) and the AGN point source, which is modelled as a Gaussian. The lower limit of Figure 3.3 is set by the *HST* resolution and the upper limit is due to the search region of $2\farcs$.

Since the selection process requires a point-like AGN source, we only select quasar mode BHs from the simulation ($\chi > 0.01$) when mimicking this sample in H-AGN. The minimum detected redshift from [279] is $z_{\rm min} = 3.0 \times 10^{-3}$, which corresponds

to a luminosity distance of $d_{\rm L} = 13\,{\rm Mpc}$. Converting the minimum detected flux (at $5\,{\rm GHz}$) of $F_\nu = 1\,{\rm mJy}$ [288] to a luminosity gives

$$L_{\rm min} = 4\pi d_L^2 F_\nu = 2.0 \times 10^{26}\,{\rm erg\,s^{-1}\,Hz^{-1}}. \tag{3.4}$$

This must be converted to $L_{\rm R} = \nu L_\nu$, thus we remove all systems from H-AGN with $L_{\rm R}\,(5\,{\rm GHz}) < 2.0 \times 10^{28}\,{\rm erg\,s^{-1}\,Hz^{-1}} \times 5\,{\rm GHz} = 1.0 \times 10^{38}\,{\rm erg\,s^{-1}}$. To calculate the radio luminosity, we first calculate the X-ray luminosity, $L_{\rm X}$, from the bolometric luminosity, L, using [289, 290]

$$L = 35.0 L_{\rm X}. \tag{3.5}$$

Utilising the best-fitting fundamental plane of BHs [291] we obtain the radio luminosity at $5\,{\rm GHz}$

$$\log\left(\frac{L_{\rm R}\,(5\,{\rm GHz})}{{\rm erg\,s^{-1}}}\right) = \xi_{\rm RX}\log\left(\frac{L_{\rm X}}{{\rm erg\,s^{-1}}}\right) + \xi_{\rm RM}\log\left(\frac{M_{\rm BH}}{{\rm M_\odot}}\right) + b_{\rm R}, \tag{3.6}$$

where $\xi_{\rm RX} = 0.60 \pm 0.11$, $\xi_{\rm RM} = 0.78^{+0.11}_{-0.09}$ and $b_{\rm R} = 7.33^{+4.05}_{-4.07}$. It should be noted that radio-loud and radio-quiet AGN obey different fundamental plane relations [292]. The value of $\xi_{\rm RX}$ above is less than the values reported in Wang *et al.* [292] of 1.39 for radio-loud and 0.85 for radio-quiet AGN. Our estimates of $L_{\rm R}\,(5\,{\rm GHz})$ are therefore smaller than if we used these alternative values of $\xi_{\rm RX}$, and we thus reject more systems here than in these alternative cases. We will find in Table 3.3 that rejecting no systems based on luminosity (an extreme case of using larger $\xi_{\rm RX}$) has little effect on our results. Therefore, for simplicity we will apply the Merloni *et al.* relation for both radio-quiet and radio-loud AGN as this is the most stringent cut.

3.3.4 Kim et al. 2016 (K16)

Kim *et al.* [280] looked for recoiling SMBHs from the $z < 0.25$ quasi-stellar objects in SDSS DR7 by targeting objects with broad lines that are blueshifted relative to the systemic velocity. They excluded those with highly asymmetric and widely separated broad-line velocity profiles (disc emitters) or those with double-peak emission lines (possible binary SMBHs). Performing a spectral decomposition of the Hα and Hβ lines, their final sample of 26 have kinematic offsets in Hα of at least $69\,{\rm km\,s^{-1}}$ (the instrument dispersion) and the Hα and Hβ velocities must agree within to 50 per cent. Assuming an AGN age of $\tau = 10\,{\rm Myr}$ [293], the recoil velocity, $v_{\rm rel}$, can be converted to a physical offset as

$$r_{\rm GB} = v_{\rm rel}\tau. \tag{3.7}$$

We include these offsets in Figure 3.3, with the upper and lower limits set by the fibre radius and instrument dispersion respectively. We do not, however, use this data in the rest of the analysis since the spatial offsets are only estimates and depend on the chosen τ. It should be noted that the magnitude of the inferred offsets appear consistent with the other data.

3.3.5 Orosz et al. 2016 (O16)

These data are currently unpublished, however were briefly presented by Orosz *et al.* [281]. The procedure for obtaining the offsets is identical to OF13, however the matching is between SDSS DR12 [294] and mJIVE-20 [295]. As before, we select objects classified by SDSS as galaxies, so obtain 1327 objects out of a total 2066. We note that the offset AGN candidates were not subject to follow-up observations to exclude spurious sources (e.g. jets, lensing systems etc.). It is therefore possible that this sample has a higher level of contamination than the other datasets. To quantify this, a Monte Carlo procedure similar to Orosz and Frey [278] was performed (Orosz private communication) to determine the probability of false identification. The probability is twice as high in this sample than for OF13, although both are ~ 0.1 per cent for a match radius of 500 mas, which is what we adopt.

3.3.6 Barrows et al. 2016 (B16)

Barrows *et al.* [271] searched for X-ray AGN by cross-matching sources from SDSS DR7 [296] with the $z < 0.2$ OSSY catalogue [297]. These sources were then cross-correlated with *Chandra* [298] and only sources containing i- or z-band SDSS images registered with *Chandra* were kept. Sources with dust lanes or multiple emission peaks were removed, since these give false centroid positions. To reject stellar-mass objects, they required the difference between the observed luminosity in the range $2-10\,\mathrm{keV}$ and that expected from star formation, to be $> 3\sigma$ and exceed $10^{42}\,\mathrm{erg\,s^{-1}}$. Further, the hardness ratio must be $\mathrm{HR} > -0.1$. This provides a sample of 48 type-II AGN. The detection radius is less than $5'$ from the observation aim point and the AGN must be within 20 kpc of centre of galaxy, which gives the upper limit in Figure 3.3. The lower limit is given by the *Chandra* resolution of $0\farcs6$. We only plot the 'offset sample' and not the full parent sample.

To select H-AGN systems that are similar to these observations, we first cut on X-ray luminosity, L_X, such that $L_\mathrm{X} > 10^{42}\,\mathrm{erg\,s^{-1}}$, where we calculate L_X using Equation 3.5. We then choose the closest remaining BH to the centre of the galaxy

to be the central BH. The hardness ratio cut is designed to select only AGN, so we use the quasar mode criterion $\chi > 0.01$.

3.3.7 Skipper & Browne 2018 (SB18)

Skipper and Browne [282] adopted a similar selection method to Barrows *et al.* for radio AGN. They cross-matched sources from SDSS with the $z < 0.2$ OSSY catalogue and compared[2] this with the Cosmic Lens All-Sky Survey (CLASS) [299, 300] to find sources within $10''$, giving the upper limit of offsets in Figure 3.3. CLASS is a radio survey with the VLA and the images have an angular resolution of $200 - 250\,\mathrm{mas}$ at an observing frequency of $8.46\,\mathrm{GHz}$. We plot the lower limit as twice the size of the point spread function (PSF) of SDSS as offsets smaller than this would unlikely be resolved. Matches outside the visible bulge of the galaxy were removed and so are galaxies in the starburst region of the BPT diagram [301], which compares the flux ratios $F_{\mathrm{O\,III}}\lambda5007/F_{\mathrm{H}\alpha}$ and $F_{\mathrm{N\,II}}\lambda6583/F_{\mathrm{H}\alpha}$. Those known to be starburst galaxies and those missing optical emission line data products in the OSSY database were also removed. This results in a final sample of 345 radio selected systems.

To mimic this selection process, we start by selecting the closest BH to the centre of each galaxy, as this is how they select their central BH. We do not impose a minimum galaxy mass cut at this stage but keep our BH mass cut. We reject all systems with $\chi \leq 0.01$. Finally, Skipper and Browne only select candidates from the CLASS catalogue which have a flux density of at least $8\,\mathrm{mJy}$ at $8.46\,\mathrm{GHz}$. We must therefore select AGN that obey this criterion. We are not given the radio luminosities of the BHs in H-AGN, so we do the following to make the correct cut. From the X-ray luminosity, L_{X}, (Equation 3.5), we find the radio luminosity at $5\,\mathrm{GHz}$ with Equation 3.6. The radio luminosity at 8.46GHz is

$$L_{\mathrm{R}}\left(8.46\,\mathrm{GHz}\right) = L_{\mathrm{R}}\left(5\,\mathrm{GHz}\right)\left(\frac{\nu}{5\,\mathrm{GHz}}\right)^{-\alpha_{\mathrm{R}}}, \tag{3.8}$$

for some spectral index α_{R}. We choose $\alpha_{\mathrm{R}} = 0.4$ as this corresponds to the crossover between flat and steep spectra [291]. Since the cutoff for CLASS is given as a flux, we wish to convert this to a minimum detectable luminosity. The minimum detected redshift is $z_{\mathrm{min}} = 8.9 \times 10^{-3}$, which corresponds to a luminosity distance of $d_{\mathrm{L}} = 38\,\mathrm{Mpc}$. Converting the minimum detected flux of $F_\nu = 8\,\mathrm{mJy}$ to a luminosity gives

$$L_{\mathrm{min}} = 4\pi d_L^2 F_\nu = 1.4 \times 10^{28}\,\mathrm{erg\,s^{-1}\,Hz^{-1}}. \tag{3.9}$$

[2] `http://www.jb.man.ac.uk/research/gravlens/class/class.html`

Converting to $L_{\rm R} = \nu L_{\nu}$, we remove all systems from H-AGN with $L_{\rm R}\,(8.46\,{\rm GHz}) < 1.4 \times 10^{28}\,{\rm erg\,s^{-1}\,Hz^{-1}} \times 8.46\,{\rm GHz} = 1.2 \times 10^{38}{\rm ergs^{-1}}$.

The AGN cut reduces the number of systems to 533 and the luminosity cut, using $\xi_{\rm RX}$, $\xi_{\rm RM}$ and $b_{\rm R}$ from subsection 3.3.3, further reduces this to 530. Note that if we chose $\alpha_{\rm R} = 0$ (i.e. a completely flat spectrum) then none of the 533 systems would be removed. Further, if we use $\xi_{\rm RX} = 0.76$, $\xi_{\rm RM} = 0.71$ and $b_{\rm R} = 1.31$, as found by Merloni *et al.* for flat-spectrum sources, we would not remove any system.

We also note that the flat radio spectra indicate that the emission from these sources is likely to be beamed. Our assumption of isotropic emission would then result in $L_{\rm min}$ being over-predicted, making our luminosity cut too severe. As in subsection 3.3.3, we choose the most stringent selection cuts and again note that we will find that our results would not change significantly if we did not make any luminosity cuts. Therefore, for simplicity, we assume isotropic emission in the luminosity cut and note than any anisotropy would not cause significant bias.

3.3.8 Barrows et al. 2019 (B19)

Barrows *et al.* [283] used an almost identical selection procedure to B16 in a search for hyper-luminous X-ray sources by again comparing SDSS and *Chandra* data. In the 2019 sample, the search radius is changed from within the $3''$ fibre to within two Petrosian radii. Further, the latest sample imposes a much stricter cut on the compactness of the source and the errors, specifically the X-ray source errors, are estimated differently. Although some galaxies are common to this sample and B16 we plot them separately in Figure 3.3 as the offset may differ between the two. To compare to H-AGN we make the same selection cuts as for B16.

3.3.9 Shen et al. 2019 (S19)

Shen *et al.* [284], using the technique of Vastrometry [302], were able to put upper limits on the magnitude of the offsets from 8210 AGN in *Gaia* DR2 [303, 304]. This technique utilises the astrometric jitter caused by stochastic variability of the AGN to place upper limits on the BH displacement. *Gaia* cannot resolve separations $\lesssim 1\,{\rm kpc}$ in the desired redshift range of $0.3 < z < 0.8$, but this method enables upper limits to be determined in the range $\sim 5\,{\rm pc}$ to $\sim 1\,{\rm kpc}$, as indicated in Figure 3.3.

Since this method does not measure the projected off-nucleus distance, we do not compare Shen *et al.*'s results directly to H-AGN, but we do note that they find that

AGN are well centred at these redshifts, with 99, 90 and 40 per cent of AGN within 1 kpc, 500 pc and 100 pc respectively.

3.3.10 Reines et al. 2020 (R20)

Reines *et al.* [285] cross-correlated galaxies from the NASA-Sloan Atlas[3] ($z < 0.055$) with a sub-sample of dwarf galaxies from the Very Large Array (VLA) Faint Images of the Radio Sky and Twenty-centimetres (FIRST) Survey [305], requiring a match radius $\leq 5''$, which is approximately the resolution of the 1.4 GHz radio observations. This gives 186 matches, after enforcing a maximum stellar mass of $M_\star \leq 3 \times 10^9\,\mathrm{M}_\odot$ and absolute magnitude cuts of $M_g, M_r > -20$ to prevent spurious mass estimates. Sources clearly not dwarf galaxies were removed and 111 objects were observed with the VLA at higher frequencies ($\sim 8 - 12$ GHz) and therefore higher resolution. The match radius and the typical angular resolution of the VLA follow-up observations of $0.25''$ define the shaded region in Figure 3.3. Only those with $> 3\sigma$ radio detections were retained, and those with optical counterparts which appear to come from unrelated background point sources were removed. Finally, the VLA detections were used to eliminate samples with emission from thermal H II regions, individual supernova remnants or young supernovae. This results in a sample of 13 compact radio sources, which are almost certainly AGN.

The VLA analysis ensures that only AGN are selected, hence we cut the H-AGN data such that $\chi > 0.01$. The minimum mass of these galaxies is $6.5 \times 10^9\,\mathrm{M}_\odot$, which is small compared to typical masses in H-AGN. Since this cut was only introduced to ensure reasonable mass estimates we do not introduce a mass cut here. We explore the impact of galaxy mass on the BH offsets in subsubsection 3.6.1.1.

3.3.11 Spingola & Barnacka 2020 (SB20)

Recent work by Spingola and Barnacka [286] demonstrates how to exploit the non-linear magnification of gravitational lensing to access parsec scales at cosmological redshifts when an AGN lies close to a caustic curve. Using VLBI radio observations, they detected one object ($z = 1.34$) with optical and radio centres within ~ 40pc of each other, and another ($z = 1.39$) with a radio-optical offset of 214 ± 137 pc.

Currently this sample size is too small for comparison with H-AGN. Nonetheless, the astrometric precision achieved is impressive. This methodology can be applied to

[3] `http://www.nsatlas.org/documentation`

future optical and radio surveys of gravitational lensing systems, which will allow to detect offset BHs at high redshifts.

3.3.12 Other observations

The remaining observations in Figure 3.3 are of single systems, hence we cannot make statistical comparisons of these to H-AGN. We include them for visual comparison with the other data, and summarise them here.

- **B08** [306] is the offset Seyfert-2 nucleus in the system NGC 3341. This system consists of three nuclei, but the optical emission lines suggest only one of these harbours an AGN, however this is not the primary nucleus. Further optical, radio and X-ray observations [307] support this conclusion.
- **C10** [308] is one of the best studied offset AGN, CXOC J100043.1+020637 (CID-42), with optical, high resolution X-ray [309] and radio [310, 311] observations consistent with an AGN ejected from a separate compact region. Simulations [312] also support this conclusion, as opposed to a dual-AGN system. The offset used in Figure 3.3 is the measured displacement between the two compact optical sources in the *HST*/ACS image.
- **B10** [313] is the displaced SMBH in M87*. The offset is measured using archival *HST* data, with the offset between the photocentre of the galaxy and AGN point source.
- **J10** [314] is the source CXO J122518.6+144545: an X-ray source from *Chandra* offset from a galaxy from SDSS DR7. It is unknown whether the source is a recoiling SMBH, bright ultra-luminous X-ray source with a bright optical counterpart, or a very blue Type IIn supernova. A candidate optical counterpart to the X-ray source is found in archival *HST* data.
- **M14** [315] is an AGN in NGC 3115 observed to be off-centred from the photometric centre, using the Gemini Multi-Object Spectrograph mounted on the Gemini-South telescope.
- **M16** [316] is a Seyfert 2-like source offset from the central stellar cluster in NGC 3621, observed using the Gemini Multi-Object Spectrograph. An X-ray source found at the centre of the stellar cluster suggests this may not in fact be an offset BH.
- **C17** [317] is a SMBH in the cluster ZwCl 8193, which is offset from both the brightest cluster galaxy ($\sim 8.5\,\mathrm{kpc}$) and its host galaxy, a small, optically faint radio galaxy ($\sim 0.1\,\mathrm{kpc}$).

- **K17** [318] is the source CX0 J101527.2+625911, a recoiling or dual SMBH discovered from the *Chandra* Source and SDSS Cross-Match Catalogue. No X-ray source is observed at the galaxy centre, supporting the recoiling SMBH hypothesis.

We note that many other catalogues are used to investigate of optical-radio offsets, such as *Gaia* DR1 [319, 320] and DR2 [272, 321] combined with ICRF2, the Rio catalogue [322] or the catalogue of Zacharias and Zacharias [323]. Although some of the conclusions of these studies are similar to this work, the lack of photometry means we cannot guarantee that these offsets are the same as are of interest here, so we do not use them. In future work, we hope to be able to use upcoming *Gaia* data releases, once photometric classifications are included, to expand our sample size.

3.4 Comparing Horizon-AGN with the data

In the previous section we described observational searches for offset AGN, which should be interpreted as finding 'candidate' rather than necessarily 'true' offset systems. Despite this limitation, we believe it is important to study the exact nature of any discrepancy between these and the simulations in order both to understand it fully and to improve the simulations. However, due to H-AGN having a much coarser spatial resolution than the observations, we cannot simply compare the magnitudes of the offsets since the resolution makes the simulated values appear much larger. Ideally, one would compare simulations of different resolutions and extrapolate the results to predict the properties of a simulation with perfect resolution. Unfortunately, we only have one simulation so must develop an alternative framework to quantify the effect of resolution to allow us to compare the observed and simulated offsets fairly. In this section we describe how we do this by determining the fraction of BHs which are in the tails of the galaxy–BH offset distribution.

3.4.1 Accounting for finite resolution

Since the simulations and observations have a finite resolution, we need a prescription to quantify the number of intrinsically offset BHs. Inspired by Skipper and Browne [282], we bin the d-dimensional offsets and fit to this two components:

$$p_d\left(r\right) = a_d\left(r\right) + b_d\left(r\right). \tag{3.10}$$

$p_d(r)$ is the probability of an offset r (as described by the number of counts in the corresponding histogram bin), $a_d(r)$ is the probability distribution function for apparent offsets produced by the finite spatial resolution of the observation or simulation, and $b_d(r)$ is the probability distribution function for intrinsic offsets. Assuming a Gaussian distribution for each Cartesian component, a_d takes the form

$$a_d\left(r\right) = Ar^{d-1}\exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{3.11}$$

Skipper and Browne give an explicit model for $b_d\left(r\right)$, however we do not need to know this to find the fraction of BHs that are intrinsically offset. Instead we enforce normalisation,

$$\int_0^\infty b_d\left(r\right)\mathrm{d}r = 1 - \int_0^\infty a_d\left(r\right)\mathrm{d}r, \tag{3.12}$$

so that

$$P_d\left(\text{offset}\right) = \frac{\int_0^\infty b_d\left(r\right)\mathrm{d}r}{\int_0^\infty \left(a_d\left(r\right) + b_d\left(r\right)\right)\mathrm{d}r} = 1 - 2^{d/2-1}\Gamma\left(\frac{d}{2}\right)A\sigma^d, \tag{3.13}$$

where $\Gamma(x)$ is the gamma function. We fit for A and σ in the region $r_{\mathrm{GB}} < r_{\mathrm{cutoff}}$, for some cut-off offset r_{cutoff}, with the expectation that $\sigma \sim 1\,\mathrm{kpc}$, the approximate spatial resolution of H-AGN. We find that the fitting parameters converge by $r_{\mathrm{cutoff}} \sim 3\,\mathrm{kpc}$, but reducing this to $r_{\mathrm{cutoff}} \sim 2\,\mathrm{kpc}$ does not significantly change the results. Choosing $r_{\mathrm{cutoff}} \gg 3\,\mathrm{kpc}$ produces unrealistic fits, as this tries to make a_d fit the tail too well, to the detriment of the small r part of the distribution.

For each system, i, with d-dimensional offset r_i, we can now define the probability that the offset is intrinsic as

$$w\left(r_i\right) = \frac{b_d\left(r_i\right)}{a_d\left(r_i\right) + b_d\left(r_i\right)}. \tag{3.14}$$

Hence, we need to find the functional form of $b_d(r)$, since the normalisation condition is no longer sufficient. Skipper and Browne [282] assume $b_2\left(r\right)$ takes the form of a decaying exponential. We would like to have a form for b_d such that its marginal distribution gives b_{d-1}. It is not obvious how to do this for an exponential, although the good fit obtained by Skipper and Browne shows that a functional form that decays exponentially at large r is desirable. We choose a generalised symmetric Laplace distribution [324, 325], since all marginal distributions of a multivariate Laplace distribution are also multivariate Laplace distributions.

Working in coordinates centred on the galactic centre, we assume that the mean vector offset is zero. If we assume that the components of the displacements along

each axis are independent and that there is no preferred direction, the generalised symmetric Laplace distribution for a displacement r is

$$g_d(r) = \frac{2^{1-d/2}}{\Gamma\left(\frac{d}{2}\right)} \lambda (\lambda r)^{d/2} \mathrm{K}_{1-d/2}(\lambda r), \tag{3.15}$$

with K_n the modified Bessel function of the 2$^{\mathrm{nd}}$ kind.

Unfortunately, with the resolution of H-AGN, the peak of $g_d(r)$ occurs at a similar value to the peak of $a_d(r)$ for appropriate fitting parameters if we assume $b_d \propto g_d$. This suppresses the contribution from the resolution to very low levels and pushes $\sigma < 1\,\mathrm{kpc}$. In order to suppress b_d at small r_{GB}, we multiply g_d by a sigmoid function and thus arrive at

$$b_d(r) = B r^{d/2} \mathrm{K}_{1-d/2}(\lambda r) \frac{1}{1 + \exp(-\kappa(r - \nu))}. \tag{3.16}$$

ν is an additional free parameter to be fitted, which we expect to be around 3 kpc because this is where $p_d(r)$ starts to deviate from $a_d(r)$ (Figure 3.4). Although this removes the property that b_{d-1} is the marginal of b_d, we anticipate that future higher-resolution simulations would not require the sigmoid function and thus this property would be conserved.

In Figure 3.4 we demonstrate these fits for a highly offset sample ($z = 1.0$) and one with a low fraction of offset BHs ($z = 2.5$) where we use quasar mode systems only. The results are similar for the radio mode, but we choose to only plot the quasar mode for clarity and because this is the mode used when comparing to observations. Note we assume that the shape parameter remains equal to 1 at all redshifts for simplicity. The degree to which a sample is intrinsically offset is determined by how well a_d fits the distribution.

3.4.2 Probability of consistency

After making selection cuts to match the simulated galaxies to the observed ones, we wish to find the probability of generating a dataset with up to as many intrinsically offset BHs as is observed. To do this, using the parameter vector $\mathbf{\Omega} = \{A, \sigma, B, \lambda, \kappa, \nu\}$ from the fit to Equation 3.10, we find $\{w_i\}$. For each system, we then draw a random number $R_i \in [0, 1]$ from a uniform distribution $U[0, 1]$ and define the random variable a_i such that

$$a_i = \begin{cases} 1 & \text{if} \quad w_i > R_i \\ 0 & \text{if} \quad w_i \leq R_i. \end{cases} \tag{3.17}$$

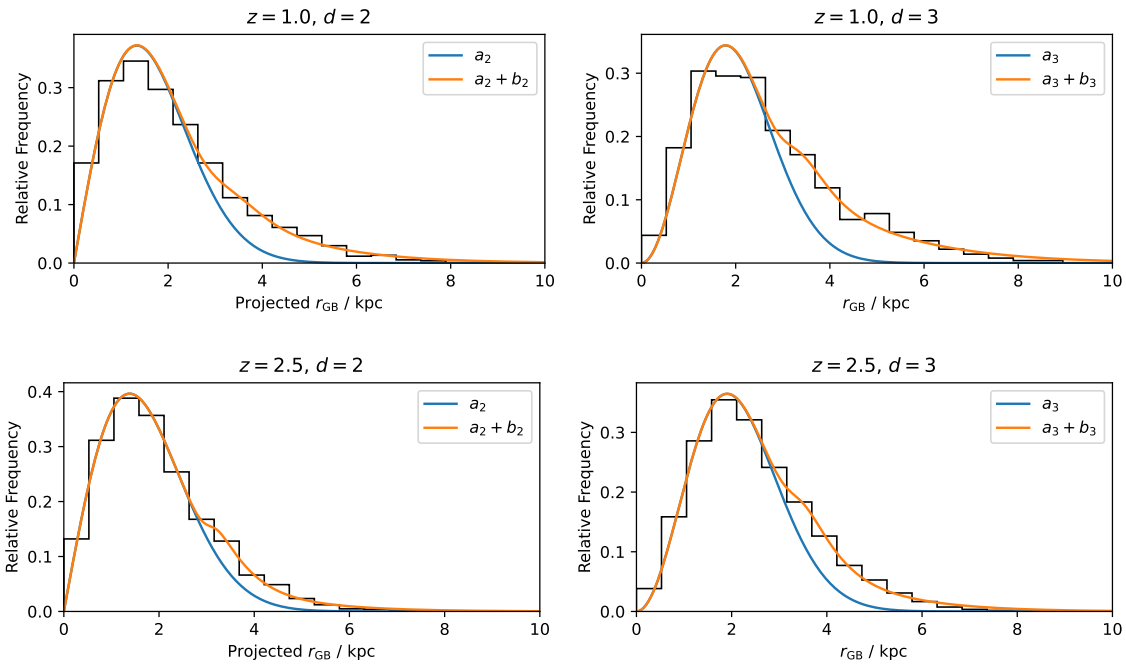


Figure 3.4: Distributions of offsets for the quasar mode systems in H-AGN and the fits for both the true ($d = 3$) and projected ($d = 2$) offsets. The a_d term captures the offsets due to the finite resolution and b_d models the intrinsically offset population. The $z = 1.0$ sample has a higher fraction of intrinsically offset BHs and hence requires a larger contribution from b_d than the $z = 2.5$ sample.

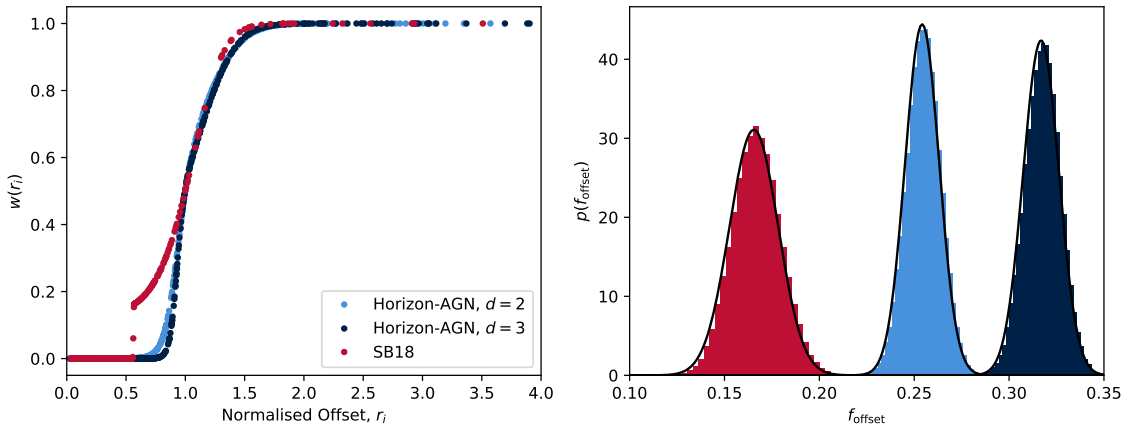


Figure 3.5: *Left*: The probability of a BH being intrinsically offset, $w\left(r_i\right)$, as a function of the offset, r_i. The points correspond to individual systems. The relation is not monotonic for SB18, as w is determined by the angular offset, whereas here we plot the offset distance. *Right*: The probability density functions of the fraction of intrinsically offset BHs, $p\left(f_{\rm offset}\right)$, using the best fit parameters. The black lines are Gaussians with the same mean and standard deviation as the data. For all these plots we use the selection procedures outlined in subsection 3.3.7.

For $\mathcal{N}$ systems, the fraction of offset BHs is

$$f_{\rm offset} = \frac{\sum_i a_i}{\mathcal{N}}. \tag{3.18}$$

Repeating this procedure $N_{\rm MC} = 10^5$ times, we find a distribution of $f_{\rm offset}$, $p\left(f_{\rm offset}|\mathbf{\Omega}\right)$.

Using the means and errors on the fits for $\mathbf{\Omega}$, we draw $N_{\rm dist}$ random values of $\mathbf{\Omega}$ from a Gaussian distribution and repeat the above analysis for each of those samples. Since $\mathbf{\Omega_B} = \{B, \lambda, \kappa, \nu\}$ is determined after finding $\mathbf{\Omega_A} = \{A, \sigma\}$, we draw $\mathbf{\Omega_A}$ from its multivariate Gaussian distribution first, fit the residuals to obtain means and covariances for $\mathbf{\Omega_B}$ and then draw $\mathbf{\Omega_B}$ from the resulting multivariate Gaussian distribution. We reject any of the iterations if $\omega < 0$ for any $\omega \in \mathbf{\Omega}$ since these are unphysical, or if the fit for $\mathbf{\Omega}_B$ is unsuccessful for the given $\mathbf{\Omega}_A$. We combine these $N_{\rm dist}$ samples to find the distributions of the parameters describing $p\left(f_{\rm offset}|\mathbf{\Omega}\right)$.

In Figure 3.5 we plot $w\left(r_i\right)$ and $p\left(f_{\rm offset}\right)$ for the best-fitting $\mathbf{\Omega}$ for the SB18 sample and the H-AGN sample designed to mimic these observations (subsection 3.3.7). We see that $p\left(f_{\rm offset}\right)$ is well approximated by a Gaussian, hence we will characterise the distributions by their mean, $\mu\left(\mathbf{\Omega}\right)$, and their standard deviation, $s\left(\mathbf{\Omega}\right)$.

The results for the fits for the distributions in Figure 3.4 are plotted in Figure 3.6. Since the $z = 1.0$ sample is more offset than at $z = 2.5$, we see that it has a larger

value of μ. We also note that σ is comparable between the two fits, since the resolution of H-AGN is constant with redshift at$\sim 1\,$kpc.

For each of the $N_{\rm dist}$ runs, we find the probability of generating a sample from H-AGN with up to a fraction $f_{\rm crit}$ of offset BHs, which is

$$\begin{aligned} P_{\rm crit} \equiv P\left(f_{\rm offset} \leq f_{\rm crit}\right) &= \int_{-\infty}^{f_{\rm crit}} \frac{1}{\sqrt{2\pi}s} \exp\left(-\frac{(f-\mu)^2}{2s^2}\right) \mathrm{d}f \\ &= \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{f_{\rm crit}-\mu}{s\sqrt{2}}\right)\right], \end{aligned} \tag{3.19}$$

since $p\left(f_{\rm offset}|\mathbf{\Omega}\right)$ is approximately Gaussian. We can then find the mean and error of $P_{\rm crit}$ from the $N_{\rm dist}$ iterations.

To prevent poor fits to the tail, we enforce $\nu \leq 4\,$kpc (Equation 3.16), although our results are not sensitive to this choice; relaxing this to $\nu \leq 8\,$kpc changes μ for the quasar-mode sample at $z = 0$ with $d = 3$ from 0.40 ± 0.06 to 0.36 ± 0.12. It is also necessary, for $d = 2$, to demand $\nu \geq \nu_{\rm min}$, where we set $\nu_{\rm min} = 2.3\,$kpc. Once again, our results are not sensitive to this choice; if we remove this constraint for the quasar-mode sample at $z = 0$ with $d = 2$, then μ changes from 0.25 ± 0.05 to 0.26 ± 0.05.

3.5 Results

In this section we start by analysing the H-AGN systems, by looking at the impact of the mode of feedback (subsection 3.5.1) and the correlations between offsets and the galaxy and halo properties (subsection 3.5.2). We then assess the compatibility of the observations with each other and with H-AGN in subsection 3.5.3 and subsection 3.5.4 respectively.

3.5.1 Quasar vs radio mode

To compare the effect of selecting quasar or radio mode BHs at each redshift we start by assigning the closest BH to the centre of the galaxy as the central BH. At each redshift, we find the minimum luminosity of the quasar mode BHs and only consider BHs with luminosities greater than this. At $z = 0.1$ this corresponds to a luminosity cut of $2.6 \times 10^{43}\,\mathrm{erg\,s^{-1}}$. We split the sample into two sets, those in the quasar mode ($\chi > 0.01$) and those in the radio mode ($\chi \leq 0.01$). We fit the distributions of offsets and plot the resulting P_d(offset) for the two samples as a function of redshift in Figure 3.7.

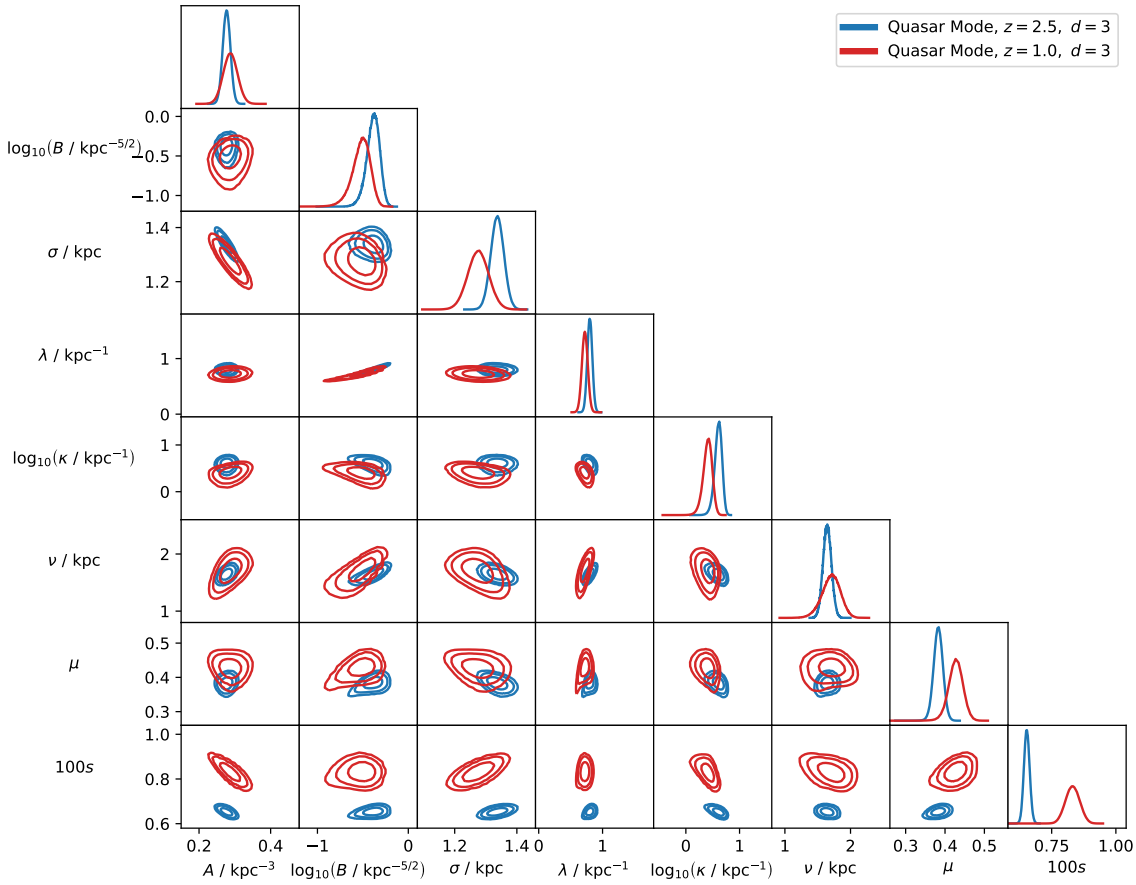


Figure 3.6: The distributions of fitting parameters A, B, σ, λ, κ and ν (Equations 3.10, 3.11 and 3.16) and the parameters describing the distribution of f_{offset}: the mean, μ and standard deviation, s, where we consider the true offsets ($d = 3$) of the quasar mode systems at the given redshifts. We fit the offsets to the sum of a Gaussian of width σ and the product of a sigmoid function and generalised symmetric Laplace distribution, with typical scale λ. κ and ν parametrise the steepness and position of the sigmoid respectively. A controls the weight of the Gaussian term, and B has the same role but for the non-Gaussian term (see Equation 3.11 and Equation 3.16). The contours show the 1, 2 and 3σ confidence intervals.

At all redshifts, the radio mode quasars are more offset than those in the quasar mode. This is to be expected since by definition the quasar mode BHs operate at a higher fraction of the Eddington rate, so these will tend to lie closer to the centre of the galaxy where the gas density is higher. We note that this could also arise due to the different implementations of AGN feedback between the two samples. For $z \gtrsim 3$ the two modes have comparable values of $P_d(\text{offset})$, but this is also where the errors on the radio mode sample become large.

It is also clear in Figure 3.7 that BHs are more central at earlier epochs. This is consistent with the picture that BHs initially reside at the centres of galaxies and then move off-centre later due to interactions with other galaxies [237]. It is hard to test this prediction of redshift evolution observationally, since the majority of current observations are at $z < 0.2$ (Figure 3.3). We therefore look forward to the results of applying the methods of Spingola and Barnacka [286] to the CLASS sample of lensed sources, since this will probe $z = 0.6 - 3.6$.

The most physically interesting parameter from the fits to the offset distributions is λ, which tells us about the scale to which offset BHs extend. In Figure 3.8 we plot the redshift evolution of $\lambda d_{\rm A}$, for angular diameter distance $d_{\rm A}$. Fitting this to a power law in the cosmological scale factor, we find

$$\lambda d_{\rm A} \propto (1+z)^{\alpha}, \quad \alpha = \begin{cases} 0.39 \pm 0.07, & \text{Quasar Mode} \\ 1.06 \pm 0.20, & \text{Radio Mode.} \end{cases} \tag{3.20}$$

A larger value of λ indicates that the intrinsically offset BHs reside closer to the galactic centre. Since $\lambda d_{\rm A}$ increases with redshift, we conclude that BHs are more localised to their host's centre at earlier epochs, in terms of observed projected angular offset. This supports the picture of Figure 3.7 that BHs initially reside near galaxies' centres.

We note that, although the qualitative trends are the same for $d = 2$ and $d = 3$, the three-dimensional analysis tends to give a higher probability of a BH being offset than when we use the projected offsets. We find that this is due to a preference of our fitting procedure to obtain larger σ for $d = 2$ compared to $d = 3$. Since our procedure is designed to match the observational technique, this is not a concern provided we only directly compare the observations to the $d = 2$ model.

3.5.2 Correlation of offsets with galaxy and halo properties

To investigate which parameters besides redshift affect the magnitude of the galaxy-BH offset, we train a Random Forest regressor [170, 326] on the combined quasar plus

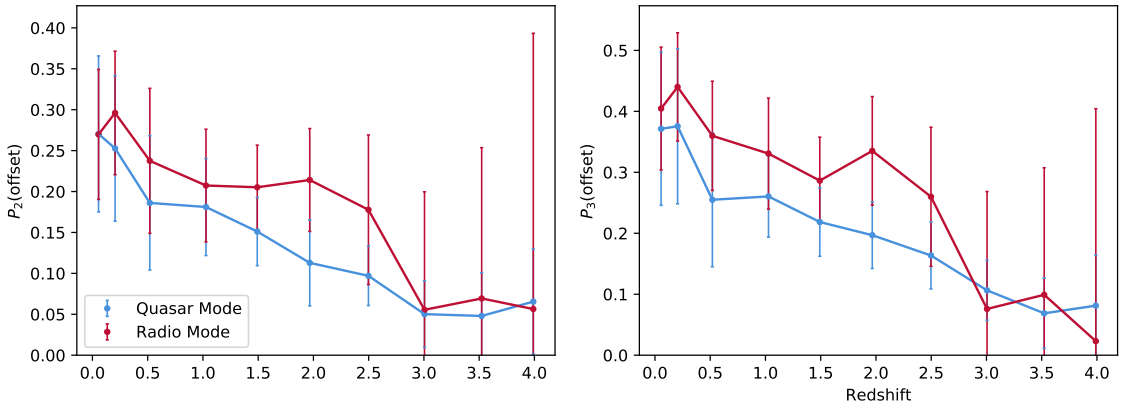


Figure 3.7: Fraction of offset BHs (see Equation 3.13) as a function of redshift in H-AGN. The BHs are selected as the closest BH to the centre of their galaxy. The 'Quasar Mode' BHs have $\chi > 0.01$ and the 'Radio Mode' BHs have the same luminosity range as the 'Quasar Mode' BHs, but with $\chi \leq 0.01$. The lower panel uses the full 3D offset while the upper panel projects to 2D using an observer at the centre of the simulation box. We find Radio Mode BHs to be more intrinsically offset than Quasar Mode ones at $0 < z < 3$.

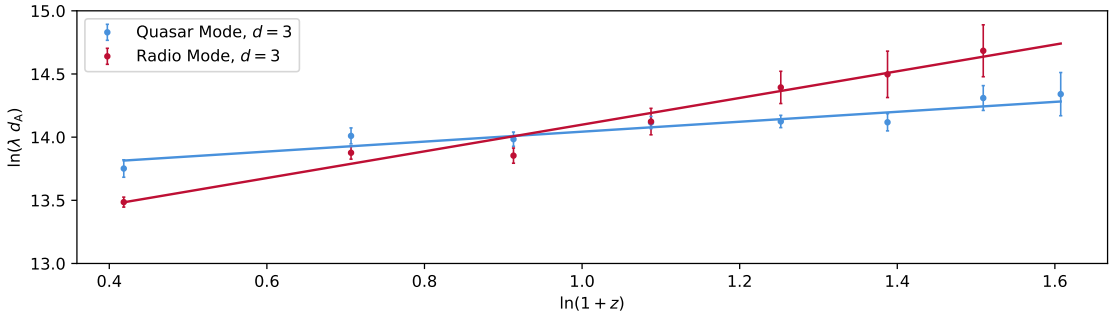


Figure 3.8: Evolution of $\lambda d_{\rm A}$ as a function of redshift, for both the quasar mode and radio mode samples, where λ^{-1} characterises the length scale of BH offsets and $d_{\rm A}(z)$ is the angular diameter distance to redshift z. Thus λd_A gives the reciprocal of the angular scales over which BHs are intrinsically offset. We plot the best fits to $\lambda d_{\rm A} \propto (1+z)^{\alpha}$.

radio mode sample at $z = 0.1$, optimising the regressor's hyperparameters using 5-fold cross-validation. This regressor fits nonlinear decision trees to predict components of the data from others. These recursively split the training data from the "root" node (containing the entire data set) to more homogeneous "child" nodes by minimising a loss function at each split. The Random Forest injects randomness by randomly selecting the number of features considered at each split and by training each tree in the ensemble with bootstrap aggregated samples. The returned prediction for a given input is the mean prediction of the ensemble. Random Forests can provide an estimate of the variables which are most important for the regression by assigning each a "feature importance." This is proportional to the number of times the variable is used in a split, weighted by the number of samples it splits, such that the features which are the best predictors have the largest feature importances. Before outlining the results, we describe the features we choose and how we calculate them.

3.5.2.1 Chosen features

From the halo and galaxy finder we obtain the masses of the halo and galaxy, $M_{\rm H}$ and $M_{\rm G}$ respectively, the halo virial radius, $r_{\rm vir}$, and the galaxy's effective radius, $r_{\rm eff}$. We use the black hole mass, $M_{\rm B}$, and Eddington ratio, χ, as directly output from the simulation. From this output we also retrieve the properties of the gas surrounding the BH: the average relative velocity between the BH and gas, $\bar{u}$, the gas density, $\rho_{\rm gas}$, and the average sound speed, $\bar{c}_s$, calculated as described in subsection 3.2.2. We calculate the mean velocity of the particles in each galaxy, $\boldsymbol{v}_{\rm G}$, and hence the relative velocity between the BH and galaxy

$$\boldsymbol{v}_{\rm GB} \equiv \boldsymbol{v}_{\rm B} - \boldsymbol{v}_{\rm G}, \tag{3.21}$$

where $\boldsymbol{v}_{\rm B}$ is the velocity of the BH. This velocity is decomposed into components parallel, $v_{\rm GB\parallel}$, and perpendicular, $v_{\rm GB\perp}$, to $\boldsymbol{r}_{\rm GB}$. We find the angular momentum of the galaxy, $\boldsymbol{J}_{\rm G}$, and halo, $\boldsymbol{J}_{\rm H}$, in their centre of mass frames, and the corresponding spin parameters [327], $\lambda_{\rm G}$ and $\lambda_{\rm H}$. The angular momentum of the BH about the galactic centre in the galaxy's centre of mass frame, $\boldsymbol{J}_{\rm B}$, is also calculated.

We produce merger trees for all galaxies back to $z = 7$, with an average time-step of $\sim 50\,$Myr. Following Martin *et al.* [328], we identify major mergers to be those with mass ratios, $q_{\rm merge}$, greater than 1:4, where we define the time of the merger, $t_{\rm merge}$, to be the point of coalescence according to the halo-finder, and we calculate the mass ratio at the point at which the less massive galaxy begins to lose mass. The most massive galaxy in the merger has mass M_1 at the time $q_{\rm merge}$ is measured. If

we do not identify any mergers for a given galaxy fulfilling our mass ratio criterion, then we do not include that system in this part of the analysis. The conclusions of the regression are the same for the other parameters whether or not these systems are included.

To investigate the impact of the external environment we calculate the gravitational field at the BH, $\boldsymbol{g}$. We find the density field by applying a cloud-in-cell algorithm to the DM and star particles in the simulation, adding this to the gas density field, and solve Poisson's equation on a 512^3 grid. This corresponds to a minimum spatial resolution of $\sim 200h^{-1}\,\mathrm{kpc}$.

Our final features describe the geometry of the system. We include the magnitude of the offset between the halo and galaxy centres, $|\boldsymbol{r}_{\mathrm{HG}}|$, and the angles between various vectors described above: $\hat{J}_{\mathrm{G}} \cdot \hat{J}_{\mathrm{B}}$, $\hat{J}_{\mathrm{H}} \cdot \hat{J}_{\mathrm{G}}$, $r_{\mathrm{GB}} \cdot \hat{J}_{\mathrm{G}}$, $v_{\mathrm{GB}} \cdot \hat{J}_{\mathrm{G}}$, $r_{\mathrm{GB}} \cdot \hat{g}$, $v_{\mathrm{GB}} \cdot \hat{g}$ and $r_{\mathrm{GB}} \cdot v_{\mathrm{GB}}$, where we denote $\hat{x}$ as the unit vector parallel to $\boldsymbol{x}$.

3.5.2.2 Correlation results

The resulting feature importances are plotted in Figure 3.9, where we also plot the two-dimensional distributions of the offsets with the three most importance features: $v_{\mathrm{GB},\perp}$, r_{HG} and $\hat{J}_{\mathrm{G}} \cdot \hat{J}_{\mathrm{B}}$.

We see that the strongest correlation is with $v_{\mathrm{GB},\perp}$, such that BHs with larger offsets have a large $v_{\mathrm{GB},\perp}$. To understand this, in the lower panel of Figure 3.10 we plot the distribution of angles between $\boldsymbol{J}_{\mathrm{G}}$ and $\boldsymbol{J}_{\mathrm{B}}$, which we also know is an important feature. Since these tend to align, we conclude that the BHs move on prograde orbits. In particular, 65 per cent of systems are aligned within 60° at $z = 0.1$. Consequently, $v_{\mathrm{GB},\perp}$ gives the orbital velocity of the BH, making it an important parameter for deducing r_{GB}.

The two-dimensional distribution of $\cos^{-1}(\hat{J}_{\mathrm{G}} \cdot \hat{J}_{\mathrm{B}})$ and r_{GB} demonstrates that a wide range of orbital radii are possible if the angular momenta perfectly align, with only smaller offsets permitted as the level of aligned decreases.

In Figure 3.10 we also plot the distributions of the angles between $\boldsymbol{r}_{\mathrm{GB}}$ and $\boldsymbol{v}_{\mathrm{GB}}$ for $z = 0.1$, 0.5 and 1.5. We plot the mean of these distributions across a wider range of redshift in Figure 3.11. We see that for $z \lesssim 2$ the velocity of the BH tends to oppose the direction of its offset, i.e. the BHs are, on average, returning to the centres of their host galaxies. The distribution is more uniform at $z = 0.1$ than $z = 1.5$. The anti-alignment of $\boldsymbol{r}_{\mathrm{GB}}$ and $\boldsymbol{v}_{\mathrm{GB}}$ can be interpreted as the orbital decay due to dynamical friction, which stalls at low z, and the uniform part of this distribution is due to the range of eccentricities of the BH orbits. The positive values of $\langle \hat{v}_{\mathrm{GB}} \cdot \hat{r}_{\mathrm{GB}} \rangle$ for $z = 3-4$

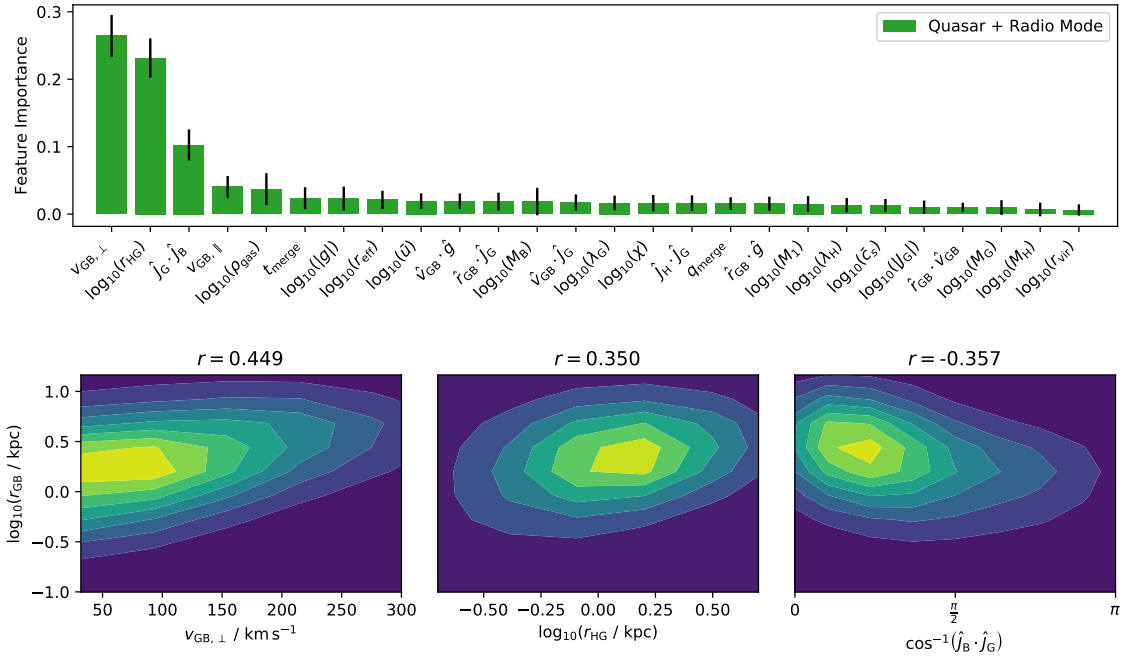


Figure 3.9: Feature importances for $\log_{10}(r_{\rm GB})$, for BH offset $r_{\rm GB}$, from optimised Random Forests for the combined quasar and radio mode sample from H-AGN at $z = 0.1$. The relative velocity between the BH and its host galaxy perpendicular to the offset, $v_{\rm GB,\perp}$, is the most important feature, followed by $r_{\rm HG}$ and $\hat{J}_{\rm G} \cdot \hat{J}_{\rm B}$, with all other parameters relatively unimportant and thus not strongly correlated with $r_{\rm GB}$. The two-dimensional histograms for the best three features are shown in the lower panel, with the Spearman's rank correlation coefficients as the titles.

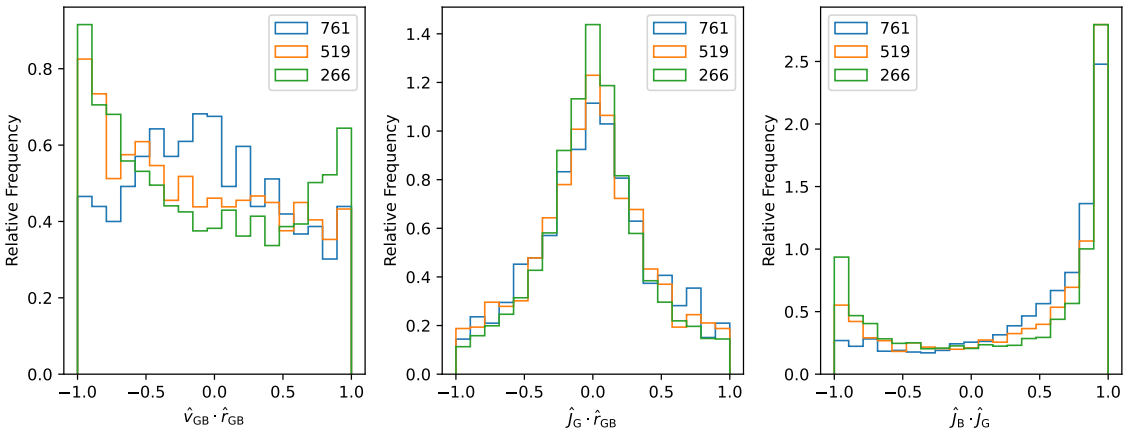


Figure 3.10: Distributions of *Left*: The cosine of the angle between the BH offset, $\boldsymbol{r_{GB}}$, and the velocity of the BH relative to its host galaxy $\boldsymbol{v_{GB}}$; *Centre*: The cosine of the angle between $\boldsymbol{r_{GB}}$ and the angular momentum of the galaxy about its centre in its centre of mass frame, $\boldsymbol{J}_G$; *Right*: The cosine of the angle between the angular momentum of the BH about the galactic centre in the galaxy's rest frame, $\boldsymbol{J}_B$, and $\boldsymbol{J}_G$. There is a slight propensity for $\boldsymbol{r_{GB}}$ and $\boldsymbol{v_{GB}}$ to be anti-aligned, so the BH tends to move back towards the galactic centre. This preference is stronger at higher redshift. The BH offsets tend to be perpendicular to $\boldsymbol{J}_G$, and hence to lie in the plane of the galaxy. $\boldsymbol{J}_B$ and $\boldsymbol{J}_G$ are preferentially aligned at all z considered, indicating that BHs move on prograde orbits.

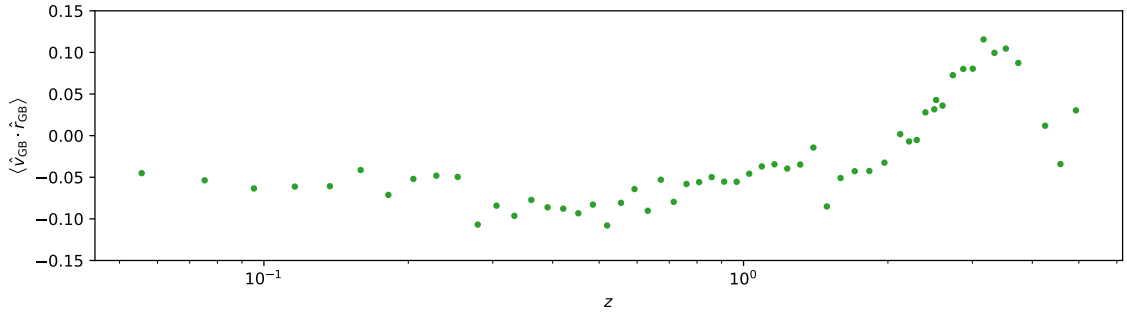


Figure 3.11: The mean of $\hat{v}_{GB} \cdot \hat{r}_{GB}$ as a function of redshift, where $\hat{v}_{GB}$ and $\hat{r}_{GB}$ are unit vectors parallel to the relative velocity and position of the BH relative to its host galaxy respectively, for the combined quasar and radio mode sample. For the lowest redshifts, there is a net anti-alignment, so BHs are on average returning back to the galactic centre. The positive value of $\langle \hat{v}_{GB} \cdot \hat{r}_{GB} \rangle$ at $z = 3 - 4$ corresponds to BHs moving away from the centres of their hosts on average.

are due to BHs being kicked away from the centres of their galaxies. This is why the fraction of offset BHs is higher today than at the earliest epochs.

We further observe in Figure 3.10 a strong preference for $\boldsymbol{r}_{\rm GB}$ to be orthogonal to $\boldsymbol{J}_{\rm G}$. If we use $\boldsymbol{J}_{\rm G}$ as a proxy for the normal to the plane of the galaxy, we infer that BHs have a propensity to lie in a plane parallel to the galactic plane. We find that 73 per cent of BHs in the combined quasar and radio mode sample have displacements parallel to $\boldsymbol{J}_{\rm G}$ within $\pm 1\,$kpc. Given that the resolution of H-AGN is $\sim 1\,$kpc, we therefore conclude that BHs tend to lie in the galactic plane. H-AGN does not contain a procedure for the asymmetric gravitational wave (GW) emission upon BH coalescence and the subsequent recoil, and only contains Schwarzschild BHs. For two Kerr SMBHs in a galaxy merger, if there is a gas rich environment, the spins align with each other and the circumbinary disc's angular momentum [329]. Since gravitational wave emission results in a kick preferentially perpendicular to the BH orbital plane [330], we would thus expect the distribution to become more isotropic if GW effects were included.

The second most important feature was found to be the magnitude of the halo-galaxy offset, $r_{\rm HG} \equiv |\boldsymbol{r}_{\rm HG}|$. This could occur if the halo and BH were tightly bound, and the galaxy is the offset member of the set. However, this is not the case since, as can be seen in the central panel of Figure 3.9, the median halo-galaxy offset is half the median galaxy-BH offset, and the halo-BH and galaxy-BH offsets are well correlated, with a Spearman's rank regression coefficient of $r = 0.81$. Instead, we interpret $r_{\rm HG}$ as a measure of how disturbed the system is; more disturbed systems have larger $r_{\rm HG}$ and thus greater galaxy-BH offsets.

We repeated the analysis using an Extra Trees regressor, separately using the quasar and radio mode samples, using different mass ratios to define a major merger (1:2 and 1:3), choosing the most massive or most recent merger irrespective of $q_{\rm merge}$, and altering when we measure quantities from the merger. In all cases, the results were qualitatively similar, with $v_{\rm GB,\perp}$ always being the most important feature. We also separately analysed the quasar and radio mode samples, and for an uncut H-AGN sample, where we choose the closest BH to the centre of the galaxy, but do not make any cuts on galaxy mass, halo mass, BH luminosity or accretion rate. In all cases, the BH closest to the centre of the galaxy preferentially moves on a prograde orbit in the galactic plane. One difference is that the full H-AGN sample exhibits a slight propensity for $\boldsymbol{r}_{\rm GB}$ and $\boldsymbol{v}_{\rm GB}$ to be orthogonal at the lowest redshifts, consistent with more circular orbits for the complete sample.

At $z = 0.1$, Figure 3.9 suggests that the magnitude of the BH offset is independent of $\boldsymbol{g}$. We also find that the distribution of the angle between the offset and $\boldsymbol{g}$ is consistent with isotropic, with a p-value of 0.50 for the two-sided KS test that the distribution is drawn from an isotropic distribution.

Figure 3.9 indicates that the quantities describing the merger history of the galaxy are relatively unimportant. Perhaps counter-intuitively, the Spearman correlation coefficient between $\log_{10}(r_{\rm GB})$ and $t_{\rm merge}$ is 0.184, so there is a slight positive correlation. In fact, this is true only for small $t_{\rm merge} \lesssim 5\,{\rm Gyr}$. Since we define the start of the merger to be the point at which the halo-finder can no longer identify two separate galaxies, there may be some delay between this point and the BHs being dislodged from their centres. Further, if a BH merger event occurs, there is an additional delay due to the finite timescales associated with the binary's evolution [230], although only a limited period of this evolution will be resolved in H-AGN. Thus, an initial increase of $r_{\rm GB}$ with $t_{\rm merge}$ is not too surprising, and indicates the time required for the central BH to become perturbed. For $t_{\rm merge} \gtrsim 5\,{\rm Gyr}$ there is little correlation with the offset.

3.5.3 Compatibility of observational datasets

Before comparing to H-AGN, it is important to investigate the level at which the independent observational datasets agree with one another. To do this we will restrict our attention to samples with more than two systems.

All the datasets that we use include the offsets greater than 3σ, so an initial test of compatibility is to compare the fraction of systems which have an offset $> 3\sigma$. Dividing the systems into two groups (those which have offsets $> 3\sigma$ and those which do not), and assuming a Binomial distribution, with probability Λ of a single BH being offset at $> 3\sigma$, we can estimate Λ and its error using a maximum likelihood estimate.

We plot the calculated Λ and the errors in Figure 3.12, alongside the values calculated using different cuts from H-AGN. The errors for the H-AGN values are calculated using the errors on the value of σ, whereas we assume a fixed σ for the data. If the sample contains a sufficiently large number of systems, we fit the observational data to Equation 3.11 (as described in subsection 3.4.1) to obtain σ. Otherwise we used the quoted uncertainty on each offset.

The considerable difference between using the fitted σ or the published errors is expected from B00, since the errors are given as lower bounds, so the offsets relative to these errors will be larger. The inconsistency between the two values of Λ for B19 suggest that the reported errors may be too conservative, as they give a much lower

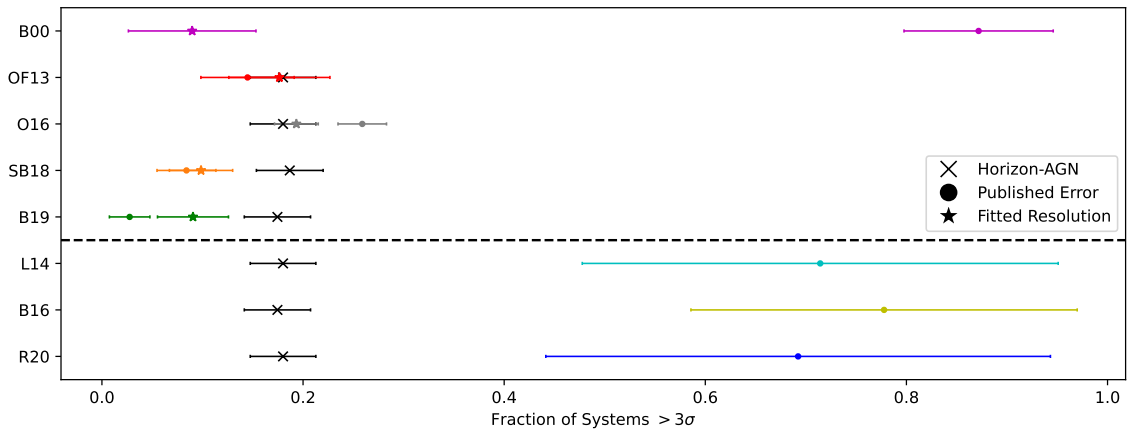


Figure 3.12: The fraction of systems with offsets $> 3\sigma$ from the various datasets (coloured points), compared to different cuts from H-AGN (black points). The errors are from the maximum likelihood estimate and show the 95 per cent confidence intervals. For the circular markers, we set σ to be the published error on each offset, whereas for the points marked with stars we use the fitted value of σ from Equation 3.10. The three datasets below the dashed line have very few systems, hence the large errors, and are only included for completeness; we make no further quantitative comparison between these and H-AGN. The reader should therefore focus mainly on the points above the dashed line.

Λ than the fitted offset. To mitigate these issues, henceforth we choose σ to be the fitted value for the samples where this is possible (B00, OF13, O16, SB18 and B19). Otherwise we use the published errors.

Due to their small sample sizes, the L14, B16 and R20 values have very large errors, so it is hard to say whether these are discrepant with the other values. We make no further quantitative comparison using these data, such as assigning a probability of consistency with H-AGN; these are only included for completeness. We see that B00, SB18 and B19 give similar values of Λ of 0.09, 0.10 and 0.09 respectively. These values are lower than those calculated from H-AGN. OF13 has 18 per cent of systems having offsets greater than 3σ, which is consistent with the 18 per cent of systems obeying this criterion in H-AGN.

The only data to have a larger fraction of offset BHs than H-AGN is O16, with 19 per cent of systems having offsets above 3σ, although the values are consistent within the errors.

We now perform a KS test on the normalised angular offset, $r_{\rm GB}/\sigma$, for all the datasets, to test the null hypothesis that the normalised offsets for each sample of galaxies are drawn from the same population distribution. The results are plotted in

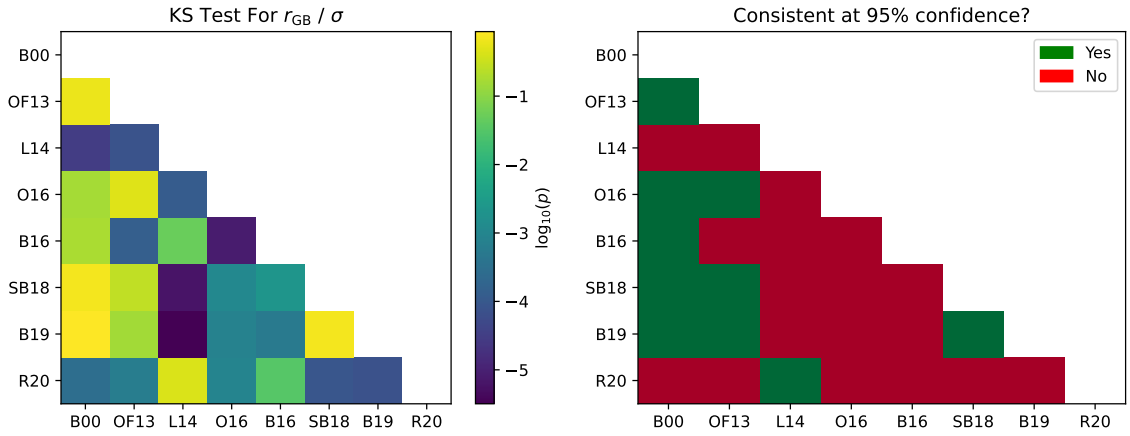


Figure 3.13: *Left:* The p-values of KS tests between the different datasets' angular offsets, normalised by the resolution, $r_{\rm GB}/\sigma$, to see if they could be drawn from the same distribution. $p = 1$ indicates that the underlying distributions are identical. *Right:* A summary of the KS test, where green indicates $p > 0.05$ and red indicates $p < 0.05$; the datasets in green are consistent with each other.

Figure 3.13. Since B16 only contains the $> 3\sigma$ offsets, we cut the other datasets to include only the $> 3\sigma$ offsets when comparing to B16.

From this we can make the following observations

- B00, OF13, SB18, B19 are mutually consistent with each other.
- The $> 3\sigma$ distribution in B00 is consistent with B16.
- L14 is consistent with R20.
- O16 is consistent with OF13 and B00.

This means that each dataset is consistent with at least one of the others, but the datasets are not all consistent with each other. It is interesting that the tail of B19 is not consistent with B16, indicating that the change in selection criteria has altered their results.

3.5.4 Comparison of Horizon-AGN with observations

In Table 3.3 we give the results for fitting a_2 to the various datasets in terms of the fraction of offset BHs in the data, $f_{\rm crit}$, and from H-AGN, $P_2\,(\text{offset})$, using Equation 3.13. We also calculate the probability of generating datasets with the fraction of offset BHs up to $f_{\rm crit}$, $P_{\rm crit}$, using Equation 3.19 and $N_{\rm dist} \sim 2 \times 10^6$.

As noted in subsection 3.5.1, we find that the systems appear more offset in three dimensions than if they are projected onto the sky, so for consistency we only consider $d = 2$ here. Even using the two dimensional offsets, we find that H-AGN

Table 3.3: Comparison of observations to H-AGN samples ($z = 0.1$) designed to mimic their selection criteria. P_2(offset) is the fraction of offset BHs (Equation 3.13), calculated by fitting the 2D projected offsets to Equation 3.11. It is independent of the form of the intrinsically offset distribution. $f_{\rm crit}$ is the corresponding quantity for the observations. $P_{\rm crit}$ is the probability of generating a fraction of up to $f_{\rm crit}$ offset BHs. The selection criteria have little effect on P_2(offset); for the quasar-mode sample with no further cuts this is 0.27 ± 0.14.

	$f_{\rm crit}$	P_2 (offset)	$P_{\rm crit}$ $(d = 2)$
OF13	0.18 ± 0.03	0.27 ± 0.14	0.07
O16	0.24 ± 0.02	0.27 ± 0.14	0.48
SB18	0.17 ± 0.04	0.28 ± 0.11	0.05
B19	0.09 ± 0.02	0.27 ± 0.10	0.003

predicts a higher fraction of offset BHs, although the large errors obtained by just using a_2 to determine a probability of being offset results in P_2(offset) and $f_{\rm crit}$ being approximately consistent.

In most cases, $P_{\rm crit}$ is never greater than a few per cent, and is only 0.3 per cent for B19, indicating that it is unlikely to create these observations given the offset distributions from H-AGN. The one exception is for O16, where there is a 48 per cent chance of producing up to this fraction of offset BHs. From Figures 3.12 and 3.13 we know that O16 does not agree well with most other observations. Furthermore, as noted in subsection 3.3.5, these data did not undergo follow-up observations, so it is likely that the increased fraction of offset BHs is due to spurious sources. Given that H-AGN overpredicts the fraction of offset BHs compared to the other seven studies considered here, we conclude that the fraction of offset BHs in H-AGN is larger than observed. This result is strengthened by the argument that the calculated value of $f_{\rm crit}$ should be treated as an upper limit (see section 3.3).

3.6 Discussion

3.6.1 Systematic uncertainties

3.6.1.1 Effect of galaxy mass

A potentially important consideration is whether the stellar masses of the observed galaxies are compatible with those of the H-AGN sample. The relatively small box size of H-AGN means that it is dominated by lower mass objects. The observational masses are obtained from the MPA-JHU analysis of SDSS data, based on the methods of Brinchmann *et al.* [331], Kauffmann *et al.* [332] and Tremonti *et al.* [333].

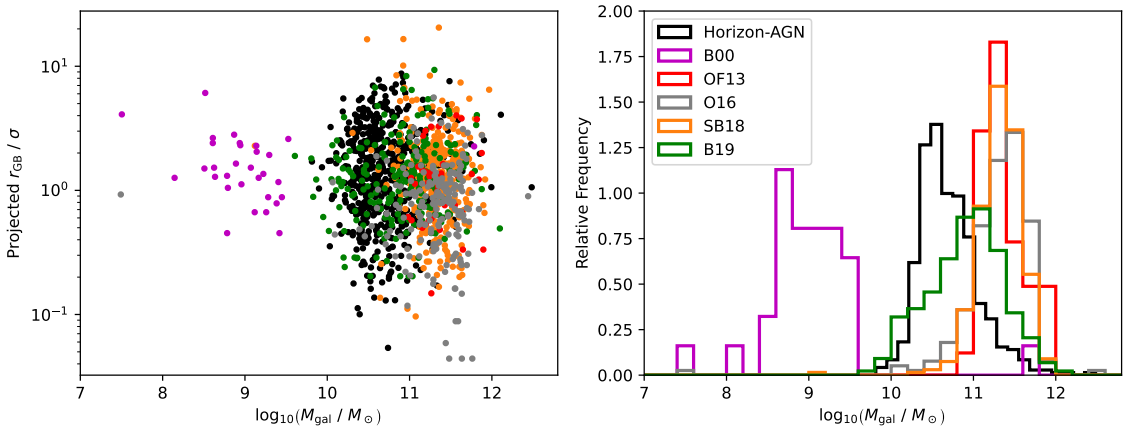


Figure 3.14: The distributions of host galaxy masses considered in B00, O13, O16, SB18, B19 and the quasar mode sample from H-AGN. In the left panel we plot the mass-offset plane, where the offsets are normalised by the fitted resolution, σ.

Table 3.4: Spearman correlation coefficients, r, between the BH offset, $r_{\rm GB}$, and galaxy mass, $M_{\rm G}$. The p value is for a two-sided hypothesis test, with the null hypothesis that $r_{\rm GB}$ and $M_{\rm G}$ are uncorrelated. Most datasets show no significant correlation.

Sample	Spearman r	p-value
B00	-0.35	0.06
OF13	0.16	0.32
O16	0.03	0.67
SB18	-0.07	0.18
B19	0.15	0.03
H-AGN (Quasar Mode)	0.02	0.60

In Figure 3.14 we plot the mass distributions from the four samples and the mass-offset plane, normalised by σ. The Spearman regression coefficients and p values for the test that the mass and offset are uncorrelated are given in Table 3.4. For all samples except B19 there is very little overlap in mass between the observations and H-AGN. However, we find very little variation of offset with galaxy mass, although B19 does exhibit a slight positive correlation with $p = 0.03$.

We repeated the Monte Carlo sampling for the SB18-like H-AGN sample, but now imposing that the galaxy mass satisfies $M_{\rm G} > 4.3 \times 10^{10} M_\odot$, which gives the most massive 50 per cent of the sample. We find that this mass cut has very little effect on the fitting parameters. In particular, we find that μ and s are consistent between the two samples; μ changes from 0.40 ± 0.06 to 0.39 ± 0.06 and $10^3 s$ slightly increases from 14 ± 2 to 19 ± 3. This further indicates that, for the mass range considered here,

the mass of the galaxy is unimportant.

3.6.1.2 Galactic vs projected luminosity centre

When considering the offsets from H-AGN, so far we have defined the centre of the galaxy to be the position of the densest star particle within a galaxy, where we use the full three-dimensional information. This clearly is not how the centre is defined in observations, so it is important to check that any discrepancy between observations and simulations is not due to the definition of the centre of the galaxy. To do this we must compare the true position of the galactic centre to that inferred from the projected distribution of star particles.

For every galaxy we find all star particles within a box, $\mathcal{B}_{\rm G}$, of side length $4r_{\rm eff}$, centred on the galactic centre. The projected coordinates of the $i^{\rm th}$ star particle are

$$x_i = (\alpha_i - \alpha_{\rm G}) \cos \delta_{\rm G}, \qquad y = \delta_i - \delta_{\rm G} \tag{3.22}$$

where the projected true centre has right ascension $\alpha_{\rm G}$ and declination $\delta_{\rm G}$ and the star particle has right ascension α_i and declination δ_i, measured by an observer at the centre of the box. Assuming these are distributed according to a Sérsic [334] distribution, such that the probability of being at some coordinates $(\tilde{x}_i, \tilde{y}_i) = (x_i - x_0, y_i - y_0)$, is

$$p\left(\tilde{x}_i, \tilde{y}_i\right) = I_0 \exp\left(-b_n \left[\left(\frac{R_i}{R_{\rm eff}}\right)^{\frac{1}{n}} - 1\right]\right), \tag{3.23}$$

where $\tilde{x}_i = R_i \cos\varphi$ and $\tilde{y}_i = R_i \left(1 - \epsilon\right) \sin\varphi$, for polar angle $\varphi \in [0, 2\pi)$ relative to the major axis, which is at an angle θ relative to the x axis, and ellipticity $\epsilon \in [0, 1)$. We normalise the probability distribution,

$$I_0 = \frac{b_n^{2n} e^{-b_n}}{2\pi n R_{\rm eff}^2 \Gamma\left(2n\right)\left(1 - \epsilon\right)}, \tag{3.24}$$

and define $R_{\rm eff}$ such that half of the probability lies within $R_{\rm eff}$,

$$\frac{\gamma\left(b_n, 2n\right)}{\Gamma\left(2n\right)} = \frac{1}{2}, \tag{3.25}$$

where $\gamma\left(z, a\right)$ is the incomplete lower gamma function. We thus must fit for 6 parameters ($\{R_{\rm eff}, n, x_0, y_0, \epsilon, \theta\}$), which we do by maximising the likelihood

$$\log \mathcal{L}_{\rm G}\left(R_{\rm eff}, n, x_0, y_0, \epsilon, \theta\right) = \sum_{i \in \mathcal{B}_{\rm G}} \log p\left(\tilde{x}_i, \tilde{y}_i\right). \tag{3.26}$$

We impose uniform priors on all parameters, so that the maximum likelihood is also the maximum of the posterior. We run the optimisation 5 times for each galaxy,

with a different start point each time, generated randomly from the priors and adopt the maximum likelihood of these 5 as the true maximum likelihood. We find this number to be sufficient for several of the endpoints of the optimiser to be coincident at a likelihood value above the remainder, indicating that they have reached the maximum-likelihood point and therefore that 5 repeats is sufficient.

Using the fitted projected luminosity centres (x_0, y_0), and assuming that the radial distance to the galactic centre is the same in this method as before, we can thus find the coordinates of the observed luminosity centre, $\boldsymbol{r}_{\mathrm{G,lum}}$.

Working with the quasar mode sample at $z = 0.1$, in 71 per cent of cases, the luminosity centre shifts by less than 1kpc. With both $d = 2$ and $d = 3$, the mean BH offset shifts by less than 0.08kpc. Fitting the results to a_d, we find that probabilities of being offset is

$$P_2\,(\text{offset},\ \text{lum}) = 0.30 \pm 0.12, \quad P_3\,(\text{offset},\ \text{lum}) = 0.40 \pm 0.16, \tag{3.27}$$

which are consistent with the values from the galactic centre in Table 3.3.

The effects are slightly greater for the radio mode sample, where 56 per cent of cases shift by less than 1 kpc and the mean difference is 0.18 kpc. This is still smaller than the resolution of H-AGN. The probabilities of being offset in two and three dimensions for the radio mode BHs are

$$P_2\,(\text{offset},\ \text{lum}) = 0.31 \pm 0.09, \quad P_3\,(\text{offset},\ \text{lum}) = 0.402 \pm 0.10, \tag{3.28}$$

which are again consistent with Table 3.3.

We therefore conclude that the offset fraction in H-AGN does not change significantly if we use the luminosity, rather than galactic, centre.

3.6.2 Comparison with other simulations

The majority of cosmological hydrodynamical simulations besides H-AGN do not include a prescription for dynamical friction on BHs. For example, Illustris-TNG [335], MassiveBlack-II (Khandai *et al.* 336, Khandai private communication) and Simba [337] artificially return their BHs back to the potential minimum of their host. The EAGLE simulations [261] also do this for BHs with masses below 100 times the gas particle mass.

The Magneticum simulations, however, do attempt to keep their BHs near the galactic centres using physical processes [338]: they impose strict momentum conservation during gas accretion and BH mergers and they include Chandrasekhar friction

[257]. The peak of their distribution of the BH offsets relative to the host's potential minimum occurs at $0.7h^{-1}$ kpc. Assuming this distribution takes the form of a_d (Equation 3.11) with $d = 3$, we would envisage the peak to be at $r_{\rm GB} = \sqrt{(d-1)}\sigma$. Taking σ to be half the minimum gravitational softening length ($2.0h^{-1}$ kpc), as is approximately true for H-AGN, this is value we would predict. Thus the distributions in offsets between the simulations are qualitatively similar and dominated by the resolution.

In subsection 3.5.2, we found that offset BHs tend to lie within the galactic plane. This is in contrast with Tremmel *et al.* [339], who found that wandering BHs are preferentially found outside of galactic discs, with $> 4\sigma$ confidence, for Milky Way (MW)-type halos. To make a more meaningful comparison with Tremmel *et al.* [339], we define a MW-type halo as having a mass $5 \times 10^{11} \leq M_{\rm H}/{\rm M}_\odot \leq 2 \times 10^{12}$. We consider all MW-type halos which are in a galaxy+halo structure and find all wandering BHs within 10 kpc of the centre of the halo. We use Tremmel *et al.* [339]'s definition of a wandering BH as one further than 0.7 kpc from the halo centre. Defining the galactic plane such that the normal of the plane is aligned with $\boldsymbol{J}_{\rm G}$, we find that 50 per cent of BHs lie within 30° of the galactic plane (which is the value expected for a uniform distribution), compared to 20 ± 7 per cent for Tremmel *et al.*'s sample. Comparing this distribution to an isotropic distribution with a KS-test, we obtain a p-value of $p = 0.50$, indicating that wandering BHs in MW-type halos in H-AGN are distributed isotropically.

These results are not necessarily in tension with those of Tremmel *et al.*, since we analyse a very different population of BHs: the minimum mass BH of Tremmel *et al.* is $10^6\,{\rm M}_\odot$, whereas we cannot use BHs below $2 \times 10^7\,{\rm M}_\odot$. Since the wandering BH population only dominates the mass budget of BHs below $\sim 10^5\,{\rm M}_\odot$ [237], we do not expect to capture the behaviour of the full wandering BH population in H-AGN, but only the most massive cases. This could also explain why we find far fewer BHs within the virial radius of a MW-type halo than Tremmel *et al.*: 1.4 ± 0.7 compared to 12.2 ± 8.4.

As noted in subsection 3.4.1, the large value of σ for H-AGN means that we have to introduce a sigmoid function to suppress the contribution of b_d at small offsets. The upcoming New Horizon simulation (Dubois et al., in preparation) is a zoom-in of a 'field' environment of H-AGN, with a comoving radius of 10 Mpc. The simulation has reached $z = 0.7$ at the time of writing and has a much higher resolution than H-AGN, with a maximum spatial resolution corresponding to a physical scale of 40 pc at $z = 0$. It will be interesting to see how our results change with this increased

resolution, potentially removing the need for the sigmoid suppression and thus making the marginal distribution of b_d equal to b_{d-1}.

3.6.3 Interpretation of results

The dynamical friction model used in H-AGN results in a resolution-dominated distribution of offsets of BHs from the centres of their host galaxies. At large r_{GB} we find an exponentially decaying tail, as observed. The fraction of offset BHs in H-AGN, however, is larger than is observed, with a probability of only a few per cent of generating the observations from the simulated distributions (Table 3.3).

The dynamical friction model used (Equation 3.3) only depends on the gas parameters, and not on the star or DM particles. In order to resolve the force from these particles, the resolution should obey [340]

$$\Delta x < \frac{GM_{\mathrm{B}}}{\sigma_v^2} = 1\,\mathrm{pc}\left(\frac{M_{\mathrm{B}}}{10^7\,\mathrm{M}_\odot}\right)\left(\frac{\sigma_v}{200\,\mathrm{km\,s^{-1}}}\right)^{-2}, \tag{3.29}$$

where σ_v is the velocity dispersion of the particles of interest. Remembering that $\Delta x \sim 1\,\mathrm{kpc}$ for H-AGN, we see that it is not inconceivable that we would fail to obey this criterion, so that to improve the offset prediction a sub-grid model for dynamical friction from these particles would be needed.

This was investigated by Pfister *et al.* [341], who found that the stellar component of dynamical friction is more stabilising than its gaseous counterpart, ensuring that BHs remain centralised post-merger. At higher redshifts, however, the irregular stellar distribution prevents stars from providing this constant acceleration.

Although the current dynamical friction model has the desired effect of providing a restoring force such that the BH velocity tends to oppose its displacement (Figure 3.10), this orbital decay appears to stall at low redshift. Given that this is where Pfister *et al.* find stellar dynamical friction to be most effective, our results indicate that such a model should be included to match the observed BH offset distribution more closely.

It would be more concerning if the fraction of offset BHs in H-AGN was smaller than observed, since reasonable alterations must increase the dynamical friction force and thus reduce the magnitudes of the offsets. The contributions from stars and DM could not push a BH to larger r_{GB}.

3.7 Conclusions

We have studied the statistical properties of the offsets between BHs and the centres of their host galaxies at a range of redshifts in the H-AGN simulation, and compared to a set of observations. The H-AGN simulation is almost unique among cosmological hydrodynamical simulations in employing a subgrid model of dynamical friction for BHs rather than artificially advecting them to galaxies' centres at each timestep.

We described the distribution of these offsets as the combination of a Gaussian and a generalised symmetric Laplace distribution, where a more intrinsically offset BH has a larger contribution from the latter. The fraction of offset BHs was compared to observations from the literature, where we tested for consistency and investigated the relative mass dependences of the offsets. From a feature importance analysis we determined the properties of the halo-galaxy-BH system that most strongly affect the offset in the simulation and hence derived a physical picture for the system's evolution.

Our key findings are as follows:

- *The fraction of intrinsically offset BHs is higher in the simulation than in most observations.* Although both distributions are dominated by the resolution, we find ~ 27 per cent of BHs in H-AGN are intrinsically offset compared to the upper limits of $\sim 10 - 20$ per cent in the observations. We believe this to be due to the unmodelled dynamical friction from stars and DM in the simulation.
- *A higher fraction of the simulated BHs are intrinsically offset today than at earlier epochs.* This suggests that BHs form near the centres of galaxies at early times, but are then displaced by interaction and mergers, before slowly migrating back. This is expected in the hierarchical structure formation paradigm, where galaxies and their BHs interact and merge to form larger ones.
- *Offset BHs in H-AGN exist on prograde orbits in the plane of the galaxy with orbital radii that decay over time.* Although the total fraction of offset BHs increases with time, a given BH moves back towards the centre of its host galaxy due to dynamical friction, unless something causes it to become more offset. This orbital decay stalls at low redshift.

To prevent the orbital decay of BHs from stalling at low redshift, and hence align the fraction of offset BHs from the simulation more closely with the observations, future simulations should model the dynamical friction from stars and DM as well as gas. With observations planned to probe higher redshifts [286], it will soon be possible to compare the offset distributions at an epoch when dynamical friction from stars

is truly subdominant to that from gas, providing a further test of the BH physics that is implemented in cosmological simulations. Ultimately, this will be necessary to understand fully the role played by SMBHs in shaping the galaxy population over cosmic time.

Chapter 4

Galileons and Supermassive Black Holes

4.1 Introduction

Now that we have developed methods to produce Monte Carlo-based forward models to constrain fundamental physics in the astrophysical regime, we now adapt this approach to an observational test of Vainshtein screened modified gravity theories. This test was introduced by Hui and Nicolis [342], and is based on the same physical principle as the test presented in Sakstein *et al.* [2] and Asvathaman *et al.* [3]: we investigate whether the offset between a galaxy's centre and its central SMBH is preferentially aligned along the local gravitational field, as would be expected by the SEP violation in this class of modified gravity theories.

Specifically, we forward model the magnitude and direction of the BH–galaxy offset for each galaxy in the samples collated in [99]. The measured BH–galaxy offset for these galaxies is determined by cross-matching the optical centre to observations at a different wavelength, which provides the location of the active galactic nucleus. We use the `CSiBORG` suite of constrained N-body simulations of the local universe discussed in Section 1.3.2 to map out the large-scale Galileon field. Combining this with models for galaxy and halo structure allows us to make predictions for the BH–galaxy offsets. Marginalising over uncertainties in the Galileon field and galaxy properties, as well as parameters describing the noise due to non-fifth-force contributions to the signal, we compare our predictions to the observed offsets via a MCMC algorithm. We find that the strength of the fifth force relative to gravity, $\Delta G/G_{\rm N}$, is constrained to be < 0.16 at 1σ confidence; this bound is applicable to Galileons with crossover scales $r_{\rm C} \gtrsim H_0^{-1}$.

In Section 4.2 we discuss this phenomenon in the context of the cubic Galileon model, although the effect occurs more generally in Galileon theories due to their BH no hair theorem [343]. We present the observational data used in this work in Section 4.3. Section 4.4 details our inference methods and the results are presented in Section 4.5. We discuss systematic uncertainties and compare our constraints to previous work in Section 4.6 and conclude in Section 4.7.

4.2 Galaxy–black hole offsets in Galileon gravity

We consider a theory containing a single scalar field, φ, which respects the Galileon symmetry $\varphi \to \varphi + b + c_\mu x^\mu$ [344] and is Vainshtein screened [48] on small scales. A common example is the cubic Galileon, which has the action

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{R}{16\pi G} - \frac{1}{2}c_2\mathcal{L}_2 - \frac{1}{2}c_3\mathcal{L}_3 - \mathcal{L}_\mathrm{m}\right], \tag{4.1}$$

where R is the Ricci scalar, g is the determinant of the metric $g_{\mu\nu}$, c_3 and c_4 are constants, $\mathcal{L}_\mathrm{m}$ is the matter Lagrangian, and

$$\mathcal{L}_2 = \nabla_\mu\varphi\nabla^\mu\varphi, \quad \mathcal{L}_3 = \frac{2}{\mathcal{M}^3}\Box\varphi\nabla_\mu\varphi\nabla^\mu\varphi, \tag{4.2}$$

where $\mathcal{M}^3 = M_\mathrm{Pl}H_0^2$. Note that these Lagrangians are required to result in second-order equations of motion and are thus special cases of Horndeski theory [345]: the most general theory of gravity in four dimensions with second-order equations of motion constructed from the metric tensor and a scalar field. Equation 4.1 contains the usual Einstein-Hilbert term (Equation 1.1); the frame which contains this term is referred to as the Einstein frame.

There are two branches to the cubic Galileon, depending on the sign of the kinetic term. If, using the mostly minus signature, $c_2 > 0$ (normal branch) we have a scalar field with a canonical kinetic term; these cannot self-accelerate and are simply models of fifth-forces. On the other hand, if $c_2 < 0$ (self-accelerating branch) then the field can self-accelerate and does not necessarily require a cosmological constant [346]. This self-acceleration can most easily be seen in the tracker solution (see Section 4.6.2), where the Galileon density parameter is $\Omega_\varphi = \mathcal{A}c_2\xi^2$, where $\mathcal{A} < 0$ is a constant which depends on whether one considers the cubic, quartic or quintic Galileon model [347] and ξ is a real constant. For Ω_φ to be a positive constant and thus take the role of Ω_Λ, one requires $c_2 < 0$. One finds that the condition $c_2 < 0$ is more generally required for self-acceleration in Galileon models, even when one no longer considers the tracker solution [348].

One may be concerned that models with $c_2 < 0$ will have a field with negative kinetic energy, introducing ghost instabilities. However, the non-linear kinetic terms in the action make the condition for no ghosts to be less trivial than requiring $c_2 > 0$. In fact, it has been known since the introduction of Galileon theories that such models admit self-accelerating de Sitter solutions without ghost instabilities [344]. Constraints on model parameters have been derived to avoid such instabilities for both scalar and tensor perturbations [349], allowing theoretically reasonable self-accelerating solutions.

If we work in the quasi-static approximation and neglect terms suppressed by the Newtonian potentials and their spatial derivatives, we obtain [350] the equation of motion for perturbations, φ, about a background, $\bar{\varphi}$,

$$\nabla^2\varphi + \frac{1}{3\beta_1 a^2\mathcal{M}^3}\left[\left(\nabla^2\varphi\right)^2 - \nabla_i\nabla_j\varphi\nabla^i\nabla^j\varphi\right] = \frac{M_{\rm Pl}}{3\beta_2}8\pi G_{\rm N}a^2\bar{\rho}\Delta, \tag{4.3}$$

where $i \in \{1, 2, 3\}$ and

$$\beta_1 = \frac{1}{6c_3}\left[-c_2 - \frac{4c_3}{\mathcal{M}^3}\left(\ddot{\bar{\varphi}} + 2H\dot{\bar{\varphi}}\right) + \frac{2\kappa c_3^2}{\mathcal{M}^6}\dot{\bar{\varphi}}^4\right], \quad \beta_2 = 2\frac{\mathcal{M}^3 M_{\rm Pl}}{\dot{\bar{\varphi}}^2}\beta_1, \tag{4.4}$$

with $\kappa \equiv 8\pi G$. Our test will depend only on Equation 4.3 which holds for both signs of c_2; our constraints therefore apply to both branches. We can rewrite Equation 4.3 in a more familiar form,

$$\nabla^2\varphi + \frac{r_{\rm C}^2}{3}\left[\left(\nabla^2\varphi\right)^2 - \nabla_i\nabla_j\varphi\nabla^i\nabla^j\varphi\right] = 8\pi\alpha G_{\rm N}\bar{\rho}\Delta, \tag{4.5}$$

where α describes the strength of the coupling of the Galileon to matter, and $r_{\rm C}$, called the 'crossover scale', parameterises the new kinetic terms. We note that, using Equation 4.4, these parameters are functions of time. In this work we use low-redshift observations so will ignore this temporal evolution and consider only their present day values. Since the coupling in self-accelerating models tends to dramatically increase as we approach the present day [350], the constraints we find imply a bound on α over the history of the Universe for this branch.

To remain agnostic to the details of the Galileon model, we assume that α and $r_{\rm C}$ are independent. This is not true for all models: in the Dvali-Gabadadze-Porrati (DGP) model [351], for example, α is related to $r_{\rm C}$ and the Hubble parameter, $H(t)$, as [352]

$$\alpha_{\rm DGP}(t) = \frac{1}{3}\left[1 \pm 2Hr_{\rm C}\left(1 + \frac{\dot{H}}{3H^2}\right)\right]^{-1}. \tag{4.6}$$

An overdot denotes a derivative with respect to cosmic time t, the $+$ sign refers to the normal branch and the $-$ sign to the self-accelerating branch. Although the DGP model is a five-dimensional model (where our Universe is the four-dimensional boundary of a five-dimensional spacetime), by integrating out the bulk degrees of freedom, one finds that the four-dimensional effective action consists of a graviton and an extra scalar degree of freedom describing the extrinsic curvature of the boundary [353, 354]. This action contains the cubic term in Equation 4.2 and was the original inspiration for the Galileon family of models. To test a specific model, one should compare the model's trajectory in the $\alpha - r_{\rm C}$ plane to the constraints obtained in this work.

Far outside the Vainshtein radius, $r_{\rm V}$, the new kinetic terms are negligible and we recover Poisson's equation. For a source of mass M, this transition occurs at

$$r_{\rm V} = \left(\frac{4}{3}\alpha G_{\rm N} M r_{\rm C}^2\right)^{\frac{1}{3}}. \tag{4.7}$$

Within the Vainshtein radius the fifth force has magnitude

$$a_5 = -\alpha\nabla\varphi = \frac{\Delta G}{G_{\rm N}}\frac{G_{\rm N} Q M}{r^2}\left(\frac{r}{r_{\rm V}}\right)^q \tag{4.8}$$

where $q = 3/2$ for the cubic Galileon, Q is the scalar charge of the object given in terms of its stress-energy tensor as

$$Q = \int T^{\mu}{}_{\mu} {\rm d}^3x, \tag{4.9}$$

and

$$\frac{\Delta G}{G_{\rm N}} \equiv 2\alpha^2. \tag{4.10}$$

The suppression of a_5 for $r \ll r_V$ is what constitutes Vainshtein screening.

For a non-relativistic object, Q is equivalent to the object's mass, but for compact objects $Q < m$ because T does not include gravitational binding energy. The limiting case is a black hole, for which $Q = 0$.

Due to the Galileon symmetry, by adding a term with a linear gradient, one can always generate a new solution $\varphi \to \varphi + \varphi_{\rm ext}$, where $\nabla\varphi_{\rm ext}$ is a constant. Therefore, although stars in galaxies tend to reside within their host's Vainshtein radius, this does not mean they cannot feel a fifth force. Rather, they interact with the field sourced by large scale structure [342], which has a wavelength long compared to the Vainshtein radius and hence has approximately constant gradient on the scale of the

galaxy. Cosmological simulations have confirmed this prediction [355] and indicate that φ obeys linear dynamics on scales $\gtrsim 10\,\mathrm{Mpc}$ for $r_\mathrm{C} \simeq 6\,\mathrm{Gpc}$ [356–359].

In conjunction with the no-hair theorem described above, this property of the Galileon symmetry can lead to an offset between the centre of a galaxy and its central BH. This occurs since, if a galaxy is falling down a scalar field potential, the non-relativistic matter feels the attractive fifth force, whereas the BH does not ($Q = 0$). Therefore the BH lags behind the galaxy. The offset is stabilised by the gravitational force between the BH and the galaxy and its DM halo, which can lead to a constant displacement in equilibrium.

4.3 Observational data

In this chapter we use the four largest datasets collated and summarised in Section 3.3 [99] which contain measurements of the offsets between an AGN and its host galaxy's centre: OF13 [278], O16 [281], SB18 [282] and B19 [283]. Each of these cross-match the optical centres of galaxies from the Sloan Digital Sky Survey (SDSS) [294] to observations at a different wavelength, where the latter provides the position of the AGN. OF13, O16 and SB18 search for radio counterparts, using the International Celestial Reference Frame (ICRF2) [360], mJIVE-20 [295] and the Cosmic Lens All-Sky Survey (CLASS) [299, 300] respectively. In B19 AGN positions are obtained from *Chandra* X-ray data [298]. The distributions of offsets from these samples are dominated by a Gaussian component describing the spatial resolution of the measurements, with width $\sigma_\mathrm{obs} \sim 50\,\mathrm{mas}$ for the radio samples and $\sigma_\mathrm{obs} \sim 150\,\mathrm{mas}$ for B19. Approximately 10-30 per cent of the probability density can be attributed to a non-Gaussian component (e.g. a Laplace distribution) which is dominant in the tails of the distribution [99]. The degree to which an AGN is intrinsically offset is given by the ratio of the non-Gaussian to Gaussian terms.

We plot the physical and angular offsets as a function of redshift for the galaxies used in this chapter in Figure 4.1. The galaxies are typically at redshift $z \sim 0.1$, such that a $3\sigma_\mathrm{obs}$ offset for a galaxy from the radio samples corresponds to a physical offset of $\sim 340\,\mathrm{pc}$.

For information on the halo structures (which determine the restoring force), we cross-correlate these data with the Nasa Sloan Atlas (NSA)[1] to find the closest source within $0.5'$. The NSA contains measured and derived quantities for nearby galaxies using state of the art sky subtraction and photometric determinations [361] in the

[1] `www.nsatlas.org`

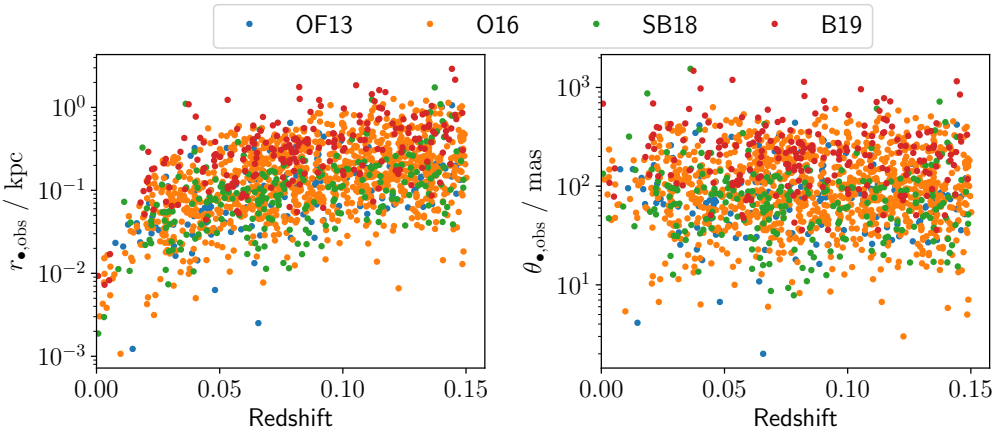


Figure 4.1: Observed BH–galaxy offset as a function of redshift for the galaxies used in this work. The left panel gives the physical offsets and the right panel gives the angular offsets. For redshift we use `zdist` from the NSA.

optical and near-infrared, largely from the Sloan Digital Sky Survey. Approximately 10 per cent of galaxies are discarded due to not having an NSA counterpart, and we retain 144, 1328, 230 and 214 galaxies for the OF13, O16, SB18 and B19 samples respectively.

For distances, we use `zdist`, which is determined using the peculiar velocity model of Willick *et al.* [362]. Since, as described in Section 4.4.4, we use a Sérsic profile to determine the central baryonic surface density, we use quantities relevant to such a profile: stellar mass $M_\star$ = `sersic_mass`, apparent Sérsic minor-to-major axis ratio $(b/a)_{\mathrm{obs}}$ = `sersic_ba`, Sérsic index $n_\star$ = `sersic_n`, and half-light radius along the major axis r_{eff} = `sersic_th50`.

4.4 Methods

Despite the different theoretical background and observational signals, our approach is similar to [363, 364]; we forward model the offset, $\boldsymbol{r}_\bullet$, for the galaxies in our samples and, in conjunction with an empirical noise model describing astrophysical contributions to $\boldsymbol{r}_\bullet$, derive a likelihood function for the observed offsets for a given $\Delta G/G_{\mathrm{N}}$ and r_{V}. For a fixed r_{V}, we then constrain $\Delta G/G_{\mathrm{N}}$ by MCMC. As detailed in Section 4.4.3, we take r_{V} to be a universal free parameter instead of using a different Vainshtein radius for each galaxy. This will make our constraint on $\Delta G/G_{\mathrm{N}}$ at a given r_{C} conservative as we will systematically under-estimate the magnitude of the Galileon field.

We derive the offset expected in Galileon gravity in Section 4.4.1 and the gravitational field required to determine this in Sections 4.4.2 and 4.4.3. In Section 4.4.4 we convert this to a predicted offset for each galaxy. Using Monte Carlo sampling, we obtain a distribution of offsets, which is modelled as in Section 4.4.5. We utilise one of the empirical noise models outlined in Section 4.4.6 to calculate the likelihood function in Section 4.4.7. The parameters which are fixed in this section are summarised in Table 4.1.

4.4.1 Offset expected from a fifth force

In this section we briefly summarise the derivation of [342] for the predicted offset of a BH from the galactic centre in Galileon gravity.

In the rest frame of the galaxy, the black hole equation of motion is

$$\ddot{r} = -\frac{G_N M\,(<r)}{r^2} + a_{\rm BH}, \tag{4.11}$$

where $M\,(<r)$ is the mass enclosed at a distance r from the galaxy's centre. Since the galaxy receives an additional acceleration from the fifth force, the restoring force on the BH must be the same for equilibrium. Using Equation 4.8, this is

$$\boldsymbol{a}_{\rm BH} = -\alpha\nabla\varphi_{\rm ext} = -\left(\frac{\Delta G}{G_N}\right)\nabla\Phi_{\rm lss} = \left(\frac{\Delta G}{G_N}\right)\boldsymbol{g}_{\rm lss}, \tag{4.12}$$

where $\Phi_{\rm lss}$ and $\boldsymbol{g}_{\rm lss}$ are the gravitational potential and acceleration sourced by large scale structure respectively. The Galileon force is therefore proportional to the regular gravitational force. Note that this is only true in the linear regime of the Galileon, where it satisfies a Poisson equation identical to the gravitational potential up to a normalisation factor of 2α.

Since the mean predicted offset is $\mathcal{O}(10\,\mathrm{pc})$ for $\Delta G/G_{\rm N} = 1$, we are only interested in the very central regions of our galaxies. We therefore assume a constant density, ρ_0, giving an enclosed mass $M(<r) = 4\pi\rho_0 r^3/3$. At equilibrium $\ddot{r} = 0$, so the offset between the BH and the centre of the galaxy, $r_\bullet$, is

$$r_\bullet = \frac{3}{4\pi}\frac{\Delta G}{G_N}\frac{|\boldsymbol{g}_{\rm lss}|}{G_{\rm N}\rho_0}. \tag{4.13}$$

We define the right ascension (J2000; RA; α) and declination (Dec; δ) directions on the plane of the sky and decompose the offset into these components. Since the observations are two dimensional, from now on we will use $\boldsymbol{r}_\bullet$ to refer to $r_{\bullet,\alpha}\hat{\alpha} + r_{\bullet,\delta}\hat{\delta}$ and we define the angular offset

$$\boldsymbol{\theta}_\bullet \equiv \frac{\boldsymbol{r}_\bullet}{d_{\rm A}} = \theta_{\bullet,\alpha}\hat{\alpha} + \theta_{\bullet,\delta}\hat{\delta}, \tag{4.14}$$

for angular diameter distance $d_{\rm A}$.

Table 4.1: The fixed parameters used to convert dynamical information from the NSA to a predicted offset as described in the text. Above the horizontal line we give the parameters used in the fiducial analysis, and the remainder are used in Section 4.6.1.1 to test for systematics. In the final column we give the value chosen for each parameter, although we show in Section 4.6 that our results are unchanged for reasonable alternative values.

Parameter	**Description**	**Value**
`ba_min`	The minimum allowed minor-to-major axis ratio.	0.15
$\sigma_{\rm M}$	Scatter (dex) in $M_\star - M_{\rm gas}$ relation (Equation 4.32).	0.3
$\sigma_{\rm R}$	Scatter (dex) in $R_{\rm eff,gas} - R_{\rm eff}$ relation (Equation 4.33).	0.25
$\sigma_{\rm D}$	Additional scatter (dex) in dynamical surface density.	0.5
$n_{\rm gas}$	Sérsic index for gas component	1
`nsim`	The number of constrained simulations used to reconstruct the gravitational field.	106
N	The number of grid points per side length used to reconstruct the gravitational field.	512
ℓ	The size of the box used to create artificial long wavelength modes for the gravitational field, in units of the box length of the constrained simulations.	6
`N_MC`	The number of Monte Carlo runs to get the distribution of offsets for the template signal.	500,000
`N_AM`	The number of abundance matching realisations.	200
n	Slope of halo density profile in the central regions of the halos.	$0 \leq n < 1$

4.4.2 Modelling the gravitational field

In this section we describe how we model the gravitational field using the `CSiBORG` ΛCDM N-body simulations and the matter power spectrum for a ΛCDM cosmology. Although for given initial conditions the presence of a Galileon increases power on large scales [365], our use of concordance cosmological parameters reproduces the observed matter power spectrum [178], justifying our use of these as an approximation to the density field for a GR+Galileon cosmology. Nevertheless, if the power was enhanced on large scales then we would be systematically underestimating the magnitude of the gravitational (and hence Galileon) field and therefore overestimating the strength of the coupling. This makes our constraints on $\Delta G/G_{\rm N}$ conservative. This is further justified post-hoc by our tight constraint on fifth-force strength, which ensures that differences between the ΛCDM and 'true' density fields cannot be large.

We model the density fields produced in the `CSiBORG` simulations, $\Delta(\boldsymbol{x})$, by applying a cloud-in-cell algorithm to the DM particles and solve Poisson's equation

$$\Phi\left(\boldsymbol{k}\right) = -\frac{4\pi G_{\rm N}\bar{\rho}}{k^2}\Delta\left(\boldsymbol{k}\right), \tag{4.15}$$

on a grid with $N = 512$ grid points per side, where $\Delta(\boldsymbol{k})$ is the Fourier transform of $\Delta(\boldsymbol{x})$. We have checked that using a coarser resolution ($N = 256$) does not affect our results.

Due to the finite size of the box, we do not have information about the gravitational field all the way down to $k = 0$; we can only construct modes with $k \geq \pi/L$ for box length L. To fully reconstruct the gravitational field, we must therefore add in long wavelength modes [366]. Unlike the modes captured by the constrained simulation, we have no constraints on the direction of these, so each mode is added as a noise term with a random orientation.

We start by generating a continuous matter power spectrum, $P(k)$, using `CLASS` [367], assuming a ΛCDM cosmology with the same parameters as in Section 1.3.2. We then construct a grid of size $L' = \ell L$ for $\ell > 1$ with N' grid points, such that the maximum k obeys $k_{\rm max} \geq \pi/L$. This ensures that the modes added here begin where those in the simulation boxes end.

A Gaussian random field, $\psi(\boldsymbol{k})$, is generated on the grid, with the condition $\psi(\boldsymbol{k}) = \psi^*(-\boldsymbol{k})$ to ensure the density contrast is real. To determine this, we must adapt the continuous power spectrum for the discrete case [368, 369]; we must account for the normalisation in our Fourier convention, the change in measure and the units of the power spectrum. After doing this, we can obtain the potential by solving Equation 4.15.

We filter out all modes which overlap between the large and small boxes to prevent double counting and inverse Fourier transform to obtain $\boldsymbol{g}(\boldsymbol{r})$ from these large scale modes. These are added to the field obtained from the constrained simulation. Since the magnitude and direction of the modes added in the above procedure are not constrained, we must marginalise over the direction and magnitude by incorporating this into our Monte Carlo sampling. The addition of long wavelength modes increases the root mean square of $|\boldsymbol{g}|$ by 5% and scatters each Cartesian component by $\sim 17\%$ compared to their uncorrected values.

When adding in long wavelength modes to account for scales not captured by the constrained simulations, one should ensure that the size of the larger box, $L' = \ell L$, is sufficiently large to accurately capture all small-k modes. To find a suitable value for this, for simplicity we assume that the modes are continuous (so we are integrating rather than summing on a grid) and denote these by the subscript 'cts'. In this case, the expectation value of the square of the gravitational field at position $\boldsymbol{x}$ is

$$\begin{aligned}\left\langle \left|\boldsymbol{g}_{\mathrm{cts}}(\boldsymbol{x})\right|^2 \right\rangle &= (4\pi G\bar{\rho})^2 \int \frac{\mathrm{d}^3\boldsymbol{k}}{(2\pi)^3}\frac{\mathrm{d}^3\boldsymbol{k}'}{(2\pi)^3}\frac{\boldsymbol{k}\cdot\boldsymbol{k}'}{k^2k'^2} e^{i(\boldsymbol{k}-\boldsymbol{k}')\cdot\boldsymbol{x}} \left\langle \Delta_{\mathrm{cts}}(\boldsymbol{k})\,\Delta^*_{\mathrm{cts}}(\boldsymbol{k}')\right\rangle \\ &= 8\,(G\bar{\rho})^2 \int_0^\infty P(k)\,\mathrm{d}k,\end{aligned} \tag{4.16}$$

since the power spectrum is as defined in Equation 1.20. For simplicity, we assume the matter power spectrum can be described by a broken power law

$$P(k) \approx \begin{cases} P_{\mathrm{eq}}\left(\frac{k}{k_{\mathrm{eq}}}\right), & k < k_{\mathrm{eq}} \\ P_{\mathrm{eq}}\left(\frac{k}{k_{\mathrm{eq}}}\right)^{-3}, & k > k_{\mathrm{eq}}. \end{cases} \tag{4.17}$$

For a finite box size, one cannot calculate the integral in Equation 4.16, since one can only integrate from some finite k_0 to obtain $\langle|\boldsymbol{g}_{\mathrm{cts}}|^2\rangle_0$, such that

$$\left\langle \left|\boldsymbol{g}_{\mathrm{cts}}\right|^2 \right\rangle = \left\langle \left|\boldsymbol{g}_{\mathrm{cts}}\right|^2 \right\rangle_0 + \delta g^2_{\mathrm{cts}}(k_0), \tag{4.18}$$

where the correction from long wavelength modes is

$$\delta g^2_{\mathrm{cts}}(k_0) \equiv 8\,(G\bar{\rho})^2 \int_0^{k_0} P(k)\,\mathrm{d}k. \tag{4.19}$$

In the case where $k_0 < k_{\mathrm{eq}}$ we have

$$\delta g^2_{\mathrm{cts}} = (2G\bar{\rho}k_0)^2 \frac{P_{\mathrm{eq}}}{k_{\mathrm{eq}}}, \tag{4.20}$$

whereas for $k_0 > k_{\rm eq}$ we have

$$\delta g^2_{\rm cts} = \delta g^2_{\rm cts}\left(k_{\rm eq}\right) + 8\left(G\bar{\rho}\right)^2 \int_{k_{\rm eq}}^{k_0} P\left(k\right) {\rm d}k = \left(2G\bar{\rho}\right)^2 P_{\rm eq} k_{\rm eq} \left(2 - \left(\frac{k_{\rm eq}}{k_0}\right)^2\right). \quad (4.21)$$

By sending $k_0 \to \infty$, we see that

$$\left\langle \left|\boldsymbol{g}_{\rm cts}\right|^2 \right\rangle = 8\left(G\bar{\rho}\right)^2 P_{\rm eq} k_{\rm eq}, \quad (4.22)$$

and so we arrive at, defining $L_0 \equiv 2\pi/k_0$ and $L_{\rm eq} \equiv 2\pi/k_{\rm eq}$,

$$\frac{\delta g^2_{\rm cts}\left(L_0\right)}{\left\langle \left|\boldsymbol{g}_{\rm cts}\right|^2 \right\rangle} = \begin{cases} \frac{1}{2}\left(\frac{L_{\rm eq}}{L_0}\right)^2, & L_0 > L_{\rm eq} \\ 1 - \frac{1}{2}\left(\frac{L_0}{L_{\rm eq}}\right)^2, & L_0 < L_{\rm eq}. \end{cases} \quad (4.23)$$

Using $L_{\rm eq} \approx 450\,{\rm Mpc}$, we find that for $L' = 6\,{\rm Gpc}$ ($\ell \approx 6$), the correction is $\sqrt{\delta g^2_{\rm cts} / < |\boldsymbol{g}_{\rm cts}|^2 >} \approx 0.05$, or for $L' = 14\,{\rm Gpc}$ ($\ell \approx 14$), this is ≈ 0.02. This is small compared to the ~ 30 per cent variation in $|\boldsymbol{g}|$ across the constrained simulations, so using either of these values is acceptable. We investigate the choice of ℓ further in Section 4.6.1.3.

4.4.3 Modelling the Galileon field

In Equation 4.12 we assume that the fifth force is proportional to the gravitational field sourced by large scale structure, while on small scales the Galileon is screened. As a model for this, we take the field calculated in Section 4.4.2 and apply a low-pass filter, such that we remove all k modes with $|\boldsymbol{k}| > k_{\rm V}$, corresponding to a Vainshtein radius for large scale structure of

$$r_{\rm V} \equiv \frac{2\pi}{k_{\rm V}}. \quad (4.24)$$

We choose a constant $r_{\rm V}$ for all galaxies, instead of filtering at the scale of each galaxy's Vainshtein radius, $r_{\rm V}^{(g)}$, as a function of its mass and $r_{\rm C}$ (Equation 4.7). This essentially corresponds to an average over all galaxies, and simplifies the analysis because we do not have to apply a different filter for each galaxy or re-filter each time we change $\Delta G/G_{\rm N}$ in our inference which would require us to re-derive the likelihood by running the Monte Carlo sampling again.

To convert our constraint on $\Delta G/G_{\rm N}$ as a function of $r_{\rm V}$ to one on α as a function of $r_{\rm C}$, we require $r_{\rm C}$ as a function of $r_{\rm V}$ and α. To determine this in a cosmological

context, we consider the mass enclosed within radius r due to cosmological perturbations:

$$M(r) = \bar{\rho} \int_{|\boldsymbol{x}|<r} \Delta(\boldsymbol{x}) \, \mathrm{d}^3 x. \tag{4.25}$$

We only consider the contribution from perturbations and not the background because the gravitational effects of the latter are encoded in the evolution of the Hubble parameter [370]. The mean square value is determined by the matter power spectrum,

$$\left\langle M^2(r) \right\rangle = (4\pi\bar{\rho})^2 \int \frac{\mathrm{d}^3 k}{(2\pi)^3} \frac{P(k)}{k^4} \left(\sin(kr) - kr\cos(kr)\right)^2, \tag{4.26}$$

where we use `CLASS` to compute the non-linear $P(k)$ for a ΛCDM cosmology. Using Equation 4.7, we can thus determine r_C as

$$r_\mathrm{C} \simeq \frac{1}{3} \left(\frac{3 r_\mathrm{V}^3}{4\alpha G \left\langle M^2(r_\mathrm{V}) \right\rangle^{\frac{1}{2}}} \right)^{\frac{1}{2}}, \tag{4.27}$$

where the arbitrary factor of 1/3 is included so that $r_\mathrm{V} = 10\,\mathrm{Mpc}$ corresponds to $r_\mathrm{C} \sim 6\,\mathrm{Gpc}$, as found in simulations [359]. We run our inference for a range of $r_\mathrm{V} > 1\,\mathrm{Mpc}$, with $r_\mathrm{V} = 100\,\mathrm{Mpc}$ as our fiducial case.

Using the power spectrum to obtain the covariance of $\Delta(\boldsymbol{k})$, we can find the expectation value of the square of the gravitational field from large scale structure[2], δg^2_cts, by squaring Equation 4.15 and only keeping modes $|\boldsymbol{k}| < k_\mathrm{V}$. Assuming $P(k) \propto k$ for $k < k_\mathrm{eq}$ where $L_\mathrm{eq} = 2\pi/k_\mathrm{eq} \approx 450\,\mathrm{Mpc}$, for two values of $r_\mathrm{V} < L_\mathrm{eq}$, r_1 and r_2, we find the fractional difference in the field is

$$\frac{\left(\delta g_\mathrm{cts}(r_1)^2\right)^{\frac{1}{2}} - \left(\delta g_\mathrm{cts}(r_2)^2\right)^{\frac{1}{2}}}{\left(\delta g_\mathrm{cts}(r_1)^2\right)^{\frac{1}{2}}} = 1 - \sqrt{\frac{1 - \frac{1}{2}\left(\frac{r_2}{L_\mathrm{eq}}\right)^2}{1 - \frac{1}{2}\left(\frac{r_1}{L_\mathrm{eq}}\right)^2}}. \tag{4.28}$$

This fractional difference is only 1 per cent between $r_\mathrm{V} = 1\,\mathrm{Mpc}$ and $r_\mathrm{V} = 100\,\mathrm{Mpc}$ or 5 per cent between $r_\mathrm{V} = 1\,\mathrm{Mpc}$ and $r_\mathrm{V} = 200\,\mathrm{Mpc}$. This implies that the calculated field is relatively insensitive to the choice of r_V provided $r_\mathrm{V} < L_\mathrm{eq}$.

In reality the separation between screened and unscreened modes will be more gradual than this step-function filter. However, this insensitivity to r_V suggests that a smoother filter will not dramatically change our results.

This model neglects the impact of the non-linear regime and requires the Galileon to be linear. This, combined with the insensitivity of the fifth force field to $r_\mathrm{V} < L_\mathrm{eq}$ means our constraints are valid for $10\,\mathrm{Mpc} \lesssim r_\mathrm{V} \lesssim 450\,\mathrm{Mpc}$.

[2]This definition is identical to Equation 4.19 and so δg^2_cts is given by Equation 4.23.

4.4.4 Calculating the offset

In order to the calculate the magnitude of the offset, $r_\bullet$, we need to know the total enclosed mass within separation $r_\bullet$, $M(< r_\bullet)$. In the absence of central kinematic data in our observational datasets, we must either attempt to fit a density profile to the galaxies using empirical methods such as abundance matching (AM) or employ empirical scalings between mass and light at the centre of the galaxies. We will find that $r_\bullet \ll$ size of the galaxy, so that the latter is more reliable as the former requires an integral over the galaxy's full luminosity profile. We therefore use this method.

Assuming a cored density profile, we wish to find the (constant) central density, ρ_0, using the information obtained from the NSA. To do this we must determine the major and minor axis lengths from the observed minor-to-major axis ratio, $(b/a)_{\rm obs}$, effective radius, $r_{\rm eff}$, and redshift, $z_{\rm dist}$. We also use the measured stellar mass, $M_\star$, and intensity profile to determine the central surface density, utilising observed correlations to estimate the contributions from DM and gas. Combining these two results gives us ρ_0.

Using the observed minor-to-major axis ratio $(b/a)_{\rm obs}$ from the NSA, we assign a random inclination, i, to the galaxy and, using simple geometry [371], estimate the true axis ratio to be

$$\left(\frac{b}{a}\right)^2 = 1 - \frac{1-(b/a)^2_{\rm obs}}{\sin^2 i}, \tag{4.29}$$

with the condition that $(b/a) \geq$ `ba_min`, where `ba_min` $= 0.15$ since this is the lowest axis ratio recorded in the NSA. If one knew (b/a) *a priori*, this step would be unnecessary and one could simply calculate the inclination by comparing (b/a) to $(b/a)_{\rm obs}$. However, we do not have this information so pursue the maximally agnostic strategy of drawing random inclinations to reflect our lack of knowledge of this quantity. Through Monte Carlo sampling, this uncertainty is propagated through to the final constraints.

We use `zdist` from the NSA catalogue to determine the angular diameter distance d_A to each galaxy and calculate the major-axis length as $R_{\rm eff}^{\rm maj} \equiv d_A r_{\rm eff}$. This is related to circularised, $R_{\rm eff}$, and minor-axis, $R_{\rm eff}^{\rm min}$, half-light radii as

$$R_{\rm eff} = \left(\frac{b}{a}\right)^{\frac{1}{2}} R_{\rm eff}^{\rm maj}, \qquad R_{\rm eff}^{\rm min} = \left(\frac{b}{a}\right) R_{\rm eff}^{\rm maj}. \tag{4.30}$$

We now find the stellar surface density in the plane of the galaxy [372, 373]

$$\Sigma_\star = \frac{M_\star}{2\pi b_{n_\star}^{-2n_\star} \Gamma\left(2n_\star\right) R_{\rm eff}^2}, \tag{4.31}$$

where $b_{n_\star} \equiv 2n_\star - 1/3 + 0.009876/n_\star$, which is different to the observed surface density as we have now corrected for randomly the assigned inclination. Using the reverse of the method from [363], we can estimate the gas mass from the stellar mass using [374]

$$\log_{10}\left(\frac{M_\star}{M_\odot}\right) = 1.89\log_{10}\left(\frac{M_{\rm gas}}{M_\odot}\right) - 8.12, \tag{4.32}$$

with scatter $\sigma_{\rm M} = 0.3\,{\rm dex}$. Assuming an exponential disk with effective radius given by

$$\log_{10}\left(\frac{R_{\rm eff,gas}}{\rm kpc}\right) = \log_{10}\left(\frac{0.92R_{\rm eff}}{\rm kpc}\right), \tag{4.33}$$

with scatter $\sigma_{\rm R} = 0.25\,{\rm dex}$, we can calculate the central gas surface density, $\Sigma_{\rm gas}$, using Equation 4.31 with the appropriate mass and radius and with $n_{\rm gas} = 1$ instead of $n_\star$. We now have the central baryonic surface density

$$\Sigma_B = \Sigma_\star + \Sigma_{\rm gas}. \tag{4.34}$$

To convert this to the central dynamical surface density, $\Sigma_{\rm D}$, we use the empirical relation [375–377]

$$\Sigma_{\rm D} = \Sigma_{\rm M} S\left(\frac{\Sigma_{\rm B}}{\Sigma_{\rm M}}\right), \tag{4.35}$$

where

$$S\left(y\right) = \frac{y}{2} + y^{\frac{1}{2}}\left(1 + \frac{y}{4}\right)^{\frac{1}{2}} + 2\sinh^{-1}\left(\frac{y^{\frac{1}{2}}}{2}\right), \tag{4.36}$$

with $\Sigma_{\rm M} = 1.37 \times 10^8 M_\odot {\rm kpc}^{-2}$. Despite the already large scatter due to uncertainties on the input quantities, one may expect the scaling relations to provide good fits only to a subset of the galaxy population. Therefore, to ensure our constraints on $\Delta G/G_{\rm N}$ are conservative, we impose an additional scatter of $\sigma_{\rm D} = 0.5\,{\rm dex}$. For scale height h, which we assume is equal to $R_{\rm eff}^{\rm min}$, $\Sigma_{\rm D}$ is related to the central density as

$$\Sigma_{\rm D} \equiv 2h\rho_0, \tag{4.37}$$

which we can substitute into Equation 4.13 to determine $r_\bullet$.

4.4.5 Gaussian mixture model

From Equation 4.13, we see that the offset is proportional to $\Delta G/G_{\rm N}$. We therefore construct a template signal with $\Delta G/G_{\rm N} = 1$ containing $N_{\rm MC}$ realisations of our probabilistic model.

For a given $r_{\rm V}$, we must convert the $N_{\rm MC}$ samples of predicted $\theta_{\bullet,\alpha}$ and $\theta_{\bullet,\delta}$ for each galaxy into a distribution. We model the samples as a GMM [170] where the likelihood function for some galaxy g is

$$\mathcal{L}_{\rm g}\left(\theta_{\bullet,\alpha}|\Delta G, r_{\rm V}\right) = \sum_i \frac{w_{{\rm g},\alpha}^{(i)}}{\sqrt{2\pi}\sigma_{{\rm g},\alpha}^{(i)}} \exp\left[-\frac{\left(\theta_{\bullet,\alpha} - \mu_{{\rm g},\alpha}^{(i)}\right)^2}{2{\sigma_{{\rm g},\alpha}^{(i)}}^2}\right], \tag{4.38}$$

where

$$\sum_i w_{{\rm g},\alpha}^{(i)} = 1, \quad w_{{\rm g},\alpha}^{(i)} \geq 0, \tag{4.39}$$

and $\{w^{(i)}, \sigma^{(i)}, \mu^{(i)}\}$, the weights, standard deviations and means of the Gaussians, are implicit functions of $r_{\rm V}$. There is an analogous definition for the declination component. The sum runs over the number of Gaussian components. The number of components is chosen to minimise the BIC (Equation 1.43). We find an independent set of Gaussians for each galaxy and component. For a different value of $\Delta G/G_{\rm N}$, we must transform the means and widths of the Gaussians in the GMM

$$\tilde{\mu}_{{\rm g},\alpha}^{(i)} = \left(\frac{\Delta G}{G_{\rm N}}\right)\mu_{{\rm g},\alpha}^{(i)}, \quad \tilde{\sigma}_{{\rm g},\alpha}^{(i)} = \left(\frac{\Delta G}{G_{\rm N}}\right)\sigma_{{\rm g},\alpha}^{(i)}. \tag{4.40}$$

Treating the orthogonal RA and Dec components as independent, the overall likelihood $\mathcal{L}_{\rm g}(\theta_{\bullet,\alpha}, \theta_{\bullet,\delta})$ for a test galaxy g to have $\boldsymbol{\theta}_\bullet$ components $\theta_{\bullet,\alpha}$ and $\theta_{\bullet,\delta}$ is

$$\mathcal{L}_{\rm g}\left(\theta_{\bullet,\alpha}, \theta_{\bullet,\delta}|\Delta G, r_{\rm V}\right) = \mathcal{L}_{\rm g}\left(\theta_{\bullet,\alpha}|\Delta G, r_{\rm V}\right)\mathcal{L}_{\rm g}\left(\theta_{\bullet,\delta}|\Delta G, r_{\rm V}\right). \tag{4.41}$$

4.4.6 Modelling the noise

Galileons are not the only type of physics that can lead to BH–galaxy offsets, requiring us to develop a model for astrophysical noise. Ideally we would construct a model for the observed offsets in the absence of a fifth force by using cosmological hydrodynamical simulations. Unfortunately, as discussed in Chapter 3, the majority of these simulations do not include a prescription for dynamical friction on BHs and instead pin the BHs at the centre of the galaxy [260–262]. Those that do allow the BH to move over-predict the fraction of offset BHs compared to observations. We therefore must construct an empirical noise model based on the global distribution of offsets in the various datasets.

We consider three different noise models: a Gaussian distribution, the sum of a Gaussian and Laplace distribution, and an Edgeworth expansion. We outline these below, before describing how we discriminate between them. In the following, we define the observed offsets to be $\theta_{\bullet,\alpha,{\rm obs}}$ and $\theta_{\bullet,\delta,{\rm obs}}$, and the true offsets – which are to be compared to the fifth force prediction – are $\theta_{\bullet,\alpha}$ and $\theta_{\bullet,\delta}$.

4.4.6.1 Gaussian Noise Model

In this model we assume that the observed value is Gaussian distributed about the predicted value due to a fifth force, such that the observed value has some uncertainty $\sigma_{\rm obs}$. We assume this is equal for both the RA and Dec components. We therefore have

$$\mathcal{L}_{\rm g}\left(\theta_{\bullet,\alpha,\rm obs}|\theta_{\bullet,\alpha},\mathbf{\Omega}\right) = \frac{1}{\sqrt{2\pi}\sigma_{\rm obs}}\exp\left(-\frac{\left(\theta_{\bullet,\alpha,\rm obs}-\theta_{\bullet,\alpha}\right)^2}{2\sigma_{\rm obs}^2}\right), \tag{4.42}$$

with $\mathbf{\Omega} = \{\sigma_{\rm obs}\}$.

4.4.6.2 Gaussian plus Laplace distribution

It is known that the distribution of observed BH–galaxy offsets is non-Gaussian (see Chapter 3), so we should also consider non-Gaussian noise models. Inspired by Chapter 3, our first non-Gaussian model is the sum of a Gaussian and Laplace distribution,

$$\mathcal{L}_{\rm g}\left(\theta_{\bullet,\alpha,\rm obs}|\theta_{\bullet,\alpha},\mathbf{\Omega}\right) = \frac{f}{\sqrt{2\pi}\sigma_{\rm obs}}\exp\left(-\frac{\left(\theta_{\bullet,\alpha,\rm obs}-\theta_{\bullet,\alpha}\right)^2}{2\sigma_{\rm obs}^2}\right)+\frac{1-f}{2\nu}\exp\left(-\frac{|\theta_{\bullet,\alpha,\rm obs}-\theta_{\bullet,\alpha}|}{\nu}\right), \tag{4.43}$$

with $\mathbf{\Omega} = \{\sigma_{\rm obs}, \nu, f\}$. As with the Gaussian model, $\sigma_{\rm obs}$ describes the uncertainty in the observed value. The Laplace term dominates in the tails of the distribution, hence ν tells us about the scale to which offset BHs extend, while f is the fraction of the probability in the Gaussian component.

4.4.6.3 Edgeworth Expansion

We now consider another way to incorporate non-Gaussianity, namely through the Edgeworth expansion [378]

$$\mathcal{L}_{\rm g}\left(\theta_{\bullet,\alpha,\rm obs}|\theta_{\bullet,\alpha},\mathbf{\Omega}\right) = \frac{1}{\sqrt{2\pi}\sigma_{\rm obs}}\exp\left(-\frac{\left(\theta_{\bullet,\alpha,\rm obs}-\theta_{\bullet,\alpha}\right)^2}{2\sigma_{\rm obs}^2}\right)\sum_{n=0}^{F}\alpha_n \mathrm{H}_n\left(\frac{\theta_{\bullet,\alpha,\rm obs}-\theta_{\bullet,\alpha}}{\sigma_{\rm obs}\sqrt{2}}\right), \tag{4.44}$$

where $\mathrm{H}_n(x)$ are Hermite Polynomials. The parameters $\{\alpha_n\}$ are related, since the probability density must be non-negative.

The non-zero coefficients α_n in the Edgeworth expansion are (up to $F = 8$)

$$\begin{gathered}\alpha_0 = 1, \quad \alpha_3 = \frac{\kappa_3}{2^{3/2}\times 3!}, \quad \alpha_4 = \frac{\kappa_4}{2^2\times 4!}, \quad \alpha_5 = \frac{\kappa_5}{2^{5/2}\times 5!},\\ \alpha_6 = \frac{10\kappa_3^2+\kappa_6}{2^3\times 6!}, \quad \alpha_7 = \frac{35\kappa_3\kappa_4+\kappa_7}{2^{7/2}\times 7!}, \quad \alpha_8 = \frac{35\kappa_4^2+56\kappa_3\kappa_5+\kappa_8}{2^4\times 8!},\end{gathered} \tag{4.45}$$

where κ_n is the n^{th} cumulant. The powers of $2^{n/2}$ are due to us using the Physicist's and not the Statistician's Hermite Polynomials. For notational convenience, we define the following parameters

$$\gamma \equiv \frac{\kappa_3}{2^{3/2}}, \quad \tau \equiv \frac{\kappa_4}{2^2}, \quad \eta \equiv \frac{\kappa_5}{2^{5/2}}, \quad \zeta \equiv \frac{\kappa_6}{2^3}, \quad \xi \equiv \frac{\kappa_7}{2^{7/2}}, \quad \iota \equiv \frac{\kappa_8}{2^4}, \tag{4.46}$$

so the coefficients are

$$\begin{gathered} \alpha_0 = 1, \quad \alpha_3 = \frac{\gamma}{3!}, \quad \alpha_4 = \frac{\tau}{4!}, \quad \alpha_5 = \frac{\eta}{5!}, \\ \alpha_6 = \frac{1}{6!}\left(\zeta + 10\gamma^2\right), \quad \alpha_7 = \frac{1}{7!}\left(\xi + 35\gamma\tau\right), \quad \alpha_8 = \frac{1}{8!}\left(\iota + 56\gamma\eta + 35\tau^2\right). \end{gathered} \tag{4.47}$$

We need to impose restrictions on the parameters since they must describe a probability distribution. Following [379], we define

$$\langle a, b\rangle \equiv \langle ab\rangle - \langle a\rangle \langle b\rangle, \tag{4.48}$$

and, since the probability density is non-negative, we use the Cauchy-Schwarz Inequality

$$\langle a, b\rangle^2 \leq \langle a, a\rangle \langle b, b\rangle, \tag{4.49}$$

which for some random variable x gives

$$\langle x^p, x^q\rangle^2 \leq \langle x^p, x^p\rangle \langle x^q, x^q\rangle. \tag{4.50}$$

These can be trivially expressed in terms of cumulants, and thus we have necessary conditions for the parameters $\mathbf{\Omega} = \{\sigma_{\text{obs}}, \gamma, \tau, \eta, \zeta, \xi, \iota\}$ to describe a probability distribution. We impose this constraint for all $p < q \leq 4$ and consider the cases $3 \leq F \leq 8$.

4.4.6.4 Physical offsets

So far the assumed noise models only contain contributions due to observational effects, so that all noise parameters are angular quantities. This corresponds to the assumption that any noise contribution with a fixed physical scale is subdominant. To account for this possibility, for each noise model we consider the corresponding model with

$$\sigma_{\text{obs}} \to \sqrt{\sigma_{\text{obs}}^2 + \left(\frac{\sigma_{\text{int}}}{d_{\text{A}}}\right)^2}, \tag{4.51}$$

which assumes that BHs are isotropically, Gaussian distributed with width σ_{int}. This adds one more parameter to the noise model and means that each galaxy has a different width of its noise Gaussian.

4.4.6.5 Choosing the model

To decide which noise model to use, for each dataset we find the maximum likelihood estimates, $\hat{\mathcal{L}} \equiv \max \mathcal{L}(d|\Delta G, r_{\rm V}, \mathbf{\Omega})$, for each model. We then choose the model that minimises the BIC (Equation 1.43) with $\mathcal{N} = 2\mathcal{N}_{\rm gal}$ for $\mathcal{N}_{\rm gal}$ galaxies in the dataset. We find that in all cases the sum of the Gaussian and Laplace distribution without a physical offset best describes the data, so this is the noise model we use below. In Figure 4.1 we see that the observed angular offsets are independent of redshift, suggesting that these are dominated by observational effects. It is therefore unsurprising that the addition of an intrinsic offset is not required.

4.4.7 Likelihood Model

Now we have a distribution for obtaining an observed value given a predicted one, $\mathcal{L}_{\rm g}(\theta_{\bullet,\alpha,{\rm obs}}|\theta_{\bullet,\alpha}, \mathbf{\Omega})$, after accounting for the noise contribution to the signal. The resulting likelihood for an observed offset $\theta_{\bullet,\alpha,{\rm obs}}$ is

$$\mathcal{L}_{\rm g}(\theta_{\bullet,\alpha,{\rm obs}}|\Delta G, r_{\rm V}, \mathbf{\Omega}) = \int \mathcal{L}_{\rm g}(\theta_{\bullet,\alpha,{\rm obs}}|\theta_{\bullet,\alpha}, \mathbf{\Omega})\, \mathcal{L}_{\rm g}(\theta_{\bullet,\alpha}|\Delta G, r_{\rm V})\, {\rm d}\theta_{\bullet,\alpha}. \tag{4.52}$$

For example, for the Gaussian noise model, this is

$$\mathcal{L}_{\rm g}(\theta_{\bullet,\alpha,{\rm obs}}|\Delta G, r_{\rm V}, \mathbf{\Omega}) = \sum_i \frac{w^{(i)}_{{\rm g},\alpha}}{\sqrt{2\pi\left(\sigma^2_{\rm obs} + \tilde{\sigma}^{(i)\,2}_{{\rm g},\alpha}\right)}} \exp\left(-\frac{\left(\theta_{\bullet,\alpha,{\rm obs}} - \tilde{\mu}^{(i)}_{{\rm g},\alpha}\right)^2}{2\left(\sigma^2_{\rm obs} + \tilde{\sigma}^{(i)\,2}_{{\rm g},\alpha}\right)}\right). \tag{4.53}$$

We treat each galaxy as independent to obtain the likelihood of our dataset d to be

$$\mathcal{L}(d|\Delta G, r_{\rm V}, \mathbf{\Omega}) = \prod_{\rm g} \mathcal{L}_{\rm g}(\theta_{\bullet,\alpha,{\rm obs}}|\Delta G, r_{\rm V}, \mathbf{\Omega})\, \mathcal{L}_{\rm g}(\theta_{\bullet,\delta,{\rm obs}}|\Delta G, r_{\rm V}, \mathbf{\Omega}). \tag{4.54}$$

Finally, given some prior on ΔG and $r_{\rm V}$, $P(\Delta G, r_{\rm V}, \mathbf{\Omega})$, we use Bayes' theorem to obtain

$$P(\Delta G, r_{\rm V}, \mathbf{\Omega}|d) = \frac{\mathcal{L}(d|\Delta G, r_{\rm V}, \mathbf{\Omega})\, P(\Delta G, r_{\rm V}, \mathbf{\Omega})}{P(d)}, \tag{4.55}$$

where $P(d)$ is the constant probability of the data for any $\{\Delta G, r_{\rm V}, \mathbf{\Omega}\}$. We are now in a position to derive posteriors on $\Delta G/G_{\rm N}$ and the noise model parameters at fixed $r_{\rm V}$, for which we use the EMCEE sampler [62]. We impose the improper prior $\Delta G \geq 0$, flat in ΔG. The priors for all inferred parameters are given in Table 4.2.

Table 4.2: Inferred parameters describing the predicted signal and the empirical noise models. In all cases the Gaussian plus Laplace distribution is the preferred noise model, and the parameters for this are given in the top part of the table. Below the horizontal line we give the parameters for the Edgeworth expansion (Equations 4.44 and 4.46) and the models containing intrinsic offsets, which we use in Section 4.6.1.2. All priors are uniform in the range given. We fix $r_{\rm V}$ in the inference and allow all other parameters relevant to the chosen noise model to vary. The (p, q) values refer to Equation 4.50.

Parameter	Prior/Constraint
$\Delta G/G_{\rm N}$	≥ 0
$\log_{10}(r_{\rm V}$ / Mpc)	≥ 0.5
$\sigma_{\rm obs}$	> 0
ν	> 0
f	$[0, 1]$
$\sigma_{\rm int}$	≥ 0
γ	-
τ	$p = 1$, $q = 2$
η	-
ζ	$p = 2$, $q = 3$
ξ	-
ι	$p < q$, $q = 4$

4.5 Results

In Figure 4.2 we show the corner plot from the inference with $r_{\rm V} = 100\,{\rm Mpc}$ for each dataset, using the empirical noise model consisting of a Gaussian plus a Laplace distribution. We see that each dataset is consistent with $\Delta G/G_{\rm N} = 0$. Assuming each sample is independent, we multiply the likelihoods and, giving each dataset a different set of noise parameters, find the joint constraint of

$$\frac{\Delta G}{G_{\rm N}} < 0.16 \tag{4.56}$$

at 1σ confidence for this value of $r_{\rm V}$, or < 0.36 at 2σ confidence. We find that the constraint is driven by O16 due to its large size. If we did not include the O16 data, our strongest constraint would come from SB18 and would be $\Delta G/G_{\rm N} < 0.65$ at 1σ confidence.

For the radio samples (OF13, O16 and SB18) we find that $\sigma_{\rm obs} \sim 50\,{\rm mas}$ and for B19 we find $\sigma_{\rm obs} \sim 150\,{\rm mas}$, as expected in Section 4.3. Further, as in [99], we find that O16 has a much higher contribution from the non-Gaussian component than the other datasets, as shown by the smaller value of f.

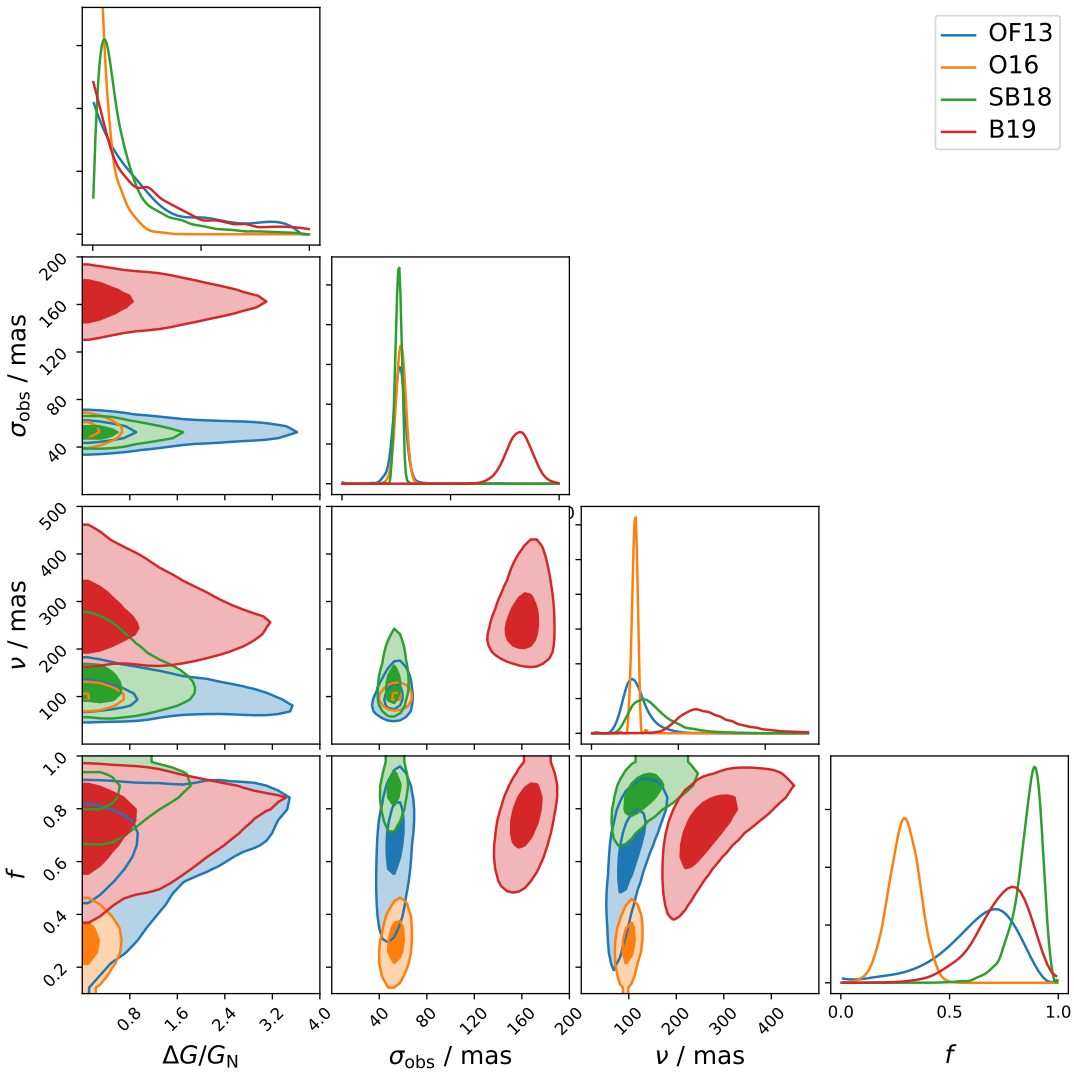


Figure 4.2: Corner plot of the constraints on the strength of the coupling to the Galileon field, $\Delta G/G_{\rm N}$, and the noise parameters at $r_{\rm V} = 100\,{\rm Mpc}$. The contours show the 1 and 2σ confidence intervals. Each dataset is consistent with $\Delta G/G_{\rm N} = 0$, with $\Delta G/G_{\rm N} > 0.16$ ruled out at 1σ confidence when we combine the datasets. Note that while $\Delta G/G_{\rm N}$ is assumed universal, $\sigma_{\rm obs}$, ν and f are sample-specific.

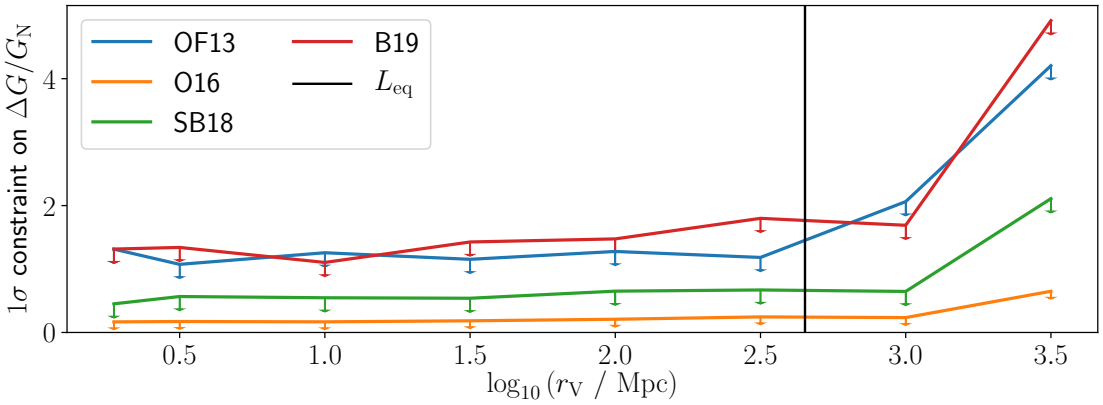


Figure 4.3: 1σ constraint on $\Delta G/G_{\rm N}$ as a function of average Vainshtein radius, $r_{\rm V}$. We expect $r_{\rm V} \sim 10\,{\rm Mpc}$ for crossover scales $\sim H_0^{-1}$. $r_{\rm V} > L_{\rm eq}$ corresponds to crossover scales much larger than the observable universe (see Figure 4.5).

We repeat the inference at different values of $r_{\rm V} > 1\,{\rm Mpc}$ and plot the 1σ constraint as a function of $r_{\rm V}$ for each dataset in Figure 4.3. As anticipated in Section 4.4.3, we find that the constraint is relatively independent of $r_{\rm V}$ for $r_{\rm V} < L_{\rm eq}$, where the level of bumpiness for these $r_{\rm V}$ indicates the noise level of this method, due to the finite number of Monte Carlo realisations used to determine the likelihood. To quantify this, we run the end-to-end inference a further 6 times for the OF13 data with $r_{\rm V} = 100\,{\rm Mpc}$ and find an unbiased sample variance of the 1σ and 2σ constraints of 8% and 7% respectively. This shows that the number of Monte Carlo realisations is sufficiently large.

The lack of dependence of the constraint on $r_{\rm V}$ means we expect our constraints will not change for a broad transition from screened to unscreened in this regime, as opposed to the step-function we currently use.

For larger values of $r_{\rm V}$ we find that, although the posteriors are still consistent with $\Delta G/G_{\rm N} = 0$, the constraint weakens. The smaller magnitude of the fifth force field at these $r_{\rm V}$ (more modes of the $\boldsymbol{g}$ field are excluded) means a given offset requires a larger value of $\Delta G/G_{\rm N}$ (Equation 4.13), and thus a worse constraint. $r_{\rm V} = L_{\rm eq}$ corresponds to the largest crossover scales in Figure 4.5, which are already much larger than H_0^{-1}.

4.6 Discussion

4.6.1 Systematic uncertainties

In this section we vary several parameters of the analysis which could contribute systematic error if kept fixed. For computational convenience, throughout this section we use the OF13 data as this is the smallest dataset, although the conclusions apply equally to all the datasets. Hence the constraint for the fiducial case in this section is $\Delta G/G_{\rm N} < 1.28$, as opposed to the full joint constraint of $\Delta G/G_{\rm N} < 0.16$.

4.6.1.1 Density profile

Although halos in N-body simulations are typically well fit by NFW profiles, the situation is less clear observationally and in the presence of baryons. While the process of adiabatic contraction steepens the central DM density during the process of galaxy formation [380, 381], subsequent stellar feedback can inject energy into the halo and cause it to expand (e.g. [243, 382]). Previously we assumed that this results in an approximately uniform density at the centre of the halo, however this need not be true.

To test whether our constraints are sensitive to the assumed density profile, we now suppose that the inner density can be described by a power law. To remain agnostic as to the slope of the DM density profile over the extent of the galaxy, we parameterise it as

$$\rho\left(r\right) = \begin{cases} \rho_{\rm s}\left(\frac{r_{\rm s}}{r}\right)^n, & r \leq r_{\rm s} \\ \frac{4\rho_{\rm s}}{r/r_{\rm s}(1+r/r_{\rm s})^2}, & r > r_{\rm s}, \end{cases} \tag{4.57}$$

with n a free parameter that would equal 1 for NFW. With most observational evidence favouring profiles somewhat shallower than NFW (e.g. [383–385]) we take a fiducial value for n of 0.5, although we will check explicitly the result of varying it within the a priori plausible range $0 \leq n < 1$. r_s is the scale radius of the halo, while ρ_s is the density at that radius.

For $n = 0.5$, the mean predicted offset is $\sim 50\,\mathrm{pc}$ for $\Delta G/G_{\rm N} = 1$. This is much less than $r_{\rm s}$ so that the relevant profile is the power law, and the NFW profile is only required to determine $\rho_{\rm s}$. We therefore do not need to consider the case where the restoring force is too small to balance the fifth force, which could lead to the BH being ejected from the galaxy [2].

Within the power-law region the enclosed mass is

$$M\left(<r\right) = \frac{4\pi}{3-n}\rho_{\rm s} r^3 \left(\frac{r_{\rm s}}{r}\right)^n, \quad r < r_{\rm s}. \tag{4.58}$$

Substituting this into Equation 4.11, the equivalent of Equation 4.13 is

$$r_\bullet = \left(\frac{\Delta G}{G_N}\right)^{\frac{1}{1-n}} \left(\frac{3-n}{4\pi}\frac{|\boldsymbol{g}_{\rm lss}|}{G_{\rm N}\rho_{\rm s}}\frac{1}{r_{\rm s}^n}\right)^{\frac{1}{1-n}}, \tag{4.59}$$

Note that this offset diverges for an NFW profile ($n = 1$). We also clearly need $n < 3$, however this is already required for the halo to have a finite mass within any finite radius. If $n < 1$, then the largest offsets will be for the least dense galaxies, whereas the converse is true if $n > 1$. The case $n > 1$ is interesting in that a larger $\Delta G/G_{\rm N}$ will actually shrink the offset between the BH and galactic centre. However we do not consider this case further since, by considering small perturbations about $r_\bullet$, the equilibrium offset is found to only be stable for $n < 1$. As mentioned above, $n < 1$ is often a better fit for the density profile, so it is not unreasonable to only consider these values.

The values of $r_{\rm s}$ and $\rho_{\rm s}$ for each test galaxy are estimated using the technique of halo abundance matching (AM) [386, 387]. The technique assumes a near-monotonic relation between the absolute magnitude of a galaxy and a halo 'proxy', typically a combination of virial mass and concentration. We use the best-fit AM model of [388] applied to the ROCKSTAR halo catalogue of the DarkSky-400 N-body simulation [389] and the Sérsic r-band luminosity function of [390]. We generate `N_AM` $= 200$ mock galaxy catalogues, where each catalogue is a different random realisation of the noise from the intrinsic scatter in the galaxy–halo connection implied by the model. We draw values of $r_{\rm s}$ and $\rho_{\rm s}$ for each galaxy from a randomly chosen catalogue for each Monte Carlo realisation of our model, where we use the halo from that catalogue which is closest in magnitude to that galaxy. By iterating this procedure many times, we marginalise over the stochasticity in the galaxy–halo connection. We re-run the end-to-end inference with $n = 0.5$ and only 100 catalogues and find the constraint to be unchanged (Figure 4.4), indicating that `N_AM` is sufficiently large to sample the distributions of $\rho_{\rm s}$ and $r_{\rm s}$.

As before, we obtain $N_{\rm MC}$ samples from our model of offsets in Galileon gravity and convert this into a GMM for the case $\Delta G/G_{\rm N} = 1$. For a power law profile, the offset is no longer proportional to $\Delta G/G_{\rm N}$, and instead the relation is

$$r_\bullet \propto \left(\frac{\Delta G}{G_{\rm N}}\right)^{\frac{1}{1-n}}. \tag{4.60}$$

Thus, to convert the GMM to a different value of $\Delta G/G_{\rm N}$, Equation 4.40 is changed to

$$\tilde{\mu}^{(i)}_{{\rm g},\alpha} = \left(\frac{\Delta G}{G_{\rm N}}\right)^{\frac{1}{1-n}} \mu^{(i)}_{{\rm g},\alpha}, \quad \tilde{\sigma}^{(i)}_{{\rm g},\alpha} = \left(\frac{\Delta G}{G_{\rm N}}\right)^{\frac{1}{1-n}} \sigma^{(i)}_{{\rm g},\alpha}, \tag{4.61}$$

with the rest of the analysis unchanged from Section 4.4. We again use the Gaussian plus Laplace distribution for our noise model.

In Figure 4.4 we plot the 1σ constraints on $\Delta G/G_{\rm N}$ for different power law indices. Fitting the constraint to an exponential, as would be expected from Equation 4.59, we find the constraint weakens like $e^{\beta n}$ where $\beta \approx 4.3$. Increasing n increases the density at a given radius which increases the restoring force and thus a larger $\Delta G/G_{\rm N}$ is necessary for a given offset.

Comparing to Figure 4.2, we see that the constraint using a power law profile is an order of magnitude tighter than when we use the scaling relations from Section 4.4.4 for the smallest values of n, and comparable when $n \lesssim 1$. The stronger constraints can be understood in terms of the pivot scale at which the density profile changes from NFW to a power law. In Equation 4.57 we chose $r_{\rm s}$ as the pivot scale however, since we use $n < 1$, if this transition occurred at a smaller radius, the density at a given radius in the power law region would be greater. Changing the pivot scale to $\sim 0.01 r_{\rm s}$ ($\sim 360\,{\rm pc}$ for a typical galaxy from OF13) would provide constraints similar to our previous prediction. There is no reason *a priori* why we would choose $0.01 r_{\rm s}$ as a pivot scale, but it is reassuring that it is not an unreasonable choice.

Our fiducial case is the halo density profile which gives the most conservative constraint of those considered, hence we report the value derived from scaling relations with a cored profile as our final constraint.

We also assumed that the gas density profile has a Sérsic index $n_{\rm gas} = 1$, i.e. an exponential disk. We repeated the inference with $n_{\rm gas} = 0.5$ and 2.0. Increasing $n_{\rm gas}$ slightly widens the posterior on $\Delta G/G_{\rm N}$. This is expected as increasing $n_{\rm gas}$ increases the central gas density, which decreases the predicted offset for a given $\Delta G/G_{\rm N}$ so slightly larger values of $\Delta G/G_{\rm N}$ are permitted. Given that the gas mass is sub-dominant compared to the other mass components, we would expect the change in the constraint to be small, as indeed it is.

4.6.1.2 Noise model

Our previous work on BH offsets [99] demonstrated that the distribution of offsets is non-Gaussian for these data, so it is unsurprising that the addition of a Laplace distribution is favoured by the BIC. To ensure that our results are insensitive to our choice of non-Gaussianity, we re-run the inference using an Edgeworth expansion for our empirical noise mode, truncating the sum at $F = 4$ in Equation 4.44 as this has the minimum BIC for $3 \leq F \leq 8$. Using the OF13 data, we find that the Edgeworth noise model gives a result which is consistent with zero, with $\Delta G/G_{\rm N} < 1.89$ at 1σ

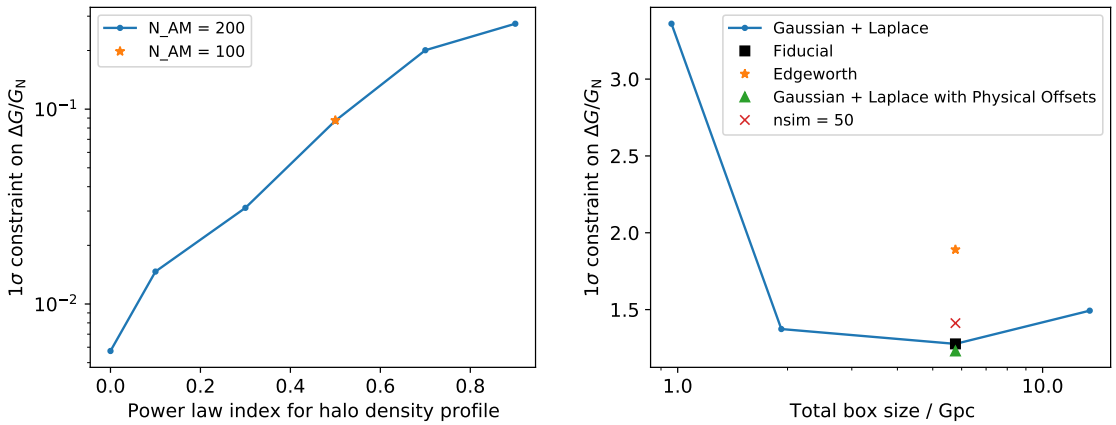


Figure 4.4: Constraints on $\Delta G/G_{\rm N}$ at $r_{\rm V} = 100\,{\rm Mpc}$ at 1σ confidence using the OF13 data. *Left:* A power law halo density profile of index n is assumed. To determine the density profile, we use `N_AM`=200 mock catalogues obtained through abundance matching. We re-run the analysis at $n = 0.5$ with `N_AM`=100 and see that the constraint is unchanged. *Right:* Using our fiducial density profile, we vary the box size, L', used to add in long wavelength modes when reconstructing the Galileon filed. $L' \sim 1\,{\rm Gpc}$ corresponds to no extra long wavelength modes. We also show the constraints from different noise models at $L' \sim 6\,{\rm Gpc}$, and if we reduce the number of constrained simulations to 50.

confidence, compared to $\Delta G/G_{\rm N} < 1.28$ for the Gaussian plus Laplace distribution with this data.

Although disfavoured by the BIC, we now explore the effect of adding a contribution from physical offsets to the Gaussian plus Laplace distribution noise model (Section 4.4.6.4). We find that our constraint is slightly tightened to $\Delta G/G_{\rm N} < 1.23$ for the OF13 data. As the difference is within the uncertainty on our constraint of 8% (see Section 4.5), we conclude that our constraint is insensitive to this change. We find that $\sigma_{\rm int} < 22\,{\rm pc}$ at 1σ confidence, showing that this contribution is small if it exists at all. We plot the constraints from both of these alternative noise models in Figure 4.4.

We conclude that our results are not sensitive to the noise model, provided that we include non-Gaussianity.

4.6.1.3 Galileon field

When reconstructing the Galileon field, one should check that the resolution is sufficiently high such that the maximum k is determined by $r_{\rm V}$ and not the resolution, i.e. that $L/N < r_{\rm V}$.

Our minimum $r_{\rm V}$ equals L/N for $N = 512$. We have checked that our constraints are unchanged if we use $N = 256$ for $r_{\rm V} \geq 10\,{\rm Mpc}$ (where we require $N > 100$).

Repeating the analysis with 34 constrained simulations did not change the constraint (see Figure 4.4), indicating that we have a sufficient number of constrained simulations to sample the distribution of the Galileon field at each galaxy.

In Section 4.4.2 we discussed how we should add Fourier modes to the Galileon field with wavelengths longer than the box length, L, of our constrained simulations. This was achieved by creating unconstrained Gaussian random fields in a box of side length $L' = \ell L$. We chose $\ell = 6$, however to check that our constraints are independent of this choice, we re-run the inference with different values of ℓ and plot the results in Figure 4.4. Comparing $\ell = 6$ to $\ell = 14$ for the OF13 data, we find the results are consistent, justifying our choice of $\ell = 6$. We also find the constraint if we do not include this additional long wavelength information ($\ell = 1$) and find that it is weakened from $\Delta G/G_{\rm N} < 1.28$ to $\Delta G/G_{\rm N} < 3.36$ at 1σ confidence for the OF13 data. This is to be expected as adding in long wavelength modes increases the magnitude of the Galileon field, which tightens the $\Delta G/G_{\rm N}$ constraint.

Removing all of the modes from the constrained simulations is equivalent to setting $r_{\rm V} \sim 1\,{\rm Gpc}$. We do this in Figure 4.3 and find little change in our constraints. This is to be expected since the main contributors to the magnitude of the Galileon field are modes with $k \sim 2\pi/L_{\rm eq}$, so the majority of the modes from the constrained simulation have a negligible impact on this. Even though the constrained simulations turn out to be relatively unimportant in setting our constraints, they would be necessary to obtain a detection because they contain the information on the direction of $\boldsymbol{g}$. In the future it will be interesting to repeat the inference using simulations with initial conditions constrained in the much larger SDSS volume [86].

4.6.1.4 Other potential systematics

Although we use the BIC to determine how many components to fit in the GMM (Section 4.4.5), which should penalise components which fit outliers of the distribution, it is important to check that our constraints are not driven by unlikely realisations in our Monte Carlo sampling. Still minimising the BIC, but restricting ourselves to no more than 15 Gaussian components, we find that our constraints are unchanged. We find that all galaxies across all runs require $\sim$10 GMM components even without a maximum number of components, showing that the imposed maximum is not important.

We re-run the end-to-end inference with $\sigma_{\rm D} = 1.0, 2.5$ and $5.0\,$dex, and find little variation of the constraint with this parameter for the smaller values of $\sigma_{\rm D}$. To understand why increasing the scatter in all of the quantities has little impact on our $\Delta G/G_{\rm N}$ constraint, we fit a single Gaussian instead of a GMM to the distribution of offsets for each galaxy. As expected, the log-normal scatter increases the magnitude of the mean and the width of the Gaussian. Increasing the mean and the covariance have competing effects in $\log \mathcal{L}$ and, until we reach relatively large values of scatter, the two effects cancel and thus the constraint on $\Delta G/G_{\rm N}$ has little dependence on $\sigma_{\rm D}$. Increasing $\sigma_{\rm M}$ to 0.6 or $\sigma_{\rm R}$ to 0.5 also has a negligible impact on our constraint.

The final parameter in the inference is `ba_min`, the minimum allowed minor-to-major axis ratio allowed. We set this to 0.15 as this is the lowest axis ratio recorded in the NSA. Changing this to 0.20 was found not to change the constraint.

4.6.2 Comparison with the literature

Previous attempts to constrain Galileons using the polarisation of BH–galaxy offsets relative to the direction of a partially unscreened Galileon field have targeted galaxies in massive galaxy clusters. Since Equation 4.5 is a total derivative for a spherically symmetric mass distribution, this equation becomes a modified non-linear Gauss' law; only the mass within some radius sources the field at that point. This results in the Vainshtein mechanism being less efficient inside extended mass distributions, hence why Galileons with sub-Gpc values of $r_{\rm C}$ can be constrained in these environments. Constraints $\alpha \lesssim \mathcal{O}(1)$ were obtained [2, 3] using the central BH in M87 for Galileons with $r_{\rm C} \lesssim 1\,$Gpc.

By considering galaxies in more rarefied environments, we study the situation where the Galileon field is sourced by large scale structure as opposed to a cluster. This allows us to probe larger values of $r_{\rm C}$, since we no longer require that the mass in the vicinity of the BH sources a partially unscreened Galileon field.

Constraints on Galileons can also be found using the technique of lunar laser ranging [10], which currently sets the bound [1, 11][3]

$$r_{\rm C}\alpha^{-\frac{3}{2}} \gtrsim 150\,{\rm Mpc}. \tag{4.62}$$

Slightly weaker constraints are obtained by studying the precession of planetary orbits in the Solar System [392].

[3]Note that this constraint is sensitive to the rotation vector of the Moon [391], which is set to its GR value. This could introduce some model dependence in the result.

In Figure 4.5 we plot the constraints on α as function of $r_{\rm C}$ for the cubic Galileon from the literature and compare to the constraints obtained in this work. As discussed in Section 4.4.3, we convert $r_{\rm V}$ to $r_{\rm C}$ by using Equations 4.26 and 4.27. From this we see that our constraints are applicable to the region $r_{\rm C} \gtrsim H_0^{-1}$. As anticipated, this is complementary to previous work, which constrains $r_{\rm C} \sim 0.01 - 1\,{\rm Gpc}$, and is comparable in strength.

Our conversion from $r_{\rm V}$ to $r_{\rm C}$ in a cosmological context is based on the non-linear matter power spectrum from `CLASS`; for smaller values of $r_{\rm V}$ within the 1-halo term this approximation breaks down and an alternative conversion would be required. We anticipate that if we were to improve the modelling of the Galileon field to incorporate the non-linear regime, our constraint on α would remain relatively unchanged as we moved to smaller $r_{\rm C}$. Our test will then become competitive with Lunar Laser Ranging and constrain self-accelerating Galileons, providing an alternative to the Integrated Sachs-Wolfe probe [393].

We note that the strength of our constraints are similar to those from M87 [2], despite our sample containing 1916 galaxies and their one. This is due to the interplay of three effects. First, the observations used in this work have lower resolution, with $\sigma \sim 50\,{\rm mas}$ for the radio data, whereas the galaxy–BH offset for M87 is measured to be $< 30\,{\rm mas}$. Second, we marginalise over an empirical noise model, whereas [2] assume that the entire offset is due to a fifth force, which would make their constraint tighter but also more prone to systematics to do with astrophysical contributions to the offset. Finally, we only consider the Galileon field sourced by large scale structure, which is smaller than that near a massive cluster and hence allows larger $\Delta G/G_{\rm N}$ values for a given offset. Combining cosmic and cluster fields in future work will therefore afford much tighter constraints.

To enable easy comparison to the literature, we also convert our constraint on $r_{\rm V}$ and $\Delta G/G_{\rm N}$ to one on the Horndeski parameters c_2 and c_3. To do this, we consider the tracker solution [394]

$$\dot{\bar{\varphi}} H = \xi H_0^2 = \text{constant}, \tag{4.63}$$

where ξ is related to the dimensionless Horndeski parameters for cubic Galileons as

$$\xi = -\frac{c_2}{6c_3}. \tag{4.64}$$

From Equations 4.3 to 4.5 we see that today

$$\beta_1 = \frac{\xi}{3}\left[c_3\xi^3 - 1 + 2\frac{\dot{H}}{H_0^2}\right], \quad \beta_2 = \frac{2}{\xi^2}\beta_1, \tag{4.65}$$

where

$$\beta_1 = (H_0 r_{\rm C})^{-2}, \quad \beta_2 = \frac{1}{3\alpha}. \tag{4.66}$$

From this conversion and Equation 4.27, we see that lines of constant $r_{\rm V}$ are transformed to lines of constant $c_3\xi^3$, where ξ is proportional to α on these curves. In Figure 4.6 we plot our constraints in the $c_3 - \xi$ plane. In this plot we demonstrate the regions of parameter space that can be probed by our test, and ways of further constraining this region. It is clear from the plot that further work should target smaller $r_{\rm V}$, as our constraint already lies close to the line corresponding to $r_{\rm V} \to \infty$.

We also plot the $\alpha - r_{\rm V}$ curves in Figure 4.5 for the normal and self-accelerating branches of the DGP model, evaluated using Equation 4.6 and assuming the matter density and H_0 is the same as in the constrained simulations. We see that the normal branch is not yet constrained by our test. For our smallest value of $r_{\rm V}$, we would require a constraint $\alpha < 0.04$ to do this, corresponding to an improvement of a factor of ~ 7.

Due to the assumption of linearity, our constraints are insensitive to the specific Galileon model; using the Vainshtein radius for a quartic Galileon in Figure 4.5 changes the conversion to $r_{\rm C}$ by a numerical factor of $\mathcal{O}(1)$ [36]. Our bounds are therefore equally applicable to quartic and quintic Galileons. The results of GW170817 [148] already severely constrain the self-accelerating branches of these models, with constraints on the Horndeski parameters [395]

$$|c_4| \lesssim 2.8 \times 10^{-17} \left(\frac{2}{\xi}\right)^4, \quad |c_5| \lesssim 3.8 \times 10^{-17} \left(\frac{2}{\xi}\right)^5, \tag{4.67}$$

where $\xi = H(t)\dot{\varphi}/H_0^2$. To convert these to bounds on α and $r_{\rm C}$ using variants of Equation 4.4 relevant to these models, one would also need to know c_3. Since this is not constrained by GW170817, we do not perform this comparison explicitly, but note that our results provide independent stringent constraints on the quartic and quintic models.

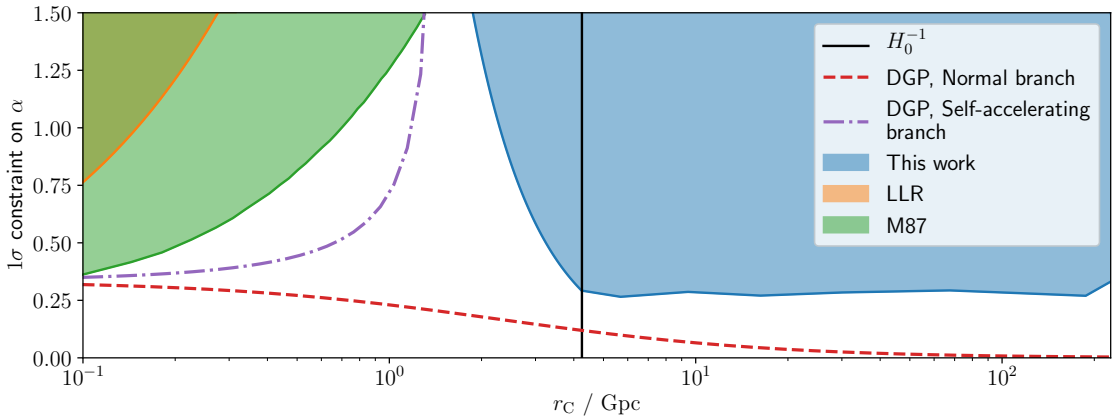


Figure 4.5: Constraints on the coupling of a cubic Galileon to matter, α, as a function of the crossover scale, r_C, from lunar laser ranging (LLR) [1], the BH at the centre of M87 [2, 3] and this work. The shaded regions are excluded at 1σ confidence. We also plot the $\alpha - r_C$ curves for the normal and self-accelerating branches of the DGP model (Equation 4.6).

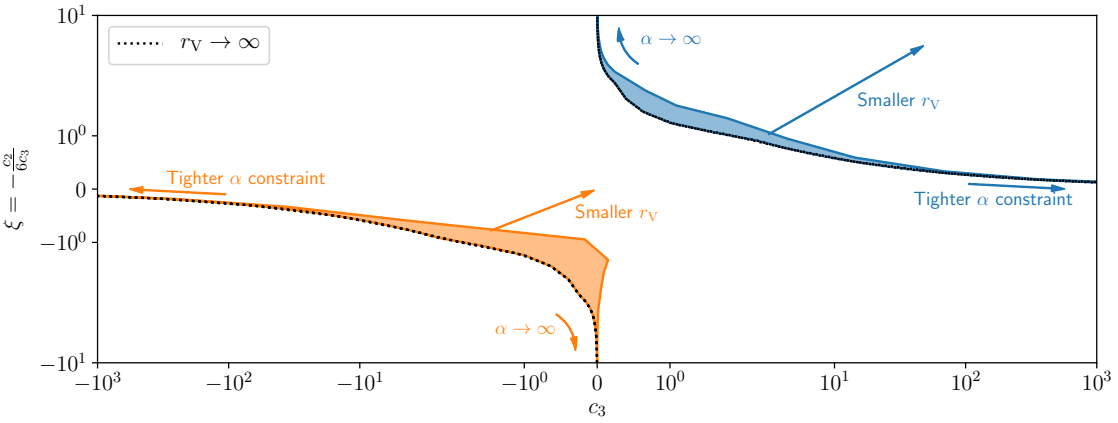


Figure 4.6: Constraints on the Horndeski parameters c_2 and c_3 for the cubic Galileon tracker solution $\dot{\varphi}H = \xi H_0^2$ at 1σ confidence. The blue region corresponds to $\xi > 0$ and the orange region is for $\xi < 0$. Lines of constant r_V are given by the curves $\xi^3 c_3 =$ constant, where ξ is proportional to α along these curves. Probing smaller r_V in future work will push the constraint to the upper right, as shown by the arrows.

4.7 Conclusions

Galileons are scalar field theories which obey the Galileon symmetry $\varphi \to \varphi + b + c_\mu x^\mu$ [344] and are parameterised by their crossover scale, r_C, and coupling to matter, α. If $r_\mathrm{C} \sim H_0^{-1}$ then the Galileon is said to be cosmologically relevant and could explain the late-time accelerated expansion of the Universe. Here we present a test of Galileon gravity by comparing the predicted equilibrium offsets between the centre of a galaxy and its central supermassive BH to observational data.

Using similar methods to those in Chapter 2, we construct a galaxy-by-galaxy Bayesian forward model of these offsets based on dynamical information about these galaxies, a reconstruction of the local gravitational field and a non-Gaussian empirical noise model. In doing so we utilised `CSiBORG`, a suite of constrained N-body simulations of the local Universe using initial conditions from the `BORG` algorithm. Marginalising over noise parameters and propagating uncertainties on input quantities via Monte Carlo sampling, we derive constraints on the magnitude of the fifth force arising from cosmologically relevant Galileons, $\Delta G/G_\mathrm{N} \equiv 2\alpha^2$.

We rule out $\Delta G/G_\mathrm{N} > 0.16$ at 1σ confidence for $r_\mathrm{C} \gtrsim H_0^{-1}$. We find our constraints to be robust to the assumed halo density profile, the choice of empirical noise model and parameters used to infer physical properties from the dynamical information available on these galaxies. These constraints are complementary to previous constraints from galaxy–BH offsets [2, 3]: we probe larger values of r_C because we consider the Galileon field sourced by large scale structure as opposed to a massive galaxy cluster.

Improved modelling of the Galileon field to enter the non-linear regime should make our constraints competitive with lunar laser ranging (LLR) [1] at smaller values of r_C. Our approach necessarily requires a large number of galaxies to fit the parameters of the empirical noise model accurately, so targeted observations at a small number of galaxies would not be particularly useful in tightening our constraints. Improved modelling of dynamical friction in future cosmological hydrodynamical simulations [99] could remove our reliance on an empirical noise model, making such simulations useful for future constraints on Galileons and similar theories.

Chapter 5

Dark matter annihilation and decay

5.1 Introduction

As alluded to in Section 1.1.2.2, the thermal relic cross-section for DM particles of mass $\sim 0.1 - 1\,\mathrm{TeV}$ is suspiciously similar to that expected if the coupling were approximately equal to the electroweak coupling scale. DM annihilation or decay for particles at these masses could be detected through the emission of gamma rays by their decay products. Despite the theoretically low interaction rates - one would expect only a few in 10^{15} particles to annihilate per Hubble time in the present Universe [57] - the vast quantities of DM on cosmological scales makes these processes potentially detectable in the state-of-the-art gamma ray measurements from the *Fermi* Large Area Telescope (*Fermi*-LAT) [396]. Indeed, the excess of observed gamma rays towards the galactic centre (Galactic Centre Excess; GCE) [397–399] can be fitted well by the annihilation of DM [400–406]. There is debate over whether other explanations could also explain the emission. Several studies [407–412] argue that the GCE can be explained by a population of unresolved point sources, although this is contested [405, 413–415], and some groups [416–419] find both DM annihilation and other models can fit the data. A combination of these two processes is of course possible, and is plausible given the spatial variation of the GCE [420]. Other astrophysical processes at the centre of the galaxy [421–426] could also be responsible.

Given these conflicting explanations of the GCE, in order to unambiguously detect or rule out DM annihilation or decay models, one should determine if an excess of gamma rays is detected from other sources or across the full sky. Previous studies have placed constraints on $\langle \sigma v \rangle$ through cross-correlation between *Fermi*-LAT data and galaxy [e.g. 8, 427] or lensing [428] catalogues, or by studying nearby dwarf

galaxies [7, 429, 430] or groups [431]. Moreover, such emission should increase the kinetic energy of baryons, so by considering the impact on the CMB [432] or galaxy formation [433], one can rule out a velocity-independent cross-section for thermal relic DM particles less massive than $\sim 30\,\mathrm{GeV}$.

Instead of focusing on a few nearby or massive objects, the aim of this work is to search for the signature of DM decay and annihilation across the full sky by forward-modelling the observed gamma ray sky, as first suggested in [434]. As proposed in [435], anisotropies in the cosmic gamma-ray background could be a signature of DM annihilation or decay. This has previously been studied through the two-point correlation function [e.g. 436], which is calibrated with *unconstrained* N-body simulations (e.g. [437] use the Millennium-II and Aquarius simulations [438–440]). Instead, we utilise the `CSiBORG` suite of *constrained* N-body simulations [90–93]. The ICs for these simulations are chosen to produce final three-dimensional DM density fields which are consistent with the observed positions of galaxies in the 2M++ galaxy catalogue. The ICs are inferred using the `BORG` (Bayesian Origin Reconstruction from Galaxies) algorithm [4, 82–85], a fully Bayesian forward model. We use the particle positions in the simulations to produce maps of the expected gamma ray flux from halos down to $4.38 \times 10^{11}\,\mathrm{M}_\odot$ in mass (although we also model smaller substructures), as well as from DM not identified as belonging to halos. The halos are assumed to have Navarro-Frenk-White (NFW) profiles [441] and we explicitly model unresolved substructure in a probabilistic manner, since the signal from DM annihilation is sensitive to the peaks in the density field.

We do not include a contribution from the Milky Way halo or Local Group dwarf galaxies (which are unresolved in our simulations) in our templates so as to produce constraints entirely from large scale structure. These constraints will be free from many of the systematics affecting searches in particular objects, and will reveal the amount of constraining power for DM annihilation and decay to be found in various parts of the cosmic web. To account for non-DM effects we include templates for emission due to point sources, galactic emission and an isotropic background. We marginalise over the amplitudes of these templates, as well as the realisations in the `CSiBORG` suite which sample the full `BORG` posterior in ICs of the 2M++ volume. We then compare to *Fermi*-LAT observations via a MCMC algorithm. A full-sky field-level inference allows us to capture not only the two-point statistics, but implicitly all higher orders too.

In this work we rule out the thermal relic cross-section at 95% confidence for annihilations which produce Z bosons, gluons or quarks less massive than the bottom

quark if DM has a mass $m_\chi \lesssim 7\,\mathrm{GeV}/c^2$. We find a contribution to the gamma ray sky with the same spatial distribution as expected from DM decay (flux proportional to DM density) at 3.3σ confidence, with a decay rate $\Gamma \approx 3 \times 10^{-28}\,\mathrm{s}^{-1}$ for these channels. However, a power law spectrum with an index $p = -2.75^{+0.71}_{-0.46}$ provides a better fit to the data, suggesting a non-DM origin. In our fiducial analysis, we do not rule out the thermal relic annihilation cross-section at any mass for production of top or bottom quarks; we obtain upper bounds which are half the size if we marginalise over the contribution proportional to the DM density, but we do not include this contribution in the fiducial analysis. Our constraints on DM decay to leptons are approximately an order of magnitude less stringent than decay to quarks.

This chapter is structured as follows. We discuss DM annihilation and decay models in Section 5.2 and introduce the gamma ray data that we use to constrain these models in Section 5.3. Our inference and template construction methods are outlined in Section 5.4. The results are presented in Section 5.5 and discussed in Section 5.6, including the potential systematic uncertainties and a comparison to the literature. We conclude in Section 5.7. Equations throughout the chapter use units $\hbar = c = 1$.

5.2 Theoretical background

We start by assuming that DM is made of a single particle, χ, of mass m_χ, whose antiparticle is itself (e.g. Majorana fermions). This particle is assumed to be able to both decay

$$\chi \to A\bar{A}, \tag{5.1}$$

and annihilate

$$\chi\chi \to A\bar{A}, \tag{5.2}$$

to a standard model particle, A, and its antiparticle, $\bar{A}$. The annihilation of the produced particles would lead to gamma ray emission at some energy E_γ, which one could detect. If these processes occur via channel i with branching ratio Br_i, then the photon flux for annihilation per unit density squared at redshift z is [442]

$$\frac{\mathrm{d}\Phi_{\mathrm{pp}}^{\mathrm{ann}}}{\mathrm{d}E_\gamma} = \frac{\langle\sigma v\rangle}{8\pi m_\chi^2} \sum_i \mathrm{Br}_i \left.\frac{\mathrm{d}N_i}{\mathrm{d}E'_\gamma}\right|_{E'_\gamma = E_\gamma(1+z)}, \tag{5.3}$$

and for decay per unit density

$$\frac{\mathrm{d}\Phi_{\mathrm{pp}}^{\mathrm{dec}}}{\mathrm{d}E_\gamma} = \frac{\Gamma}{4\pi m_\chi} \sum_i \mathrm{Br}_i \left.\frac{\mathrm{d}N_i}{\mathrm{d}E'_\gamma}\right|_{E'_\gamma = E_\gamma(1+z)}, \tag{5.4}$$

where $\langle\sigma v\rangle$ is the thermally averaged cross section, $\tau = 1/\Gamma$ is the lifetime of the particle, and $\mathrm{d}N_i/\mathrm{d}E_\gamma$ is the photon energy distribution for channel i. Throughout this work we assume s-wave annihilation so the parameter $\langle\sigma v\rangle$ is assumed to be a constant, i.e. independent of v. If χ is not its own antiparticle (e.g. Dirac fermions), the annihilation flux is half of this value, provided there is no matter-antimatter asymmetry. Since we do not *a priori* know which branching ratios to use, in this work we assume that the annihilation or decay occurs via a single channel, however our analysis can be trivially extended to multiple channels.

Since these results apply at unit density, we must now take into account the integrated DM density along the line of sight. By introducing the J factor

$$\frac{\mathrm{d}J}{\mathrm{d}\Omega} = \int \rho^2_{\mathrm{DM}}\left(s, \Omega\right) \mathrm{d}s, \tag{5.5}$$

and D factor

$$\frac{\mathrm{d}D}{\mathrm{d}\Omega} = \int \rho_{\mathrm{DM}}\left(s, \Omega\right) \mathrm{d}s, \tag{5.6}$$

we arrive at the total photon fluxes per unit solid angle

$$\frac{\mathrm{d}^2\Phi^{\mathrm{ann}}}{\mathrm{d}E_\gamma \mathrm{d}\Omega} = \frac{\mathrm{d}\Phi^{\mathrm{ann}}_{\mathrm{pp}}}{\mathrm{d}E_\gamma}\frac{\mathrm{d}J}{\mathrm{d}\Omega}, \tag{5.7}$$

$$\frac{\mathrm{d}^2\Phi^{\mathrm{dec}}}{\mathrm{d}E_\gamma \mathrm{d}\Omega} = \frac{\mathrm{d}\Phi^{\mathrm{dec}}_{\mathrm{pp}}}{\mathrm{d}E_\gamma}\frac{\mathrm{d}D}{\mathrm{d}\Omega}, \tag{5.8}$$

where we note that we have assumed that the cosmological redshift variation across the source is negligible, so we can factor out Equations 5.3 and 5.4 from the line of sight integral.

5.3 Gamma ray data

In this work we use gamma ray observations from *Fermi*-LAT, and analyse these using the Fermi Tools[1] and FermiPy [443]. To mitigate the effect of contamination from cosmic rays whilst ensuring a sufficiently high photon acceptance rate at high energy to enable one to constrain emission models, we consider photons within the event class SOURCEVETO. We select all photons in this event class of energy $500\,\mathrm{MeV} - 50\,\mathrm{GeV}$ between mission weeks 9 and 634 which are flagged as belonging to the upper quartile of angular resolution (PSF3) and set the maximum zenith angle to be 90°. We subdivide these data into 9 logarithmically space energy bins, then bin spatially

[1] `https://fermi.gsfc.nasa.gov/ssc/data/analysis/software/`

onto HEALPix [2] [182, 183] maps. Although the angular resolution of the data corresponds to `nside` ≈ 1024, we compare it to the theoretical maps at `nside` $= 256$ for computational efficiency. In Section 5.6.2.1 we find that out results are not significantly affected by this choice.

Due to the high density and close proximity of the centre of our own galaxy, one would expect that a DM annihilation or decay signal would be dominated by this region. However, the constraint or detection one would obtain from studying this region would be sensitive to the modelling of the Milky Way density profile, and one would have to ensure that such a signal could not arise due to potentially incorrect modelling of the galactic diffuse or isotropic components, or through processes not captured by these models, such as an unresolved population of millisecond pulsars. This is a complicated yet feasible task [see e.g. 405], but is beyond the scope of this work; here we wish to produce constraints on DM annihilation and decay which are independent of the GCE so we simply mask the galactic plane, with the aim that any constraint or detection is driven by the density fields inferred in Section 5.4.1. We therefore mask the region with galactic latitude $|\lambda| < 30°$.

5.4 Methods

In this section we detail how we construct the full-sky templates for dark matter annihilation and decay and how these are compared to the gamma ray data to constrain the annihilation cross-section and decay rate. In Section 5.4.1 we describe the constrained simulations used to generate these templates and in Section 5.4.2 we describe how the J and D factors are computed from the DM particles in these simulations. These templates are combined with those from Section 5.4.3 which describe non-DM contributions to the gamma ray sky, and we compare these to the data using the likelihood model in Section 5.4.4.

5.4.1 Bayesian large-scale structure inference

To compute the J and D factors, we use the set of DM-only constrained simulations, `CSiBORG` , introduced in Section 1.3.2. Therefore, our results are conditioned on the `BORG` cosmology; a study of the cosmology-dependence of our results is beyond the scope of this work. To improve the effective resolution of our calculation in high-density regions we run a halofinder on the `CSiBORG` particles and use analytic formulae for a specified density profile. Specifically, we use the watershed halofinder `PHEW` [444]

[2] `http://healpix.sf.net`

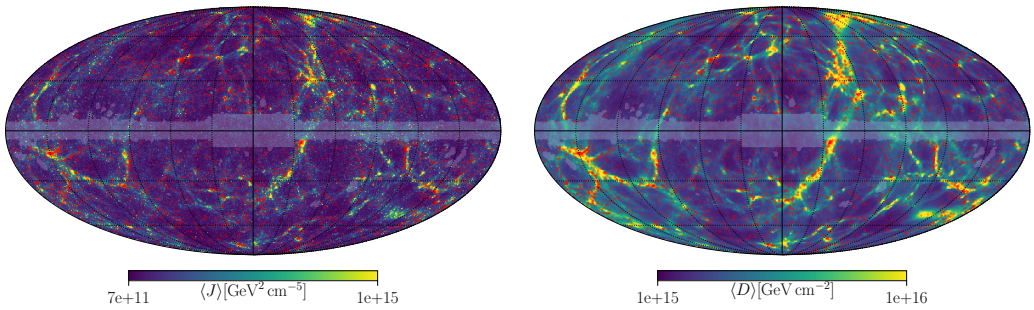


Figure 5.1: Mollweide projection in galactic coordinates of the ensemble mean J and D factors over the `CSiBORG` realisations, alongside the brightest ~ 6000 galaxies from the 2M++ dataset (red points). One can see that the calculated J and D factors, i.e. the underlying DM distribution, trace the observed galaxies well. We overplot the mask on completeness used in the `BORG` inference of the initial conditions [4, 5] (faded region near the galactic plane masked out).

which runs on the fly as a patch to `RAMSES`. This splits the negative density field into basins which share a common local minimum via steepest descent. These basins are then combined according to user-defined thresholds on the density of saddle points between basins to merge low-mass subhalos into their parents; we use the standard threshold value from Bleuler *et al.* [444] of $200\rho_c$. The halo catalogues are publicly available for the full `CSiBORG` suite [93].

5.4.2 Computing the J and D factors

The halofinding allows us to split `CSiBORG` particles into two types: those that belong to halos and those that do not. Since the J factor depends on the square of the density, it is more sensitive to the small-scale matter distribution, and thus we must treat halos separately from the background density field in order to account for structures below the resolution of the `CSiBORG` simulations. The D factor is less sensitive to these small scales, and thus we treat all particles equally in this case. D is computed using the procedure outlined in Section 5.4.2.1, where we use all particles. We compute J as in Section 5.4.2.1, but only considering non-halo particles, and add this to the contribution from halos, which are treated as in Section 5.4.2.2. We plot the resulting ensemble mean J and D factor maps in Figure 5.1.

For the J factor maps, we generate the templates on a higher resolution HEALPix grid than that on which we perform the inference (`nside` $= 2048$ instead of `nside` $= 256$) and subsequently degrade them. Due to the nonlinear dependence of J on ρ, this allows for a more faithful representation of the density field than if J was

initially calculated at `nside` = 256, which is especially important for the regions of the sky corresponding to halos produced in `CSiBORG` . We concluded the HEALPix resolution of `nside` = 2048 was sufficient by comparing the total J factor to those with increasing HEALPix resolution (`nside` = 4096, 8192) since the change in total J factor was at the sub-percent level with increasing `nside`.

Regarding the D factor calculation, we directly calculate the line-of-sight integral of the density field within a given HEALPix pixel at the selected resolution. The convergence of total all-sky D factor with this procedure is of course present for all considered resolutions, since it should be simply proportional to the total mass within the `CSiBORG` volume.

5.4.2.1 Smoothed density field

We wish to determine the density of DM particles which do not belong to a halo on a regular Cartesian grid with $N_{\rm grid} = 1024$ grid points per side. To do this, we adopt a procedure based on smooth particle hydrodynamics (SPH) [445] as described in [446] and outlined below. Using the SPH algorithm over e.g. a cloud-in-cell (CIC) approach allows us to better capture the peaks of the matter density field, since the SPH kernel will adapt to the local density of matter, in contrast to the CIC approach which has a fixed kernel corresponding to a trilinear interpolation scheme. We compare the results of using a SPH kernel to a CIC algorithm in Section 5.6.2.1.

First, we determine the number of particles, $N_{\rm p}$, within the cell corresponding to each grid point (i, j, k). We then define

$$N_{\rm X} = \max\left(N_{\rm p}, N_{\rm SPH}\right), \tag{5.9}$$

where $N_{\rm SPH} = 32$. The choice for this number of neighbour is partly motivated by the typical number of edges linking a node to its neighbours in a Delaunay tesselation. That number is approximately 16 for a Euclidean three dimensional vector space [447, 448]. We pick a value twice as big as we intend the filter to have a larger reach than the first layer of neighbours. We then find the mass associated with this grid point by considering the nearest $N_{\rm X}$ particles to be

$$\tilde{m}_{ijk} = \frac{1}{R_{ijk}^3} \sum_{l=1}^{N_{\rm X}-1} m_l W_l \mathcal{S}\left(\frac{d_l}{R_{ijk}}\right), \tag{5.10}$$

where R_{ijk} is half the distance to the furthest of the $N_{\rm X}$ particles from the grid point, m_l is the mass of particle l, which is at a distance d_l from the grid point, W_l is the

weight for particle l

$$W_l = \left(\sum_{ijk} \frac{1}{R_{ijk}^3} \mathcal{S}\left(\frac{d_l}{R_{ijk}} \right) \right)^{-1} \tag{5.11}$$

and the interpolating function, $\mathcal{S}$, is chosen to be

$$\mathcal{S}(x) = \begin{cases} 1 - \frac{3}{2}x + \frac{3}{4}x^3, & 0 \leq x < 1 \\ \frac{1}{4}(2-x)^3, & 1 \leq x < 2 \\ 0, & \text{otherwise.} \end{cases} \tag{5.12}$$

If the spacing between grid points is Δr, then the density assigned to each site is

$$\tilde{\rho}_{ijk} = \frac{\tilde{m}_{ijk}}{\Delta r^3}. \tag{5.13}$$

To compute the J and D factors, we compute Equations 5.5 and 5.6, respectively, along line of sight corresponding to each HEALPix pixel at the chosen resolution. We integrate up to the edge of the simulated volume and perform trilinear interpolation of the density field onto the line of sight. The convergence of this approach was checked by increasing the resolution of the SPH kernel. The total assigned mass to the grid was consistent between all resolutions we tried ($N_{\rm grid} = 256$, 512, 1024), therefore we opted for $N_{\rm grid} = 1024$ for our final J and D factor calculations.

5.4.2.2 J factor from halos

To include the contribution from particles inside halos, we use a custom extension of the CLUMPY package[3] [449–451]. In this section we review how the J factor is calculated in this package, and our assumptions for the halo density profiles.

We assume that all halos are spherically symmetric, and that the total density profile, $\rho_{\rm tot}$, can be described by a simple analytic form. For our fiducial case, we consider the three-parameter family of profiles [238, 452]

$$\rho_{\alpha\beta\gamma}(r) = \frac{2^{\frac{\beta-\gamma}{\alpha}} \rho_{\rm s}}{(r/r_{\rm s})^{\gamma} \left(1 + (r/r_{\rm s})^{\alpha}\right)^{\frac{\beta-\gamma}{\alpha}}}, \tag{5.14}$$

where $r_{\rm s}$ is the scale radius, $\rho_{\rm s}$ is the density at $r_{\rm s}$, and α describes the sharpness of the transition between the inner (γ) and outer (β) logarithmic slopes. For a NFW profile, $\alpha = 1$, $\beta = 3$, $\gamma = 1$. In Section 5.6.2.1, we also consider the Einasto profile [440, 453]

$$\rho_{\rm EINASTO}(r) = \rho_{-2} \exp\left(-\frac{2}{\alpha} \left(\left(\frac{r}{r_{-2}} \right)^{\alpha} - 1 \right) \right). \tag{5.15}$$

[3] `https://clumpy.gitlab.io/CLUMPY`

In the above, r_{-2} represents the radius at which the logarithmic slope of the profile equals -2, while $\rho_{-2} \equiv \rho(r_{-2})$. For both profiles we calculate the parameters defining the profile from the total halo mass and corresponding concentration. The mass-concentration relation we use is shown in Equation 5.25. We also be note that the parameter α of the Einasto profile is determined as a function of virial peak height [454]. Given that all of our halos are at very low redshift ($z \lesssim 0.05$) and almost all have masses within $M \in [10^{12}, 10^{15}]\,\mathrm{M}_\odot$, the parametric relation is quite accurate. We use the COLOSSUS package throughout [455]. We note that, since the Einasto parameters are fitted to halos produced in DM-only simulations, this captures only the uncertainty in N-body modelling. Baryons induces a potentially larger effect, which is however harder to model reliably. We discuss this further in Section 5.6.2.2.

One would also expect that a halo contains a large number of 'clumps' or subhalos, such that the true smooth component of the density profile is [456]

$$\rho_{\rm sm}\left(r\right) = \rho_{\rm tot}\left(r\right) - \langle \rho_{\rm subs}\left(r\right) \rangle, \tag{5.16}$$

where $\langle \rho_{\rm subs}(r) \rangle$ gives the average contribution from the substructure. If clump i has density profile $\rho^i_{\rm cl}$, then it contributes to the total J factor value within a HEALPIX pixel, p, as

$$J_p = \int_{p,\Delta_{\rm halo}} \left(\rho_{\rm sm}(s,\Omega) + \sum_i \rho^i_{\rm cl}(s,\Omega) \right)^2 {\rm d}s{\rm d}\Omega = J_{{\rm sm},p} + J_{{\rm subs},p} + J_{{\rm cross},p}, \tag{5.17}$$

where

$$J_{{\rm sm},p} = \int_{p,\Delta_{\rm halo}} \rho^2_{\rm sm}(s,\Omega){\rm d}s{\rm d}\Omega, \tag{5.18}$$

$$J_{{\rm subs},p} = \int_{p,\Delta_{\rm halo}} \left(\sum_i \rho^i_{\rm cl}(s,\Omega) \right)^2 {\rm d}s{\rm d}\Omega, \tag{5.19}$$

$$J_{{\rm cross},p} = 2\int_{p,\Delta_{\rm halo}} \rho_{\rm sm}(s,\Omega) \left(\sum_i \rho^i_{\rm cl}(s,\Omega) \right) {\rm d}s{\rm d}\Omega, \tag{5.20}$$

and $\Delta_{\rm halo}$ represents the intersection of the halo volume with the cone spanned by the pixel p. Our task is therefore to determine the distribution of clumps for a given halo and to calculate these integrals. Here, we readily use the solution provided by the CLUMPY package and describe it briefly below.

Assuming that a given halo has $N_{\rm tot}$ independent clumps, we factorise the distribution for the number of clumps with some mass M, concentration c, in some region ${\rm d}V = {\rm d}^3r$ as [450, 457]

$$\frac{{\rm d}N}{{\rm d}V{\rm d}M{\rm d}c} = N_{\rm tot} \frac{{\rm d}\mathcal{P}_V\left(r\right)}{{\rm d}V} \frac{{\rm d}\mathcal{P}_M\left(M\right)}{{\rm d}M} \frac{{\rm d}\mathcal{P}_c\left(M,c\right)}{{\rm d}c}. \tag{5.21}$$

Since the clumps form before the host halos within ΛCDM, their spatial distribution will follow the host dark-matter density profile. This has been shown to be a good assumption in simulations of galaxy-sized halos [440, 458]. Given the self-similar nature of collapse of ΛCDM halos, we extrapolate this conclusion to halos from our `CSiBORG` ensemble. We assume that the distribution of masses is a power-law

$$\frac{\mathrm{d}\mathcal{P}_M}{\mathrm{d}M} \propto M^{-\alpha_M}, \tag{5.22}$$

in the range $M \in [10^{-6}M_\odot, 10^{-2}M_\mathrm{h}]$ for a halo of mass M_h, where $\alpha_M = 1.9$ (see section 2.3 of [459] and references therein). Again, the values are motivated by numerical simulations of Milky Way sized halos, which we extrapolate to bigger halos present in our forward model.

Besides modelling the uncertainty due to the spatial and mass distribution of substructure, the CLUMPY package also allows us to include the uncertainty in the mass-concentration relation. For the substructure component, we consider two cases for the concentration distribution. In the first case, we assume that the concentration of all substructure halos is a deterministic function of the mass

$$\frac{\mathrm{d}\mathcal{P}_c}{\mathrm{d}c} = \delta\left(c - \bar{c}\left(M\right)\right), \tag{5.23}$$

where we define δ as the Dirac-delta distribution. The second possibility that we follow is that the concentration is log-normally distributed about this mean

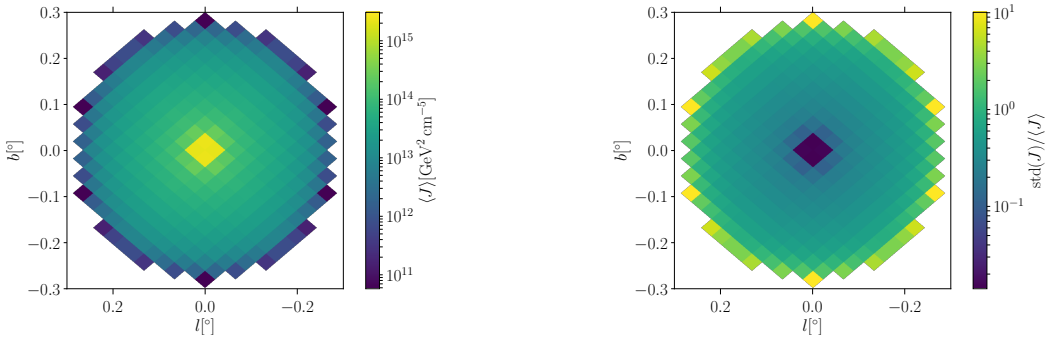
$$\frac{\mathrm{d}\mathcal{P}_c}{\mathrm{d}c} = \frac{1}{\sqrt{2\pi}c\sigma_c\left(M\right)} \exp\left(-\frac{\left(\log c - \log\left(\bar{c}\left(M\right)\right)\right)^2}{2\sigma_c^2}\right). \tag{5.24}$$

Motivated by [437], the substructure halos are assumed to have the following mass-concentration relation [460]

$$\bar{c}\left(M\right) = \sum_{j=0}^{5} C_j \left[\ln\left(\frac{M}{h^{-1}M_\odot}\right)\right]^j, \tag{5.25}$$

where $C_j = (37.5153, -1.5093, 1.636\times10^{-2}, 3.66\times10^{-4}, -2.8927\times10^{-5}, 5.32\times10^{-7})$, with $\sigma_\mathrm{c} = 0.0$, i.e. we assume all substructure halos of the same mass have the same concentration. In Section 5.6.2.3 we consider a nonzero value $\sigma_\mathrm{c} = 0.2$ in Equation 5.24, as motivated by [461–463] as a comparison.

Given that we do not resolve substructures of all the halos present with our simulation, we assume the resulting distribution for J_sm and J_subs to be a Gaussian.



(a) (b)

Figure 5.2: The figure shows (a) J factor of a typical NFW halo within `CSiBORG` ($M_{\rm h} \approx 5 \times 10^{13}\, M_\odot$) and (b) the corresponding relative fluctuations in the J factor due to the substructure contribution. The quantities $\langle J \rangle$ and std(J) are calculated according to Equations 5.35 and 5.36 respectively. As can be seen, the relative size of fluctuations in the J factor grows towards the outskirts. Qualitatively similar features are observed if we assume an Einasto profile. Note that here we placed the halo at the centre of the HEALPix grid for numerical convenience.

Therefore, we only need to find the mean and variance of these contributions in each HEALPix pixel. We define the 1-clump luminosity to be

$$L\left(M, c\right) \equiv \int_{V_{\rm subhalo}} \rho^2_{\rm cl}\left(r; M, c\right) \mathrm{d}^3 r, \tag{5.26}$$

and its moments as

$$\langle L^n \rangle \equiv \int_{M_1}^{M_2} \frac{\mathrm{d}\mathcal{P}_M}{\mathrm{d}M} \int \frac{\mathrm{d}\mathcal{P}_c}{\mathrm{d}c} L^n \mathrm{d}c\, \mathrm{d}M, \tag{5.27}$$

for a given mass range of clumps $[M_1, M_2]$, while the mean contribution of $J_{{\rm subs},p}$ from this volume is

$$\langle J_{{\rm subs},p} \rangle = N_{\rm tot} \int_{p,\mathcal{V}_\ell} \frac{\mathrm{d}\mathcal{P}_V}{\mathrm{d}V} \mathrm{d}\ell\, \mathrm{d}\Omega\, \langle L \rangle\,, \tag{5.28}$$

with $N_{\rm tot}$ representing the total number of clumps within the selected mass range $[M_1, M_2]$. Note that we assume that the clumps are non-overlapping, such that the cross-terms in Equation 5.19 can be neglected. For more details on how these quantities are defined we refer the reader to the CLUMPY related publications [449–451].

Note that since we are assuming unresolved substructures for our `CSiBORG` extragalactic halos, we are integrating over the total subhalo volume $V_{\rm subhalo}$ for the subhalo luminosity. Furthermore, since there can be many subhalos present within the line of sight determined by the given HEALPix pixel we are also accounting for the span of the host halo along this line of sight through $\mathcal{V}_\ell \equiv [\ell_{\rm min}, \ell_{\rm max}]$, with $\ell_{\rm min}$ and $\ell_{\rm max}$

being the closest and furthest points of the host halo along this line of sight. Since these integrals do not have a closed form for general DM profiles, we evaluate all numerically.

Given that the mean of some power of the distance from the observer, ℓ, to a clump which falls inside the HEALPix pixel p is

$$\left\langle \ell_p^n \right\rangle = \int_{p,\mathcal{V}_\ell} \ell^{n+2} \frac{\mathrm{d}\mathcal{P}_V}{\mathrm{d}V} \mathrm{d}\ell \, \mathrm{d}\Omega, \tag{5.29}$$

we can write the variance on $J_{\rm subs}$ as

$$\sigma^2_{J_{\rm subs,}p} = \left\langle L^2 \right\rangle \left\langle \ell_p^{-4} \right\rangle - \left\langle L \right\rangle^2 \left\langle \ell_p^{-2} \right\rangle^2, \tag{5.30}$$

since L and ℓ are independent. For the cross-term, $J_{{\rm cross},p}$, we use that its mean is

$$\left\langle J_{{\rm cross},p} \right\rangle = 2 \int_{p,\mathcal{V}_\ell} \rho_{\rm sm} \left\langle \rho_{\rm subs} \right\rangle \mathrm{d}\ell \, \mathrm{d}\Omega, \tag{5.31}$$

while its variance can be computed as

$$\sigma^2_{J_{\rm cross,}p} = 4 \int_{p,\mathcal{V}_\ell} \rho^2_{\rm sm}(l,\Omega) \sigma^2_{\rm subs}(l,\Omega) \mathrm{d}\ell \, \mathrm{d}\Omega, \tag{5.32}$$

with

$$\sigma^2_{\rm subs,p} \equiv \sigma^2_{\rm subs}(l,\Omega) = \langle \rho^2_{\rm subs}(l,\Omega) \rangle - \langle \rho_{\rm subs}(l,\Omega) \rangle^2, \tag{5.33}$$

$$\langle \rho_{\rm subs}(l,\Delta\Omega) \rangle = \int_{\Delta\Omega,\mathcal{V}_\ell} \frac{\mathrm{d}\mathcal{P}_V}{\mathrm{d}V} \int_{\mathcal{V}_M} \frac{\mathrm{d}\mathcal{P}_M}{\mathrm{d}M} \int_{\mathcal{V}_{c(M)}} \frac{\mathrm{d}\mathcal{P}_c}{\mathrm{d}c} \rho_{\rm subs}(\ell,\Omega;M,c(M)) \, \mathrm{d}\ell \mathrm{d}\Omega \mathrm{d}M \mathrm{d}c, \tag{5.34}$$

with $\mathcal{V}_M$ and $\mathcal{V}_{c(M)}$ representing the mass and corresponding concentration ranges for the subhalo distribution respectively.

We decide to only include $\sigma_{J_{\rm subs,}p}$ as it is the dominant source of uncertainty. This can intuitively be understood from the Equation 5.32. We note that the integrand is negligible both in the outskirts of the host halo, since $\rho_{\rm sm} \sim 0$ and $\sigma^2_{\rm subs}$ remains finite, and in the very centre of the host halo, since $\sigma_{{\rm subs},p} \sim 0$. Therefore, $J_{{\rm cross},p}$ contributes only at a very limited range of scales. Furthermore, for our halos, there is a clear hierarchy between the cross- and subs- term $J_{{\rm cross},p} \lesssim 0.01 - 0.1 J_{{\rm subs},p}$, therefore we focus only on the $\sigma_{J_{\rm subs,}p}$ as the dominant source of uncertainty of J factor due to substructure.

We hence write that the distribution followed by the J factor for a given halo in a given pixel is given by a Gaussian with mean

$$\left\langle J_p \right\rangle = J_{{\rm sm},p} + \left\langle J_{{\rm subs},p} \right\rangle + \left\langle J_{{\rm cross},p} \right\rangle, \tag{5.35}$$

and variance

$$\sigma_p^2 = \sigma_{J_{\mathrm{subs},p}}^2 + \sigma_{J_{\mathrm{cross},p}}^2 \approx \sigma_{J_{\mathrm{subs},p}}^2. \tag{5.36}$$

We calculate these quantities for all `CSiBORG` realisations. To distinguish between these, we introduce a second index, j, to label the simulation, i.e. $\langle J_{pj} \rangle$ is $\langle J_p \rangle$ for `CSiBORG` simulation j, and likewise σ_{pj} is σ_p for the same simulation.

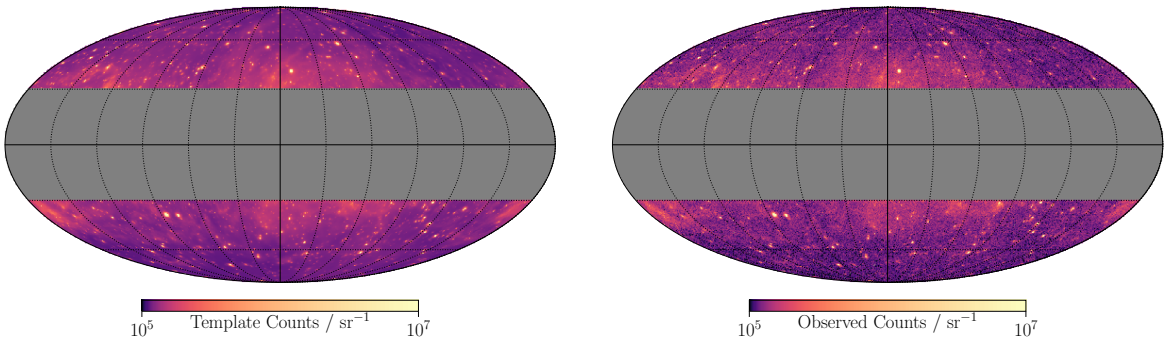
In Figure 5.2, we show the result of the model for a typical halo within `CSiBORG` with a mass of $M_h \approx 5 \times 10^{13}\,\mathrm{M}_\odot$. We see that the effects from the term in $\sigma_{J_{\mathrm{subs}}}^2$ cannot be neglected, especially in the outskirts of the halo. In the very centre, where the structure of the halo is dominated by the smooth component, the fluctuations in the J factor due to the substructure are negligible, amounting only to few percent, while in the outer parts these fluctuations become more important. This is an expected result given that the boost in J factor due to substructure becomes more important in the outer edges, where the smooth component, $J_{\mathrm{sm},p}$, is subdominant with respect to the substructure J factor, $J_{\mathrm{subs},p}$. This behaviour is identical for an Einasto profile.

Besides this, we also include the contributions of sub-subclumps to the J factor of halos, using one additional level of substructure, which is the default setting of the CLUMPY code. Due to the increased computational cost, we considered two-level substructure contribution for our halos for only one `CSiBORG` realisation. Including additional substructure levels will result in an overall change in J factor of less than $\sim 5\%$ [450], and ignoring such levels will make our constraints on $\langle \sigma v \rangle$ conservative since this will systematically underestimate the J factor.

In conclusion, to obtain the total all-sky J factor, we combine the line-of-sight calculation for the density field obtained from particles outside of halos detected within `CSiBORG` realisations with the component coming from the halo particles of the `CSiBORG` by treating the halos as presented in this section, utilising the CLUMPY code. This final template is used in the inference pipeline. We discuss the numerical convergence of these calculations in Section 5.6.2.

5.4.3 Non-DM Templates

In order to constrain the parameters describing DM annihilation or decay, one also needs to take into account other sources of gamma rays. We consider a model with three contributions: our own galaxy (gal), an isotropic background (iso) and point sources (psc). We produce separate templates, $\{T_i^t(\hat{r})\}$, $t \in \{\mathrm{iso}, \mathrm{gal}, \mathrm{psc}\}$ for each energy bin, i, and assign each template a different normalisation, which we infer from the data. The sum of the three templates is plotted in Figure 5.3 alongside the



(a) (b)

Figure 5.3: Mollweide projection in galactic coordinates of (a) the template predictions and (b) observed photon counts across the full energy range considered in this work. For visualisation, we sum the isotropic, galactic diffuse and point source templates assuming each has unit amplitude. In our inference we simultaneously infer the normalisation of each of the three components, with a different amplitude for each energy bin, and the contribution proportional to the J or D factor.

observed photon counts, from which one sees that these three templates dominate the observed gamma ray sky.

The isotropic component is designed to capture emission from unresolved extra-galactic sources, residual cosmic-rays and extra-galactic diffuse sources. It consists of a spatially constant map, with a spectral shape given by the *Fermi* Isotropic Spectral Template[4], but with an over normalisation $A_i^{\rm iso}$, which we infer separately for each energy bin, i.

For the galactic component, we use the spatial models described in [464], which are developed using spectral line surveys of HI and CO and infrared tracers of dust column density and a model of inverse Compton emission. These spatial templates describe the relative change in flux across different parts of the sky. We keep these fixed during our analysis, and infer the normalisation, $A_i^{\rm gal}$, in each energy bin, which controls the total emission from our galaxy. We note that these templates are not designed to be used for analyses which aim to fit medium or large scale diffuse structures, since the templates include a filtered residual map, which is smoothed to a few degrees. Any large-scale component not explicitly modelled when generating the templates will be absorbed by this residual and would be undetectable in our analysis. However, as can be seen in Figure 5.1, the J and D factor maps are dominated by small-scale features due to massive structures, and thus we are in the regime for which these templates are valid.

[4] `https://fermi.gsfc.nasa.gov/ssc/data/access/lat/BackgroundModels.html`

Finally, we produce a single template containing all point and extended sources from the models provided in the LAT 12-year Source Catalog (4FGL-DR3).[5] We introduce a free scaling parameter, $A_i^{\rm psc}$, for each energy bin which will be inferred. Since we find no cross-correlation between the point-source template and our own J and D factor templates, having a separate normalisation for the point-source template will not introduce new degeneracies (see Section 5.6.2.4).

5.4.4 Likelihood model

Instead of directly constraining the DM annihilation or decay parameters, we split the inference into two parts. First, we assume that there is a contribution to the gamma ray sky which is proportional to the J or D factor, i.e. for a given `CSiBORG` simulation j, the flux in energy bin i and pixel p, Φ_{ipj}, has terms

$$\Phi_{ipj} \supset \left(A_i^{\rm J} \frac{J_{pj}}{J_0} + A_i^{\rm D} \frac{D_{pj}}{D_0} \right) \Delta E_i, \tag{5.37}$$

where ΔE_i is the width of the bin and J_0 and D_0 set the units. We choose $J_0 = 10^{13}\,\mathrm{GeV^2\,cm^{-5}\,sr^{-1}}$ and $D_0 = 10^{13}\,\mathrm{GeV\,cm^{-2}\,sr^{-1}}$. We fit for the total flux of such a contribution in each energy bin to obtain a spectrum. In the second half of the inference, we fit this spectrum to a series of models (including DM annihilation and decay) in an attempt to determine the origin of such a signal.

This method has two main advantages. First, we can consider each energy bin and `CSiBORG` simulation separately in the first half of the inference. Although this involves initially computing 909 MCMC chains (one for each energy bin and for each `CSiBORG` simulation), since the problem is embarrassingly parallelisable and because we only need to infer four or five parameters for each chain ($A_i^{\rm J}$ and/or $A_i^{\rm D}$, $A_i^{\rm iso}$, $A_i^{\rm gal}$, $A_i^{\rm psc}$) compared to ~ 30 if we combined the energy bins, we find that this approach is computationally more efficient. Second, by remaining agnostic to the origin of $A_i^{\rm J}$ or $A_i^{\rm D}$ until the second step, we are able to more easily determine which energy bins drive our constraints. Hence, it becomes simpler to compare different models since we do not need to rerun the map-level inference every time that we wish to change the DM particle mass or decay channel (which is more computationally expensive).

5.4.4.1 Inferring the spectrum

We assume that photon counts in energy bin i from the J and D factor contributions, as well as each of the contributions described in Section 5.4.3 is Poisson distributed.

[5] `https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermilpsc.html`

The variation of the mean of the latter with sky position, $\hat{r}$, and energy is described by the known templates $\{T_i^t(\hat{r})\}$, where t labels the templates. For pixel p and `CSiBORG` realisation j, we then define

$$J_{ipj} \equiv \frac{J_{pj}}{J_0}\Delta E_i, \quad D_{ipj} \equiv \frac{D_{pj}}{D_0}\Delta E_i, \quad T_{ip}^t \equiv \int_p T_i^t(\hat{r})\,\mathrm{d}\Omega, \tag{5.38}$$

such that the mean number counts in pixel p and energy bin i is predicted to be

$$\lambda_{ipj} = \mathcal{F}_{ip} \times \left(A_i^{\mathrm{J}} J_{ipj} + A_i^{\mathrm{D}} D_{ipj} + \sum_t A_i^t T_{ip}^t \right), \tag{5.39}$$

and we have multiplied our templates by the *Fermi* exposure, $\mathcal{F}_{ip}$, which describes the angular variation of the sensitivity of *Fermi*. This step is performed using the Fermi Tools, where we also convolve all templates with the point spread function. The likelihood of observing n_{ip} counts in pixel p and energy bin i given the mean λ_{ipj} is

$$\mathcal{L}(n_{ip}|\lambda_{ipj}) = \frac{\lambda_{ipj}^{n_{ip}} \exp(-\lambda_{ipj})}{n_{ip}!}. \tag{5.40}$$

As discussed in Section 5.4.2, we do not know the exact DM distribution for a given `CSiBORG` simulation due to unresolved substructure in halos, although we did not include this uncertainty in Equation 5.40. We model the uncertainty on the substructure contribution to the J factor as a truncated Gaussian. This choice allows us to marginalise analytically over the substructure uncertainty, such that the conditional probability for a given λ_{ipj} given the model parameters is

$$\begin{aligned}\mathcal{L}\left(\lambda_{ipj}|A_i^{\mathrm{J}}, A_i^{\mathrm{D}}, \{A_i^t\}, j\right) = &\frac{1}{\sigma_{ipj}}\sqrt{\frac{2}{\pi}} \\ &\times \left(1 + \mathrm{erf}\left(\frac{\mu_{ipj}}{\sigma_{ipj}\sqrt{2}}\right)\right)^{-1} \exp\left(-\frac{(\lambda_{ipj} - \mu_{ipj})^2}{2\sigma_{ipj}^2}\right),\end{aligned} \tag{5.41}$$

for $\lambda_{ipj} > 0$, and zero otherwise, where

$$\mu_{ipj} = \mathcal{F}_{ip} \times \left(A_i^{\mathrm{J}} \langle J_{ipj} \rangle + A_i^{\mathrm{D}} \langle D_{ipj} \rangle + \sum_t A_i^t T_{ip}^t \right), \tag{5.42}$$

and

$$\sigma_{ipj} = A_i^{\mathrm{J}} \frac{\sigma_{pj}}{J_0} \Delta E_i. \tag{5.43}$$

The likelihood for observing n_{ip} counts in pixel p and energy bin i is then

$$\begin{aligned}
&\mathcal{L}\left(n_{ip}|A_i^{\mathrm{J}}, A_i^{\mathrm{D}}, \{A_i^t\}, j\right) \\
&\qquad = \int \mathcal{L}_{ipj}\left(n_{ip}|\lambda_{ipj}\right) \mathcal{L}\left(\lambda_{ipj}|A_i^{\mathrm{J}}, A_i^{\mathrm{D}}, \{A_i^t\}, j\right) \mathrm{d}\lambda_{ipj} \\
&\qquad = \sqrt{\frac{2^n}{\pi}} \frac{\sigma_{ipj}^{n_{ip}}}{n_{ip}!} \exp\left(-\frac{\mu_{ipj}^2}{2\sigma_{ipj}^2}\right) \left(1 + \mathrm{erf}\left(\frac{\mu_{ipj}}{\sigma_{ipj}\sqrt{2}}\right)\right)^{-1} \\
&\qquad \left(\Gamma\left(\frac{n_{ip}+1}{2}\right) {}_1F_1\left(\frac{n_{ip}+1}{2}; \frac{1}{2}; \frac{\left(\mu_{ipj} - \sigma_{ipj}^2\right)^2}{2\sigma_{ipj}^2}\right) + \right. \\
&\qquad\qquad \sqrt{2}\frac{\left(\mu_{ipj} - \sigma_{ipj}^2\right)}{\sigma_{ipj}} \Gamma\left(\frac{n_{ip}}{2} + 1\right) \\
&\qquad\qquad \left. \times {}_1F_1\left(\frac{n_{ip}+2}{2}; \frac{3}{2}; \frac{\left(\mu_{ipj} - \sigma_{ipj}^2\right)^2}{2\sigma_{ipj}^2}\right)\right), \quad (5.44)
\end{aligned}$$

where ${}_1F_1$ is the confluent hypergeometric function of the first kind. We describe how we implement this likelihood numerically in Appendix A.

Assuming that each pixel is independent, the likelihood for the observed data in energy bin i, $\mathcal{D}_i$, is

$$\mathcal{L}\left(\mathcal{D}_i|A_i^{\mathrm{J}}, A_i^{\mathrm{D}}, \{A_i^t\}, j\right) = \prod_p \mathcal{L}\left(n_{ip}|A_i^{\mathrm{J}}, A_i^{\mathrm{D}}, \{A_i^t\}, j\right). \quad (5.45)$$

Using the priors, P, given in Table 5.1, we apply Bayes' identity (Equation 1.39)

$$\mathcal{L}\left(A_i^{\mathrm{J}}, A_i^{\mathrm{D}}, \{A_i^t\}, j|\mathcal{D}_i\right) = \frac{\mathcal{L}\left(\mathcal{D}_i|A_i^{\mathrm{J}}, A_i^{\mathrm{D}}, \{A_i^t\}, j\right) P\left(A_i^{\mathrm{J}}\right) P\left(A_i^{\mathrm{D}}\right) P\left(\{A_i^t\}\right) P\left(j\right)}{\mathcal{Z}\left(\mathcal{D}_i\right)}, \quad (5.46)$$

where

$$P\left(\{A_i^t\}\right) \equiv \prod_t P\left(A_i^t\right), \quad (5.47)$$

to obtain the posterior, $\mathcal{P}\left(A_i^{\mathrm{J}}, A_i^{\mathrm{D}}, \{A_i^t\}, j|\mathcal{D}_i\right)$, where $\mathcal{Z}\left(\mathcal{D}_i\right)$ is the evidence and where we consider each energy bin and `CSiBORG` simulation separately. We use the EMCEE sampler [62] and terminate the chain when the estimate of the autocorrelation length changes by less than 1 per cent per iteration and the chain is at least 100 autocorrelation lengths long in all of the parameters.

We apply a Monte Carlo estimate to the likelihood of the `CSiBORG` samples, such that the one-dimensional posterior for the amplitude A_i^{X} (X being J or D) is

$$\mathcal{L}\left(A_i^{\mathrm{J}}|\mathcal{D}_i\right) = \frac{1}{N_{\mathrm{sim}}} \sum_j \int \mathrm{d}\{A_i^t\} \mathrm{d}A_i^{\mathrm{D}} \mathcal{L}\left(A_i^{\mathrm{J}}, A_i^{\mathrm{D}}, \{A_i^t\}, j|\mathcal{D}_i\right). \quad (5.48)$$

where $Y = J$ if $X = D$ and vice-versa. In practice, we compute the average over `CSiBORG` realisations by first fitting the one-dimensional posteriors $\mathcal{L}\left(A_i^{\mathrm{X}}|\mathcal{D}_i, j\right)$ with a spline using the `GetDist` package [176] and then computing the mean of the resulting functions. This is equivalent to concatenating the Markov Chains if each chain had the same length.

5.4.4.2 Constraining DM parameters

We now have a posterior, $\mathcal{L}\left(A_i^{\mathrm{J}}|\mathcal{D}_i\right)$, describing the gamma ray spectrum from sources which have the same spatial distribution as the J factor. We wish to fit this spectrum to a model, $f_i\left(\boldsymbol{\theta}\right)$, for these sources and infer the model parameters $\boldsymbol{\theta}$. We assume that our model is deterministic, such that

$$\mathcal{L}\left(A_i^{\mathrm{J}}|\boldsymbol{\theta}\right) = \delta\left(A_i^{\mathrm{J}} - f_i\left(\boldsymbol{\theta}\right)\right), \tag{5.49}$$

and therefore we obtain the likelihood for the observed gamma ray sky by incorporating Equation 5.48,

$$\mathcal{L}\left(\mathcal{D}_i|\boldsymbol{\theta}\right) = \int \mathrm{d}A_i^{\mathrm{J}}\, \mathcal{L}\left(\mathcal{D}_i|A_i^{\mathrm{J}}\right) \mathcal{L}\left(A_i^{\mathrm{J}}|\boldsymbol{\theta}\right) = \int \mathrm{d}A_i^{\mathrm{J}}\, \frac{\mathcal{L}\left(A_i^{\mathrm{J}}|\mathcal{D}_i\right) \mathcal{Z}\left(\mathcal{D}_i\right)}{P\left(A_i^{\mathrm{J}}\right)} \delta\left(A_i^{\mathrm{J}} - f_i\left(\boldsymbol{\theta}\right)\right). \tag{5.50}$$

We assume that all energy bins are independent such that the likelihood of $\boldsymbol{\theta}$ given the full dataset $\mathcal{D}$ is

$$\mathcal{L}\left(\mathcal{D}|\boldsymbol{\theta}\right) = \prod_i \mathcal{L}\left(\mathcal{D}_i|\boldsymbol{\theta}\right), \tag{5.51}$$

and so with a final application of Bayes' identity we obtain the posterior for our model parameters

$$\mathcal{L}\left(\boldsymbol{\theta}|\mathcal{D}\right) = \frac{\mathcal{L}\left(\mathcal{D}|\boldsymbol{\theta}\right) P\left(\boldsymbol{\theta}\right)}{\mathcal{Z}\left(\mathcal{D}\right)}. \tag{5.52}$$

If $f_i\left(\boldsymbol{\theta}\right)$ comprises exclusively of DM annihilation, then, at fixed DM mass and annihilation channel, the transformation from the posterior distribution of A_i^{J} to that of $\langle\sigma v\rangle$ is trivial. For more complicated models (i.e. where $\boldsymbol{\theta}$ consists of more than one parameter), we again calculate the posterior on $\boldsymbol{\theta}$ using the EMCEE package.

For DM annihilation and decay, prompt production, decays, hadronisation and radiative processes associated with the resulting SM products produce a variety of stable species, including gamma rays. For a given channel, one must know the energy spectrum of the intermediate SM particles and the resulting branching ratios and energies of the subsequently produced particles. One then has, for each channel, a model for the spectrum of gamma rays as a function of DM particle mass and

Table 5.1: Priors on DM properties and template amplitudes ($A_i^t \in \{A_i^{\rm gal}, A_i^{\rm iso}, A_i^{\rm psc}\}$), as defined in the text. All priors are uniform in the range given, except from the DM particle mass, m_χ, since we constrain the cross section, $\langle\sigma v\rangle$, and decay rate, Γ, at fixed m_χ. The priors on $A_i^{\rm J}$ and $A_i^{\rm D}$ depend on the minimum energy of the energy bin, E_i, although in all cases the prior is much wider than the posterior. For DM decay we also ensure that m_χ is at least twice the rest mass of the final decay products.

Parameter	Prior
m_χ / GeV/c^2	$[2, 500]$
$\langle\sigma v\rangle$ / $10^{-26}\,\mathrm{cm^3 s^{-1}}$	$[0, 10^3]$
Γ / $10^{-30}\,\mathrm{s^{-1}}$	$[0, 10^3]$
A_i^t	$[0.5, 1.5]$
$A_i^{\rm J}$ / $10^{-16}\,\mathrm{cm^{-2} s^{-1} MeV^{-1}}$	$[0, (300\,\mathrm{GeV}/E_i)^2]$
$A_i^{\rm D}$ / $10^{-16}\,\mathrm{cm^{-2} s^{-1} MeV^{-1}}$	$[0,\ 0.5 \times (300\,\mathrm{GeV}/E_i)^2]$

annihilation cross-section or decay rate. In this work we utilise the pre-computed spectra provided by the *Fermi* collaboration[6] which are calculated as described by Jeltema and Profumo [465].

5.5 Results

In Figure 5.4, we show the corner plot for the first stage of our inference, where we infer $A_i^{\rm J}$ and $A_i^{\rm D}$ simultaneously. We emphasise that we fit a different $A_i^{\rm J}$ and $A_i^{\rm D}$ to each energy bin, i, and `CSiBORG` simulation. In this example, we consider simulation number 7444 (as given in [93]) and the energy range $30 - 50\,\mathrm{GeV}$ (energy bin 9). We see that the parameters corresponding to the isotropic, galactic and point-source contributions are all approximately unity, as one would expect. For this energy bin and `CSiBORG` simulation we see that there is no evidence for a contribution to the gamma ray flux proportional to either the J or D factor. We note that $A_i^{\rm J}$ and $A_i^{\rm D}$ are highly degenerate, such that a large value of $A_i^{\rm J}$ corresponds to a small $A_i^{\rm D}$. For our fiducial analysis, we therefore choose to set one of these parameters equal to zero at a time; i.e the inference to place constraints on $\langle\sigma v\rangle$ will assume $\forall i\ A_i^{\rm D} = 0$ and for Γ we assume $\forall i\ A_i^{\rm J} = 0$. This will make our constraints conservative (see Section 5.6.1.1).

We note that $A_i^{\rm iso}$ is strongly degenerate with $A_i^{\rm gal}$, which is to be expected since both describe large-scale features across the sky. If we used exactly the same selection criteria as the *Fermi* analysis which produced the non-DM templates, then $A_i^{\rm gal}$, $A_i^{\rm iso}$ and $A_i^{\rm psc}$ would all have a mean of unity. This is not true here because the isotropic

[6] `https://fermi.gsfc.nasa.gov/ssc/data/analysis/scitools/source_models.html`

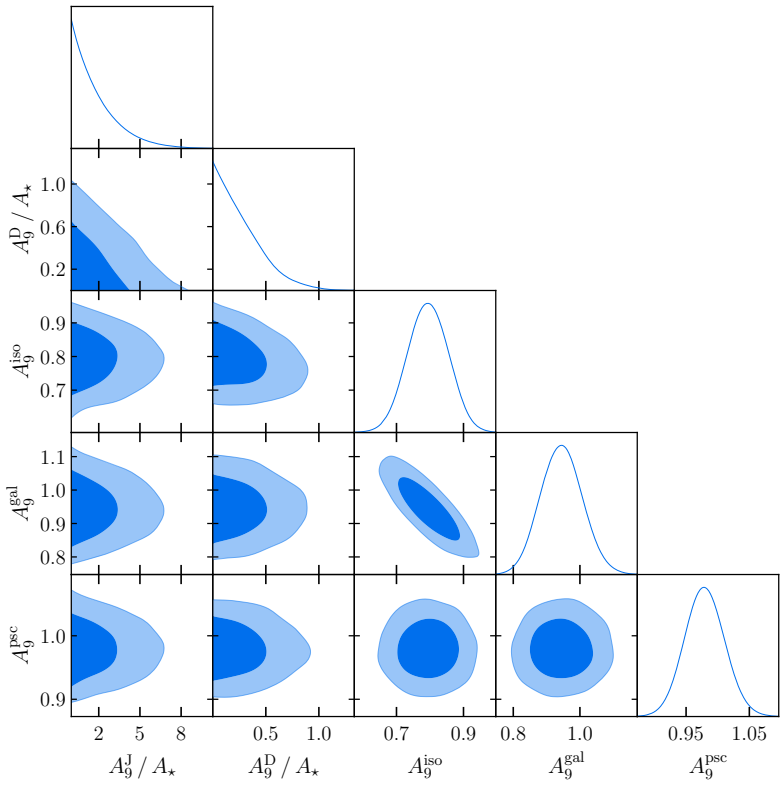
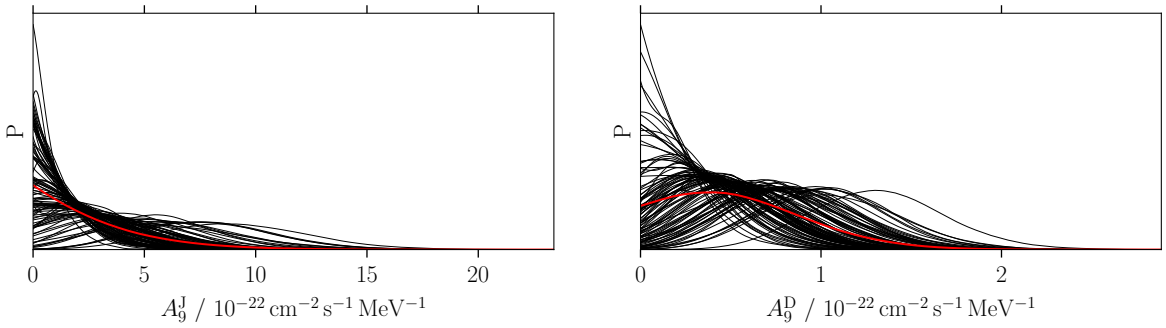


Figure 5.4: Posterior distributions for `CSiBORG` simulation 7444 of the parameters describing the gamma ray flux in the energy range $30 - 50\,\mathrm{GeV}$. We include templates proportional to the J factor ($A_9^{\rm J}$) and D factor ($A_9^{\rm D}$), as well as an isotropic ($A_9^{\rm iso}$), galactic diffuse ($A_9^{\rm gal}$) and point source ($A_9^{\rm psc}$) contributions, and define $A_\star \equiv 10^{-22}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}\mathrm{MeV}^{-1}$. The contours show the 1 and 2σ confidence intervals.



(a) (b)

Figure 5.5: One-dimensional posterior distributions on the coefficients describing the flux proportional to the (a) J factor and (b) D factor in the energy range $30-50\,\mathrm{GeV}$. Each black line gives the posterior distribution for an individual `CSiBORG` simulation, and the red line is the mean of these, i.e. the final posterior distribution.

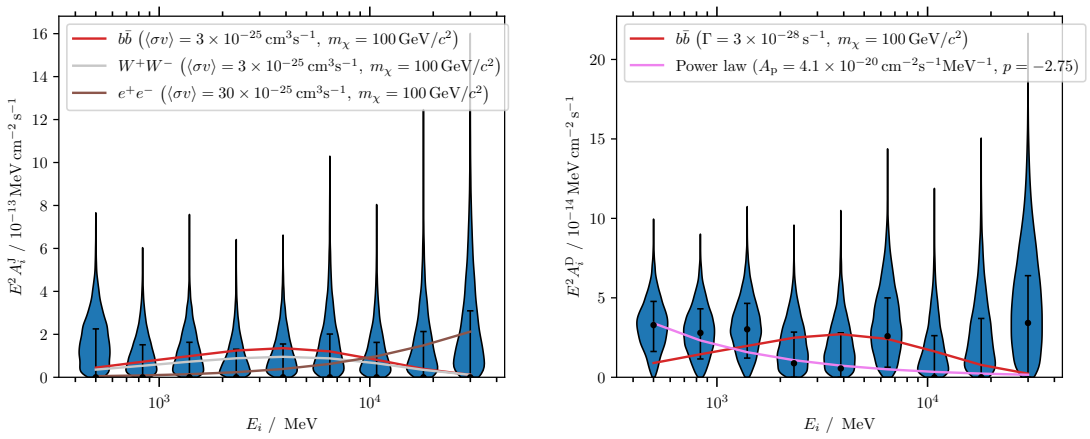


Figure 5.6: One-dimensional posterior distributions on the coefficients describing the flux in each energy bin, i, proportional to (Left) the J factor and (Right) the D factor. The black points correspond to the maximum likelihood points and the error bars show the 1σ confidence interval. For reference, we plot the expected A_i^{J} and A_i^{D} for DM annihilation and decay, respectively, via the $b\bar{b}$ channel for a particle of mass $m_\chi = 100\,\mathrm{GeV}/c^2$ with a thermally averaged cross-section of $\langle\sigma v\rangle = 3\times10^{-25}\,\mathrm{cm^3 s^{-1}}$ and decay rate $\Gamma = 3\times10^{-28}\,\mathrm{s^{-1}}$. We also plot A_i^{D} if the spectrum was due to a power law of amplitude $A_\mathrm{p} = 4.1\times10^{-20}\,\mathrm{cm^{-2} s^{-1} MeV^{-1}}$ and index $p = -2.75$, and A_i^{J} for annihilation of $m_\chi = 100\,\mathrm{GeV}/c^2$ particles which occurs via the W^+W^- ($\langle\sigma v\rangle = 3\times10^{-25}\,\mathrm{cm^3 s^{-1}}$) and e^+e^- ($\langle\sigma v\rangle = 3\times10^{-24}\,\mathrm{cm^3 s^{-1}}$) channels.

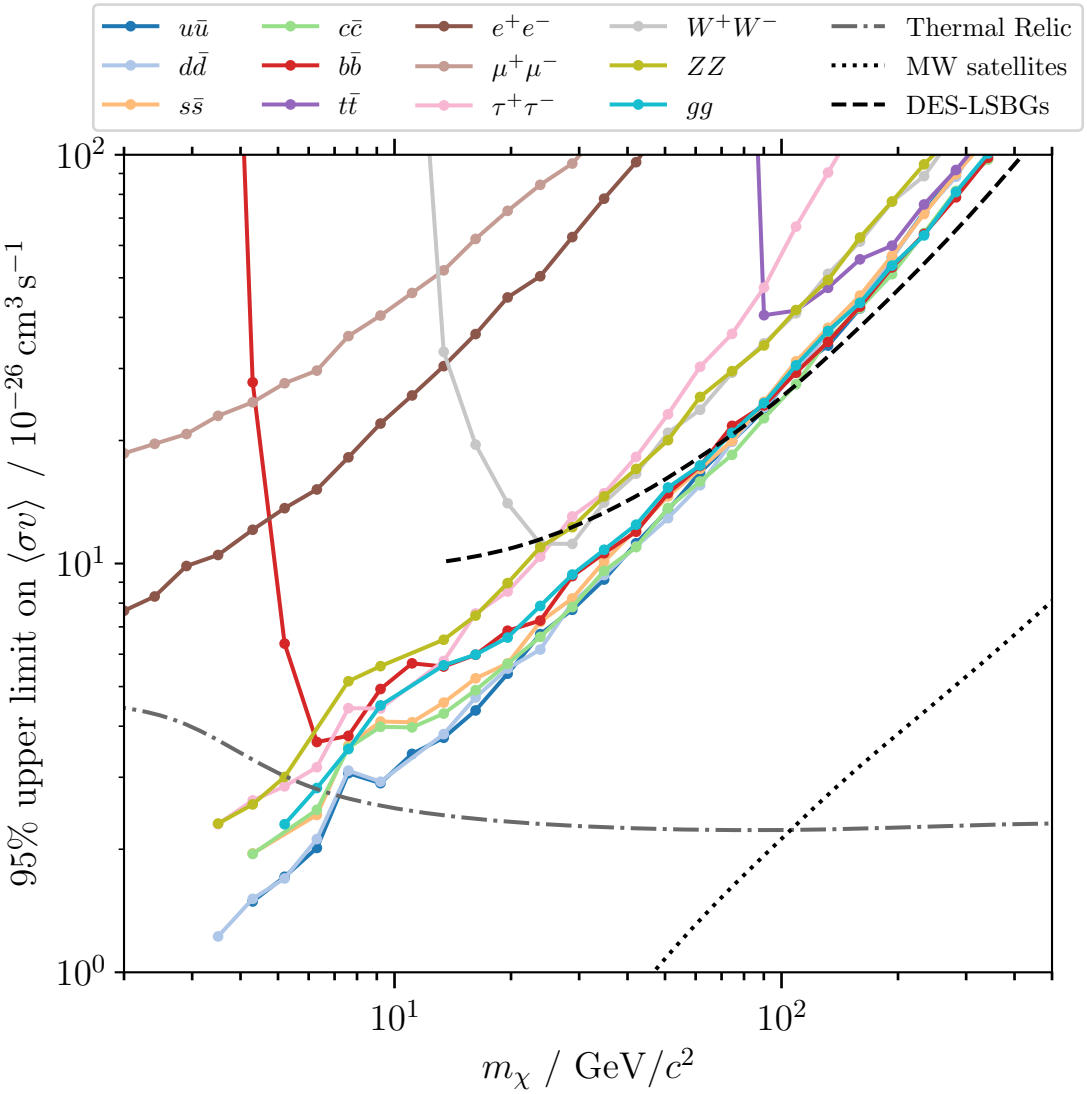


Figure 5.7: Constraints on DM annihilation cross-section, $\langle\sigma v\rangle$, as a function of particle mass, m_χ, for different annihilation channels. The dot-dashed grey line is the expectation for a thermal relic, $\langle\sigma v\rangle_{\rm th}$, as calculated by [6]. All points below this line rule out the thermal relic cross-section at 95% confidence for the corresponding mass and channel. The dotted black line is the constraint obtained by Albert *et al.* [7] from Milky Way satellites for the $b\bar{b}$ channel; we see our constraints are approximately an order of magnitude less stringent. The dashed black line shows the constraints for the $b\bar{b}$ channel derived from the cross-correlation between *Fermi*-LAT and the Dark Energy Survey Y3 low surface brightness galaxy sample (DES-LSBGs) [8]. Our field-level inference improves the constraints from large scale structure by approximately a factor of 2 at $m_\chi = 10\,\mathrm{GeV}/c^2$.

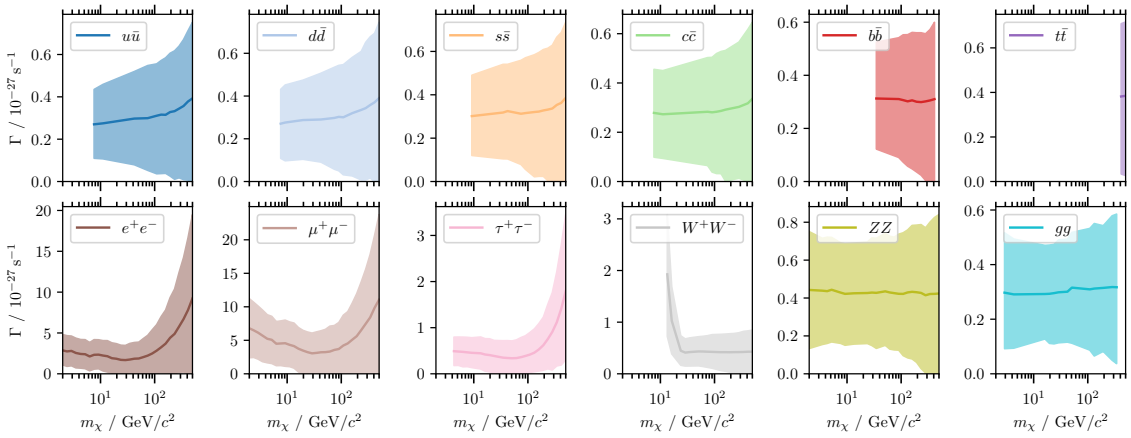


Figure 5.8: Constraints on DM decay rate, Γ, as a function of particle mass, m_χ, for different decay channels. The solid lines are the median values and the bands show the 95% confidence regions. We do not infer m_χ, which means that every constraint is conditioned on the corresponding particle mass. For some decay channels and some masses, we see that Γ is inferred to be non-zero, however (as shown in Section 5.6.1.2) we find that a power-law model better describes the spectrum, suggesting that this flux is not in fact due to DM decay. We note that for some channels we cannot probe the full mass range due to the requirement that m_χ is at least as large as the sum of the masses of the decay products.

template is calibrated for latitudes $10° < |\lambda| < 60°$, whereas we fit our template to $|\lambda| > 30°$. In general, we find $A_i^{\rm iso}$ to be slightly smaller than 1. This is more prominent in the higher energy bins; we find that $A_i^{\rm iso}$ is closer to unity at lower energy. We verify that this not due to the addition of the J or D factor templates by rerunning the analysis with $A_i^{\rm D} = A_i^{\rm J} = 0$ and find that $A_i^{\rm iso}$ remains less than one.

We generate such MCMC chains for each of the 101 `CSiBORG` simulations, and plot the resulting one-dimensional posterior distributions for $A_i^{\rm J}$ and $A_i^{\rm D}$ in Figure 5.5. Knowing that each `CSiBORG` simulation is a fair Monte Carlo sample, the final posterior distribution on $A_g^{\rm X}$ is simply the average of each individual probability distribution, which yields the red lines in the figure. When marginalised over the `BORG` chain, we again find $A_i^{\rm J}$ and $A_i^{\rm D}$ are consistent with zero for this energy bin.

This process is repeated for each energy bin to determine the posterior for a given $A_i^{\rm J}$ or $A_i^{\rm D}$, marginalised over all other contributions to the gamma ray sky and over the uncertainties involved in producing maps of the J and D factors. These spectra are displayed in Figure 5.6, where we indicate the maximum posterior points by the circles and 1σ confidence intervals by the error bars. For a given DM mass and channel, these posteriors can be trivially transformed into constraints on $\langle\sigma v\rangle$ or Γ

for a given energy bin. We then simply multiply the posteriors from each bin to determine our final constraint on these parameters.

In Figure 5.7, we plot the 95% upper limit on $\langle\sigma v\rangle$ as a function of DM particle mass, m_χ, for an annihilation which solely produces particle-antiparticle pairs of a single type, but for any standard-model quark, charged lepton or gauge boson (except photons). We compare these constraints to the thermal relic cross-section ($\langle\sigma v\rangle_{\rm th} \approx 3\times10^{-26}\,{\rm cm}^3\,{\rm s}^{-1}$), such that any point below this value in Figure 5.7 rules out DM being a thermal relic for the corresponding mass and annihilation channel at 95% confidence. For all annihilations producing Z bosons, gluons or quarks less massive than the bottom quark, we see that, if DM is a thermal relic, it should be more massive than $\sim 7\,{\rm GeV}/c^2$ since we rule out smaller masses. We are unable to rule out the thermal relic cross-section at any mass for production of bottom quarks, top quarks or W bosons. Our constraints for lepton production are much weaker at a given particle mass, such that our constraints for electron or muon production do not cross $\langle\sigma v\rangle_{\rm th}$. We rule out τ production for $m_\chi \lesssim 6\,{\rm GeV}/c^2$ at this cross-section.

Turning our attention to DM decay, Figure 5.8 shows the inferred decay rate, Γ, for different decay channels as a function of m_χ. Contrary to our analysis of DM annihilation, we find that for the majority of channels we infer a non-zero Γ at over 2σ confidence for at least some m_χ (corresponding to non-zero $A_i^{\rm D}$ in Figure 5.6). The results are relatively insensitive to the DM particle mass, provided m_χ is above the threshold for production. For the $b\bar{b}$ channel, we find the inferred Γ is $\sim 3\times10^{-28}\,{\rm s}^{-1}$, which corresponds to approximately one decay per Hubble time in a volume $\sim 560\,{\rm km}^3$ at mean cosmological density if $m_\chi = 100\,{\rm GeV}/c^2$. This is around the smallest Γ that has been constrained by any previous study (see Section 5.6.3.2). The inferred Γ for decay to the lightest charged leptons is approximately an order of magnitude larger than this.

To determine the overall detection significance, we compute the coefficient for the total flux across all energy bins which multiplies the D factor

$$A_{\rm tot}^{\rm D} \equiv \sum_i A_i^{\rm D}\Delta E_i = 1.02^{+0.24}_{-0.28}\times10^{-16}\,{\rm cm}^2{\rm s}^{-1}. \tag{5.53}$$

Simply dividing the best fit value by the lower error would suggest that our detection of a contribution to the gamma ray sky proportional to the D factor has a significance of 3.6σ when averaged over all available energies. Since our posterior is non-Gaussian, we wish to compute this significance through other methods. We compute the maximum likelihood for each $A_i^{\rm D}$ and, since each energy bin is treated as independent, the maximum likelihood for $A_{\rm tot}^{\rm D}$ is the product of these values. We compare this to

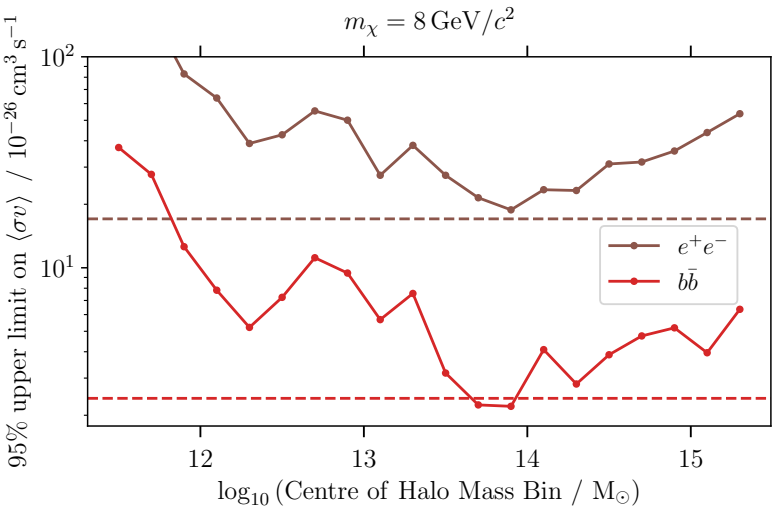


Figure 5.9: Constraints on the s-wave self-annihilation cross-section, $\langle\sigma v\rangle$, to either bottom quarks or electrons from halos within a given mass range, where we consider bins in halo mass of width $\Delta \log_{10}(M_\mathrm{h}/\mathrm{M}_\odot) = 1$. Here we only use a single `CSiBORG` simulation (9844); the 95% upper limits on $\langle\sigma v\rangle$ for this simulation if we use all halos are indicated by the dashed horizontal lines. Our constraints are dominated by halos of mass $\sim 10^{13.5} - 10^{14.5}\,\mathrm{M}_\odot$.

the likelihood for $A_i^\mathrm{D} = 0$ and find the change in log-likelihood between these two cases is $\Delta\chi^2 \equiv 2\Delta \ln \hat{\mathcal{L}} = 11.2$, which is equivalent to 3.3σ for a Gaussian likelihood or a change in the Bayesian Information Criterion (BIC) [58] of 9.0, if one takes the $A_\mathrm{tot}^\mathrm{D} \neq 0$ model as having one more parameter. In Section 5.6.1.2 we ask whether this is due to DM decay, finding that a non-DM spectrum is preferred by the data.

5.6 Discussion

In this section, we discuss the possible origin of our results, their limitations and a comparison to existing results in the scientific literature.

5.6.1 Interpretation of results

In this section we investigate which objects and observations drive our results and whether there are non-DM explanations for the signal proportional to the D factor.

5.6.1.1 Annihilation

For each channel and m_χ, we compute the change in log-likelihood between $\langle\sigma v\rangle = 0$ and the 1σ constraint on $\langle\sigma v\rangle$ separately for each energy bin to determine which

energy range is dominating our constraint. For DM particle masses $m_\chi \lesssim 10\,\mathrm{GeV}/c^2$, our constraints are driven by the first two energy bins. For these masses, we expect there to be very few photons produced at high energies at fixed $\langle\sigma v\rangle$, so these bins are unable to constrain $\langle\sigma v\rangle$ as very large values are required to produce an appreciable flux. As we move to higher masses, we notice the effect of the data at higher energy, such that at the highest masses we find that the sixth energy bin ($6.5 - 10.8\,\mathrm{GeV}$) is the most constraining. We find a similar trend for the constraints on Γ and if we compare using the 95% confidence limit instead of the 1σ constraint.

To determine which halos drive our constraints, we produce several J-factor maps where each one is only due to objects in a given mass range; we create separate maps for halos in moving bins of width $\Delta \log_{10}(M_\mathrm{h}/\mathrm{M}_\odot) = 1$. We rerun the inference for a single representative `CSiBORG` simulation (9844) separately for each of these mass bins, i.e. we assume that only a single mass bin contributes to the total J factor. We plot the constraint on $\langle\sigma v\rangle$ as a function of halo mass in Figure 5.9 and observe that the tightest constraints are obtained for halos in the range $\sim 10^{13.5} - 10^{14.5}\,\mathrm{M}_\odot$. If one studied a single object at a fixed distance, then the most massive halo would give the tightest constraints since it has the largest J factor. However, such massive objects are rare, so there is a compromise between having many objects of a given mass across the sky and those objects having a large J factor. Given the tight constraints one can obtain with dwarf galaxies in the Local Group, it is perhaps not surprising that the inclusion of lower mass objects can lead to an improvement in the upper limit on $\langle\sigma v\rangle$. The inclusion of these structures in this work was possible due to the use of constrained simulations, which provide plausible realisations of these halos given the ICs which are constrained on large scales. As a result, Figure 5.9 shows the types of objects in the nearby universe one should target to extract maximum information about DM annihilation.

Since we found a non-zero flux proportional to the D factor, we rerun the analysis for all `CSiBORG` simulations but infer both A_i^D and A_i^J simultaneously. In this way, when making our constraint on $\langle\sigma v\rangle$, we now marginalise over this contribution. Note that in this marginalisation we do not assume a spectral form for the D factor template, such that we marginalise over any source whose spatial distribution is proportional to the local DM density, which may or may not be DM decay. As anticipated in Section 5.5, we find that our constraints become tighter, such that our upper limit on $\langle\sigma v\rangle$ is typically over a factor of two smaller. This is as expected: the negative degeneracy between A_i^J and A_i^D (Figure 5.4) means that, if we allow $A_i^\mathrm{D} > 0$, we must reduce A_i^J so that the total flux from these two contributions is

approximately constant. When marginalising over $A_i^{\rm D}$, this will result in a posterior on $A_i^{\rm J}$ which is necessarily narrower. Since these results are tighter than when we set $A_i^{\rm D} = 0$, we choose to report the latter as our fiducial results so that our conclusions are conservative.

5.6.1.2 Decay

In Figure 5.8 we found a non-zero DM decay rate is compatible with the observed gamma ray sky at over 2σ confidence for a range of DM masses and decay channels. More conservatively, one would say that we find a signal which is proportional to the D factor, i.e. the emitted flux from some region appears to be proportional to the local density, and is compatible with the spectrum of DM decay. This source does not necessarily need to be DM decay, which can be investigated by fitting the inferred spectrum to a different model. For this we choose a power-law profile, such that the parameter $A_i^{\rm D}$ arises from integrating the spectrum

$$\frac{{\rm d}N}{{\rm d}E_\gamma} = A_{\rm p} \left(\frac{E_\gamma}{E_0} \right)^p , \tag{5.54}$$

across energy bin i, where we normalise to $E_0 \equiv 1\,{\rm GeV}$. We place broad, uniform priors on $A_{\rm p}$ and p in the range $[0, 10^{-18}]\,{\rm cm}^{-2}{\rm s}^{-1}{\rm MeV}^{-1}$ and $[-5, 2]$, respectively, and find

$$\begin{aligned} A_{\rm p} &= (4.1 \pm 1.5) \times 10^{-20}\,{\rm cm}^{-2}{\rm s}^{-1}{\rm MeV}^{-1}, \\ p &= -2.75^{+0.71}_{-0.46}, \end{aligned} \tag{5.55}$$

where the limits are at 1σ confidence. To enable a comparison, we plot the spectrum for this model and a DM decay model in Figure 5.6, where we see that the power law fits better at most energies.

To assess the relative goodness of fit of the two models we compute the BIC. Since we have set deliberately wide priors on our model parameters, ratios of the Bayesian evidence are difficult to interpret. For all channels and masses, we find that the BIC prefers the power law spectrum, with BIC ≥ 1.3 (the bound is saturated for the W^+W^- channel). Decays via the $t\bar{t}$ channel are least preferred by the data, with BIC ≥ 6.0. We therefore conclude that, although we do find an excess of gamma ray flux which traces the density of DM, its spectrum is fit marginally better by a power law, so this is not evidence for DM decay.

This conclusion is consistent with previous works studying the origin of the residual gamma ray flux. The cross-correlation of the gamma ray sky with galaxy catalogues has been detected at $2 - 4\sigma$ [466], and the spectral index of the 1-halo contribution was found to be -2.7 if a single power law is assumed, which is consistent with

our inferred index and has a similar significance of detection. Moreover, analysis of the angular power spectrum of the gamma ray sky [467] suggests a component which can be modelled as a double power law with an exponential cutoff, with power law indices -2.55 ± 0.23 and -1.86 ± 0.15, the former of which is consistent with our result. Ackermann *et al.* [467] note that this spectrum is compatible with blazar-like sources being the dominant component at these energies. Since we detect a non-zero contribution with the same spatial variation as our D factor maps, our work suggests that these excesses could be due to sources with a linear bias with respect to the local dark matter density. An analysis by Colavincenzo *et al.* [468] of the cross-correlation between *Fermi* data and cluster catalogues finds a peak in the cross-correlation function at $\sim 0.7^\circ$. Our field-level inference means that it is more challenging to determine the exact scales driving our results, although the cross-correlation function between the D factor map and the residuals between the observations and non-DM templates does extend beyond $\ell \sim 100$, suggesting that such scales are important in our analysis.

5.6.2 Systematic uncertainties

In this section we investigate potential systematic errors in our analysis by changing some of the analysis choices in Section 5.4. For computational convenience, throughout this section we only use one `CSiBORG` realisation (simulation 7444 as given in [93]) unless otherwise stated.

5.6.2.1 Computing the J and D factors

In Section 5.4.2.1 we computed the D factor and non-halo contribution to the J factor by smoothing simulation particles onto a grid with a kernel inspired by SPH. Besides this kernel, we also consider the CIC density assignment in order to quantify the impact of the kernel choice on our constraints. The median change in the constraint on $\langle \sigma v \rangle$ is 2% for the $b\bar{b}$ channel if we change to this kernel. For the DM decay inference, we find the median and 95% upper limit on Γ change by a median of 9% and 7.5% respectively for this channel. The inferred value of p only changes by 0.01 when we change to the CIC kernel. If the low density regions were driving our constraints then one would expect large differences between the two procedures, since these regions have the fewest simulation particles and the two kernels have different noise properties for low particle counts. However, we do not see this since the expected flux is highest in the high density regions and the low density regions are relatively unconstraining.

For computational convenience, for our fiducial analysis we chose a HEALPix resolution of `nside`=256. We rerun the analysis at coarser resolution (`nside`=128) and find that our constraints on $\langle\sigma v\rangle$ weaken by a median change of 42% across all masses for the $b\bar{b}$ channel. The median and 95% upper limit on Γ change by a median of 19% and 14% respectively for the $b\bar{b}$ channel, and the inferred value of p changes by only 0.03 when we lower the HEALPix resolution. It is unsurprising that the D factor analysis is less affected by this choice; for the J factor our template is dominated by high density peaks in the DM density field, since the flux is proportional to the square of the density. By using a higher resolution map, one can localise these peaks better to obtain tighter constraints if these are not aligned with peaks in the observed gamma ray sky.

The J and D factor maps were calculated for each of the 101 `CSiBORG` simulations. By utilising the full suite, we marginalise over both the uncertainties in the constrained density modes from the `BORG` algorithm and the unconstrained, small-scale modes. To verify that we have a sufficiently large number of simulations to achieve this, we rerun our analysis one hundred times for the $b\bar{b}$ channel using fifty randomly selected simulations to determine a bootstrap uncertainty on our constraints. The standard deviation of the 95% upper limit on $\langle\sigma v\rangle$ has a median value of 20% when considering all masses. The inferred Γ has a median bootstrap uncertainty of 4%. The uncertainty on the inferred power law index, p, is 0.05 and the fractional bootstrap uncertainty on $A_{\rm p}$ is 4%, which are small compared to the uncertainties we quote in Equation 5.55. We therefore conclude that the number of constrained simulations is adequate.

5.6.2.2 Halo density profile

After identifying halos within the `CSiBORG` simulations, we assumed that all halos have NFW profiles with masses as given by the halofinder and concentrations given by the mass-concentration relation of [460]. To determine the sensitivity of our constraints to the assumed profile, we rerun the analysis but assuming that all halos are described by Einasto profiles, as calculated in Section 5.4.2.2. We find our constraints on $\langle\sigma v\rangle$ can be up to 80% tighter if one uses an Einasto profile compared to a NFW. For small mass halos we find that the Einasto profile leads to larger densities near the centre than the NFW profile and, given the importance of these lower mass objects for our constraint (Section 5.6.1.1), this leads to smaller values of $\langle\sigma v\rangle$. Note that a similar tightening of the constraints was observed in [469] when the Milky Way

profile is changed from a NFW to an Einasto profile. We choose to report the most conservative constraints, hence the choice of NFW profiles in our fiducial analysis.

Both the form and parameters of the NFW and Einasto profiles are inspired by N-body, DM-only simulations. Observations and hydrodynamical simulations suggest that these may not accurately describe the true density profiles. In the presence of baryons, the DM profile could be steeper due to adiabatic contraction during galaxy formation [380, 381], or shallower due to the subsequent stellar feedback (e.g. [243, 382]). For steeper slopes of the density profile, the J factor near the centre would also be larger and hence one would expect tighter constraints on $\langle \sigma v \rangle$. Ideally one would use profiles motivated by hydrodynamical simulations, however common parameterisations such as [470] apply primarily to sub-Milky Way mass halos while most of those produced in `CSiBORG` are in the group and cluster regime. Rather than perform a large extrapolation, we leave it to future work to implement a robust baryonification scheme on these scales.

5.6.2.3 Substructure uncertainties

Since the gamma ray flux from DM annihilation is proportional to the square of the density, the substructure of DM halos is an important contribution that one must consider; if one computes the angular power spectrum for the J factor, C_ℓ, one finds that $\ell\,(\ell+1)\,C_\ell$ is continuing to rise at the smallest scales considered in this work. Usually this substructure is modelled as a mass-dependent multiplicative boost factor [e.g. 471] and uncertainties captured by looking at the extreme values of the boost for given masses [437]. We on the other hand capture substructure and its uncertainty through CLUMPY 's probabilistic approach to substructure modelling. This led us to use a non-Poisson likelihood, since we introduced uncertainties, σ_{jp}, on the Poisson means. To evaluate the impact of this choice, we rerun the analysis assuming a Poisson likelihood by setting $\sigma_{jp} = 0$. We find that our constraints typically change by a few per cent across all channels and masses, indicating that the impact of this uncertainty is negligible. However, we note that this could not have been known *a priori*. Although the fractional uncertainties are small near the centres of halos, this is not true in the outskirts, motivating our thorough treatment of uncertainties.

For simplicity, we previously neglected the uncertainty which arises due to stochasticity in the mass-concentration relation. We find our constraints are not very sensitive to the scatter in this relation. For the $b\bar{b}$ channel across all masses, the median change in the constraint on $\langle \sigma v \rangle$ is 1.4% if this uncertainty is included.

Another source of systematics due to substructure modelling might be driven by the breakdown of our assumption that the uncertainty in the J factor is Gaussian. Namely, as the considered mass of the clump grows, the total number of such clumps within the host halo decreases. Therefore, it is expected that at some point we transition from the Gaussian into a Poisson regime [472]. Furthermore, it is not obvious that the contribution to the J factor from these more massive clumps will not outshine the cumulative contribution of the lower mass clumps. In order to check for this, we use the `-H5` module of the CLUMPY package to explicitly draw substructure realisations for a typical halo ($M_{\rm h} \approx 5 \times 10^{13}\, M_\odot$) from the `CSiBORG` simulations. We modify Equation 5.35 such that

$$\langle J_p \rangle = \langle J_{{\rm cont},p}(M_{\rm th}) \rangle + \langle J_{{\rm drawn},p}(M_{\rm th}) \rangle, \tag{5.56}$$

and

$$\langle J_{{\rm cont},p}(M_{\rm th}) \rangle = J_{{\rm sm},p} + \langle J_{{\rm subs},p}(M_{\rm th}) \rangle + \langle J_{{\rm cross},p}(M_{\rm th}) \rangle \,, \tag{5.57}$$

where we introduce

$$\langle J_{{\rm drawn},p}(M_{\rm th}) \rangle = \frac{1}{N_{\rm ds}} \sum_i^{N_{\rm ds}} J_{{\rm drawn},pi}(M_{\rm th}), \tag{5.58}$$

with $N_{\rm ds}$ being the total number of explicit realisations of the clumps with a mass above a given mass threshold, $M_{\rm th}$. The quantities $\langle J_{{\rm subs},p}(M_{\rm th}) \rangle$ and $\langle J_{{\rm cross},p}(M_{\rm th}) \rangle$ from Equation 5.57 are obtained by replacing the upper limit of the clump mass distribution by $M_{\rm th}$, i.e. replacing M_2 with $M_{\rm th}$ in Equation 5.27. To estimate $\langle J_{{\rm drawn},p}(M_{\rm th}) \rangle$, we run $N_{\rm ds} \approx 1000$ explicit realisations of substructure clumps for a typical halo, requiring that we capture fluctuations in the value of $J_{{\rm sm},p}$ – the leading contribution to the total J factor of the halo – at the percent level. In other words, any clump whose contribution to the given pixel will induce a fluctuation to the value of $J_{{\rm sm},p}$ of the order of $\sim 1\%$ will be explicitly drawn onto the HEALPix grid. This is equivalent to taking one sample from Equation 5.21, but with a modified mass range of the mass function, and setting the lower limit for this draw to be $M_{\rm th}$. For more details see section 2.4.3 of [449].

For this experiment, we selected a HEALPix resolution of `nside` $= 1024$, corresponding to the *Fermi*-LAT angular resolution. The corresponding threshold mass for this setup translates to $M_{\rm th} = 5.3 \times 10^9\, M_\odot$ for our chosen halo. We find that

$$\langle J_{{\rm drawn},p}(M_{\rm th}) \rangle \sim 0.04\, \langle J_{{\rm cont},p}(M_{\rm th}) \rangle, \tag{5.59}$$

which justifies our starting assumption of treating the substructure contribution to the total J factor as Gaussian, since the "drawn" (Poisson) component is subdominant compared to the "continuous" (Gaussian) contribution.

Note, however, that choosing a smaller $M_{\rm th}$, i.e. looking at even smaller fluctuations of $J_{{\rm sm},p}$, would lead to probing the even lower mass end of the substructure mass function, which would of course alter the ratio of the "drawn" and "continuous" component. However, going below this limit would already enter into a regime where drawing $10^4 - 10^5$ clumps from the corresponding version of Equation 5.21 will be necessary, which is computationally expensive and well within the validity of the Gaussian approximation. As a comparison, using $M_{\rm th} \approx 5.3 \times 10^9\, M_\odot$ required around $\sim 10^3$ draws. Throughout this section we assumed the same DM profile parameterisation and mass-concentration relation as in the fiducial inference (see Section 5.6.2.2). The conclusions are unchanged for the same halo using the Einasto profile.

5.6.2.4 Non-DM Templates

The point source template is designed to remove small-scale emission so that large-scale variations in the gamma-ray sky can be modelled more robustly. Since this template is derived from the data, there is a risk that a subset of the point sources could be due to annihilating or decaying regions of high DM density. Modelling these as a non-DM component would therefore be incorrect. To assess this, we compute the angular cross-correlation function between halos identified from the `CSiBORG` simulations and the positions of point sources detected by *Fermi*-LAT. We find no significant correlation at any scale, justifying our modelling assumptions. This would be expected from the lack of degeneracy between the amplitude of the point-source template, $A_i^{\rm psc}$, and the amplitudes of the J and D factor templates, $A_i^{\rm J}$ and $A_i^{\rm D}$, in Figure 5.4. This suggests that our constraints are not driven by point sources, so the precise model we use for these is not important.

Although in Figure 5.4 we see there is little degeneracy between the parameters describing galactic diffuse emission and the DM annihilation or decay parameters, one should verify that the results are robust to reasonable variations in these non-DM templates. In our fiducial analysis we used the most recent galactic diffuse model provided by the *Fermi* collaboration (`gll_iem_v07`). We rerun the analysis with an older model (`gll_iem_v02`) and find our constraints on $\langle \sigma v \rangle$ are slightly weaker, with a median change of 10% across all mass bins for the $b\bar{b}$ channel, and that the inferred value of Γ can vary by $\sim 30\%$. Although there is some variation as we change the

model, these are at a similar level to other systematic effects. The inferred power law index, p, only changes by 0.01 so is insensitive to this choice.

5.6.2.5 High redshift sources and optical depth

The gamma rays emitted from either DM annihilation or decay would interact with the extragalactic background light (EBL) or CMB photons [473, 474]. This interaction manifests itself through pair production and therefore can cause signal attenuation, which can be modelled through an energy and redshift dependent optical depth coefficient $\tau(E, z)$. Since both the EBL and CMB are approximately isotropic, the optical depth will not have a directional dependence. For the redshift range considered in this chapter ($z \lesssim 0.05$), the attenuation of the photon flux due to interaction with background photons will not be significant, except at very high energies ($\sim$ TeV), which lie well above the maximum photon energies we consider here ($\sim 50\,\mathrm{GeV}$), and thus we neglected this contribution.

Although this is the case for the very nearby Universe, there is also a contribution to J and D from sources outside the `CSiBORG` volume. The expected contribution to the differential photon flux from annihilation is [57]

$$\left\langle \frac{\mathrm{d}^2\Phi^{\mathrm{ann}}}{\mathrm{d}E_\gamma \mathrm{d}\Omega} \right\rangle = \frac{\langle \sigma v \rangle \bar{\rho}^2_{\mathrm{DM},0}}{8\pi m_\chi^2} \int \mathrm{d}z \left(\frac{\mathrm{d}N_\gamma}{\mathrm{d}E'_\gamma} \right)\Big|_{E'_\gamma = E_\gamma(1+z)} \frac{(1+z)^3}{H(z)} e^{-\tau(E'_\gamma, z)} \langle (1+\delta(z,\Omega))^2 \rangle, \tag{5.60}$$

where $\bar{\rho}_{\mathrm{DM},0}$ represents present-day DM density and $\delta(z, \Omega)$ is the density fluctuation. This can be directly computed from the nonlinear matter power-spectrum (see for example [475]) or using the halo model approach [435, 476]. Within the halo model, this factor comes directly from averaging the one-halo annihilation luminosity over the halo mass function. This is equivalent to marginalising over plausible realisations of the DM distribution in our Universe by utilising the Press-Schechter [477] formalism, or any other halo-formation model.

As in [476] (see Fig. 10 in their Appendix B), we estimate the integrand at a given E_γ and integrate between $z = 0.05$ and $z = 10$ to determine the ratio of this contribution to that explicitly modelled from `CSiBORG` . We find that this ratio is approximately unity at $E_\gamma = 5\,\mathrm{GeV}$ for $m_\chi = 10\,\mathrm{GeV}/c^2$ for the $b\bar{b}$ channel. One may be concerned that this is an important contribution, however since our chosen HEALPix resolution of `nside` $= 256$ corresponds to a physical scale of $\sim 0.6 h^{-1}\mathrm{Mpc}$ at the edge of the `CSiBORG` volume, one would expect that the extragalactic sources beyond $z \gtrsim 0.05$ are unresolved, and therefore this contribution will almost entirely be absorbed into the isotropic template. Of course, clustering of sources at redshifts

beyond the CSiBORG volume would lead to an anisotropy in this unresolved emission. Our constraints are completely independent of how one models the isotropic part of the high-redshift component, and we leave it to further work to model the fluctuations about this, for example by including constrained realisations of the density field for larger volumes. Since this contribution can only increase the J factor, we always underestimate the J factor in our templates, making our constraints conservative.

We note that in Equation 5.60 one must correct for the redshift of emission, i.e. the spectrum should be evaluated at $E_\gamma\,(1+z)$ for a source at redshift z if we observe a photon at energy E_γ. Since we only considered sources at $z \lesssim 0.05$, we neglected this effect. If we consider the extreme case where all our sources were actually at $z = 0.05$, we would find that our estimate of the flux at a given $\langle\sigma v\rangle$ is correct to within 5% in the five lowest energy bins for the $b\bar{b}$ channel at $m_\chi = 100\,\mathrm{GeV}/c^2$ (similar effects are seen for other channels and masses). Since this is comparable to the size of other reasonable variations to the model and this is an unrealistically extreme case, we are justified in making this assumption.

5.6.3 Comparison to literature

To enable a comparison between the constraints on $\langle\sigma v\rangle$ and Γ obtained in this work using large scale structure and those from the literature, we now briefly summarise other methods for inferring these parameters and the results they produce.

5.6.3.1 Annihilation

The release of energy by annihilating DM has the potential to affect several observables over the Universe's history. The earliest important observable is the element yield from Big Bang Nucleosynthesis (BBN): a $100\,\mathrm{GeV}/c^2$ DM particle with the thermal relic cross-section would release $\sim$1 MeV of energy for every baryon in the Universe per Hubble time during the BBN era. This has the potential to alter subdominant nuclear reactions, although the effect is not strong enough to lead to stringent constraints [478, 479]. The next important epoch is recombination, where annihilating DM has the potential to ionise a non-negligible fraction of the hydrogen in the Universe. This would absorb CMB photons after recombination, to which the CMB angular power spectra and power spectra are acutely sensitive [480]. This allows thermal relic DM with a velocity-independent cross-section to be ruled out for masses below $\sim 10 - 30\,\mathrm{GeV}/c^2$ depending on the annihilation channel [432, 481]. Of course, these bounds could be evaded by a large branching fraction into neutrinos or

other particles with no electromagnetic interaction. The CMB constraints are particularly important for light DM ($m_\chi \lesssim 1\,\mathrm{GeV}/c^2$) where the effective area and angular resolution of telescopes like *Fermi*-LAT are poor. At even later times ($2 \lesssim z \lesssim 6$), observations of the Lyman-α forest constrain the gas temperature (e.g. [482, 483]) which would be increased by annihilations, although this has not been used to set quantitative constraints.

These bounds were derived purely by considering the effects of energy injection into the Universe, but more information is available from observations of the potential annihilation products themselves. This is most often done by means of high-energy photons (a common final product regardless of annihilation channel), and forms the context for our own analysis. Of course, the spectra of the final-state photons depend crucially on the channel, as we have described previously. In the local Universe, the most promising targets are the galactic centre, nearby groups or clusters, and dwarf galaxies in the Local Group. The former is the greatest nearby concentration of DM, but also suffers from large astrophysical backgrounds, and the expected signal depends sensitively on the poorly-known DM density profile of the Milky Way. Nevertheless, there are claims for a gamma-ray excess which could be due to annihilating DM [400]. In particular, Hooper and Goodenough [400] claim the excess is well fit for $m_\chi \sim 7 - 10\,\mathrm{GeV}/c^2$ with $\langle\sigma v\rangle \sim (0.5 - 5) \times 10^{-26}\,\mathrm{cm^3 s^{-1}}$ annihilating via the $\tau^+\tau^-$ channel. Although we cannot rule out the lower values of $\langle\sigma v\rangle$ for this mass range, we do find that $\langle\sigma v\rangle < 4.4 \times 10^{-26}\,\mathrm{cm^3 s^{-1}}$ at 95% confidence for these masses and this channel, which is incompatible with the larger values of $\langle\sigma v\rangle$ reported. Clusters are also massive accumulations of DM and permit a statistical analysis, but also suffer from potentially significant backgrounds. Dwarf galaxies, although smaller and less dense, have a much lower baryonic mass and hence the lowest contribution from degenerate astrophysical effects, affording a cleaner test. However, one is limited to a small sample size and thus one has to assume that the objects do not have peculiarities, e.g. unusual boost factors. By looking at a larger number of sources, as we do here, one can average over a more representative sample of substructure.

Fermi-LAT has been used to set limits on the annihilation cross-section using dwarf galaxies, the Milky Way halo [440, 484, 485], and galaxy groups [431]. The strongest constraints come from the dwarfs, which, due to their lower distances, offer higher peak signals than clusters [486]. These have been used to rule out the thermal relic cross-section for masses below $\sim 100\,\mathrm{GeV}/c^2$ assuming annihilation to b quarks [7, 429], as depicted in Figure 5.7 (although see [487]). Even stronger constraints, ruling out the thermal relic scenario to $\mathcal{O}$(TeV) mass scales for annihilation to $b\bar{b}$, have

been claimed from a radio search of the Large Magellanic Cloud [488]. Somewhat weaker constraints have also been obtained using dwarf irregular galaxies [430] and by cross-correlating *Fermi*-LAT data with the positions of nearby galaxies without knowledge of those galaxies' distances (dashed line in Figure 5.7) [8, 427]. Further information can be gleaned by cross-correlating gamma-ray flux with a tracer of density such as gravitational lensing [428].

Data from the ground-based air Cherenkov telescopes VERITAS, MAGIC, HAWC and H.E.S.S. have also been used to set constraints from dwarfs, which dominate those from *Fermi*-LAT for $m_\chi \gg 1\,\mathrm{TeV}/c^2$ [489–492]. H.E.S.S. has also been applied to the galactic centre, achieving stronger constraints at very high energies at the cost of increased systematic uncertainty due to astrophysical backgrounds [469, 493]. It is worth noting also that constraints on both annihilation and decay can be set by direct detection laboratory experiments, although these are considerably weaker than astrophysical constraints [494].

Annihilating DM produces other cosmic rays besides photons, most notably positrons and antiprotons. The AMS-02 instrument has provided data on the spectrum of a wide range of cosmic ray species [495–497]. Despite uncertainties due to cosmic ray propagation and the impact of the Sun's magnetic field, antiproton observations have been used to set bounds that beat those from *Fermi*-LAT in some cases, for example constraining the $\mu^+\mu^-$ channel to $m_\chi \sim 100\,\mathrm{GeV}/c^2$ at the thermal relic cross-section [496]. It is also possible to search for gamma-ray lines, which are generically expected to be weak but may be prominent if the DM particle decays to charged particles similar to it in mass. Line limits from *Fermi*-LAT and H.E.S.S. are presented in [469, 498] and [499, 500] respectively.

5.6.3.2 Decay

Similar considerations to those of Sec. 5.6.3.1 allow cosmological constraints to be placed on DM decay. These constraints are stronger at lower redshifts, where a greater fraction of DM decays per Hubble time for fixed decay rate. This allows BBN to test decay lifetimes around 10^{18}s, the CMB 10^{25}s and the Lyman-α forest $10^{25} - 10^{26}$s [57].

Constraints can also be derived from gamma-ray and neutrino telescopes. In particular, data from *Fermi*-LAT, AMS-02, PAO, KASCADE and CASAMIA have been used to constrain the DM lifetime at the $10^{27} - 10^{28}$s level for $10^2 < m_\chi < 10^{17}\,\mathrm{GeV}/c^2$ [501, 502]. For lower-mass DM decaying primarily leptonically, bounds

at the $10^{25} - 10^{26}$s level can be set from X-ray and gamma-ray telescopes, the spectrometer on board the Voyager I spacecraft and the heating of gas-rich dwarf galaxies, as well as the Lyman-α forest and CMB as described above [433, 503]. These constraints imply that over a very large DM mass range, only a tiny fraction of the total DM can decay during the lifetime of the Universe. Decaying DM can also be constrained using the masses and abundances of Milky Way satellites in case the decay gives momentum to the DM particle, which provides a constraint of order the age of the Universe (e.g. [504]). The inferred values of Γ in this work are compatible with these constraints.

5.6.4 Future directions

5.6.4.1 Including additional mass

Our analysis deliberately targets large scale structure as a source of annihilation or decay flux in order to be fully complementary to studies of particular objects while avoiding their systematics. This has made our constraints conservative because significant J and D factor contributions come from the Milky Way halo and dwarf spheroidals in the Local group. Incorporating these into our mass model would therefore produce the most constraining results possible, modelling flux from all mass in the local Universe. For the Milky Way this could be done by detailed modelling of the properties of our host halo along the lines of [484, 485] but using the latest data from *Gaia* [505, 506]; this will be the subject of future work. A separate likelihood component could be added for local dwarf galaxies (cf. [7, 429]). We note that inferring ICs which could produce such structures with the correct masses and locations using a process similar to the SIBELIUS simulations [507, 508] would be a computationally demanding task. A more feasible approach may be to populate larger halos with such objects *a posteriori* in a manner similar to how one paints galaxies onto a N-body simulation.

5.6.4.2 Velocity dependence

In this work we have assumed that σv is independent of energy. One can generalise this such that the cross-section is multiplied by a function S of the relative velocity between two DM particles, v_r, i.e. $(\sigma v) = (\sigma v)_0 \, S\,(v_\mathrm{r}/c)$. The velocity-dependent term is commonly modelled as $S(x) = x^n$, where in this work we have considered $n = 0$ (s-wave) scattering. Other velocity dependencies are theoretically interesting: in models with minimal flavour violation, $n = 2$ (p-wave) annihilation dominates for

Majorana fermions forming Standard Model fermion-antifermion pairs, since the s-wave is chirality-suppressed [509]. A null-result for p-wave annihilation in the galactic centre is presented in [510]. Similarly, $n = 4$ (d-wave) dominates in such models if DM is instead a real scalar [511, 512]. Due to the small DM velocities within halos, one would expect these signals to be harder to detect than s-wave scattering. However, if DM has long-range self-interactions, then the annihilation is Sommerfeld enhanced [513], corresponding to $n = -1$ and thus a high, potentially detectable annihilation rate.

When studying these velocity-dependent cross sections, one cannot use the J factor given in Equation 5.5, but instead [514]

$$J_S = \int \mathrm{d}s\, \mathrm{d}^3 v_1\, \mathrm{d}^3 v_2\, S\left(\frac{|\boldsymbol{v}_1 - \boldsymbol{v}_2|}{c}\right) f\left(\boldsymbol{r}, \boldsymbol{v_1}\right) f\left(\boldsymbol{r}, \boldsymbol{v_2}\right), \tag{5.61}$$

where $f\left(\boldsymbol{r}, \boldsymbol{v}\right)$ is the distribution function of DM particles. This thus requires one to know or model the velocities of DM particles within halos, and is therefore left to future work, although we note that J_S has recently been calculated for a range of DM density profiles [515].

5.7 Conclusions

Indirect detection of DM annihilation or decay through gamma ray emission has previously typically involved inference from a small number of nearby, DM rich objects (the Milky Way, dwarf spheroidals in the Local Group or local groups and clusters) or by cross-correlating the gamma ray background with other catalogues. Instead, in this work we utilise the `CSiBORG` suite of constrained simulations of the local $155\,h^{-1}$Mpc to forward-model the predicted gamma ray sky for s-wave DM annihilation or decay due to large scale structure. We marginalise over uncertainties in the density field reconstruction, unresolved substructure and non-DM contributions to the signal, and compare to data from *Fermi*-LAT via a MCMC algorithm.

We rule out the thermal relic cross-section at 95% confidence for DM particles of mass $m_\chi \lesssim 7\,\mathrm{GeV}/c^2$ whose annihilation produces Z bosons, gluons or quarks less massive than the bottom quark. Our constraints for the production of charged leptons are approximately an order of magnitude less stringent and we are unable to rule out the thermal relic cross section for the production of top or bottom quarks in our fiducial analysis. We infer at 3.3σ a non-zero contribution to the gamma ray sky with the same spatial distribution as predicted by DM decay. For the decay to quarks, this corresponds to a decay rate of $\Gamma \approx 3 \times 10^{-28}\,\mathrm{s}^{-1}$. However, we find

that a power-law spectrum is preferred by the data, and we infer that the power-law index is $p = -2.75^{+0.71}_{-0.46}$. If we marginalise over the contribution with the same spatial distribution as DM decay, we obtain constraints on $\langle \sigma v \rangle$ which are twice as tight as our fiducial analysis.

Our constraints on the annihilation cross section are less stringent than those obtained by studying other objects, such as the GCE or dwarf galaxies in the Local Group. Given the sensitivity of the dwarf spheroidal analysis to the prior on galaxy mass [487] and the conflicting explanations for the GCE, this work provides a useful independent probe of novel DM properties by forward modelling the whole gamma ray sky and will thus be sensitive to different systematics. The field-based framework we develop implicitly incorporates not just the two point correlation function – a more traditional way to constrain DM properties from large scale structure – but all other higher order statistics as well. Since both DM annihilation and decay fluxes are determined by line of sight integrals of the density field, the use of constrained simulations provides a convenient way of calculating these integrals for the observed Universe. Future work should be dedicated to a joint inference where one combines the contribution to the dark matter annihilation or decay signal from large scale structure with objects from the Local Group. As with analyses on smaller scales, there is some sensitivity to how one parametrises the halo density profiles, and thus future analysis should include procedures to (probabilistically) model baryonic effects on the dark matter density profile.

Chapter 6

Calibrating galaxy formation effects in galactic tests of fundamental physics

As demonstrated in the previous chapters, Bayesian Monte Carlo-based forward models have proven to be successful in constraining fundamental physics on galactic scales. These analyses have however assumed empirical noise models in which astrophysical contributions to the signals are assumed to be Gaussian distributed and uncorrelated with the properties of galaxies and their environments. The inferences would be biased if astrophysical effects were in fact significantly degenerate with the fundamental physics being tested. In this chapter we propose an approach for constructing reliable noise models based on correlations found in cosmological hydrodynamical simulations between relevant parameters of the system.

As a case study, we consider warping of stellar disks and offsets between the stellar and gas mass centroids of galaxies. These are important probes of screened fifth forces [9, 41, 363, 516], and have recently [364] been used to rule out astrophysically relevant Hu-Sawicki $f(R)$ gravity [517], a paradigmatic modified gravity model. A galactic disk can be warped through a plethora of physical phenomena besides modified gravity, including gas infall into the DM halo [518], DM self-interactions [519, 520], or interaction of the disk with companions [521, 522] (see [523] for a review). Furthermore, gas can be displaced from the centres of galaxies in clusters due to a combination of ram pressure stripping and tidal interactions [524]. It is therefore likely that disks will be warped and the stellar and gas mass separated even in the absence of a fifth force. This was accounted for in Refs. [9, 363, 364] by convolving the fifth force likelihood with a Gaussian noise model with a width that was either constant between galaxies or proportional to their distance. It is however unclear

that this model should be sufficiently flexible to account for baryonic physics, which will make the noise a function of galaxies' properties and environments.

In Section 6.1 we introduce our case study and in Section 6.2 we outline the criteria used to evaluate the suitability of the astrophysical noise model. We detail our measurement and modelling of the signals in the simulation in Section 6.3 and verify that we obtain a null detection of a fifth force in Section 6.4. We investigate the validity of the Gaussian noise model used to make these constraints in Section 6.5, and discuss the impact of the assumed halo density profile for this model in Section 6.6. We discuss the broader context of our work, and conclude, in Section 6.7.

6.1 Case study — screened fifth forces

While our methodology will prove to be general, we choose a specific case study to show it working in practice. We focus on aspects of galaxy morphology (gas–star offsets and warping of stellar disks) caused by thin-shell-screened fifth forces generated by a new light scalar gravitational degree of freedom. Throughout this chapter we use units in which $\hbar \equiv c \equiv 1$.

6.1.1 Theoretical background

In this chapter we focus on thin-shell screening mechanisms ($Z^{ij} = \delta^{ij}$ in Equation 1.27), where the degree of suppression of the fifth force is determined by the gravitational potential, Φ, such that an object is approximately unscreened if $|\Phi| < \chi$ and screened otherwise, where χ is the theory-dependent "self-screening parameter". The archetypal example is $f(R)$ gravity [525, 526], which is obtained from the Einstein-Hilbert action of GR by replacing the Ricci scalar, R, with $R + f(R)$, i.e.

$$S = \int \mathrm{d}^4x\sqrt{-g}\,\frac{R + f(R)}{16\pi G_\mathrm{N}} + S_\mathrm{m}, \tag{6.1}$$

for matter action S_m (for reviews of $f(R)$ gravity see [394, 527]). The propagating degree of freedom of $f(R)$ gravity is $f_R \equiv \frac{\mathrm{d}f}{\mathrm{d}R}$, with a background value today of f_{R0}. We will phrase our results in terms of the Compton wavelength, λ_c, of the scalar field (applicable to any scalar-tensor theory), or equivalently f_{R0} in the Hu-Sawicki model of $f(R)$. Astrophysical tests are relatively insensitive to the specific theory [528] and our results are applicable to all thin-shell-screened theories with an astrophysical range fifth force. We will compute Φ sourced by matter within λ_c of an

object [529–531], and use the screening cutoff

$$\chi = \frac{3}{2} f_{R0} = \frac{3}{2} \times 10^{-8} \left(\frac{\lambda_{\rm c}}{0.32\,{\rm Mpc}} \right)^2 , \tag{6.2}$$

appropriate for the $n = 1$ Hu-Sawicki model [517]. Note that a Compton wavelength of $\lambda_{\rm c} = 1\,{\rm Mpc}$ corresponds to a mass of $4 \times 10^{-29}\,{\rm eV}/c^2$.

6.1.2 Observables: gas-star offsets and galaxy warps

Main sequence stars will always be screened if $\chi \lesssim 10^{-6}$, since this is approximately the Newtonian potential at their surfaces. On the other hand, if a galaxy is in a sufficiently low density environment, the gas and DM within the galaxy can be unscreened. Therefore different components of a galaxy can experience different accelerations, and thus the Equivalence Principle is violated.

This generates two key morphological signals, as illustrated in Figure 6.1. The first is that the centre of the galaxy as measured by the gas will not coincide with the centre as measured by the stars. For an external fifth force field $\boldsymbol{a}_5$, evaluated with $\Delta G/G_{\rm N} = 1$, the displacement of the gas centre from the stellar, $\boldsymbol{r}_\star = r_\star \hat{r}_\star$, is

$$\frac{G_{\rm N} M\left(r_\star\right)}{r_\star^2} \hat{r}_\star \sim \frac{\Delta G}{G_{\rm N}} \boldsymbol{a}_5, \tag{6.3}$$

where $M(r)$ is the enclosed mass at a distance r from the halo centre.

This displacement results in a gravitational potential gradient across the stellar disk, which warps the disk in a characteristic 'U' shape. For equilibrium, we require the total acceleration to be constant along the disk. By equating the gravitational and fifth force contributions, we find the displacement, z, normal to the the major axis of the disk, x, in the plane of the sky to be [9]

$$z\left(x\right) = -\frac{\Delta G}{G_{\rm N}^2} \frac{|r|^3}{M\left(r\right)} \boldsymbol{a}_5 \cdot \hat{z} \approx -\frac{\Delta G}{G_{\rm N}^2} \frac{|x|^3}{M\left(x\right)} \boldsymbol{a}_5 \cdot \hat{z}, \tag{6.4}$$

where we approximate $z \ll x$ for the second equality. Note that near the centre of the disk this approximation does not hold, so $z\left(x = 0\right)$ is not necessarily zero, although this does not affect w_1 below. The disk therefore bends in the opposite direction to the projection of $\boldsymbol{a}_5$ onto the disk normal. The magnitude of this warp can be described by the warp statistic

$$w_1 \equiv \frac{1}{L^3} \int_{-L}^{L} |x| \left(z\left(x\right) - \langle z \rangle \right) {\rm d}x, \tag{6.5}$$

where we choose $L = 3R_{\rm eff}$, as in [9, 364]. The modulus sign in the integral picks out specifically 'U'-shaped warps, as opposed to the more commonly observed 'S'-shaped warps [532, 533].

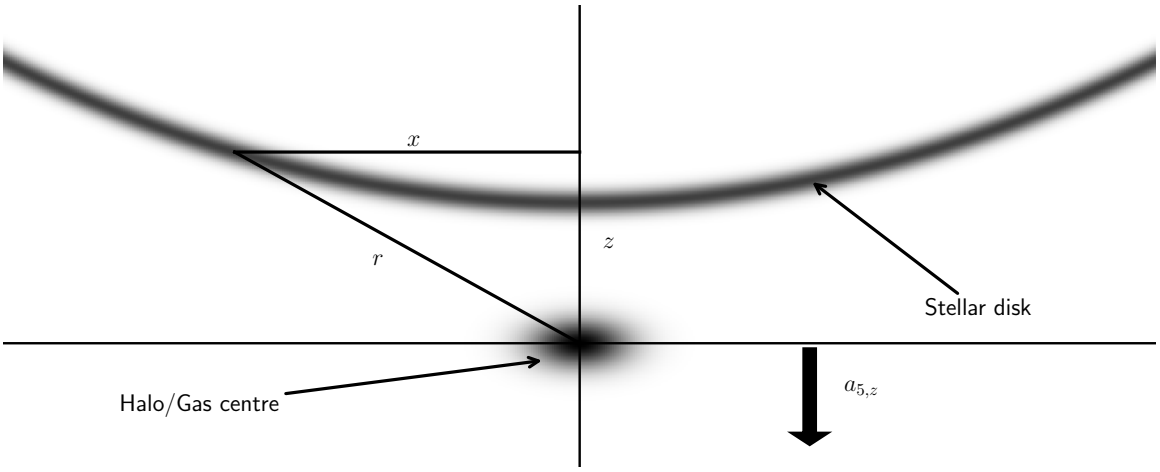


Figure 6.1: Schematic diagram of the displacement between the gas and star centres and the formation of a warped stellar disk due to a thin-shell screened fifth force, a_5. Unlike the gas and DM, the stars typically do not feel the fifth force since they are self-screened, resulting in these two morphological features.

6.2 Assessing the impact of baryons

To infer $\Delta G/G_{\rm N}$ as a function of $\lambda_{\rm c}$, we construct a galaxy-by-galaxy Bayesian forward model. This consists of two steps. First, we model our target observable (signal; $\boldsymbol{r}_\star$ or w_1) as a function of our new physics parameters ($\Delta G/G_{\rm N}$ and $\lambda_{\rm c}$) and the properties of the galaxy considered (e.g. the halo density profile) by evaluating Equation 6.3 or Equation 6.5. Any uncertainty in these properties would turn the predicted signal into a distribution, the likelihood function. We combine this with the second part of the model that describes other processes (noise) that could lead to the same observable and hence alter the prediction due to new physics. We then constrain the new physics parameters and the noise together by comparing to observations with a MCMC algorithm.

Once we have specified our model, we must check for systematic uncertainties which could bias the inference. The use of cosmological hydrodynamical simulations for this purpose offers two advantages: i) we know exactly what the theoretical parameters and implementation of baryonic physics are in the simulation, and ii) we have more information available there than we do observationally.

In particular, there are three questions to consider:

1. Are there any correlations between galactic properties and the target observable in the simulation that are not accounted for in the noise model?

2. How significant are those correlations in the inference?

3. Is the model sufficient in light of the extra information available in the simulation?

To answer point 1, we investigate whether we can predict the simulated observable from the parameters used to predict the signal in the context of the new physics model. Investigating potentially complex correlations without an a priori known functional form is best done in a machine-learning context, using algorithms to adaptively determine the functions to employ. For our case study, we train a Random Forest regressor on the simulated data; by fitting nonlinear decision trees to predict the signal from other variables, the regressor is able to assign a relative importance to each feature for determining the simulated signal [170]. Of course, to ensure our conclusions are robust to the choice of estimator, one should try multiple approaches. For our example, we repeat the analysis using an Extra Trees regressor and obtain consistent results.

If no significant correlations are found (and the real universe is similar to the simulated one), it is justified to model baryonic noise through uncorrelated random variables. Conversely, if one or more parameter is found to correlate with the simulated signal, then this model may not be sufficient. To quantify this, one may then compare the constraints obtained from the simulated data using different noise models. If a simple parameterisation can be found between the simulated signal and galactic properties due to baryonic physics, then one could allow the parameters of the noise model to vary continuously with these properties and hence marginalise over them. Alternatively, one could use one or more galactic properties to sort the sample into bins, and fit a separate noise model within each bin. If the difference in the constraint between these two methods is within some specified tolerance, then one can conclude that the simplified model is adequate; otherwise one should use the more complex one.

The extra data afforded by a simulation may be used to check other aspects of the inference method as well, for example unobservable properties of galaxies' halos. In data these must be modelled using observables that are available, which can introduce significant uncertainty. For the simulated data, however, one can compare the results of using the "true" vs model parameters to assess the accuracy of the model. In our example we consider the inner power law slope of the halo density profile: in the observational sample of [364] the absence of dynamical information at small radius makes this unobservable (it is estimated using halo abundance matching), but it can be measured directly from the DM particles in the simulation.

We explore the fifth force inferences of Refs. [9, 363, 364] in the context of H-AGN, as introduced in Section 3.2. We again produce galaxy+halo structures by matching the most massive unassigned galaxy to a halo, provided its centre is within 10 per cent of the virial radius, $r_{\rm vir}$, of the halo. Each halo is considered in turn, moving from the most to least massive. Of the initial 126,361 galaxies identified, 117,099 are partnered with a halo with this procedure.

6.3 Methods

6.3.1 Measuring the offsets and warps

6.3.1.1 Stellar warp

For every galaxy, g, with its centre at $\boldsymbol{r}_{\rm g}$ relative to the centre of the simulation volume, we find all star particles identified by the galaxy finder as belonging to that galaxy. We take the coordinates of each star particle, $\boldsymbol{r}_i$, and project these into a plane containing the angular momentum of the galaxy, $\boldsymbol{J}_{\rm g}$, to increase the probability of viewing a disk edge on. To make this projection, we define two orthogonal unit vectors for each galaxy

$$\hat{e}_1 \equiv \frac{\boldsymbol{J}_{\rm g}}{|\boldsymbol{J}_{\rm g}|}, \qquad \hat{e}_2 \equiv \frac{\boldsymbol{r}_{\rm g} \times \boldsymbol{J}_{\rm g}}{|\boldsymbol{r}_{\rm g} \times \boldsymbol{J}_{\rm g}|}, \tag{6.6}$$

and hence find the projected (angular) coordinates of the $i^{\rm th}$ star particle to be

$$\tilde{x}_i \equiv \frac{\boldsymbol{r}_i \cdot \hat{e}_1}{|\boldsymbol{r}_{\rm g}|}, \qquad \tilde{z}_i \equiv \frac{\boldsymbol{r}_i \cdot \hat{e}_2}{|\boldsymbol{r}_{\rm g}|}. \tag{6.7}$$

We fit the distribution of these particles to a Sérsic [334] profile of index n, such that the probability of having a particle at $(x_i, z_i) = (\tilde{x}_i - x_0, \tilde{z}_i - z_0)$ is

$$p\left(x_i, z_i\right) = I_0 \exp\left(-b_n \left[\left(\frac{R_i}{R_{\rm eff}}\right)^{\frac{1}{n}} - 1\right]\right), \tag{6.8}$$

where R_i is the two-dimensional distance from the centre of the distribution. The major axis of the elliptical contours, of ellipticity $\epsilon \in [0, 1)$, is at an angle θ relative to the x axis, where

$$\epsilon \equiv 1 - \frac{b}{a}, \tag{6.9}$$

for major and minor axis lengths a and b respectively. The normalisation constant, I_0, is

$$I_0 = \frac{b_n^{2n} e^{-b_n}}{2\pi n R_{\rm eff}^2 \Gamma\left(2n\right)\left(1-\epsilon\right)}, \tag{6.10}$$

where we have defined $R_{\rm eff}$ such that half of the probability lies within $R_{\rm eff}$,

$$\frac{\gamma\left(b_n, 2n\right)}{\Gamma\left(2n\right)} = \frac{1}{2}, \tag{6.11}$$

where γ is the incomplete lower gamma function. We fit for the 6 parameters of this distribution $(R_{\rm eff}, n, x_0, z_0, \epsilon, \theta)$ by maximising the likelihood

$$\log \mathcal{L}_{\rm g}\left(R_{\rm eff}, n, x_0, z_0, \epsilon, \theta\right) = \sum_{i \in {\rm g}} \log p\left(x_i, z_i\right). \tag{6.12}$$

We enforce uniform priors on all parameters in the ranges given in Table 6.1, so that the maximum likelihood is also the maximum of the posterior. The optimisation is run five times for each galaxy using Powell's method [534], with a different randomly generated start point each time. We adopt the maximum likelihood of these five as the true maximum likelihood, but require that at least two other converged points have parameters within 5 per cent of the maximum likelihood point. Otherwise, we say that the fit has not converged and we run the fit five more times. Once again we find the maximum likelihood point (of the ten) and require two different points to have parameters within 5 per cent of it. We keep adding five more fits until we obtain a converged result. After repeating this procedure five times, we find that we have successfully fitted over 99 per cent of the galaxies.

After six runs, 9 of our galaxies have zero likelihood for every iteration of the fit (i.e. this has been returned 30 times). Upon inspection of these galaxies, we find that they contain fewer than 75 star particles. It is not surprising, therefore, that we cannot fit a distribution to them. The shape measurements are unlikely to be reliable if we have too few particles, and we therefore reject galaxies with masses below $2 \times 10^9\,{\rm M}_\odot$. Given that the DM resolution is 40 times coarser than the stellar mass resolution (Section 3.2), we implement a corresponding minimum halo mass of $8 \times 10^{10}\,{\rm M}_\odot$. Both of these cuts are also implemented when generating the sample for the gas–star offset inference. We find that changing these mass cuts by ± 50 per cent does not significantly affect our results.

The observational warp statistic is determined from an image, so we must generate mock images of galaxies from H-AGN. Since the optical data in [9, 364] are from the Nasa Sloan Atlas (NSA)[1], we use the pixel size for i- and r-band images from the Sloan Digital Sky Survey (SDSS) [535] ($\Delta x = \Delta z = 0\farcs396$) as the NSA predominately contains sources from SDSS. We use the projected coordinates of all star particles

[1] `https://www.sdss.org/dr13/manga/manga-target-selection/nsa/`

Table 6.1: Priors used in the Sérsic fit to the star particles for each galaxy, where the symbols are defined in Section 6.3.1.1. All priors are uniform in the range given.

Parameter	**Description**	**Prior**
$R_{\rm eff}$	Effective radius.	> 0
n	Sérsic index.	> 0
x_0	Centre of profile along x.	-
z_0	Centre of profile along z.	-
ϵ	Ellipticity.	$[0, 1)$
θ	Angle between major axis and x.	$[0, \pi)$

within a galaxy in the $\hat{e}_1 - \hat{e}_2$ plane to determine the intensity map $I(x, z)$. From this we find the luminosity-weighted z position as a function of x,

$$\bar{z}\,(x) = \frac{\sum_{z=-L_z}^{L_z} zI(x,z)}{\sum_{z=-L_z}^{L_z} I(x,z)}, \tag{6.13}$$

where $L_z = N_{\rm eff}(b/a)R_{\rm eff}$, and the sum is over all pixels, which have x and z spacing Δx and Δz respectively. As in [9, 364], we choose $N_{\rm eff} = 3$. We calculate the mean of $\bar{z}$ across the whole image

$$\langle z \rangle = \frac{1}{n_x} \sum_{x=-N_{\rm eff}R_{\rm eff}}^{N_{\rm eff}R_{\rm eff}} \bar{z}(x), \tag{6.14}$$

where we have n_x grid points along the x axis, and we subtract this from $\bar{z}$ to have a variable of zero mean,

$$\bar{z}'(x) \equiv \bar{z}(x) - \langle z \rangle\,. \tag{6.15}$$

Finally, we use this variable to calculate the warp statistic

$$w_1 = \frac{1}{\left(N_{\rm eff}R_{\rm eff}\right)^3} \sum_{x=-N_{\rm eff}R_{\rm eff}}^{N_{\rm eff}R_{\rm eff}} |x|\bar{z}'(x)\Delta x. \tag{6.16}$$

Just creating a 2D histogram of the star particles onto the grid described above gives too many columns for which the intensity is zero, and thus Equation 6.13 is undefined. This is not a problem in the SDSS images where such zero-intensity columns are not found. To circumvent this problem we smear each point-like star particles into a Gaussian with a standard deviation equal to the pixel width. This procedure works well for the majority of galaxies, and we discard those for which we still have columns of zero intensity.

6.3.1.2 Gas-star offset

The gas centre of the galaxy is obtained by considering all gas within a box centred on the position of the densest star particle, extending to $\pm N_\star R_{\rm eff}$ in each dimension, where we choose $N_\star = 4$. If we simply calculated the centre of mass of the gas in this box, we would bias our results towards small offsets; the extreme case of a uniform gas density distribution would have a centre of mass at the origin (so zero offset) even though there is no physical justification for this. Instead, we fit the density profile to a three dimensional Gaussian,

$$\rho_{\rm gas}\left(\boldsymbol{r}\right) = \rho_{\rm g} \exp\left(-\frac{1}{2}\left(\boldsymbol{r}-\boldsymbol{\mu}\right)^{\rm T}\Sigma_{\rm g}^{-1}\left(\boldsymbol{r}-\boldsymbol{\mu}\right)\right), \tag{6.17}$$

where we fit for the central density, $\rho_{\rm g}$, mean position, $\boldsymbol{\mu}$, and the covariance matrix, $\Sigma_{\rm g}$. The fitted mean is then taken to be the gas centre, and we project the resulting offset between this and the centre of mass of the star particles into right ascension (RA), $r_{\star,\alpha}$, and declination (Dec), $r_{\star,\delta}$, components. To determine convergence, we calculate the R^2 value for the fit,

$$R^2 \equiv 1 - \frac{\sum\left(\rho_{\rm meas}\left(\boldsymbol{r}\right) - \rho_{\rm gas}\left(\boldsymbol{r}\right)\right)^2}{\sum\left(\rho_{\rm meas}\left(\boldsymbol{r}\right) - \bar{\rho}\right)^2} \tag{6.18}$$

where $\rho_{\rm meas}$ is the measured gas density, $\bar{\rho}$ is the mean of $\rho_{\rm meas}$, and the sum runs over all cells within the box of gas considered. We plot these in Figure 6.2 and see that the distribution is bimodal, suggesting that a cut in R^2 of $R^2_{\rm crit} = 0.6$ is sufficient to remove the poorly fitted density fields. We have repeated the analysis with $R^2_{\rm crit}$ in the range $0.5 - 0.8$ and find that the constraint is relatively insensitive to this parameter.

To ensure that the gas is associated with the galaxy of interest, we define a characteristic length scale of the gas

$$l_{\rm g} \equiv \left(\det \Sigma_{\rm g}\right)^{\frac{1}{6}}, \tag{6.19}$$

which is the geometric mean of the standard deviations of the density distribution along the principal axes. We then remove all galaxies where the gas–star offset is larger than $N_{\rm g} l_{\rm g}$, where we choose $N_{\rm g} = 4$. Varying $N_{\rm g}$ in the range $2 - 5$ changes the constraint by less than 50 per cent, so this choice is not important. Further, to prevent the gas associated with nearby galaxies from affecting our results, we remove all galaxies from our sample whose nearest neighbour is within $N_{\rm nn} = N_\star$ times the sum of the effective radii of the galaxies.

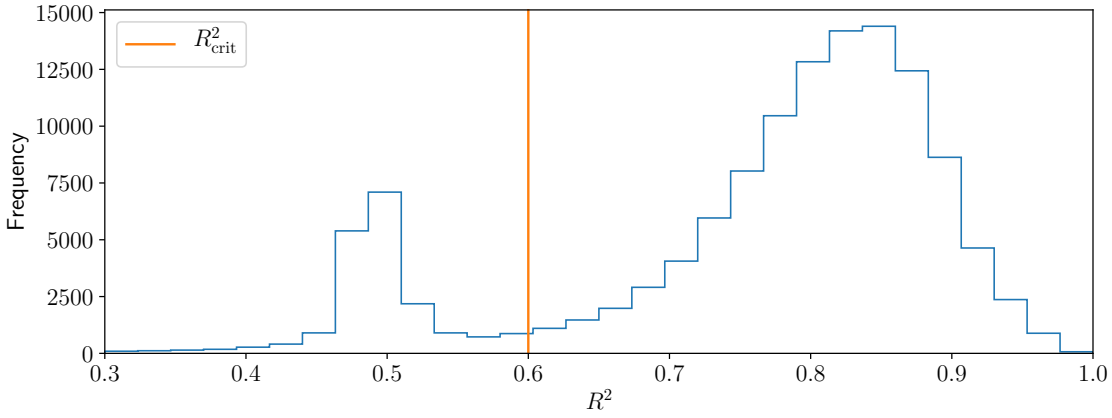


Figure 6.2: The distribution in R^2 values (Equation 6.18) for the fits to the gas density surrounding a galaxy to a Gaussian. The distribution is bimodal, indicating that a cut of $R^2_{\rm crit} = 0.6$ is sufficient to remove poor fits.

6.3.2 Halo density profile

As in [9, 363, 364], in Section 6.3.3.2 we will forward model the gas–star offsets and galaxy warps assuming that the halo density profile is a power law within some transition radius, $r_{\rm t}$,

$$\rho\left(r\right) = \rho_{\rm t}\left(\frac{r}{r_{\rm t}}\right)^{-\beta}. \tag{6.20}$$

For each galaxy, we therefore need to know $r_{\rm t}$, the density at $r_{\rm t}$ ($\rho_{\rm t}$), and the inner power law slope (β). In [363], abundance matching (AM) is used to find the Navarro-Frenk-White (NFW) [536] profile parameters for the host halo of each galaxy,

$$\rho_{\rm NFW}\left(r\right) = \frac{\rho_0}{\left(r/r_{\rm s}\right)\left(1+r/r_{\rm s}\right)^2}. \tag{6.21}$$

It is then assumed that $r_{\rm t} = r_{\rm s}$ and that $\beta = 0.5$.

As discussed in Section 3.2, the galaxies in H-AGN are already matched to halos, but we do need to fit for the NFW parameters. To do this, we use that the probability of some particle to be at radius r is

$$P\left(r|r < r_{\rm max}\right) = \frac{4\pi r^2 \rho\left(r\right)}{M\left(r_{\rm max}\right)}, \tag{6.22}$$

where

$$M\left(r_{\rm max}\right) = \int_0^{r_{\rm max}} 4\pi r^2 \rho\left(r\right) {\rm d}r. \tag{6.23}$$

From this, it is clear than any dependence on ρ_0 in $P\left(r|r < r_{\rm max}\right)$ cancels, so we must determine this at the end. To fit the profile, we maximise

$$\log \mathcal{L}_{\rm h} = \sum_i \log\left(P\left(r_i|r_i < r_{\rm max}\right)\right), \tag{6.24}$$

where the sum is over all DM particles identified as belonging to the halo within some radius $r_{\rm max}$ of the halo centre of mass. We therefore fit for $r_{\rm s}$, requiring that $r_{\rm s} > 0$. To find the second parameter of this profile (ρ_0), we enforce

$$M\left(r_{\rm max}\right) = \sum_i m_i, \tag{6.25}$$

where m_i is the mass of the $i^{\rm th}$ DM particle.

We should make an initial guess at our parameters in order to find the maximum likelihood point. We take the mass of the halo from the halo finder, $M_{\rm h}$, and estimate the concentration, $c_{\rm h} = r_{\rm vir}/r_s$, using the mass–concentration relation [537]

$$c_{\rm h,\,guess} = 57.6\left(\frac{M_{\rm h}}{M_\odot}\right)^{-0.078}. \tag{6.26}$$

We find that an initial guess of $r_{\rm s}$ equal to the $75^{\rm th}$ percentile of r_i divided by $c_{\rm h,\,guess}$ is appropriate. Finally, we need to choose a value of $r_{\rm max}$ within which we fit our NFW profile. We make the simplest choice, setting $r_{\rm max}$ to be the virial radius, $r_{\rm vir}$, as output by the halo finder.

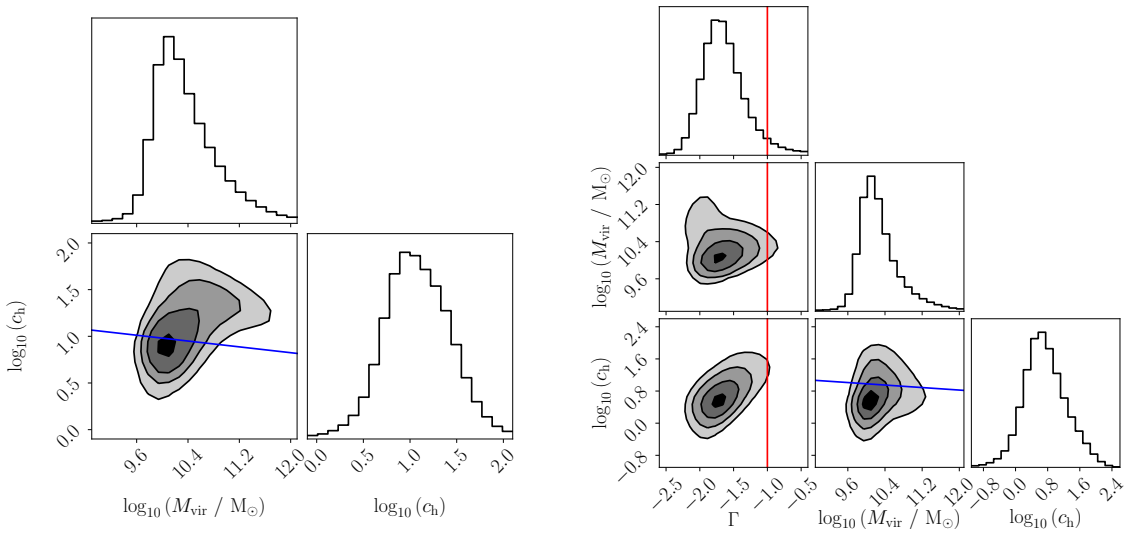
As an alternative to this method, we consider a parameterisation where the inner power law slope can vary. We consider a more general density profile

$$\rho_\Gamma\left(r\right) = \frac{\rho_0}{\left(r/r_{\rm s}\right)^{-\Gamma}\left(1 + r/r_{\rm s}\right)^{3+\Gamma}}, \tag{6.27}$$

which, like NFW, scales as r^{-3} at large radii, but is allowed to have a different inner power law slope, Γ. Comparing to Equation 6.20, we see that for $r \ll r_{\rm s}$, we have $r_{\rm t} = r_{\rm s}$, $\rho_{\rm t} = \rho_0$ and $\beta = -\Gamma$.

We fit this using the same procedure as before, first fitting for $r_{\rm s}$ and Γ, and then finding ρ_0 by considering the total mass of particles. We now have the additional requirement that $\Gamma > -3$, so that $\lim_{r\to 0} M(r) = 0$.

The distributions of the fitting parameters for both of these profiles are shown in Figure 6.3. We see that the majority of galaxies have $\Gamma < -1$, consistent with the conclusion of [538] that the halos in H-AGN have steeper density profiles near their centre than a NFW profile. We find that the mass-concentration relation of Equation 6.26 falls within the 1σ region of the distribution for the NFW profile, but the more general profile favours slightly lower concentrations.



(a) (b)

Figure 6.3: Distributions of fitting parameters to halos in H-AGN for (a) a NFW profile (Equation 6.21) and (b) a NFW-like profile with a different inner power law slope (Equation 6.27). Note that this is not a plot of posteriors, but a density map with one point per halo. The contours show the 1, 2 and 3σ levels of the distributions. The blue line is the mass-concentration relation given in Equation 6.26. This falls within the 1σ region of a, whereas the more general profile favours a smaller concentration for a given virial mass. The red line in b represents the NFW case ($\Gamma = -1$); most halos have steeper central density profiles than this ($\Gamma < -1$).

6.3.3 Modelling the offsets and warps

6.3.3.1 Gravitational and fifth force fields

H-AGN, by construction, does not include a fifth force. To check our inference method we confirm that it reconstructs no such force from the simulation, and calculate the fifth force constraints it would impose were it real data. To make the calculation computationally feasible, and to mimic the methods of [9, 363, 364], we consider two distinct contributions to the fifth force: i) a smoothed density field and ii) halos, which we assume are described by NFW profiles with the parameters obtained in Section 6.3.2. To obtain the smoothed density field, we ignore all DM particles that are assigned to halos, and project the remaining DM particles and all of the star particles onto the same grid as the gas. This grid is defined to have 2^ℓ cells per side across the full simulation volume. We assume that the density is Gaussian distributed in each cell,

$$\rho_{\rm Gauss}\left(r\right) = \frac{M}{\left(2\pi\sigma^2\right)^{\frac{3}{2}}} \exp\left(-\frac{r^2}{2\sigma^2}\right), \tag{6.28}$$

for

$$\sigma = \frac{L_{\rm box}}{2^{\ell+1}}, \tag{6.29}$$

where $L_{\rm box}$ is the simulation box length.

Using these components, we first compute the Newtonian potential at the centre of each halo sourced by all mass within $\lambda_{\rm c}$ of that point. The contribution from a grid cell of mass M of the smoothed density field at a distance r is

$$\Phi_{\rm ext,\,Gauss}\left(r\right) = -\frac{G_{\rm N}M}{r}\,{\rm erf}\left(\frac{r}{\sigma\sqrt{2}}\right). \tag{6.30}$$

The contribution from a halo of virial mass $M_{\rm vir}$ and concentration $c_{\rm h}$ at a distance r is

$$\Phi_{\rm ext,\,NFW}\left(r\right) = -\frac{G_{\rm N}M_{\rm vir}}{r}\frac{\ln\left(1+c_{\rm h}r/r_{\rm vir}\right)}{\ln\left(1+c_{\rm h}\right) - c_{\rm h}/\left(1+c_{\rm h}\right)}. \tag{6.31}$$

For each halo, we add a component due to self screening [529, 531]

$$\Phi_{\rm int} = -V_{\rm vir}^2, \tag{6.32}$$

where $V_{\rm vir}$ is the virial velocity of the halo. The potential at a given halo is then

$$\Phi = \Phi_{\rm ext} + \Phi_{\rm int} = \Phi_{\rm ext,\,Gauss} + \Phi_{\rm ext,\,NFW} + \Phi_{\rm int}, \tag{6.33}$$

such that halos with $|\Phi| < \chi$ are unscreened, where χ is given by Equation 6.2.

To calculate the fifth force at the position of each halo, we sum the contributions from the unscreened halos and the unscreened regions of the smoothed density field within $4\lambda_\mathrm{c}$ of the centre. To determine the Newtonian potential of the smoothed density field, we use a multidimensional piece-wise linear interpolator to interpolate the values of Φ_ext from the halos to the grid points (as in [364]). Assuming there is no self screening of the smoothed density field, all grid points with $|\Phi_\mathrm{ext}| < \chi$ are then unscreened. By solving the time-independent massive Klein-Gordon equation (Equation 1.27 with $Z^{ij} = \delta^{ij}$), we find the magnitudes of the contributions from the Gaussian smoothed density field and NFW halos to be

$$\begin{aligned} a_{5,\,\mathrm{Gauss}}\left(r\right) = -\frac{2\alpha^2 G_\mathrm{N} M}{r^2} e^{-r/\lambda_\mathrm{c}} \left(1 + \frac{r}{\lambda_\mathrm{c}}\right) e^{\frac{\sigma^2}{2\lambda_\mathrm{c}^2}} \left[\frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{r - \sigma^2/\lambda_\mathrm{c}}{\sigma\sqrt{2}}\right)\right) - \right. \\ \frac{1}{2}\frac{1 - r/\lambda_\mathrm{c}}{1 + r/\lambda_\mathrm{c}} e^{2r/\lambda_\mathrm{c}} \left(1 - \mathrm{erf}\left(\frac{r + \sigma^2/\lambda_\mathrm{c}}{\sigma\sqrt{2}}\right)\right) - \\ \left. \frac{2r}{1 + r/\lambda_\mathrm{c}} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(r - \sigma^2/\lambda_\mathrm{c}\right)^2}{2\sigma^2}\right)\right], \end{aligned} \tag{6.34}$$

$$\begin{aligned} a_{5,\,\mathrm{NFW}}\left(r\right) = -\frac{2\alpha^2 G_\mathrm{N} M_\mathrm{vir}}{r^2} \frac{c_\mathrm{h}\left(1 + c_\mathrm{h}\right) e^{-(r + r_\mathrm{vir}/c_\mathrm{h})/\lambda_\mathrm{c}}}{2\left(r_\mathrm{vir} + c_\mathrm{h} r\right)\left(\left(1 + c_\mathrm{h}\right)\ln\left(1 + c_\mathrm{h}\right) - c_\mathrm{h}\right)} \times \\ \left(e^{2(r + r_\mathrm{vir}/c_\mathrm{h})/\lambda_\mathrm{c}} \left(\frac{r}{\lambda_\mathrm{c}} - 1\right)\left(\frac{r_\mathrm{vir}}{c_\mathrm{h}} + r\right) \mathrm{Ei}\left(-\frac{r + r_\mathrm{vir}/c_\mathrm{h}}{\lambda_\mathrm{c}}\right) - \right. \\ \left(\frac{r}{\lambda_\mathrm{c}} + 1\right)\left(\frac{r_\mathrm{vir}}{c_\mathrm{h}} + r\right) \mathrm{Ei}\left(\frac{r + r_\mathrm{vir}/c_\mathrm{h}}{\lambda_\mathrm{c}}\right) + 2e^{r_\mathrm{vir}/(\lambda_\mathrm{c} c_\mathrm{h})} \times \\ \left. \left(e^{r/\lambda_\mathrm{c}} r + \left(\frac{r}{\lambda_\mathrm{c}} + 1\right)\left(\frac{r_\mathrm{vir}}{c_\mathrm{h}} + r\right)\left(\gamma + \ln\left(\frac{r_\mathrm{vir}}{\lambda_\mathrm{c} c_\mathrm{h}}\right)\right)\right)\right), \end{aligned} \tag{6.35}$$

where Ei is the exponential integral function and γ is Euler's gamma constant.

We choose $\ell = 7$, which corresponds to a spatial resolution of $\Delta x \sim 1.1\,\mathrm{Mpc}$, comparable to the $1.4\,\mathrm{Mpc}$ resolution used for the smoothed density field in [364]. As discussed in [363], if λ_c is less than a few Δx, the discretisation of the smoothed density field can lead to excessive shot noise. Hence, in the cases where $\lambda_\mathrm{c} < R_\mathrm{thresh}$, we evaluate the potential and acceleration from the smoothed density field at a cutoff of R_thresh and $4R_\mathrm{thresh}$ respectively, then correct our results as

$$\Phi_{\mathrm{ext},\,\mathrm{Gauss}}\left(\lambda_\mathrm{c}\right) = \left(\frac{\lambda_\mathrm{c}}{R_\mathrm{thresh}}\right)^2 \Phi_{\mathrm{ext},\,\mathrm{Gauss}}\left(R_\mathrm{thresh}\right), \tag{6.36}$$

and

$$\boldsymbol{a}_\mathrm{Gauss}\left(\lambda_\mathrm{c}\right) = \left(\frac{\lambda_\mathrm{c}}{R_\mathrm{thresh}}\right)^2 \boldsymbol{a}_\mathrm{Gauss}\left(R_\mathrm{thresh}\right), \tag{6.37}$$

where, as in [364], we choose $R_\mathrm{thresh} = 3.5\,\mathrm{Mpc}$.

6.3.3.2 Modelling the halo restoring force

We focus on the power-law region of Equation 6.20 ($r < r_\mathrm{t}$), where the mass enclosed within a radius r is given by

$$M(r) = \frac{4\pi\rho_\mathrm{t}}{3-\beta} r_\mathrm{t}^{\beta} r^{3-\beta}. \tag{6.38}$$

Using Equation 6.3, we then find the predicted offset to be

$$\boldsymbol{r}_\star = \left(a_5 \frac{\Delta G}{G_\mathrm{N}^2} \frac{3-\beta}{4\pi\rho_\mathrm{t}} r_\mathrm{t}^{-\beta} \right)^{\frac{1}{1-\beta}} \hat{a}_5, \tag{6.39}$$

and, using Equations 6.4 and 6.5, the predicted warp parameter is

$$w_1 = \frac{-\beta(3-\beta)}{(1+\beta)(2+\beta)} \frac{\Delta G}{G_\mathrm{N}^2} \frac{1}{4\pi\rho_\mathrm{t}} \frac{(3R_\mathrm{eff}/r_\mathrm{t})^{\beta}}{3R_\mathrm{eff}} \boldsymbol{a}_5 \cdot \hat{z}. \tag{6.40}$$

One may be concerned that a larger $\Delta G/G_\mathrm{N}$ would result in a smaller offset for $\beta > 1$. By considering a small perturbation about equilibrium in this case, one finds that the offset is unstable and thus the predicted signal is either zero or infinite. We therefore set $\boldsymbol{r}_\star = \boldsymbol{0}$ whenever $\beta \geq 1$. We evaluate these predicted signals for $\Delta G/G_\mathrm{N} = 1$ to create a "template" signal, which can then be multiplied by appropriate functions of $\Delta G/G_\mathrm{N}$ to obtain the corresponding prediction (see Equation 6.41 below).

6.3.4 Selection criteria

In this section we summarise the selection criteria used to obtain the samples for our inference, having justified these in the preceding sections. The fiducial values for these cuts are also shown in Table 6.2.

6.3.4.1 Warp sample

As noted in Section 6.3.1.1, the Sérsic fit converges for over 99 per cent of our galaxies, reducing the sample from 126,361 to 125,346. In Figure 6.4 we plot the distribution of ellipticity, ϵ, for these galaxies. The vertical line at $\epsilon = \epsilon_\mathrm{crit} = 0.5$ is the cut used in [9] to keep only disk-like galaxies, where we keep those with $\epsilon > 0.5$. Since the finite spatial resolution will dilate disks with scale heights below 1 kpc [539] and thus decreases their ellipticity, this criterion dramatically reduces our sample to 2,990. A further 273 galaxies are removed for having stellar masses below $2 \times 10^9\,\mathrm{M}_\odot$ and 47 more do not have finite warp values due to having columns of zero intensity in their mock images. 22 of these galaxies do not have associated halos. 54 of the remaining

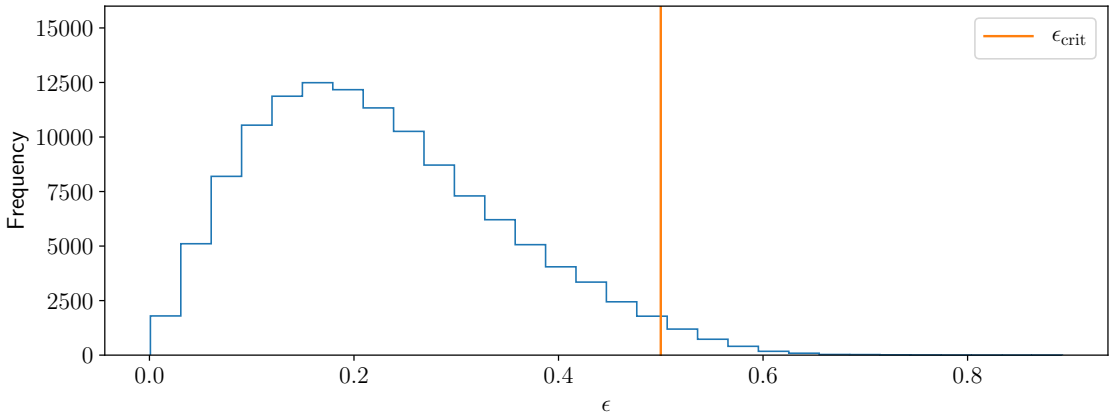


Figure 6.4: The distribution in ellipticity, ϵ, of the galaxies in H-AGN. The vertical line shows the cut at $\epsilon_{\rm crit} = 0.5$ used in [9], where only galaxies to the right of the line are used. We use the same cut here, reducing our H-AGN sample to 2,990 galaxies.

Table 6.2: The fiducial parameters used in the offset and warp inferences, as described in the text. The first six parameters are used in both inferences, whereas the next two are only used for warps, and the final four are for the offset analysis.

Parameter	**Description**	**Value**
$\lambda_{\rm c}$	Compton wavelength of fifth force field.	$\lambda_{\rm c} \in [0.4, 7.6]\,{\rm Mpc}$
Halo profile	Type of density profile used in fit.	NFW
$M_{\rm g,\,crit}$	Minimum galaxy mass.	$2 \times 10^9\,{\rm M}_\odot$
$M_{\rm h,\,crit}$	Minimum halo mass.	$8 \times 10^{10}\,{\rm M}_\odot$
$r_{\rm t}$	Where to start core.	$r_{\rm s}$
β	Power law slope of core.	0.5
$\epsilon_{\rm crit}$	Minimum ellipticity of Sérsic fit of galaxy.	0.5
$N_{\rm eff}$	Within how many $R_{\rm eff}$ to calculate warp.	3
$N_\star$	Within how many $R_{\rm eff}$ to calculate gas centre of mass.	4
$R^2_{\rm crit}$	Minimum value of R^2 for Gaussian fit to gas density to say fit has converged.	0.6
$N_{\rm g}$	Maximum number of gas length scales ($l_{\rm g}$) for which gas is associated to a galaxy.	4
$N_{\rm nn}$	Closest a nearest neighbouring galaxy can be as a multiple of the sum of their $R_{\rm eff}$.	4

galaxies have a halo mass below $8 \times 10^{10}\,\mathrm{M}_\odot$, which when rejected leaves a final sample of 2,594 galaxies. This should be compared to the 4,139 galaxies used in [364].

6.3.4.2 Offset sample

Starting with the 117,099 galaxy+halo pairs, we discard the 14,041 galaxies where the Gaussian fit to the surrounding gas does not converge, as detailed in Section 6.3.1.2. The gas–star offset is greater than $N_\mathrm{g} l_\mathrm{g}$ for 430 galaxies, where $N_\mathrm{g} = 4$, so these are also eliminated from the sample. 1,923 galaxies have nearest neighbours which are too close to make the gas–star offset measurement reliable and 44,514 have masses below $2 \times 10^9\,\mathrm{M}_\odot$, reducing our sample to 56,132. Imposing a minimum halo mass of $8 \times 10^{10}\,\mathrm{M}_\odot$ leaves a final sample of 38,042 galaxies, to be compared to 15,634 in [364].

6.3.5 Likelihood model

Now for each galaxy, i, in the offset and warp samples, we have both an "observed" (simulated), y_i, and template, $\bar{y}_i$, signal, where $y \in \{r_{\star,\alpha}, r_{\star,\delta}, w_1\}$. The likelihood for the observed signal is then

$$\mathcal{L}_i\left(y_i|\Delta G/G_\mathrm{N}, \lambda_\mathrm{c}, \sigma_{y,i}\right) = \frac{1}{\sqrt{2\pi}\sigma_{y,i}} \exp\left(-\frac{\left(y_i - \bar{y}_i\left(\Delta G/G_\mathrm{N}\right)^\tau\right)^2}{2\sigma_{y,i}^2}\right), \tag{6.41}$$

where $\tau = 1$ for the warp inference, and $\tau = (1-\beta)^{-1}$ for the offset inference. The noise parameters, $\boldsymbol{\Omega}_y = \{\sigma_{y,i}\}$, characterise the uncertainty on each of the parameters, and we use the same $\sigma_{y,i}$ for $y = r_{\star,\alpha}$ and $y = r_{\star,\delta}$. A key assumption of [364] is that these parameters are either constant for all galaxies or only depend linearly on the distance between the observer and galaxy, d_i. For our fiducial analysis we choose a constant $\boldsymbol{\Omega}_y$ for all galaxies, as the spatial resolution of H-AGN means that a physical, as opposed to angular, uncertainty is appropriate for the offset inference, and we find no systematic trend between w_1 and distance. In Section 6.6 we investigate the validity of these assumptions.

Assuming the galaxies are independent, we find the likelihood of the set of observed y to be

$$\mathcal{L}\left(y|\Delta G/G_\mathrm{N}, \lambda_\mathrm{c}, \boldsymbol{\Omega}_y\right) = \prod_i \mathcal{L}_i\left(y_i|\Delta G/G_\mathrm{N}, \lambda_\mathrm{c}, \sigma_{y,i}\right). \tag{6.42}$$

We also treat the signals as independent, such that the total likelihood of our dataset $\mathcal{D}$ is

$$\mathcal{L}\left(\mathcal{D}|\Delta G/G_{\rm N}, \lambda_{\rm c}, \mathbf{\Omega}\right) = \prod_{y} \mathcal{L}\left(y|\Delta G/G_{\rm N}, \lambda_{\rm c}, \mathbf{\Omega}_y\right), \tag{6.43}$$

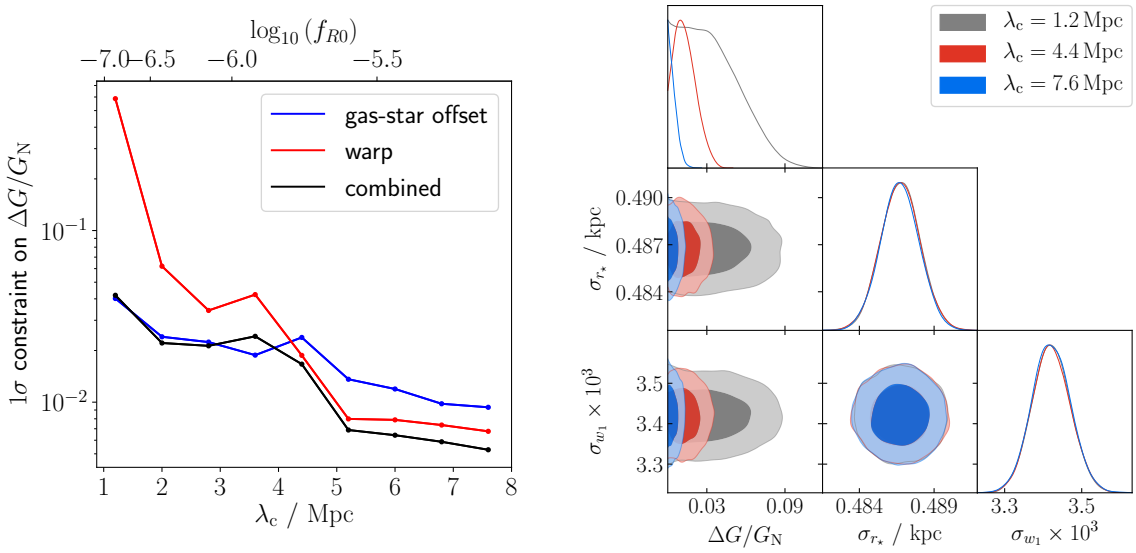
for $\mathbf{\Omega} = \{\mathbf{\Omega}_y\}$. We consider both the warp and offset samples separately and combined, so that for the above product we have three choices: $y = w_1$ (warp inference), $y \in \{r_{\star,\alpha}, r_{\star,\delta}\}$ (offset inference), and $y \in \{w_1, r_{\star,\alpha}, r_{\star,\delta}\}$ (combined inference). Finally, given some prior on ΔG, $\lambda_{\rm c}$ and $\mathbf{\Omega}$, we use Bayes' theorem (Equation 1.39) to obtain the posterior distribution of these parameters.

Imposing the priors $\Delta G/G_{\rm N} \geq 0$ and $\sigma_{y,i} > 0 \,\forall\, y, i$, and using the EMCEE sampler [62], we now derive posteriors on $\Delta G/G_{\rm N}$ and the noise model parameters at fixed $\lambda_{\rm c}$. We sample with 32 walkers and terminate the chain when the estimate of the autocorrelation length changes by less than 1 per cent per iteration and the chain is at least 50 autocorrelation lengths long.

6.4 Simulated constraints

In Figure 6.5a we plot the 1σ constraints on $\Delta G/G_{\rm N}$ as a function of $\lambda_{\rm c}$ for the warp and offset samples separately, as well as the joint constraint obtained from multiplying the likelihoods. We find the same qualitative results as [364]; $\Delta G/G_{\rm N}$ is consistent with zero, and the strength of the constraint improves with increasing $\lambda_{\rm c}$. The warp inference is weaker than the gas–star offset inference at small $\lambda_{\rm c}$ because there are far fewer galaxies in the warp sample; repeating the gas–star offset inference with the same number of galaxies as the warp sample results in comparable constraints for both signals. We find that all halos in our sample are screened for $\lambda_{\rm c} = 0.4\,{\rm Mpc}$, so we cannot achieve a constraint for this or lower $\lambda_{\rm c}$.

For $\lambda_{\rm c} = 1.2$, 4.4 and 7.6 Mpc, we plot the posterior distributions from the combined inference in Figure 6.5b. We see that the typical scale of the offsets is $\sigma_{r_\star} \sim 0.5\,{\rm kpc}$, which is approximately half the spatial resolution of H-AGN. We also see that w_1 in H-AGN is approximately five times larger than found in [364]. As previously noted, the finite spatial resolution inflates disks with scale heights below 1 kpc [539]. This leads to larger absolute fluctuations in the luminosity-weighted z position of the disk, justifying the increased w_1. It is therefore reasonable to suppose that the magnitudes of both of these noise parameters are set by the resolution of the simulation, so that these are upper limits for the true theoretical predictions. However, it is not the magnitude of the signals which we wish to determine here, but



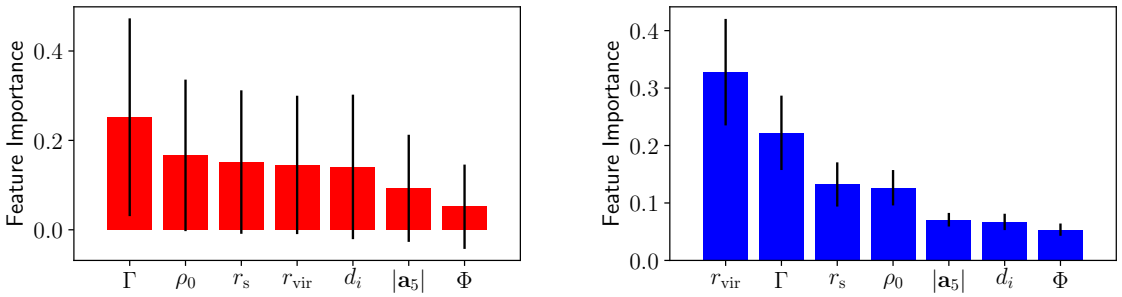
(a) (b)

Figure 6.5: (a) 1σ constraints on $\Delta G/G_{\rm N}$ as a function of $\lambda_{\rm c}$ or $n = 1$ Hu-Sawicki f_{R0} for the gas–star offset (blue) and warp (red) analyses, and their combination (black), using simulated galaxies from H-AGN. (b) Constraints on $\Delta G/G_{\rm N}$ at different $\lambda_{\rm c}$, from the combined inference of the gas–star offsets and galaxy warps in galaxies from H-AGN, along with the noise parameters $\sigma_{r_\star}$ and σ_{w_1}.

their correlations with other galaxy properties, which one would expect not to change significantly with improved resolution. Since the absolute level of noise will have an impact on the constraints in Figure 6.5b, we do not compare these directly to [364], but rather consider the relative change to our constraints if we alter the noise model.

6.5 Validity of noise model

We now are ready to address the first two questions in Section 6.2; are there unaccounted-for correlations in the noise, and, if so, do these impact the constraints? As noted in Section 6.3.5, both our fiducial analysis and [9, 364] assumed that the noise in the warp inference is a constant for all galaxies, whereas for the offset inference we assumed a constant spatial uncertainty and [363, 364] assumed a constant angular uncertainty. Assessing whether the observable is correlated with parameters used to construct the template signal due to baryonic physics (i.e. in the simulation) is analogous to asking whether the former can be predicted from the latter through some function. If not, an empirical noise model in which such correlations are absent is sufficient. Otherwise a more sophisticated noise model may be required.



(a) (b)

Figure 6.6: Feature importances for predicting (a) warps, w_1, and (b) gas–star offsets, $|\boldsymbol{r}_\star|$, in the simulation from the variables relevant to the fifth force prediction using optimised Random Forest regressors. The bars are the impurity-based feature importances (normalised so that their sum is one), and the lines give the inter-tree variability. w_1 is found to be uncorrelated with any such variable as each feature importance is consistent with zero. Conversely, the gas–star offset is correlated with several properties of the host halos.

To determine the halo parameters to use, we fit each simulated halo with a NFW and the NFW-like profile with the inner power law slope as a free parameter, and choose whichever fit minimises the BIC (Equation 1.43). We then have the characteristic density, ρ_0, scale radius, r_s, inner power law slope, Γ, and virial radius, r_{vir}, for each halo. We combine these with the distance of the galaxy from the centre of the box, d_i, screening potential, Φ, and magnitude of the fifth force field at a given λ_c, $|\boldsymbol{a}_5|$, to obtain the set of parameters that we will consider correlations of the noise with.

To determine the level of correlation, we calculate the feature importances from optimised Random Forests for the prediction of w_1 and $|\boldsymbol{r}_\star|$ from these parameters. The feature importance gives the total decrease in node impurity due to that feature, normalised so that the sum of feature importances is one. The most important features—those that correlate most strongly with the signal—have the largest feature importances, and the inter-tree variability (shown by the black lines in Figure 6.6) indicates the level of uncertainty on these values.

6.5.1 Correlation of warps with galaxy and halo properties

In Figure 6.6a we plot the feature importances for the prediction of w_1 from the parameters listed above that are used to make the template signal. We see that all features are equally (un)important and find that the regressor is unable to predict

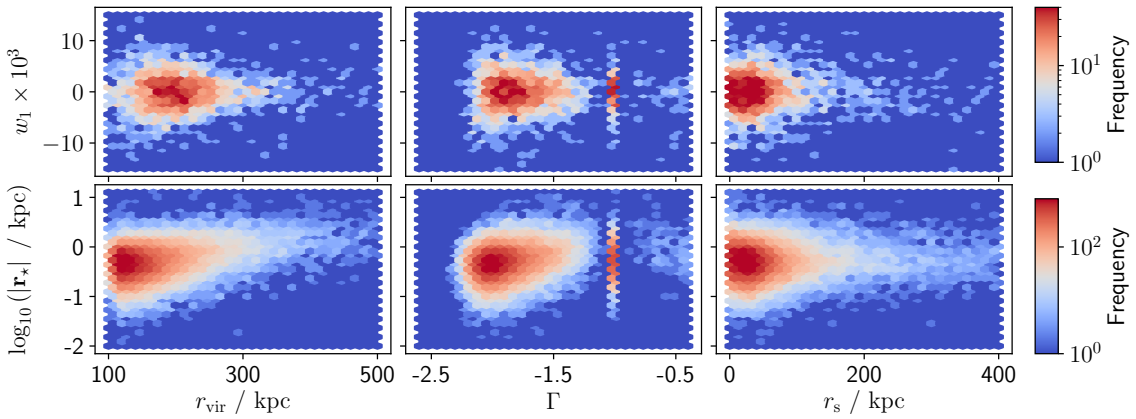


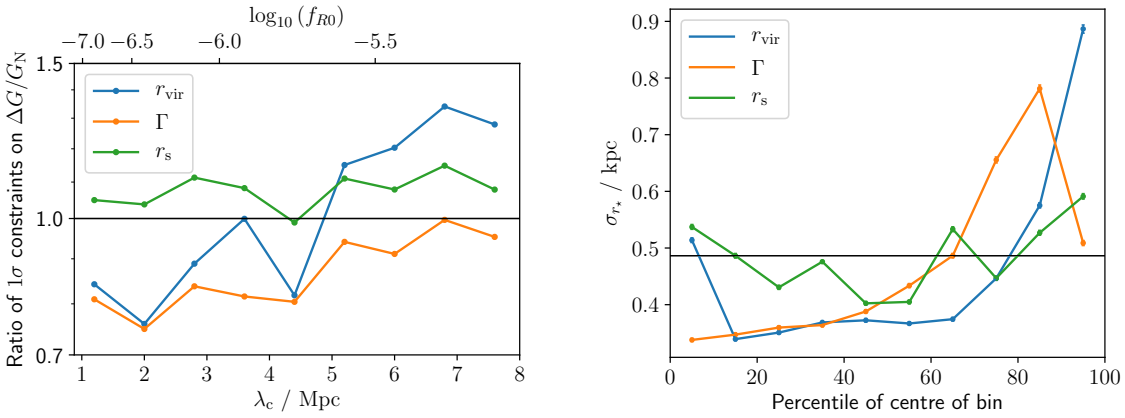
Figure 6.7: Two-dimensional histograms of the simulated signals and the most important features, as given in Figure 6.6, for determining the gas–star offset in H-AGN. We see that these features have little correlation with the warp statistic, but there is a clear correlation between $r_{\rm vir}$ and the magnitude of the gas–star offset. The high-intensity band at $\Gamma = -1$ contains the galaxies whose halos prefer a NFW fit over the more general profile according to the Bayesian Information Criterion.

w_1 reliably, with a cross-validated score of 0.02. This is also evident in the two-dimensional histograms plotted in Figure 6.7, where we see little correlation between w_1 and the parameters considered. It is therefore appropriate to assume uncorrelated noise in the inference, justifying the model of [9, 364].

6.5.2 Correlation of offsets with galaxy and halo properties

The case of gas–star offsets is more interesting, as Figure 6.6b indicates that the properties of the halo are important in predicting the measured value, and indeed we obtain a higher cross-validated score of 0.45. We find that the two most important features are $r_{\rm vir}$ and Γ, and the correlations of the signals with these parameters are clearly visible in the two-dimensional histograms of Figure 6.7. The relationship between $r_{\rm vir}$ and $|\boldsymbol{r}_\star|$ is not linear, indicating that the offset does not solely arise via scale-invariant processes.

To determine whether any correlation with these parameters affects our constraint, we now allow $\sigma_{r_\star}$ to vary with one of these parameters, which we denote p. We sort our galaxies into bins of increasing p such that each bin contains the same number of members, except in the case $p = \Gamma$, where we have one bin which is larger, containing all galaxies that are best-fit by NFW profiles ($\Gamma = -1$). We repeat our inference with a universal $\Delta G/G_{\rm N}$, but with a different $\sigma_{r_\star}$ for each bin, and find that the fitted $\sigma_{r_\star}$ are independent of both $\Delta G/G_{\rm N}$ and $\lambda_{\rm c}$.



(a) (b)

Figure 6.8: (a) Ratio of the 1σ constraints on $\Delta G/G_{\rm N}$ from the gas–star offset inference between allowing the noise parameters to vary with the halo properties vs using a single noise parameter for all galaxies. A value less than one indicates that the constraint tightens when using varying $\sigma_{r_\star}$. (b) Noise parameters, $\sigma_{r_\star}$, as a function of bin number for $\lambda_{\rm c} = 4.4\,{\rm Mpc}$, for the same bins used in Figure 6.8a. The horizontal line indicates the value obtained when a single $\sigma_{r_\star}$ is used for all galaxies. The constraint changes by $\lesssim 30$ per cent in all cases, weakening at large $\lambda_{\rm c}$ when we bin in $r_{\rm vir}$ due to the increase in $\sigma_{r_\star}$ for the largest halos.

In Figure 6.8a we compute the change in our constraints as a function of $\lambda_{\rm c}$ for 10 bins in $r_{\rm vir}$, Γ or $r_{\rm s}$, where we compare to the constraint with a single $\sigma_{r_\star}$. Binning in any other variable produces curves similar to those of $r_{\rm s}$ and Γ. We find that allowing $\sigma_{r_\star}$ to vary can either tighten, by up to ~ 25 per cent, or weaken, but by no more than ~ 30 per cent, the constraint on $\Delta G/G_{\rm N}$. It is interesting that the constraint tightens at the smallest $\lambda_{\rm c}$ when using the more sophisticated noise model; this is the region that probes that weakest fifth force and hence sets the bound on e.g. f_{R0}. We have repeated this binning procedure with only 5 bins and obtain similar results, indicating that our discretisation is sufficiently fine to capture the variation of $\sigma_{r_\star}$ with these parameters.

As one would expect, the most dramatic change to our constraint occurs when we bin in the most highly correlated property, $r_{\rm vir}$. To understand the behaviour of the constraint in this case, we plot $\sigma_{r_\star}$ as a function of bin number in Figure 6.8b. When we bin in $r_{\rm vir}$, we find that for the majority of our bins, we obtain a smaller $\sigma_{r_\star}$ than our fiducial likelihood model, whereas the positive correlation between $r_{\rm vir}$ and the observed offset causes an increased $\sigma_{r_\star}$ for the largest halos.

We now look at the effect each $r_{\rm vir}$ bin has on our constraint. We consider the

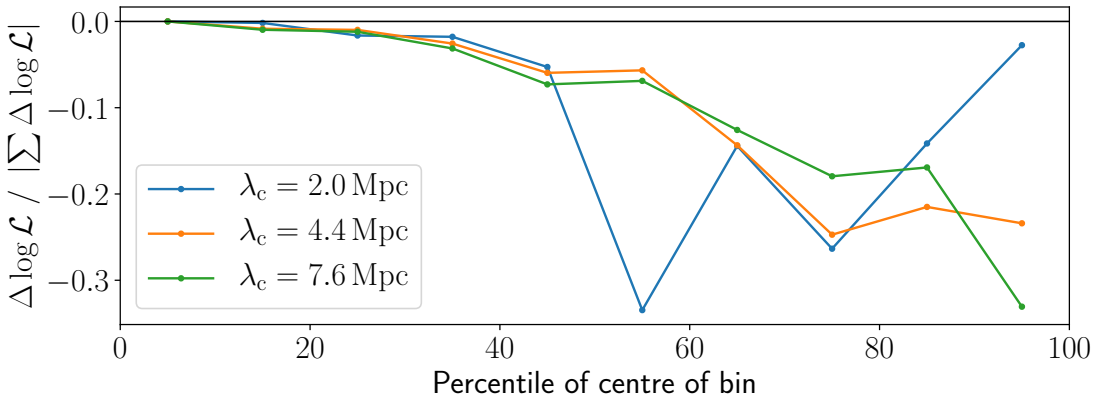


Figure 6.9: The summed change in log-likelihood, $\Delta \log \mathcal{L}$, in bins of $r_{\rm vir}$ of equal size between $\Delta G/G_{\rm N} = 0$ and the 1σ constraint, for the offset analysis with a universal noise parameter. A negative $\Delta \log \mathcal{L}$ indicates a preference for $\Delta G/G_{\rm N} = 0$. We normalise by the total change of $\log \mathcal{L}$ to show the relative contribution of each bin. For large values of $\lambda_{\rm c}$ the constraint is dominated by the largest $r_{\rm vir}$, but for smaller $\lambda_{\rm c}$ the galaxies with the largest halos are more likely to be screened, so that halos of intermediate $r_{\rm vir}$ are primarily responsible for the constraint.

change in the sum of the log-likelihood for all galaxies in a bin, $\Delta \log \mathcal{L}$, between the 1σ constraint on $\Delta G/G_{\rm N}$ and $\Delta G/G_{\rm N} = 0$, where we use a single $\sigma_{r_\star}$ for all bins, and set this to the maximum likelihood value. We plot the variation of $\Delta \log \mathcal{L}$ with bin number in Figure 6.9 for different $\lambda_{\rm c}$. For larger values of $\lambda_{\rm c}$, we see that $|\Delta \log \mathcal{L}|$ is largest for the biggest halos, i.e. our fiducial constraint is driven by the highest $r_{\rm vir}$ bins. Since these bins acquire a larger $\sigma_{r_\star}$ when we allow this to vary with $r_{\rm vir}$, the increased noise allows larger predicted signals for these galaxies. This reduces their constraining power, and hence our total constraint is weakened. For smaller values of $\lambda_{\rm c}$ the contribution is driven by intermediate $r_{\rm vir}$, since a higher fraction of the galaxies are screened in the largest bins. Now we have the opposite case, where the noise is reduced in the bins which dominate and thus we are able to achieve tighter constraints.

As will be shown in Section 6.6, the change in the constraint due to using this more complicated noise model is less than the systematic uncertainty due to the assumed halo density profile. This suggests that the simplified model used in [9, 363, 364] is adequate given other uncertainties in the model.

By choosing just one parameter to bin in, we neglect the covariance between the parameters; it is possible that a better noise model could be constructed through a multi-dimensional binning procedure. The method we outline is easily generalis-

able to higher dimensions, and we have confirmed that considering two parameters simultaneously for our case study yields similar results to just using one.

So far we have assumed that the simulated signals can be measured with perfect angular resolution. As this is not the case observationally, it is important to assess how our results are affected by the addition of a realistic angular uncertainty. To do this, we place the observer at the corner of the simulation volume to match more closely the distribution of distances of the ALFALFA survey. We remove one galaxy which is closer to the observer than 4 Mpc (the closest any galaxy is to us in [364]). We add a random angular displacement to the gas–star offset of each galaxy, drawn from a Gaussian of width $\sigma_{\rm obs}$. We consider two cases: 1) $\sigma_{\rm obs} = 18''$ to mimic the ALFALFA survey and thus the inference of [363, 364], and 2) $\sigma_{\rm obs} = 0\farcs1$, as will be achievable with the 'mid' configuration of SKA1 [540, 541]. Now we must fit for both an angular, $\sigma_{\rm obs}$, and intrinsic, $\sigma_{\rm int}$, noise component, such that the appropriate $\sigma_{r_\star}$ for galaxy i at distance d_i is

$$\sigma_{r_\star,i} = \sqrt{\sigma_{\rm int}^2 + (\sigma_{\rm obs} d_i)^2}. \tag{6.44}$$

We now re-run the inference with this additional, random angular offset 15 times for each $\lambda_{\rm c}$, fitting in each case for a universal $\Delta G/G_{\rm N}$ and $\sigma_{\rm obs}$, and either a universal $\sigma_{\rm int}$ or a different $\sigma_{\rm int}$ for each of 5 bins in $r_{\rm vir}$. The change in the constraint from using a $r_{\rm vir}$-dependent $\sigma_{\rm int}$ is shown in Figure 6.10. We see that for both the ALFALFA and SKA resolution, the mean constraint changes by $\lesssim 30$ per cent for all $\lambda_{\rm c}$.

It is unsurprising that the ALFALFA-like constraints are insensitive to whether we bin $\sigma_{\rm int}$ or not, with the 1σ regions of the uncertainty on the constraint overlapping for all $\lambda_{\rm c}$. An $18''$ angular offset at the mean distance from the observer (128 Mpc) corresponds to 11 kpc, so the angular offset dominates the intrinsic contribution of $\sim$ 0.5 kpc. One would expect that changing the model of the subdominant contribution to the noise would have little impact on the constraint, as indeed we find.

For the SKA resolution we are in the opposite regime: a 1σ angular offset now corresponds to only 60 pc at the mean distance to a galaxy, and thus our results are dominated by the intrinsic component. If SKA data were used in this test, to test the noise model it would therefore be more appropriate to convolve the observations with the simulation's resolution rather than vice versa, as we do in this section. This results in a smaller sample variance than with the ALFALFA resolution, and a practically identical variation of the change of the constraint with $\lambda_{\rm c}$ as in Figure 6.8a. Now the 1σ uncertainties on the constraints from multiple runs do not overlap, but again the constraints change by less than 30 per cent for all $\lambda_{\rm c}$.

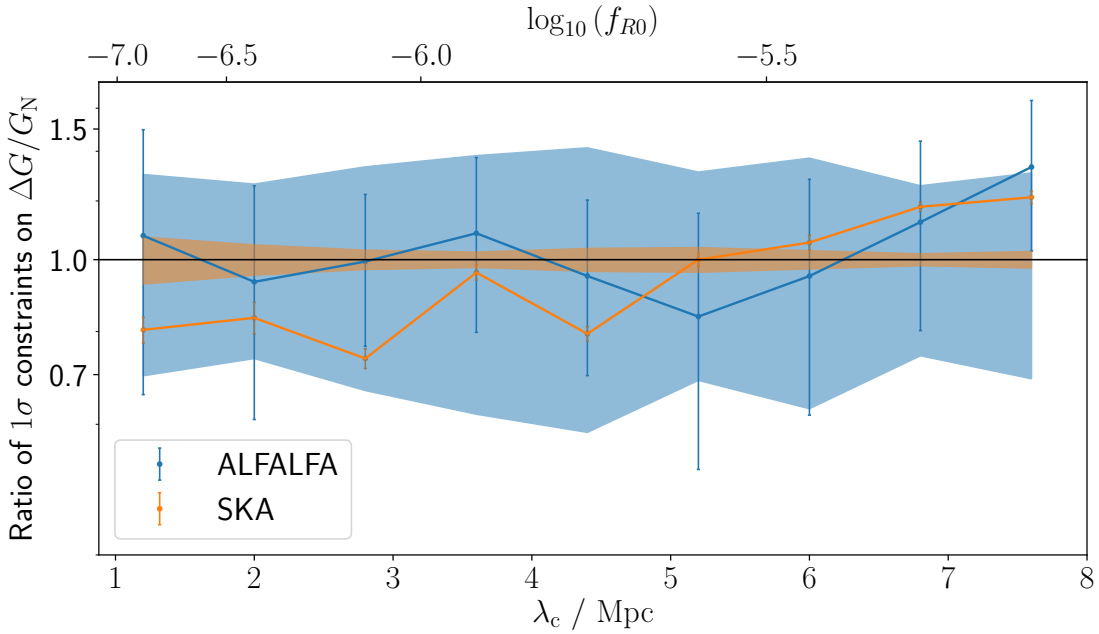


Figure 6.10: Same as Figure 6.8a, but here we add an additional random angular offset to each galaxy at either the ALFALFA (18″) or SKA (0.″1) resolution. Our binned noise model has a universal angular uncertainty combined with a different intrinsic contribution for each of 5 bins in $r_{\rm vir}$, where each bin contains the same number of galaxies. We run the un-binned and binned analyses 15 times each, and the standard deviation of the constraint between these is given by the shaded region and error bars respectively.

The uncertainty on the constraint is actually larger than the variation across these realisations, as will be discussed in Section 6.6. Given this, and that the intrinsic contribution to the offset here is expected to be an upper limit set by the simulation's resolution, we conclude that the assumption of uncorrelated Gaussian noise is justified given the other potential systematic uncertainties in the inference at both the ALFALFA and SKA resolution. This is to be expected, as we previously found this to be true even in the absence of an angular uncertainty.

6.6 Validity of halo density model

We now wish to see how the assumed form of the halo density profile affects our constraints, a test afforded by the full DM distribution available in the simulation. We consider the change in the constraint from assuming $\beta = 0.5$ with the NFW properties to using Γ as the inner power law index alongside the parameters from the more general NFW-like profile. To isolate the impact of the halo density profile from the number of galaxies used to make the constraint, when making this comparison we only use galaxies which have a non-zero template signal with both parametrisations. The resulting changes in the constraints are plotted in Figure 6.11. From this we see that, as in [364], the constraints from the warp analysis vary by less than a factor of ~ 2 for most λ_c, although we do note that the constraint tends to be weaker when using the more general density profile. The gas–star offset analysis appears to be more strongly affected by the assumed density model, with constraints up to an order of magnitude weaker with the more general density profile.

The conclusion to draw from this analysis is conditioned on the reliability of the halo density profiles in H-AGN. The inner slope of the halo is determined by the balance between adiabatic contraction [380, 381], which steepens the profile, and stellar or AGN feedback, which promotes core formation [243, 382]. While redshift-zero halos in H-AGN are found to be steeper than NFW [538], most observational evidence favours shallower profiles (e.g. [384, 385]), a manifestation of the cusp-core problem of ΛCDM [383]. It is unclear how alternative feedback prescriptions would alter the constraints in Figure 6.11, although we note that by considering only halos with non-zero template signals we have removed all those with $\Gamma < -1$ (Section 6.3.3.2).

Due to this issue, we cannot a draw a definitive conclusion about the validity of the halo density model used in [364]. We can however say that, for the H-AGN simulation, variations in the halo density model can result in a weakening of the constraint of $\Delta G/G_N$ from gas–star offsets by up to a factor ~ 10, and galaxy warps

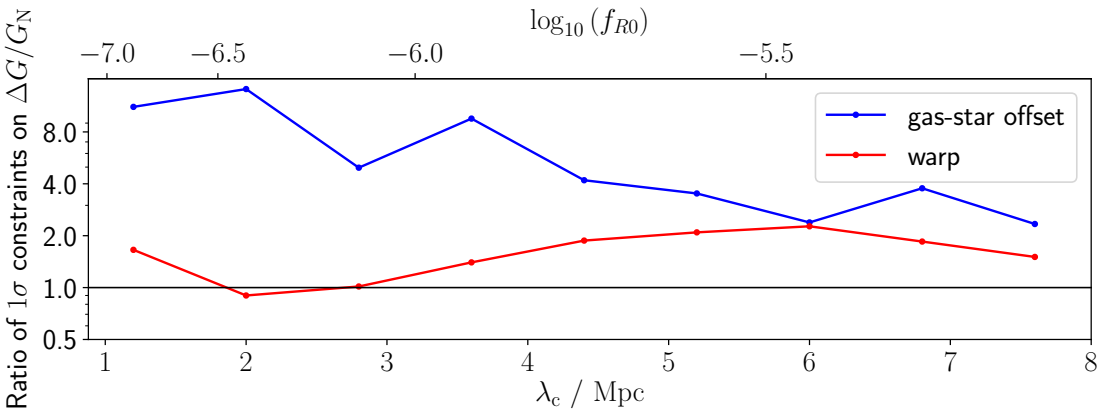


Figure 6.11: Ratio of the 1σ constraints on $\Delta G/G_N$ between assuming an inner power law of slope $\beta = 0.5$ inside the scale radius of a NFW profile, and using a more general NFW-like profile, where we explicitly fit for the inner power law slope. A value greater than one indicates a weaker constraint when using the more general density profile.

by up to a factor ~ 2. Although this is consistent with the discussion of systematic uncertainties in [364], different simulations or different halo density profiles could lead to different conclusions. It is therefore important for future work to investigate other cosmological hydrodynamical simulations with different implementations of galaxy formation physics, to determine whether this result is robust to plausible variations in the DM distribution.

For this specific example, we note that it is not the individual constraints that are important but rather the combined constraint of the two analyses. From figure 1 of [364], we see that the combined constraint on $\Delta G/G_N$ would hardly change if we removed the gas–star offset analysis and only used galaxy warps. The sensitivity of the gas–star offset inference to the halo density profile is therefore of little consequence to the conclusions of that work, which are driven mainly by the warp analysis.

6.7 Discussion and Conclusions

Galactic scale tests are capable of providing powerful constraints on new fundamental physics across a range of previously under-explored environments. Their drawback is that one must accurately model the messy astrophysical regime to capture fully, and break degeneracies with, baryonic effects that are less important in more traditional analyses.

We propose three main tests to gauge the robustness of a given model to baryonic physics:

1. Determine whether there are any unaccounted-for correlations between the target observable (signal) and galactic properties due to baryonic physics (noise). To do this, we ask whether the signal in a cosmological hydrodynamical simulation without new physics can be predicted from the parameters relevant to the new physics model. Since the functional form of any such correlation is unknown a priori, this is best addressed in a machine-learning context.

2. If one or more parameter is found to correlate with the simulated signal, then the impact of this on the new physics constraint must be established. We do this by explicitly constructing a more sophisticated noise model and repeating the inference. If the constraint changes by less than some specified tolerance then it can be concluded that the simplified noise model is satisfactory; otherwise it should be replaced with the more complex one.

3. Use the extra information available in the simulation compared to observations to assess the adequacy of the modelling of unobservable properties. With the simulated data one can compare the constraints obtained using the "true" vs model parameters to quantify the model's suitability, and improve it if necessary.

As a case study, we investigate the morphological signatures used in [364] to constrain thin-shell-screened fifth forces and rule out astrophysically relevant Hu-Sawicki $f(R)$ theories, namely offsets between the stellar and gaseous components of a galaxy and warping of the stellar disk. We work in the context of the H-AGN simulation, performed in ΛCDM and hence with no fifth force. We use a machine-learning feature importance analysis to study the correlations in the simulation between the morphological signals and the parameters used to predict them in the context of modified gravity. We find that the degree of 'U'-shaped warping of the stellar disk is independent of these parameters, justifying the assumption of uncorrelated random Gaussian noise used in [9, 364]. For the gas–star offset case, we find a positive correlation between the signal and the virial radius of the host halo, $r_{\rm vir}$. To assess the impact of this previously unaccounted-for complication, we allow the width of the Gaussian noise model to vary with $r_{\rm vir}$ by binning the galaxies in $r_{\rm vir}$ and using a different width in each bin. This changes the fifth force constraint by $\lesssim 30$ per cent, which is less than the systematic uncertainty from the assumed halo density profile. This again justifies the assumption of random noise. We find that the gas–star offset inference is more sensitive to the assumed density profile than the warp analysis, and we identify this as the largest source of systematic uncertainty in the inference of [364]. It could

be reduced in the future using dynamical information on the central regions of the test galaxies.

We anticipate the methods developed here to prove useful for validating future forward models of galactic signals used to constrain fundamental physics. In this chapter, we have only considered the implementation of galaxy formation physics in the H-AGN simulation. Other simulations have different sub-grid models, calibration methods, hydrodynamical and feedback schemes and resolutions, and any analysis must be robust to these differences. It is known that small-scale predictions for the matter power spectrum are sensitive to these effects [542], so it is natural to suspect a similar sensitivity here. Future studies should therefore apply the methods we outline to different simulations and different physics tests to assess the accuracy of constraints derived from the galactic regime.

Chapter 7

Conclusions

In this thesis we developed the use of Monte Carlo-based forward models in galactic astrophysics, with applications to tests of dark matter, dark energy, and galaxy formation. By forward-modelling observational signals on a source-by-source basis, propagating uncertainties through Monte Carlo sampling, and marginalising over a model describing other astrophysical and observational processes, we harnessed the constraining power of galaxies whilst accounting for their complexity and mitigating uncertainties due to unknown astrophysics or poorly modelled astrophysical processes. Provided one can sufficiently understand the complex, non-linear processes which govern these objects, it is possible to achieve constraints on fundamental physics with small statistical uncertainties. The practically unlimited sample size that is achievable with astrophysical objects is in stark contrast to traditional large-scale cosmological probes, where one is fundamentally limited by cosmic variance. Since astrophysical noise is the dominant source on uncertainty in such analyses, it was crucial to robustly model these contributions and assess the accuracy of the modelling. This is critically lacking in many previous studies.

To develop the inference techniques required to constrain fundamental physics on astrophysical scales, in Chapter 2 we considered the energy-dependent propagation speed for photons induced by Lorentz invariance violation in QG models or a nonzero photon mass, m_γ. In such models, photons of different energies from a distant source would arrive at different times, even if they were emitted simultaneously. By developing source-by-source, Monte Carlo-based forward models for such time delays from GRBs, and marginalising over empirical noise models describing other contributions to the time delay, we derived constraints on m_γ and the QG length scale, $\ell_{\rm QG}$, using spectral lag data from the BATSE satellite. We found $m_\gamma < 4.0 \times 10^{-5}\, h\, {\rm eV}/c^2$ and $\ell_{\rm QG} < 5.3 \times 10^{-18}\, h$ GeV^{-1} at 95% confidence, and demonstrated that these constraints are robust to the choice of noise model. The QG constraint is among the

tightest from studies which consider multiple gamma ray bursts and the constraint on m_γ, although weaker than from using radio data, provides an independent constraint which is less sensitive to the effects of dispersion by electrons.

The observed positions of galaxies in the local Universe allow one to infer the initial conditions and hence the current DM density (see Section 1.3). The use of the `BORG` algorithm to achieve this produces a Markov Chain of plausible large scale structure realisations which one can use to constrain new physics. Extending the QG and photon mass analysis, we then developed probabilistic forward models for the Shapiro time delays of GRBs by combining constrained realisations of the local density field using the `BORG` algorithm with unconstrained large-scale modes. The large Shapiro time delays of extragalactic sources allowed us to put tight constraints on differences in the PPN parameter γ between photons of different frequencies from spectral lag data, since a non-zero $\Delta\gamma$ would result in a frequency-dependent arrival time. The majority of previous constraints have assumed that the Shapiro time delay is dominated by a few local massive objects, although this is a poor approximation for distant sources, whereas we considered the cosmological context of these sources. Propagating uncertainties in the density field reconstruction and marginalising over an empirical model describing other contributions to the time delay, we utilised the same spectral lag data to constrain $\Delta\gamma < 2.1 \times 10^{-15}$ at 1σ confidence between photon energies of 25 keV and 325 keV. This is a constraint on Weak Equivalence Principle violation, since γ should be the same for all particles at all energies if this is obeyed.

Our next aim was to constrain fundamental physics with the morphology of galaxies themselves, rather than just the inferred DM density fields from their spatial distribution. For this, we considered Galileons; scalar field theories which obey the Galileon symmetry $\varphi \to \varphi + b + c_\mu x^\mu$ and are capable of self-acceleration if they have an inverted sign for the kinetic term. These theories violate the Strong Equivalence Principle, such that BHs do not couple to the Galileon field, whereas non-relativistic objects experience a fifth force with strength $\Delta G/G_\mathrm{N}$ relative to gravity. For galaxies falling down a gradient in the Galileon field, this results in an offset between the centre of the galaxy and its host supermassive BH. Before attempting to constrain the coupling of these theories, in Chapter 3 we studied this phenomenon in ΛCDM to investigate whether one could build a noise model from these considerations.

We studied the displacements between the centres of galaxies and their supermassive BHs in the cosmological hydrodynamical simulation H-AGN, and in a variety of observations from the literature. The BHs in H-AGN feel a sub-grid dynamical friction force, sourced by the surrounding gas, which prevents recoiling BHs being

ejected from the galaxy. We found that the fraction of spatially offset BHs increases with cosmic time, BHs live on prograde orbits in the plane of the galaxy with an orbital radius that decays with time but stalls near $z = 0$, and the magnitudes of offsets from the galaxy centres are substantially larger in the simulation than in observations. We attributed the stalling of the infall and excessive offset magnitudes to the fact that dynamical friction from stars and DM is not modelled in the simulation, and hence provided a way to improve the black hole dynamics of future simulations. Unfortunately, the unrealistic predictions of the simulations meant that we had to rely on empirical noise models for this test.

In Chapter 4 we reconstructed the local gravitational and Galileon fields through a suite of constrained N-body simulations and developed a Monte Carlo-based forward model for these offsets on a galaxy-by-galaxy basis. Using the measured offset between the optical centre and active galactic nucleus of 1916 galaxies from the literature, propagating uncertainties in the input quantities and marginalising over an empirical noise model describing astrophysical and observational noise, we constrained the Galileon coupling to be $\Delta G/G_{\rm N} < 0.16$ at 1σ confidence for Galileons with crossover scale $r_{\rm C} \gtrsim H_0^{-1}$.

In Chapter 5 we utilised the DM density fields inferred through galaxy positions to constrain the properties of DM itself. Decaying or annihilating DM particles could be detected through gamma-ray emission from the species they decay or annihilate into. This is usually done by modelling the flux from specific DM-rich objects such the Milky Way halo, Local Group dwarfs and nearby groups. However, these objects are expected to have significant emission from baryonic processes as well, and the analyses discard gamma-ray data over most of the sky. Here we constructed full-sky templates for gamma-ray flux from the large-scale structure within ∼200 Mpc by means of the `CSiBORG` simulation suite. Marginalising over uncertainties in this reconstruction, small-scale structure and parameters describing astrophysical contributions to the observed gamma ray sky, we compared to observations from the *Fermi* Large Area Telescope to constrain DM annihilation cross-sections and decay rates. We ruled out the thermal relic cross-section for s-wave annihilation for all $m_\chi \lesssim 7\,{\rm GeV}/c^2$ at 95% confidence if the annihilation produces Z bosons, gluons or quarks less massive than the bottom quark. We inferred a contribution to the gamma ray sky with the same spatial distribution as DM decay at 3.3σ. Although this could be due to DM decay via these channels with a decay rate $\Gamma \approx 3 \times 10^{-28}\,{\rm s}^{-1}$, we found that a power-law spectrum of index $p = -2.75^{+0.71}_{-0.46}$ is preferred by the data.

When constraining DM annihilation, we assumed that all DM halos obey NFW density profiles, however we verified that our constraints are conservative compared to those obtained using Einasto profiles. Both of these profiles are calibrated against DM-only simulations, whereas ideally one would use hydrodynamic simulations which incorporate the impact of baryons. Adiabatic contraction can steepen the inner DM density profile [380, 381], which would lead to larger predicted signals in the centres of halos and thus tighter constraints on $\langle \sigma v \rangle$. However, the subsequent stellar feedback caused by this contraction can cause the halo to expand [243, 382]. This would lead to lower DM densities and thus larger values of $\langle \sigma v \rangle$ could be permitted. Since the true DM density profiles are sensitive to the exact implementation of baryonic feedback, and due to the current lack of sufficiently massive simulated halos to calibrate against, we leave a quantitative study of the impact of baryonic processes on our constraints to future work.

The galactic scale tests outlined above proved to be powerful tools in constraining fundamental physics in previously under-explored regions of parameter space. The astrophysical regime which they probe is inherently complicated, and the inference methods used to make these constraints should be robust to baryonic effects. Previous analyses have assumed simple empirical models for astrophysical noise without detailed calibration or justification. In Chapter 6 we outlined a framework for assessing the reliability of such methods by constructing and testing more advanced baryonic models using cosmological hydrodynamical simulations. As a case study, we use the H-AGN simulation to investigate warping of stellar disks and offsets between gas and stars within galaxies, which are powerful probes of screened fifth forces. We showed that the degree of 'U'-shaped warping of galaxies is well modelled by Gaussian random noise, but that the magnitude of the gas–star offset is correlated with the virial radius of the host halo. By incorporating this correlation we confirmed recent results ruling out astrophysically relevant Hu-Sawicki $f(R)$ gravity, and identify a $\sim 30\%$ systematic uncertainty due to baryonic physics. Such an analysis must be performed case-by-case for future galactic tests of fundamental physics.

The statistical, forward-modelling techniques developed in this thesis have allowed the field of astrophysical tests of fundamental physics to become a powerful probe of how the Universe works. Through robust characterisation and assessment of astrophysical and observational noise, we have opened up an uncharted region of parameter space to answer some of the most important questions about our Universe, such as the nature of dark energy, dark matter, and gravity.

As we move into an era of increased observational power, we will probe smaller scales and thus non-Gaussian processes will become the dominant source of uncertainty and limitation to extracting information. Although in Chapter 6 we developed a framework to assess the reliability of models designed to capture this noise by constructing and testing more advanced baryonic models using cosmological hydrodynamical simulations, it would be desirable to produce these *a priori*, as opposed to justifying their validity *a posteriori*. Utilising machine learning techniques such as Bayesian Neural Networks to determine these relations, and training on a variety of cosmological hydrodynamical simulations to ensure the inference is robust to this choice, future work should incorporate machine learning into noise models and hence improve their sophistication. When constructing these noise models, one can not only improve the robustness of the results, but there is the opportunity to learn about the processes that govern galactic formation and evolution.

With the next generation of galaxy surveys – such as *Euclid* [217] or the Rubin Observatory [218] – just around the corner, we will soon have high quality data for billions of galaxies. The galactic-scale tests of fundamental physics performed in this thesis typically involved a few thousand galaxies. Future work should therefore improve the scalability of these methods, whilst still treating galaxies as more than point-like objects. This will involve adapting techniques such as DELFI to these novel probes of fundamental physics. An increased level of automation in these simulation-based inferences – for example by designing convergence criteria for the Monte Carlo sampling – will reduce the computational cost of a given analysis, allowing more tests to be run even as the sophistication of the inference increases.

A key theme throughout this thesis has been to constrain fundamental physics without using two-point statistics. Instead, objects (or pixels in Chapter 5) were forward modelled individually and directly compared to the data, which allowed us not only to capture the two-point statistics, but implicitly all higher orders too. This is advantageous since we minimise the amount of compression performed on the data and hence retain maximal constraining power. Moreover, we know that many phenomena we wish to study are non-Gaussian (e.g. the present-day matter density field is not a Gaussian Random Field, as demonstrated by the filamentary structures), so a two-point statistic does not fully describe the statistical properties of the data in these scenarios. The themes and methods discussed in this work should therefore be applied to more traditional cosmological probes, with the expectation that these will produce tighter constraints than resorting to power-spectrum analyses, as has already been shown by [79–81]. Importantly, we have developed the requisite methodologies to

propagate uncertainties and noise within this object-by-object (or field-level) inference framework and assess their appropriateness, which could naturally be incorporated in an inference pipeline.

Although primarily applied to the astrophysical regime in this thesis, the techniques introduced were designed to detect correlations in extremely noisy data with potentially unknown noise characteristics. The techniques developed can therefore be applied to any situation when one wishes to analyse a population for which one has a large amount of (potentially uncertain) data per member, making this approach applicable to other areas of astronomy. Future applications of these methods to maximally extract information from under-explored regions of parameter space and to more traditionally studied phenomena will be invaluable in the attempt to learn how galaxies form and evolve, and uncover the fundamental physics of the Universe. In achieving this goal, there is the tantalising opportunity to obtain competitive constraints on fundamental physics in a completely new way.

Appendix A

Numerical implementation of dark matter likelihood

When evaluating the likelihood in Equation 5.44, we often care about small fractional errors on the model, such that the parameter

$$x \equiv \frac{\left(\mu - \sigma^2\right)^2}{2\sigma^2} \tag{A.1}$$

is large, where we have omitted indices for clarity. We therefore must evaluate ${}_1F_1(a, c, x)$ for $x \gg 1$, although this is problematic since ${}_1F_1$ has the asymptotic expansion [543]

$${}_1F_1\left(a, c, x\right) \sim \frac{\Gamma\left(c\right)}{\Gamma\left(a\right)} e^x x^{a-c}, \tag{A.2}$$

and thus diverges as $x \to \infty$. We therefore define the function

$$f\left(a, c, x\right) \equiv e^{-x} x^{c-a} {}_1F_1\left(a, c, x\right), \tag{A.3}$$

which is finite as $x \to \infty$. Thus the likelihood for a single pixel becomes

$$\begin{aligned} \mathcal{L}\left(n\right) = \frac{\left(\mu - \sigma^2\right)^n}{\sqrt{\pi} n!} \exp\left(\frac{\sigma^2}{2} - \mu\right) \left(1 + \operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\right)^{-1} \\ \times \left(\Gamma\left(\frac{1+n}{2}\right) g_n\left(x\right) + 2\Gamma\left(1 + \frac{n}{2}\right) h_n\left(x\right)\right), \end{aligned} \tag{A.4}$$

where we have defined the functions

$$g_n\left(x\right) \equiv f\left(\frac{1+n}{2}, \frac{1}{2}, x\right), \tag{A.5}$$

and

$$h_n\left(x\right) \equiv f\left(1 + \frac{n}{2}, \frac{3}{2}, x\right). \tag{A.6}$$

Given the asymptotic expansion of ${}_1F_1$, it is clear that Equation A.4 is equivalent to the Poisson distribution in the limit $\sigma \to 0$ at fixed μ.

If n is even, then $g_n(x)$ can we expressed as

$$g_n(x) = \sum_{m=0}^{n/2} \alpha_m^{(n)} x^{-m}, \tag{A.7}$$

where the coefficients obey the recurrence relation

$$\alpha_m^{(n+2)} = \frac{2}{n+1}\left(\left(n - m + \frac{3}{2}\right)\alpha_{m-1}^{(n)} + \alpha_m^{(n)}\right), \tag{A.8}$$

and

$$\alpha_0^{(0)} = 1. \tag{A.9}$$

Similarly, if n is odd, then

$$g_n(x) = \sum_{m=0}^{(n-1)/2} \beta_m^{(n)} x^{-m}, \tag{A.10}$$

for

$$\beta_m^{(n+2)} = \frac{2}{n+2}\left(\left(n - m + \frac{3}{2}\right)\beta_{m-1}^{(n)} + \beta_m^{(n)}\right), \tag{A.11}$$

and

$$\beta_0^{(1)} = 1. \tag{A.12}$$

For odd n, the function $g_n(x)$ has a slightly more complicated expression

$$g_n(x) = \frac{e^{-x}}{\sqrt{x}} \sum_{m=0}^{(n-1)/2} \gamma_m^{(n)} x^{-m} + \sqrt{\pi}\,\mathrm{erf}\left(\sqrt{x}\right) \sum_{m=0}^{(n-1)/2} \delta_m^{(n)} x^{-m}, \tag{A.13}$$

where

$$\gamma_m^{(n+2)} = \frac{2}{n+1}\left(\delta_m^{(n)} + (n - m + 1)\,\gamma_{m-1}^{(n)}\right), \tag{A.14}$$

$$\delta_m^{(n+2)} = \frac{2}{n+1}\left(\delta_m^{(n)} + \left(n - m + \frac{3}{2}\right)\delta_{m-1}^{(n)}\right), \tag{A.15}$$

$$\gamma_0^{(1)} = \delta_0^{(1)} = 1. \tag{A.16}$$

Finally, the function $h_n(x)$ can be expressed in a similar form for even n,

$$h_n(x) = \frac{e^{-x}}{\sqrt{x}} \sum_{m=0}^{n/2} \epsilon_m^{(n)} x^{-m} + \sqrt{\pi}\,\mathrm{erf}\left(\sqrt{x}\right) \sum_{m=0}^{n/2} \zeta_m^{(n)} x^{-m}, \tag{A.17}$$

where

$$\epsilon_m^{(n+2)} = \frac{2}{n+2}\left(\zeta_m^{(n)} + (n-m+1)\,\epsilon_{m-1}^{(n)}\right), \tag{A.18}$$

$$\zeta_m^{(n+2)} = \frac{2}{n+2}\left(\zeta_m^{(n)} + \left(n-m+\frac{3}{2}\right)\zeta_{m-1}^{(n)}\right), \tag{A.19}$$

$$\epsilon_0^{(0)} = 0, \tag{A.20}$$

$$\zeta_0^{(0)} = \frac{1}{2}. \tag{A.21}$$

Whenever $x > 1$, we evaluate the likelihood using Equation A.4 and the series expansions for $g_n(x)$ and $h_n(x)$ given above. Otherwise, the likelihood is evaluated using Equation 5.44 directly, using the usual series expansion for ${}_1F_1$

$${}_1F_1\left(a, c, x\right) = \sum_{m=0}^{\infty} \frac{(a)_m}{(c)_m}\frac{x^m}{m!}. \tag{A.22}$$

References

[1] J. Khoury, arXiv e-prints , arXiv:1312.2006 (2013).

[2] J. Sakstein, B. Jain, J. S. Heyl, and L. Hui, ApJ **844**, L14 (2017).

[3] A. Asvathaman, J. S. Heyl, and L. Hui, MNRAS **465**, 3261 (2017).

[4] G. Lavaux and J. Jasche, MNRAS **455**, 3169 (2016).

[5] J. Jasche and G. Lavaux, A&A **625**, A64 (2019).

[6] G. Steigman, B. Dasgupta, and J. F. Beacom, Phys. Rev. D **86**, 023506 (2012).

[7] A. Albert *et al.*, ApJ **834**, 110 (2017).

[8] D. Hashimoto *et al.*, arXiv e-prints , arXiv:2202.01400 (2022).

[9] H. Desmond *et al.*, Phys. Rev. D **98**, 083010 (2018).

[10] K. Nordtvedt, Phys. Rev. **170**, 1186 (1968).

[11] J. Murphy, T. W. *et al.*, Classical and Quantum Gravity **29**, 184005 (2012).

[12] Planck Collaboration, A&A **641**, A6 (2020).

[13] The LIGO Scientific Collaboration, Phys. Rev. D **103**, 122002 (2021).

[14] A. Friedman, Zeitschrift für Physik **10**, 377 (1922).

[15] A. Friedmann, Zeitschrift für Physik **21**, 326 (1924).

[16] G. Lemaître, MNRAS **91**, 483 (1931).

[17] H. P. Robertson, ApJ **82**, 284 (1935).

[18] H. P. Robertson, ApJ **83**, 187 (1936).

[19] H. P. Robertson, ApJ **83**, 257 (1936).

[20] A. G. Walker, Proceedings of the London Mathematical Society **42**, 90 (1937).

[21] A. G. Riess *et al.*, AJ **116**, 1009 (1998).

[22] S. Perlmutter *et al.*, ApJ **517**, 565 (1999).

[23] J. A. Frieman, M. S. Turner, and D. Huterer, ARA&A **46**, 385 (2008).

[24] E. M. Lifshitz, Zhurnal Eksperimentalnoi i Teoreticheskoi Fiziki **16**, 587 (1946).

[25] V. Mukhanov *et al.*, Physics Reports **215**, 203 (1992).

[26] C.-P. Ma and E. Bertschinger, ApJ **455**, 7 (1995).

[27] V. C. Rubin, J. Ford, W. K., and N. Thonnard, ApJ **238**, 471 (1980).

[28] D. Collaboration, Phys. Rev. D **105**, 023520 (2022).

[29] E. Di Valentino *et al.*, Astroparticle Physics **131**, 102605 (2021).
[30] E. Di Valentino *et al.*, Astroparticle Physics **131**, 102604 (2021).
[31] J. S. Bullock and M. Boylan-Kolchin, Ann. Rev. Astron. Astrophys. **55**, 343 (2017).
[32] A. Del Popolo and M. Le Delliou, Galaxies **5**, 17 (2017).
[33] A. Padilla, arXiv e-prints , arXiv:1502.05296 (2015).
[34] S. Weinberg, Phys. Rev. **138**, B988 (1965).
[35] A. Joyce *et al.*, Ann. Rev. Nucl. Part. Sci. **66**, 95 (2016).
[36] T. Baker *et al.*, Rev. Mod. Phys. **93**, 015003 (2021).
[37] J. Solà Peracaula *et al.*, ApJ **886**, L6 (2019).
[38] N. Frusciante *et al.*, Phys. Rev. D **101**, 064001 (2020).
[39] H. Desmond, B. Jain, and J. Sakstein, Phys. Rev. D **100**, 043537 (2019), [Erratum: Phys.Rev.D 101, 069904 (2020), Erratum: Phys.Rev.D 101, 129901 (2020)].
[40] H. Desmond and J. Sakstein, Phys. Rev. D **102**, 023007 (2020).
[41] B. Jain and J. VanderPlas, JCAP **2011**, 032 (2011).
[42] P. Brax *et al.*, arXiv e-prints , arXiv:2201.10817 (2022).
[43] C. Burrage and J. Sakstein, Living Reviews in Relativity **21**, 1 (2018).
[44] B. Jain and J. Khoury, Annals of Physics **325**, 1479 (2010).
[45] A. Joyce, B. Jain, J. Khoury, and M. Trodden, Phys. Rep. **568**, 1 (2015).
[46] J. Khoury, arXiv e-prints , arXiv:1011.5909 (2010).
[47] E. Babichev *et al.*, International Journal of Modern Physics D **18**, 2147 (2009).
[48] A. I. Vainshtein, Physics Letters B **39**, 393 (1972).
[49] J. Khoury and A. Weltman, Phys. Rev. Lett. **93**, 171104 (2004).
[50] J. Khoury and A. Weltman, Phys. Rev. D **69**, 044026 (2004).
[51] K. Hinterbichler and J. Khoury, Phys. Rev. Lett. **104**, 231301 (2010).
[52] P. Brax *et al.*, Phys. Rev. D **82**, 063519 (2010).
[53] J. Sakstein, H. Desmond, and B. Jain, Phys. Rev. D **100**, 104035 (2019).
[54] G. Steigman and M. S. Turner, Nuclear Physics B **253**, 375 (1985).
[55] M. Kamionkowski, in *High Energy Physics and Cosmology, 1997 Summer School*, Vol. 14 (1998) p. 394.
[56] G. Bertone, D. Hooper, and J. Silk, Phys. Rep. **405**, 279 (2005).
[57] T. R. Slatyer, arXiv e-prints , arXiv:2109.02696 (2021).
[58] G. Schwarz, Annals of Statistics **6**, 461 (1978).
[59] W. Handley and P. Lemos, Phys. Rev. D **100**, 023512 (2019).
[60] W. Handley, The Journal of Open Source Software **4**, 1414 (2019).

[61] R. Wilcox, in *Introduction to Robust Estimation and Hypothesis Testing (Third Edition)*, Statistical Modeling and Decision Science, edited by R. Wilcox (Academic Press, Boston, 2012) third edition ed., pp. 137–213.
[62] D. Foreman-Mackey *et al.*, PASP **125**, 306 (2013).
[63] R. M. Neal, *Probabilistic Inference Using Markov Chain Monte Carlo Methods*, Tech. Rep. CRG-TR-93-1 (University of Toronto, 1993).
[64] J. Skilling, Bayesian Analysis **1**, 833 (2006).
[65] F. Feroz and M. P. Hobson, MNRAS **384**, 449 (2008).
[66] F. Feroz, M. P. Hobson, and M. Bridges, MNRAS **398**, 1601 (2009).
[67] F. Feroz *et al.*, The Open Journal of Astrophysics **2**, 10 (2019).
[68] J. Buchner, Statistics and Computing **26**, 383 (2016).
[69] J. Buchner, Publications of the Astronomical Society of the Pacific **131**, 108005 (2019).
[70] J. Buchner, The Journal of Open Source Software **6**, 3001 (2021).
[71] J. Lintusaari *et al.*, Systematic Biology **66**, e66 (2016).
[72] M. U. Gutmann and J. Corander, Journal of Machine Learning Research **17**, 1 (2016).
[73] F. Leclercq, Phys. Rev. D **98**, 063511 (2018).
[74] F. Leclercq, W. Enzi, J. Jasche, and A. Heavens, MNRAS **490**, 4237 (2019).
[75] N. Jeffrey and B. D. Wandelt (2020) p. arXiv:2011.05991.
[76] G. Papamakarios and I. Murray, in *NIPS* (2016).
[77] J. Alsing, B. Wandelt, and S. Feeney, MNRAS **477**, 2874 (2018).
[78] J. Alsing *et al.*, MNRAS **488**, 4440 (2019).
[79] N. Porqueres *et al.*, MNRAS **509**, 3194 (2022).
[80] A. Andrews *et al.*, arXiv e-prints , arXiv:2203.08838 (2022).
[81] F. Leclercq and A. Heavens, MNRAS **506**, L85 (2021).
[82] J. Jasche and B. D. Wandelt, MNRAS **425**, 1042 (2012).
[83] J. Jasche and B. D. Wandelt, MNRAS **432**, 894 (2013).
[84] J. Jasche *et al.*, MNRAS **406**, 60 (2010).
[85] J. Jasche, F. Leclercq, and B. D. Wandelt, JCAP **2015**, 036 (2015).
[86] G. Lavaux, J. Jasche, and F. Leclercq, arXiv e-prints , arXiv:1909.06396 (2019).
[87] A. Nusser and A. Dekel, ApJ **391**, 443 (1992).
[88] M. Crocce and R. Scoccimarro, Phys. Rev. D **73**, 063520 (2006).
[89] S. Tassev, M. Zaldarriaga, and D. J. Eisenstein, JCAP **2013**, 036 (2013).
[90] D. J. Bartlett *et al.*, Phys. Rev. D **103**, 023523 (2021).
[91] H. Desmond *et al.*, MNRAS **511**, L45 (2022).

[92] M. L. Hutt *et al.*, arXiv e-prints , arXiv:2203.14724 (2022).
[93] M. L. Hutt *et al.*, `https://doi.org/10.5281/zenodo.5851241` (2022).
[94] G. Lavaux and M. J. Hudson, MNRAS **416**, 2840 (2011).
[95] R. Teyssier, A&A **385**, 337 (2002).
[96] T. Baker, D. Psaltis, and C. Skordis, ApJ **802**, 63 (2015).
[97] D. J. Bartlett *et al.*, Phys. Rev. D **104**, 084025 (2021).
[98] D. J. Bartlett *et al.*, Phys. Rev. D **104**, 103516 (2021).
[99] D. J. Bartlett *et al.*, MNRAS **500**, 4639 (2021).
[100] D. J. Bartlett *et al.*, arXiv e-prints , arXiv:2205.12916 (2022).
[101] D. J. Bartlett *et al.*, Phys. Rev. D **103**, 123502 (2021).
[102] J.-J. Wei and X.-F. Wu, Frontiers of Physics **16**, 44300 (2021).
[103] J. Wheeler and K. W. Ford, *Geons, black holes, and quantum foam: a life in physics* (Norton, New York, 1998).
[104] G. Amelino-Camelia *et al.*, Nature **393**, 763 (1998).
[105] J. Ellis *et al.*, Phys. Rev. D **99**, 083009 (2019).
[106] G. D. Moore and A. E. Nelson, Journal of High Energy Physics **2001**, 023 (2001).
[107] V. A. Kostelecký and J. D. Tasson, Physics Letters B **749**, 551 (2015).
[108] F. R. Klinkhamer and M. Schreck, Phys. Rev. D **78**, 085026 (2008).
[109] J. Ellis *et al.*, A&A **402**, 409 (2003).
[110] J. Ellis *et al.*, Astroparticle Physics **25**, 402 (2006).
[111] J. Ellis *et al.*, Astroparticle Physics **29**, 158 (2008).
[112] M. Rodríguez Martínez, T. Piran, and Y. Oren, JCAP **2006**, 017 (2006).
[113] J. Bolmont *et al.*, ApJ **676**, 532 (2008).
[114] R. Lamon *et al.*, General Relativity and Gravitation **40**, 1731 (2008).
[115] V. Vasileiou *et al.*, Phys. Rev. D **87**, 122001 (2013).
[116] MAGIC Collaboration, Phys. Rev. Lett. **125**, 021301 (2020).
[117] J.-J. Wei and X.-F. Wu, ApJ **851**, 127 (2017).
[118] J.-J. Wei *et al.*, ApJ **834**, L13 (2017).
[119] J.-J. Wei *et al.*, ApJ **842**, 115 (2017).
[120] H. Tajima, arXiv e-prints , arXiv:0907.0714 (2009).
[121] Fermi LAT Collaboration, Nature **462**, 331 (2009).
[122] Z. Chang, Y. Jiang, and H.-N. Lin, Astroparticle Physics **36**, 47 (2012).
[123] R. J. Nemiroff *et al.*, Phys. Rev. Lett. **108**, 231103 (2012).
[124] S. Zhang and B.-Q. Ma, Astroparticle Physics **61**, 108 (2015).
[125] L. Shao, Z. Xiao, and B.-Q. Ma, Astroparticle Physics **33**, 312 (2010).

[126] H. Xu and B.-Q. Ma, Astroparticle Physics **82**, 72 (2016).
[127] H. Xu and B.-Q. Ma, Physics Letters B **760**, 602 (2016).
[128] H. Xu and B.-Q. Ma, JCAP **2018**, 050 (2018).
[129] Y. Liu and B.-Q. Ma, European Physical Journal C **78**, 825 (2018).
[130] R. Agrawal, H. Singirikonda, and S. Desai, JCAP **2021**, 029 (2021).
[131] L. Bonetti *et al.*, Physics Letters B **757**, 548 (2016).
[132] X.-F. Wu *et al.*, Phys. Rev. D **94**, 024061 (2016).
[133] L. Shao and B. Zhang, Phys. Rev. D **95**, 123010 (2017).
[134] J.-J. Wei and X.-F. Wu, JCAP **2018**, 045 (2018).
[135] J.-J. Wei and X.-F. Wu, Research in Astronomy and Astrophysics **20**, 206 (2020).
[136] N. Xing *et al.*, ApJ **882**, L13 (2019).
[137] B. Zhang *et al.*, Journal of High Energy Astrophysics **11**, 20 (2016).
[138] B. E. Schaefer, Phys. Rev. Lett. **82**, 4964 (1999).
[139] L. Bonetti *et al.*, Physics Letters B **768**, 326 (2017).
[140] M. J. Bentum *et al.*, Advances in Space Research **59**, 736 (2017).
[141] D. Alonso, Phys. Rev. D **103**, 123544 (2021).
[142] C. M. Will and J. Nordtvedt, Kenneth, ApJ **177**, 757 (1972).
[143] I. I. Shapiro, Phys. Rev. Lett. **13**, 789 (1964).
[144] B. Bertotti, L. Iess, and P. Tortora, Nature **425**, 374 (2003).
[145] K. C. Wong *et al.*, MNRAS **498**, 1420 (2020).
[146] S. Birrer *et al.*, A&A **643**, A165 (2020).
[147] Z.-X. Luo *et al.*, Journal of High Energy Astrophysics **9**, 35 (2016).
[148] LIGO Scientific Collaboration and Virgo Collaboration, Phys. Rev. Lett. **119**, 161101 (2017).
[149] A. Nusser, ApJ **821**, L2 (2016).
[150] O. Minazzoli *et al.*, Phys. Rev. D **100**, 104047 (2019).
[151] O. Minazzoli, arXiv e-prints , arXiv:2203.11215 (2022).
[152] J. Hakkila *et al.*, ApJS **169**, 62 (2007).
[153] H. Yu, S.-Q. Xi, and F.-Y. Wang, ApJ **860**, 173 (2018).
[154] D. L. Band, The Astrophysical Journal **486**, 928 (1997).
[155] D. Yonetoku *et al.*, ApJ **609**, 935 (2004).
[156] R.-F. Shen, L.-M. Song, and Z. Li, MNRAS **362**, 59 (2005).
[157] R. J. Lu *et al.*, MNRAS **367**, 275 (2006).
[158] F. Daigne and R. Mochkovitch, MNRAS **342**, 587 (2003).
[159] Z. Y. Peng *et al.*, Astronomische Nachrichten **332**, 92 (2011).

[160] S.-S. Du *et al.*, ApJ **882**, 115 (2019).
[161] R.-J. Lu *et al.*, ApJ **865**, 153 (2018).
[162] Z. L. Uhm and B. Zhang, ApJ **825**, 97 (2016).
[163] U. Jacob and T. Piran, JCAP **2008**, 031 (2008).
[164] M. Biesiada and A. Piórkowska, Classical and Quantum Gravity **26**, 125007 (2009).
[165] Y. Pan *et al.*, ApJ **808**, 78 (2015).
[166] X.-B. Zou *et al.*, Physics Letters B **776**, 284 (2018).
[167] Y. Pan *et al.*, ApJ **890**, 169 (2020).
[168] SciPy 1.0 Contributors, Nature Methods **17**, 261 (2020).
[169] A. Goldstein *et al.*, ApJS **208**, 21 (2013).
[170] F. Pedregosa *et al.*, Journal of Machine Learning Research **12**, 2825 (2011).
[171] A. A. Meiksin, Reviews of Modern Physics **81**, 1405 (2009).
[172] G. D. Becker *et al.*, MNRAS **410**, 1096 (2011).
[173] J. M. Shull, B. D. Smith, and C. W. Danforth, ApJ **759**, 23 (2012).
[174] Ž. Bošnjak and P. Kumar, MNRAS **421**, L39 (2012).
[175] F. Gao and L. Han, Computational Optimization and Applications **51**, 259 (2012).
[176] A. Lewis, arXiv e-prints , arXiv:1910.13970 (2019).
[177] P. J. E. Peebles, ApJ **185**, 413 (1973).
[178] Planck Collaboration, A&A **641**, A1 (2020).
[179] R. Reischke, S. Hagstotz, and R. Lilow, MNRAS **512**, 285 (2022).
[180] N. E. Chisari *et al.*, ApJS **242**, 2 (2019).
[181] P. Lemos, A. Challinor, and G. Efstathiou, JCAP **2017**, 014 (2017).
[182] A. Zonca *et al.*, Journal of Open Source Software **4**, 1298 (2019).
[183] K. M. Górski *et al.*, ApJ **622**, 759 (2005).
[184] D. Kocevski and E. Liang, ApJ **594**, 385 (2003).
[185] B. E. Schaefer, ApJ **602**, 306 (2004).
[186] F. Ryde, A&A **429**, 869 (2005).
[187] J. Hakkila and R. D. Preece, ApJ **740**, 104 (2011).
[188] C. Sivaram, Bulletin of the Astronomical Society of India **27**, 627 (1999).
[189] H. Gao, X.-F. Wu, and P. Mészáros, ApJ **810**, 121 (2015).
[190] Y. Sang, H.-N. Lin, and Z. Chang, MNRAS **460**, 2282 (2016).
[191] J.-J. Wei *et al.*, JCAP **2016**, 031 (2016).
[192] J.-J. Wei *et al.*, Phys. Rev. Lett. **115**, 261101 (2015).
[193] S. J. Tingay and D. L. Kaplan, ApJ **820**, L31 (2016).

[194] D. Wang, Z. Li, and J. Zhang, Physics of the Dark Universe **29**, 100571 (2020).
[195] M. J. Longo, Phys. Rev. Lett. **60**, 173 (1988).
[196] L. M. Krauss and S. Tremaine, Phys. Rev. Lett. **60**, 176 (1988).
[197] E. O. Kahya and S. Desai, Physics Letters B **756**, 265 (2016).
[198] I. LIGO Scientific Collaboration and Virgo Collaboration, Fermi Gamma-ray Burst Monitor, ApJ **848**, L13 (2017).
[199] M. Liu, Z. Zhao, X. You, J. Lu, and L. Xu, Physics Letters B **770**, 8 (2017).
[200] J.-J. Wei *et al.*, JCAP **2017**, 035 (2017).
[201] I. M. Shoemaker and K. Murase, Phys. Rev. D **97**, 083013 (2018).
[202] S. Boran *et al.*, Phys. Rev. D **97**, 041501 (2018).
[203] L. Yao, Z. Zhao, Y. Han, J. Wang, T. Liu, and M. Liu, ApJ **900**, 31 (2020).
[204] S.-C. Yang, W.-B. Han, and G. Wang, MNRAS **499**, L53 (2020).
[205] Z.-Y. Wang, R.-Y. Liu, and X.-Y. Wang, Phys. Rev. Lett. **116**, 151101 (2016).
[206] J.-J. Wei *et al.*, ApJ **818**, L2 (2016).
[207] S. Boran *et al.*, European Physical Journal C **79**, 185 (2019).
[208] R. Laha, Phys. Rev. D **100**, 103002 (2019).
[209] J.-J. Wei *et al.*, JCAP **22**, 1 (2019).
[210] Y.-P. Yang and B. Zhang, Phys. Rev. D **94**, 101501 (2016).
[211] Y. Zhang and B. Gong, ApJ **837**, 134 (2017).
[212] S. Desai and E. Kahya, European Physical Journal C **78**, 86 (2018).
[213] C. Leung *et al.*, ApJ **861**, 66 (2018).
[214] H. Wang *et al.*, ApJ **851**, L18 (2017).
[215] S. B. Lambert and C. Le Poncin-Lafitte, A&A **499**, 331 (2009).
[216] S. B. Lambert and C. Le Poncin-Lafitte, A&A **529**, A70 (2011).
[217] R. Laureijs *et al.*, arXiv e-prints , arXiv:1110.3193 (2011).
[218] LSST Science Collaboration, arXiv e-prints , arXiv:0912.0201 (2009).
[219] L. Giani and E. Frion, JCAP **2020**, 008 (2020).
[220] D. Rubin, I. Szapudi, B. J. Shappee, and G. S. Anand, ApJ **890**, L6 (2020).
[221] W. Hu and A. Cooray, Phys. Rev. D **63**, 023504 (2000).
[222] P. Li, S. Dodelson, and W. Hu, Phys. Rev. D **100**, 043502 (2019).
[223] B. Zhang, Nature **575**, 448 (2019).
[224] E. Di Valentino *et al.*, Astroparticle Physics **131**, 102606 (2021).
[225] J. Kormendy, K. Gebhardt, and D. Richstone (2000) p. 21.22.
[226] K. Gebhardt *et al.*, ApJ **539**, L13 (2000).
[227] L. Ferrarese and D. Merritt, ApJ **539**, L9 (2000).
[228] J. Magorrian *et al.*, AJ **115**, 2285 (1998).

[229] F. R. Marleau, D. Clancy, and M. Bianconi, MNRAS **435**, 3085 (2013).
[230] M. C. Begelman, R. D. Blandford, and M. J. Rees, nat **287**, 307 (1980).
[231] A. Peres, Physical Review **128**, 2471 (1962).
[232] J. D. Bekenstein, ApJ **183**, 657 (1973).
[233] A. Sesana, F. Haardt, P. Madau, and M. Volonteri, ApJ **611**, 623 (2004).
[234] P. Hut and M. J. Rees, MNRAS **259**, 27P (1992).
[235] G. Xu and J. P. Ostriker, ApJ **437**, 184 (1994).
[236] P. Boldrini, R. Mohayaee, and J. Silk, MNRAS **495**, L12 (2020).
[237] M. Volonteri, F. Haardt, and P. Madau, ApJ **582**, 559 (2003).
[238] L. Hernquist and J. E. Barnes, ApJ **349**, 562 (1990).
[239] Y. Dubois *et al.*, MNRAS **444**, 1453 (2014).
[240] M. Volonteri *et al.*, MNRAS **460**, 2979 (2016).
[241] S. Kazantzidis *et al.*, Memorie della Societa Astronomica Italiana Supplementi **5**, 261 (2004).
[242] M. Tremmel *et al.*, MNRAS **451**, 1868–1874 (2015).
[243] A. Pontzen and F. Governato, MNRAS **421**, 3464 (2012).
[244] D. N. Spergel and P. J. Steinhardt, Phys. Rev. Lett. **84**, 3760 (2000).
[245] A. Burkert, ApJ **534**, L143 (2000).
[246] A. Di Cintio *et al.*, MNRAS **469**, 2845 (2017).
[247] A. Cruz *et al.*, MNRAS **500**, 2177 (2021).
[248] E. Komatsu *et al.*, ApJS **192**, 18 (2011).
[249] Y. Dubois and R. Teyssier, A&A **477**, 79 (2008).
[250] T. Kimm, Ph.D. thesis, Oxford University, UK (2012).
[251] Y. Rasera and R. Teyssier, A&A **445**, 1 (2006).
[252] C. M. Booth and J. Schaye, MNRAS **398**, 53 (2009).
[253] N. I. Shakura and R. A. Sunyaev, A&A **500**, 33 (1973).
[254] Y. Dubois, J. Devriendt, A. Slyz, and R. Teyssier, MNRAS **420**, 2662 (2012).
[255] H. Omma, J. Binney, G. Bryan, and A. Slyz, MNRAS **348**, 1105 (2004).
[256] A. Slyz, J. Devriendt, M. Jarvis, Y. Dubois, and C. Pichon (2015) p. 52.
[257] S. Chandrasekhar, ApJ **97**, 255 (1943).
[258] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition* (Princeton Univ. Press, Princeton, NJ, 2008).
[259] R. S. Beckmann, A. Slyz, and J. Devriendt, MNRAS **478**, 995–1016 (2018).
[260] P. Taylor and C. Kobayashi, MNRAS **442**, 2751 (2014).
[261] J. Schaye *et al.*, MNRAS **446**, 521 (2015).
[262] D. Sijacki *et al.*, MNRAS **452**, 575 (2015).

[263] B. D. Smith *et al.*, MNRAS **480**, 3762 (2018).
[264] E. C. Ostriker, ApJ **513**, 252 (1999).
[265] D. Chapon, L. Mayer, and R. Teyssier, MNRAS **429**, 3114 (2013).
[266] D. Aubert, C. Pichon, and S. Colombi, MNRAS **352**, 376 (2004).
[267] D. Tweed *et al.*, A&A **506**, 647 (2009).
[268] J. E. Gunn and I. Gott, J. Richard, ApJ **176**, 1 (1972).
[269] C. Power *et al.*, MNRAS **338**, 14 (2003).
[270] N. E. Chisari *et al.*, MNRAS **472**, 1163 (2017).
[271] R. S. Barrows *et al.*, ApJ **829**, 37 (2016).
[272] Gaia Collaboration, A&A **616**, A14 (2018).
[273] Y. Y. Kovalev, L. Petrov, and A. V. Plavin, A&A **598**, L1 (2017).
[274] L. Petrov and Y. Y. Kovalev, MNRAS **471**, 3775 (2017).
[275] B. Binggeli, F. Barazza, and H. Jerjen, A&A **359**, 447 (2000).
[276] B. Binggeli and L. M. Cameron, A&A **252**, 27 (1991).
[277] B. Binggeli and L. M. Cameron, A&AS **98**, 297 (1993).
[278] G. Orosz and S. Frey, A&A **553**, A13 (2013).
[279] D. Lena *et al.*, ApJ **795**, 146 (2014).
[280] D. C. Kim, A. S. Evans, S. Stierwalt, and G. C. Privon, ApJ **824**, 122 (2016).
[281] G. Orosz, S. Frey, and Z. Paragi, `http://www.astro.noa.gr/ewass/Site/FilesRepo/0515_Frey_2016_06_28_20_41_11.pdf` (2016).
[282] C. J. Skipper and I. W. A. Browne, MNRAS **475**, 5179 (2018).
[283] R. S. Barrows, M. Mezcua, and J. M. Comerford, ApJ **882**, 181 (2019).
[284] Y. Shen, H.-C. Hwang, N. Zakamska, and X. Liu, ApJ **885**, L4 (2019).
[285] A. E. Reines, J. J. Condon, J. Darling, and J. E. Greene, ApJ **888**, 36 (2020).
[286] C. Spingola and A. Barnacka, MNRAS (2020), 10.1093/mnras/staa870.
[287] C. P. Ahn *et al.*, ApJS **203**, 21 (2012).
[288] A. Capetti and B. Balmaverde, A&A **440**, 73 (2005).
[289] M. Elvis *et al.*, ApJS **95**, 1 (1994).
[290] P. F. Hopkins *et al.*, ApJS **163**, 1 (2006).
[291] A. Merloni, S. Heinz, and T. di Matteo, MNRAS **345**, 1057 (2003).
[292] R. Wang, X.-B. Wu, and M.-Z. Kong, ApJ **645**, 890 (2006).
[293] L. Blecha, T. J. Cox, A. Loeb, and L. Hernquist, MNRAS **412**, 2154 (2011).
[294] S. Alam *et al.*, ApJS **219**, 12 (2015).
[295] A. T. Deller and E. Middelberg, AJ **147**, 14 (2014).
[296] K. N. Abazajian *et al.*, ApJS **182**, 543 (2009).
[297] K. Oh *et al.*, ApJS **195**, 13 (2011).

[298] I. N. Evans *et al.*, ApJS **189**, 37 (2010).
[299] S. T. Myers *et al.*, MNRAS **341**, 1 (2003).
[300] I. W. A. Browne *et al.*, MNRAS **341**, 13 (2003).
[301] J. A. Baldwin, M. M. Phillips, and R. Terlevich, PASP **93**, 5 (1981).
[302] H.-C. Hwang, Y. Shen, N. Zakamska, and X. Liu, ApJ **888**, 73 (2020).
[303] Gaia Collaboration, A&A **595**, A1 (2016).
[304] Gaia Collaboration, A&A **616**, A1 (2018).
[305] R. H. Becker, R. L. White, and D. J. Helfand, ApJ **450**, 559 (1995).
[306] A. J. Barth, M. C. Bentz, J. E. Greene, and L. C. Ho, ApJ **683**, L119 (2008).
[307] S. Bianchi *et al.*, MNRAS **435**, 2335 (2013).
[308] F. Civano *et al.*, ApJ **717**, 209 (2010).
[309] F. Civano *et al.*, ApJ **752**, 49 (2012).
[310] J. M. Wrobel, J. M. Comerford, and E. Middelberg, ApJ **782**, 116 (2014).
[311] M. Novak *et al.*, MNRAS **447**, 1282 (2015).
[312] L. Blecha, F. Civano, M. Elvis, and A. Loeb, MNRAS **428**, 1341 (2013).
[313] D. Batcheldor *et al.*, ApJ **717**, L6 (2010).
[314] P. G. Jonker *et al.*, MNRAS **407**, 645 (2010).
[315] R. B. Menezes, J. E. Steiner, and T. V. Ricci, ApJ **796**, L13 (2014).
[316] R. B. Menezes, J. E. Steiner, and P. da Silva, ApJ **817**, 150 (2016).
[317] J. J. Condon, J. Darling, Y. Y. Kovalev, and L. Petrov, ApJ **834**, 184 (2017).
[318] D. C. Kim *et al.*, ApJ **840**, 71 (2017).
[319] L. Lindegren *et al.*, A&A **595**, A4 (2016).
[320] F. Mignard *et al.*, A&A **595**, A5 (2016).
[321] L. Lindegren *et al.*, A&A **616**, A2 (2018).
[322] M. Assafin *et al.*, MNRAS **430**, 2797 (2013).
[323] N. Zacharias and M. I. Zacharias, AJ **147**, 95 (2014).
[324] S. Kotz, K. T. J, and K. Podgórski, *The Laplace Distribution and Generalizations* (Birkhäuser, Boston, MA, 2001).
[325] T. Kozubowski *et al.*, Journal of Multivariate Analysis **113**, 59 (2013).
[326] L. Breiman, Mach. Learn. **45**, 5–32 (2001).
[327] J. S. Bullock *et al.*, ApJ **555**, 240 (2001).
[328] G. Martin *et al.*, MNRAS **476**, 2801 (2018).
[329] E. Barausse, MNRAS **423**, 2533 (2012), arXiv:1201.5888 [astro-ph.CO] .
[330] C. O. Lousto *et al.*, Phys. Rev. D **81**, 084023 (2010).
[331] J. Brinchmann *et al.*, MNRAS **351**, 1151 (2004).
[332] G. Kauffmann *et al.*, MNRAS **341**, 33 (2003).

[333] C. A. Tremonti *et al.*, ApJ **613**, 898 (2004).
[334] J. L. Sérsic, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina **6**, 41 (1963).
[335] R. Weinberger *et al.*, MNRAS **465**, 3291 (2017).
[336] N. Khandai *et al.*, MNRAS **450**, 1349 (2015).
[337] R. Davé *et al.*, MNRAS **486**, 2827 (2019).
[338] M. Hirschmann *et al.*, MNRAS **442**, 2304 (2014).
[339] M. Tremmel *et al.*, ApJ **857**, L22 (2018).
[340] H. Pfister *et al.*, MNRAS **471**, 3646 (2017).
[341] H. Pfister *et al.*, MNRAS **486**, 101 (2019).
[342] L. Hui and A. Nicolis, Phys. Rev. Lett. **109**, 051304 (2012).
[343] L. Hui and A. Nicolis, Phys. Rev. Lett. **110**, 241104 (2013).
[344] A. Nicolis, R. Rattazzi, and E. Trincherini, Phys. Rev. D **79**, 064036 (2009).
[345] G. W. Horndeski, International Journal of Theoretical Physics **10**, 363 (1974).
[346] E. Babichev and G. Esposito-Farèse, Phys. Rev. D **87**, 044032 (2013).
[347] S. Melville and J. Noller, JCAP **2022**, 031 (2022).
[348] A. Barreira *et al.*, Phys. Rev. D **87**, 103511 (2013).
[349] A. de Felice and S. Tsujikawa, Phys. Rev. Lett. **105**, 111301 (2010).
[350] A. Barreira *et al.*, JCAP **2013**, 027 (2013).
[351] G. Dvali, G. Gabadadze, and M. Porrati, Physics Letters B **485**, 208 (2000).
[352] K. Koyama and F. P. Silva, Phys. Rev. D **75**, 084040 (2007).
[353] M. A. Luty *et al.*, Journal of High Energy Physics **2003**, 029 (2003).
[354] A. Nicolis and R. Rattazzi, Journal of High Energy Physics **2004**, 059 (2004).
[355] B. Falck, K. Koyama, G.-b. Zhao, and B. Li, JCAP **07**, 058 (2014).
[356] A. Cardoso *et al.*, Phys. Rev. D **77**, 083512 (2008).
[357] F. Schmidt, Phys. Rev. D **80**, 043001 (2009).
[358] J. Khoury and M. Wyman, Phys. Rev. D **80**, 064023 (2009).
[359] K. C. Chan and R. Scoccimarro, Phys. Rev. D **80**, 104005 (2009).
[360] A. L. Fey *et al.*, AJ **150**, 58 (2015).
[361] M. R. Blanton *et al.*, AJ **142**, 31 (2011).
[362] J. A. Willick *et al.*, ApJS **109**, 333 (1997).
[363] H. Desmond *et al.*, Phys. Rev. D **98**, 064015 (2018).
[364] H. Desmond and P. G. Ferreira, Phys. Rev. D **102**, 104060 (2020).
[365] A. Barreira, S. Bose, and B. Li, JCAP **2015**, 059 (2015).
[366] S. Cole, MNRAS **286**, 38 (1997).
[367] D. Blas, J. Lesgourgues, and T. Tram, JCAP **2011**, 034 (2011).

[368] U.-L. Pen, ApJ **490**, L127 (1997).
[369] E. Bertschinger, ApJS **137**, 1 (2001).
[370] A. Lue, R. Scoccimarro, and G. D. Starkman, Phys. Rev. D **69**, 124015 (2004).
[371] E. P. Hubble, ApJ **64**, 321 (1926).
[372] P. Prugniel and F. Simien, A&A **321**, 111 (1997).
[373] L. Ciotti and G. Bertin, A&A **352**, 447 (1999).
[374] A. A. Dutton *et al.*, MNRAS **416**, 322 (2011).
[375] F. Lelli, Galaxies **2**, 292 (2014).
[376] F. Lelli *et al.*, ApJ **827**, L19 (2016).
[377] M. Milgrom, Phys. Rev. Lett. **117**, 141101 (2016).
[378] S. Blinnikov and R. Moessner, A&AS **130**, 193 (1998).
[379] A. A. Dubkov and A. N. Malakhov, Radiophysics and Quantum Electronics **19**, 833 (1976).
[380] G. R. Blumenthal *et al.*, ApJ **301**, 27 (1986).
[381] O. Y. Gnedin *et al.*, ApJ **616**, 16 (2004).
[382] A. Del Popolo and F. Pace, Astrophys. Space Sci. **361**, 162 (2016), [Erratum: Astrophys.Space Sci. 361, 225 (2016)].
[383] W. J. G. de Blok, Advances in Astronomy **2010**, 789293 (2010).
[384] S.-H. Oh *et al.*, AJ **142**, 24 (2011).
[385] P. Salucci and A. Burkert, Astrophys. J. Lett. **537**, L9 (2000).
[386] C. Conroy, R. H. Wechsler, and A. V. Kravtsov, Astrophys. J. **647**, 201 (2006).
[387] A. V. Kravtsov *et al.*, Astrophys. J. **609**, 35 (2004).
[388] B. Lehmann *et al.*, Astrophys. J. **834**, 37 (2017).
[389] S. W. Skillman *et al.*, (2014).
[390] M. Bernardi *et al.*, Mon. Not. Roy. Astron. Soc. **436**, 697 (2013).
[391] F. Hofmann and J. Müller, Classical and Quantum Gravity **35**, 035015 (2018).
[392] J. B. R. Battat *et al.*, Phys. Rev. D **78**, 022003 (2008).
[393] J. Renk *et al.*, JCAP **2017**, 020 (2017).
[394] A. De Felice and S. Tsujikawa, Living Reviews in Relativity **13**, 3 (2010).
[395] J. M. Ezquiaga and M. Zumalacárregui, Phys. Rev. Lett. **119**, 251304 (2017).
[396] M. Ackermann *et al.*, ApJ **717**, L71 (2010).
[397] L. Goodenough and D. Hooper, arXiv e-prints , arXiv:0910.2998 (2009).
[398] M. Ajello *et al.*, ApJ **819**, 44 (2016).
[399] T. Linden *et al.*, Phys. Rev. D **94**, 103013 (2016).
[400] D. Hooper and L. Goodenough, Physics Letters B **697**, 412 (2011).
[401] D. Hooper and T. Linden, Phys. Rev. D **84**, 123005 (2011).

[402] D. Hooper and T. R. Slatyer, Physics of the Dark Universe **2**, 118 (2013).
[403] B. Zhou *et al.*, Phys. Rev. D **91**, 123010 (2015).
[404] T. Daylan *et al.*, Physics of the Dark Universe **12**, 1 (2016).
[405] I. Cholis *et al.*, Phys. Rev. D **105**, 103023 (2022).
[406] R. J. J. Grand and S. D. M. White, MNRAS **511**, L55 (2022).
[407] K. N. Abazajian, JCAP **2011**, 010 (2011).
[408] R. M. O'Leary *et al.*, arXiv e-prints , arXiv:1504.02477 (2015).
[409] J. Petrović, P. D. Serpico, and G. Zaharijas, JCAP **2015**, 023 (2015).
[410] S. K. Lee *et al.*, Phys. Rev. Lett. **116**, 051103 (2016).
[411] M. Buschmann *et al.*, Phys. Rev. D **102**, 023023 (2020).
[412] A. Gautam *et al.*, Nature Astronomy **6**, 703 (2022).
[413] D. Hooper *et al.*, Phys. Rev. D **88**, 083009 (2013).
[414] I. Cholis, D. Hooper, and T. Linden, JCAP **2015**, 043 (2015).
[415] R. K. Leane and T. R. Slatyer, Phys. Rev. Lett. **123**, 241101 (2019).
[416] K. N. Abazajian and M. Kaplinghat, Phys. Rev. D **86**, 083511 (2012).
[417] C. Gordon and O. Macías, Phys. Rev. D **88**, 083521 (2013).
[418] K. N. Abazajian *et al.*, Phys. Rev. D **90**, 023526 (2014).
[419] F. Calore, I. Cholis, and C. Weniger, JCAP **2015**, 038 (2015).
[420] S. Horiuchi, M. Kaplinghat, and A. Kwa, JCAP **2016**, 053 (2016).
[421] J. Petrović, P. D. Serpico, and G. Zaharijaš, JCAP **2014**, 052 (2014).
[422] E. Carlson and S. Profumo, Phys. Rev. D **90**, 023015 (2014).
[423] I. Cholis *et al.*, JCAP **2015**, 005 (2015).
[424] D. Gaggero *et al.*, JCAP **2015**, 056 (2015).
[425] O. Macias *et al.*, Nature Astronomy **2**, 387 (2018).
[426] R. Bartels *et al.*, Nature Astronomy **2**, 819 (2018).
[427] D. Hashimoto *et al.*, arXiv e-prints , arXiv:2109.08832 (2021).
[428] S. Ammazzalorso *et al.*, Phys. Rev. Lett. **124**, 101102 (2020).
[429] M. Ackermann *et al.*, Phys. Rev. Lett. **115**, 231301 (2015).
[430] V. Gammaldi *et al.*, Phys. Rev. D **104**, 083026 (2021).
[431] M. Lisanti *et al.*, Phys. Rev. Lett. **120**, 101101 (2018).
[432] T. R. Slatyer, Phys. Rev. D **93**, 023527 (2016).
[433] D. Wadekar and Z. Wang, arXiv e-prints , arXiv:2111.08025 (2021).
[434] A. J. Cuesta *et al.*, ApJ **726**, L6 (2011).
[435] S. Ando and E. Komatsu, Phys. Rev. D **73**, 023521 (2006).
[436] M. Ackermann *et al.* (Fermi-LAT), Phys. Rev. D **85**, 083007 (2012).
[437] M. Fornasa *et al.*, Phys. Rev. D **94**, 123005 (2016).

[438] M. Boylan-Kolchin *et al.*, MNRAS **398**, 1150 (2009).
[439] J. F. Navarro *et al.*, MNRAS **402**, 21 (2010).
[440] V. Springel *et al.*, MNRAS **391**, 1685 (2008).
[441] J. F. Navarro, C. S. Frenk, and S. D. M. White, ApJ **490**, 493 (1997).
[442] M. Lisanti *et al.*, Phys. Rev. D **97**, 063005 (2018).
[443] M. Wood *et al.*, in *35th International Cosmic Ray Conference (ICRC2017)*, International Cosmic Ray Conference, Vol. 301 (2017) p. 824.
[444] A. Bleuler *et al.*, Computational Astrophysics and Cosmology **2**, 5 (2015).
[445] J. J. Monaghan, ARA&A **30**, 543 (1992).
[446] S. Colombi, M. J. Chodorowski, and R. Teyssier, MNRAS **375**, 348 (2007).
[447] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu, *Spatial tessellations: concepts and applications of Voronoi diagrams* (John Wiley & Sons, 2000).
[448] M. C. Neyrinck, MNRAS **386**, 2101 (2008), arXiv:0712.3049 [astro-ph] .
[449] A. Charbonnier *et al.*, Computer Physics Communications **183**, 656 (2012).
[450] V. Bonnivard *et al.*, Computer Physics Communications **200**, 336 (2016).
[451] M. Hütten *et al.*, Computer Physics Communications **235**, 336 (2019).
[452] H. Zhao, MNRAS **278**, 488 (1996).
[453] J. F. Navarro *et al.*, MNRAS **349**, 1039 (2004).
[454] L. Gao *et al.*, MNRAS **387**, 536 (2008).
[455] B. Diemer, The Astrophysical Journal Supplement Series **239**, 35 (2018).
[456] L. Pieri *et al.*, Phys. Rev. D **83**, 023518 (2011).
[457] J. Lavalle, Q. Yuan, D. Maurin, and X. J. Bi, A&A **479**, 427 (2008).
[458] A. D. Ludlow *et al.*, The Astrophysical Journal **692**, 931 (2009).
[459] M. Hütten, C. Combet, G. Maier, and D. Maurin, JCAP **2016**, 047 (2016).
[460] M. A. Sánchez-Conde and F. Prada, MNRAS **442**, 2271 (2014).
[461] J. S. Bullock *et al.*, MNRAS **321**, 559 (2001).
[462] M. A. Sánchez-Conde and F. Prada, MNRAS **442**, 2271 (2014).
[463] R. H. Wechsler *et al.*, The Astrophysical Journal **568**, 52 (2002).
[464] F. Acero *et al.*, ApJS **223**, 26 (2016).
[465] T. E. Jeltema and S. Profumo, JCAP **2008**, 003 (2008).
[466] E. Branchini *et al.*, ApJS **228**, 8 (2017), arXiv:1612.05788 [astro-ph.CO] .
[467] M. Ackermann *et al.*, Phys. Rev. Lett. **121**, 241101 (2018).
[468] M. Colavincenzo *et al.*, MNRAS **491**, 3225 (2020).
[469] H. Abdallah *et al.*, Phys. Rev. Lett. **120**, 201101 (2018).
[470] A. Di Cintio *et al.*, MNRAS **441**, 2986 (2014).
[471] M. Fornasa *et al.*, MNRAS **429**, 1529 (2013).

[472] J. Lavalle, J. Pochon, P. Salati, and R. Taillet, A&A **462**, 827 (2007).
[473] A. Franceschini, G. Rodighiero, and M. Vaccari, Astrophys **487**, 0805 (2008).
[474] A. Franceschini and G. Rodighiero, A&A **603**, A34 (2017).
[475] P. D. Serpico *et al.*, MNRAS: Letters **421**, L87 (2012).
[476] M. Hütten, C. Combet, and D. Maurin, JCAP **2018**, 005 (2018).
[477] W. H. Press and P. Schechter, ApJ **187**, 425 (1974).
[478] K. Jedamzik and M. Pospelov, New Journal of Physics **11**, 105028 (2009).
[479] V. Poulin and P. D. Serpico, Phys. Rev. D **91**, 103007 (2015).
[480] S. Galli *et al.*, Phys. Rev. D **80**, 023505 (2009).
[481] M. Kawasaki *et al.*, JCAP **2021**, 015 (2021).
[482] M. Walther, J. Oñorbe, J. F. Hennawi, and Z. Lukić, ApJ **872**, 13 (2019).
[483] P. Gaikwad *et al.*, MNRAS **494**, 5091 (2020).
[484] M. Ackermann *et al.*, ApJ **761**, 91 (2012).
[485] L. J. Chang *et al.*, Phys. Rev. D **98**, 123004 (2018).
[486] M. A. Sánchez-Conde *et al.*, JCAP **2011**, 011 (2011).
[487] S. Ando *et al.*, Phys. Rev. D **102**, 061302 (2020).
[488] M. Regis *et al.*, JCAP **2021**, 046 (2021).
[489] MAGIC Collaboration *et al.*, JCAP **2016**, 039 (2016).
[490] S. Archambault *et al.*, Phys. Rev. D **95**, 082001 (2017).
[491] A. Albert *et al.*, ApJ **853**, 154 (2018).
[492] H. Abdallah *et al.*, Phys. Rev. D **102**, 062001 (2020).
[493] H. Abdallah *et al.*, Phys. Rev. Lett. **117**, 111301 (2016).
[494] T. Marrodán Undagoitia, W. Rodejohann, T. Wolf, and C. E. Yaguna, Progress of Theoretical and Experimental Physics **2022**, 013F01 (2022).
[495] L. Bergström *et al.*, Phys. Rev. Lett. **111**, 171101 (2013).
[496] A. Cuoco, J. Heisig, M. Korsmeier, and M. Krämer, JCAP **2018**, 004 (2018).
[497] F. Calore *et al.*, SciPost Physics **12**, 163 (2022).
[498] A. Abramowski *et al.*, Phys. Rev. Lett. **110**, 041301 (2013).
[499] A. Albert *et al.*, JCAP **2014**, 023 (2014).
[500] M. Ackermann *et al.*, Phys. Rev. D **91**, 122002 (2015).
[501] T. Cohen *et al.*, Phys. Rev. Lett. **119**, 021102 (2017).
[502] L. A. Anchordoqui *et al.*, Astroparticle Physics **132**, 102614 (2021).
[503] H. Liu *et al.*, Phys. Rev. D **104**, 043514 (2021).
[504] S. Mau *et al.*, ApJ **932**, 128 (2022).
[505] M. S. Nitschai, M. Cappellari, and N. Neumayer, MNRAS **494**, 6001 (2020).
[506] M. S. Nitschai *et al.*, ApJ **916**, 112 (2021).

[507] T. Sawala *et al.*, MNRAS **509**, 1432 (2022).
[508] S. McAlpine *et al.*, MNRAS **512**, 5823 (2022).
[509] J. Kumar and D. Marfatia, Phys. Rev. D **88**, 014035 (2013).
[510] C. Johnson *et al.*, Phys. Rev. D **99**, 103007 (2019).
[511] F. Giacchino, L. Lopez-Honorez, and M. H. G. Tytgat, JCAP **2013**, 025 (2013).
[512] T. Toma, Phys. Rev. Lett. **111**, 091301 (2013).
[513] A. Sommerfeld, Annalen Phys. **403**, 257 (1931).
[514] K. K. Boddy *et al.*, Phys. Rev. D **100**, 063019 (2019).
[515] B. Boucher *et al.*, arXiv e-prints , arXiv:2110.09653 (2021).
[516] V. Vikram, A. Cabré, B. Jain, and J. T. VanderPlas, JCAP **08**, 020 (2013).
[517] W. Hu and I. Sawicki, Phys. Rev. D **76**, 064004 (2007).
[518] E. C. Ostriker and J. J. Binney, MNRAS **237**, 785 (1989).
[519] L. F. Secco *et al.*, Astrophys. J. **860**, 32 (2018).
[520] K. Pardo, H. Desmond, and P. G. Ferreira, Phys. Rev. D **100**, 123006 (2019).
[521] M. D. Weinberg, MNRAS **299**, 499 (1998).
[522] M. Semczuk *et al.*, MNRAS **498**, 3535 (2020).
[523] J. Binney, ARA&A **30**, 51 (1992).
[524] T. C. Scott *et al.*, MNRAS **403**, 1175 (2010).
[525] H. A. Buchdahl, MNRAS **150**, 1 (1970).
[526] S. M. Carroll *et al.*, Phys. Rev. D **70**, 043528 (2004).
[527] T. P. Sotiriou and V. Faraoni, Reviews of Modern Physics **82**, 451 (2010).
[528] J. Sakstein, arXiv e-prints , arXiv:1502.04503 (2015).
[529] A. Cabré *et al.*, JCAP **2012**, 034 (2012).
[530] G.-B. Zhao, B. Li, and K. Koyama, Phys. Rev. D **83**, 044007 (2011).
[531] G.-B. Zhao, B. Li, and K. Koyama, Phys. Rev. Lett. **107**, 071303 (2011).
[532] A. Bosma (1991) p. 181.
[533] I. García-Ruiz, R. Sancisi, and K. Kuijken, A&A **394**, 769 (2002).
[534] M. J. D. Powell, The Computer Journal **7**, 155 (1964).
[535] F. D. Albareti *et al.*, ApJS **233**, 25 (2017).
[536] J. F. Navarro, C. S. Frenk, and S. D. M. White, ApJ **462**, 563 (1996).
[537] H. L. Child *et al.*, ApJ **859**, 55 (2018).
[538] S. Peirani *et al.*, MNRAS **472**, 2153 (2017).
[539] C. Welker *et al.*, MNRAS **465**, 1241 (2017).
[540] M. Santos, D. Alonso, P. Bull, M. B. Silva, and S. Yahya (2015) p. 21.
[541] S. Yahya *et al.*, MNRAS **450**, 2251 (2015).
[542] N. E. Chisari *et al.*, The Open Journal of Astrophysics **2**, 4 (2019).

[543] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, ninth ed. (Dover, New York City, 1964) pp. 504–515.