

Hierarchical modeling of gravitational-wave populations for disentangling environmental and modified-gravity effects

Shubham Kejriwal,^{1,*} Enrico Barausse,^{2,3,†} and Alvin J. K. Chua^{1,4,‡}

¹*Department of Physics, National University of Singapore, Singapore 117551*

²*SISSA, Via Bonomea 265, 34136 Trieste, Italy and INFN Sezione di Trieste*

³*IFPU - Institute for Fundamental Physics of the Universe, Via Beirut 2, 34014 Trieste, Italy*

⁴*Department of Mathematics, National University of Singapore, Singapore 119076*

The upcoming Laser Interferometer Space Antenna (LISA) will detect up to thousands of extreme-mass-ratio inspirals (EMRIs). These sources will spend $\sim 10^5$ cycles in band, and are therefore sensitive to tiny changes in the general-relativistic dynamics, potentially induced by astrophysical environments or modifications of general relativity (GR). Previous studies have shown that these effects can be highly degenerate for a single source. However, it may be possible to distinguish between them at the population level, because environmental effects should impact only a fraction of the sources, while modifications of GR would affect all. We therefore introduce a population-based hierarchical framework to disentangle the two hypotheses. Using simulated EMRI populations, we perform tests of the null vacuum-GR hypothesis and two alternative beyond-vacuum-GR hypotheses, namely migration torques (environmental effects) and time-varying G (modified gravity). We find that with as few as ≈ 20 detected sources, our framework can statistically distinguish between these three hypotheses, and even indicate if both environmental and modified gravity effects are simultaneously present in the population. Our framework can be applied to other models of beyond-vacuum-GR effects available in the literature.

I. INTRODUCTION

Since the first detection of gravitational waves (GWs) in 2015 [1], the LIGO-Virgo-KAGRA (LVK) GW Transient Catalog (GWTC) has grown to include more than 200 transient sources [2–6] in the ~ 100 Hz band. The GWTC has been used to set tight constraints on the properties of merging compact binary populations in the universe [7] and to test general relativity (GR) [8–10]. The upcoming network of third-generation ground-based detectors, including the Einstein Telescope [11] and Cosmic Explorer [12, 13], will further improve these constraints and extend population studies to heavier binaries in the $\sim 1 - 10$ Hz band [14–16]. At even lower frequencies ~ 1 mHz, the upcoming Laser Interferometer Space Antenna (LISA) will probe the properties of supermassive black hole (MBH) populations [17, 18].

A particularly interesting class of sources in the LISA band is provided by extreme-mass-ratio inspirals (EMRIs), where a compact object (CO) of mass $\mu \sim 10^{1-2} M_\odot$ completes $\sim 10^5$ highly relativistic orbits around an MBH of mass $M \sim 10^{5-7} M_\odot$. EMRIs are expected to constrain MBH properties such as mass and spin to sub-percent precision [19], enabling tight constraints on the MBH mass function [20]. Furthermore, the large number of strong field cycles completed by the CO makes EMRIs unique probes of astrophysical environments and potential deviations from GR [21–31]. However, the sensitivity of EMRIs to such “beyond-vacuum-GR” effects may hinder their precision astronomy prospects. These effects are

secular and perturbative in nature, often manifesting as simple power-law-like corrections to the EMRI’s vacuum-GR dynamics and introducing correlations across the parameter space. If these effects are present in the signal but are not included in the analysis, they can bias the source’s inferred parameters and potentially lead to, e.g., false claims of deviations from GR [29, 32, 33].

Previous studies have employed hierarchical models to search for evidence of modified gravity effects in GW populations (see, e.g., [8–10, 34–36]), but do not consider multiple contending effects. Recently, Yuan et al. [37, 38] proposed a statistic F to distinguish modified gravity effects (with a constant universal amplitude) from environmental effects when they are present in the signal *one at a time*, but do not address the distinguishability of multiple concurrent effects in the system.

In this paper, we propose for the first time a framework to detect and distinguish between the two main classes of beyond-vacuum-GR effects using Bayesian hierarchical modeling, even when both classes of effects are present in the system simultaneously. The first class includes effects such as astrophysical environments, which introduce, e.g., corrections to the CO’s orbit through tidal resonances [39] or migration torques [22–24, 29, 31]. These effects can be independently realized across all sources, and may only be present in a fraction of the total population. We collectively refer to such effects as *local* in this paper. The second class of effects includes fundamental physics modifications that are universal in nature, i.e., present in all sources simultaneously. In many cases, these corrections are controlled by an intrinsic “strength” constant across the population, such that GR is recovered when the strength parameter assumes a null value. An example is the time-varying Newton’s constant G (see e.g. [40, 41]), which yields GR when the rate of change of G vanishes,

* shubhamkejriwal@u.nus.edu

† barausse@sissa.it

‡ alvincjk@nus.edu.sg

i.e. $\dot{G} = 0$. In some other cases, e.g. in scalar-tensor theories [42], deviations from the GR predictions may also depend on the nature of the sources and/or its mass and spin [43–45], but are still expected to affect the whole EMRI population. We refer to this class of effects as *global* in this paper.

This classification allows us to model local and global effects at different levels in a hierarchical framework. We can then perform hypothesis testing to compare the null vacuum-GR hypothesis with the local and global effect hypotheses, to test the preference of the data for one over another. Explicitly, the hypotheses that we will consider are:

- **Vacuum-GR hypothesis ($\mathcal{H}_v$):** all sources evolve according to vacuum-GR dynamics;
- **Local effect hypothesis ($\mathcal{H}_\ell$):** a nonzero fraction of the sources are affected by a local (e.g. astrophysical environment) effect, while the rest follow vacuum-GR;
- **Global effect hypothesis ($\mathcal{H}_g$):** all sources experience a global (e.g. modified gravity) correction to the vacuum-GR dynamics.

A. Executive summary of results

In this paper, we apply the above framework to specific models of local and global effects. For the local effect, we choose migration torques, which are induced on the CO due to the presence of an accretion disk around the MBH. For the global effect, we select the time-varying G model, with $\dot{G}$ [yr^{-1}] the constant intrinsic strength of the effect. We construct four different EMRI populations, P_v , P_ℓ , P_g , and P_{mix} . The first three, P_v , P_ℓ , and P_g , are generated assuming $\mathcal{H}_v$, $\mathcal{H}_\ell$, and $\mathcal{H}_g$ as the truth, respectively. The fourth “mixed” population P_{mix} is constructed such that all sources experience a time-varying G , while a fraction of them are simultaneously affected by migration torques. We also consider three different population sizes, with 100, 500, and 1000 sources, of which $\approx 17\%$ are observable. We find that even for the population with 100 sources (of which 17 are observable), the hierarchical framework can reliably recover the correct hypothesis under which the population was constructed, and even indicate the concurrent presence of local and global effects in the population.

We also gauge the robustness of our method for different choices of local and global effect parameters in the true population. We find that if the fraction f of the population with a local effect is ≥ 0.9 , the recovered intrinsic strength of time-varying G , i.e. $\dot{G}$ [yr^{-1}], may be significantly biased. Furthermore, we highlight a special case where both local and global effects are present in the population simultaneously, but they effectively “cancel” each other out. In such unique scenarios, the log-Bayes factor $\log \mathcal{B}_\ell^g$, which quantifies the data’s preference for

the global effect hypothesis over the local one, is ≈ 0.0 , and the hierarchical framework struggles to distinguish between the two effects. The results are presented in more detail in Sec. V.

B. Structure of the paper

The remainder of this paper is organized as follows. In Sec. II, we begin with an overview of Bayesian hierarchical modeling and hypothesis testing for GW data analysis; In Sec. III, we describe our source and population models under each of the three hypotheses, $\mathcal{H}_v$, $\mathcal{H}_\ell$, and $\mathcal{H}_g$; In Sec. IV, we specify our data analysis setup, including a description of the approximate methods used for feasible hypothesis testing in the hierarchical framework; In Sec. V, we present our results, applying the hierarchical framework to various examples of EMRI populations; Finally, in Sec. VI, we summarize our results, discuss their impact, and lay out potential future directions.

Throughout this paper, we assume a flat universe with Hubble’s constant $H_0 = 70$ km/s/Mpc, the matter density $\Omega_{m,0} = 0.3$, and the dark energy density $\Omega_{\Lambda,0} = 0.7$.

II. BACKGROUND

A. Bayesian hierarchical modeling

Consider a population of N_{pop} GW sources modeled by a set of hyperparameters $\vec{\lambda}$. Each source is modeled by a set of parameters $\vec{\theta}$, distributed according to some prior probability density function (pdf), $p(\vec{\theta}|\vec{\lambda}, \mathcal{H})$. Here, $\mathcal{H}$ is the assumed hypothesis. Of these N_{pop} sources, only a subset of size $N_{\text{obs}} < N_{\text{pop}}$ will satisfy a chosen detectability criterion, such that they are considered to be “observed”. The measured data $\vec{d}_i$ of the i^{th} observed sources is represented by its maximum-likelihood estimate (MLE), $\hat{\theta}_i$, where $i = 1, \dots, N_{\text{obs}}$ [46].¹ At the MLE point, the individual source likelihood $p(\vec{d}_i|\vec{\theta}_i, \mathcal{H})$ maximizes.

Given the set of measured data for the detected sources $\{\vec{d}_i\}$, the (hyper)posterior pdf $p(\vec{\lambda}|\{\vec{d}_i\}, \mathcal{H})$ of the hyperparameter set $\vec{\lambda}$ can be written according to Bayes theorem as [47]

$$p(\vec{\lambda}|\{\vec{d}_i\}, \mathcal{H}) = \frac{p(\{\vec{d}_i\}|\vec{\lambda}, \mathcal{H})\pi(\vec{\lambda}|\mathcal{H})}{\mathcal{Z}(\{\vec{d}_i\}|\mathcal{H})}. \quad (1)$$

¹ This setup, based on [46], assumes that each i^{th} source has its own datastream $\vec{d}_i$, which does not directly apply to the LISA context, where all sources will overlap in the datastream instead. We can still apply this framework if we assume $\vec{d}_i$ to be the *residual* of the i^{th} source when all other sources are subtracted from the full datastream, i.e., $\vec{d}_i = \vec{d} - \sum_{j \neq i} \vec{h}(\hat{\theta}_j)$. Here, $\vec{d}$ is the full LISA datastream and $\vec{h}(\hat{\theta}_j)$ is the single-source GW template evaluated at $\hat{\theta}_j$.

Here $p(\{\vec{d}_i\}|\vec{\lambda}, \mathcal{H})$ is the hyperlikelihood of $\vec{\lambda}$, $\pi(\vec{\lambda}|\mathcal{H})$ is its hyperprior, and $\mathcal{Z}(\{\vec{d}_i\}|\mathcal{H})$ is the evidence for the data under the hypothesis, written explicitly as

$$\mathcal{Z}(\{\vec{d}_i\}|\mathcal{H}) = \int d\vec{\lambda} p(\{\vec{d}_i\}|\vec{\lambda}, \mathcal{H}) \pi(\vec{\lambda}|\mathcal{H}). \quad (2)$$

Assuming that individual sources are realized independently in the population, and using the correspondence between $\vec{d}_i$ and $\hat{\theta}_i$, we can write the hyperlikelihood as an effective product of single-source hyperlikelihoods [46]

$$p(\{\vec{d}_i\}|\vec{\lambda}, \mathcal{H}) = \prod_{i=1}^{N_{\text{obs}}} \frac{p(\hat{\theta}_i|\vec{\lambda}, \mathcal{H})}{\alpha(\vec{\lambda}|\mathcal{H})}, \quad (3)$$

where $\alpha(\vec{\lambda}|\mathcal{H})$ encodes the selection biases in the observation of these sources. With this, the hyperposterior of $\vec{\lambda}$ can be rewritten as

$$p(\vec{\lambda}|\{\vec{d}_i\}, \mathcal{H}) = \frac{\pi(\vec{\lambda}|\mathcal{H})}{\mathcal{Z}(\{\vec{d}_i\}|\mathcal{H})} \times \prod_{i=1}^{N_{\text{obs}}} \frac{p(\hat{\theta}_i|\vec{\lambda}, \mathcal{H})}{\alpha(\vec{\lambda}|\mathcal{H})}. \quad (4)$$

B. Hypothesis testing

We are interested in testing the vacuum-GR hypothesis $\mathcal{H}_v$ against the alternate beyond-vacuum-GR hypothesis $\mathcal{H}_a$ where $a = \ell, g$. These hypotheses can be modeled in a nested framework, with $\mathcal{H}_v \subseteq \mathcal{H}_a$. The equality holds when a general subset of hyperparameters $\vec{\lambda}' \subseteq \vec{\lambda}$ under $\mathcal{H}_a$ assumes a fiducial null value $\vec{\lambda}'_0$. Then, we can write the Bayes factor for the data's preference of $\mathcal{H}_v$ over $\mathcal{H}_a$ as the Savage-Dickey ratio [47, 48],

$$\mathcal{B}_a^v = \frac{p(\vec{\lambda}' = \vec{\lambda}'_0|\{\vec{d}_i\}, \mathcal{H}_a)}{\pi(\vec{\lambda}' = \vec{\lambda}'_0|\mathcal{H}_a)}; \quad a = \ell, g. \quad (5)$$

Here, the numerator is the marginalized hyperposterior of $\vec{\lambda}'$ at $\vec{\lambda}'_0$, and the denominator is the corresponding marginalized hyperprior. Substituting the hyperposterior (4) and the evidence (2), we can rewrite $\mathcal{B}_a^v$ as

$$\mathcal{B}_a^v = \frac{\prod_i p(\hat{\theta}_i|\vec{\lambda}' = \vec{\lambda}'_0, \mathcal{H}_a)}{\int d\vec{\lambda}' \prod_i r_\alpha p(\hat{\theta}_i|\vec{\lambda}', \mathcal{H}_a) \pi(\vec{\lambda}'|\mathcal{H}_a)} \quad (6)$$

where we define $r_\alpha := \alpha(\vec{\lambda}' = \vec{\lambda}'_0|\mathcal{H}_a)/\alpha(\vec{\lambda}'|\mathcal{H}_a)$.

In the following, we set $r_\alpha = 1$, effectively treating $\alpha(\vec{\lambda}'|\mathcal{H}_a)$ as a constant. As argued more formally in Appendix A, this assumption holds when $\vec{\lambda}'$ is a set of beyond-vacuum-GR effect hyperparameters, e.g., the intrinsic strength of time-varying G . This is because such effects only perturbatively modify the detectability of a source in the population. Setting $r_\alpha = 1$ mitigates some computational expense in the calculation of the Bayes factor, as the selection function involves an integral over source parameters over the full space. The integrand

includes the detection probability $p_{\text{det}}(\vec{\theta}|\mathcal{H})$, which is typically a numerical function in the GW context and hence expensive to compute. If $p_{\text{det}}(\vec{\theta}|\mathcal{H})$ can be cheaply calculated, this factor may be restored feasibly in future studies (see, e.g., [49]).

In Eq. (6), the single source hyperlikelihood $p(\hat{\theta}_i|\vec{\lambda}', \mathcal{H}_a)$ can be expressed as a marginal over $\vec{\theta}$,

$$p(\hat{\theta}_i|\vec{\lambda}', \mathcal{H}_a) = \int d\vec{\theta} p(\hat{\theta}_i|\vec{\theta}, \mathcal{H}_a) \pi(\vec{\theta}|\vec{\lambda}', \mathcal{H}_a). \quad (7)$$

Here, under the hypothesis $\mathcal{H}_a$, $p(\hat{\theta}_i|\vec{\theta}, \mathcal{H}_a)$ is the likelihood of $\vec{\theta}$ under the data $\hat{\theta}_i$, $\pi(\vec{\theta}|\vec{\lambda}', \mathcal{H}_a)$ is the prior pdf of $\vec{\theta}$ given $\vec{\lambda}'$ used in the analysis, and the integral is over the full space of $\vec{\theta}$. While this explicit integral evaluation can make hypothesis testing expensive, we can analytically approximate it for the models of our interest, as we shall describe in Sec. IV B.

For simplicity, we additionally assume flat priors on the model hyperparameters $\vec{\lambda}$, such that $\pi(\vec{\lambda}'|\mathcal{H}_a) = \pi(\vec{\lambda}' = \vec{\lambda}'_0|\mathcal{H}_a)$. Hence, the final form of the Bayes factor used for hypothesis testing under our assumptions is

$$\mathcal{B}_a^v = \frac{1}{\pi(\vec{\lambda}' = \vec{\lambda}'_0|\mathcal{H}_a)} \frac{\prod_i p(\hat{\theta}_i|\vec{\lambda}' = \vec{\lambda}'_0, \mathcal{H}_a)}{\int d\vec{\lambda}' \prod_i p(\hat{\theta}_i|\vec{\lambda}', \mathcal{H}_a)}. \quad (8)$$

III. SETUP

A. Source model

In this paper, we model individual sources as Kerr EM-RIs evolving in circular and equatorial orbits. Our model allows for generic corrections to the vacuum-GR dynamics from migration torques and time-varying G effects. We model these effects as simple power-law corrections to the leading-order GR flux of the axial component of the angular momentum [29],

$$\dot{L} = \dot{L}_{\text{GR}} \left(1 + A_\ell \left(\frac{p}{10M} \right)^{n_\ell} + A_g \left(\frac{p}{M} \right)^{n_g} \right). \quad (9)$$

Here p is the instantaneous semi-latus rectum of the CO's orbit and M is the MBH mass; A_ℓ is the effective strength of the migration torque effect, with typical values $\sim 10^{-5}$ [24]; n_ℓ is the post-Newtonian (PN) order at which the local effect enters the evolution, and is ≈ 8 for migration torques in α -disks [24]; A_g is the strength of the time-varying G effect, with current best constraints $\lesssim 10^{-12}$ [yr $^{-1}$] [41], and n_g is its PN order in the evolution. For time-varying G , $n_g = 4$, denoting a -4 PN order correction [29, 41]. Note that the flux balance equation reduces to vacuum-GR as $A_\ell, A_g \rightarrow 0$. We emphasize that such parametric constructions are also available for other beyond-vacuum-GR effects in the literature [23–28, 30, 32].

In addition to those described above, the EMRI model has the following parameters: a , the dimensionless spin of the black hole; μ , the mass of the CO; d_L , the luminosity distance of the source, or equivalently the redshift $z \equiv z(d_L)$; (θ_S, ϕ_S) , the sky localization angles; (θ_K, ϕ_K) , the spin orientation angles; and Φ_0 , the initial azimuthal orbital phase.

When inferring sources in the population, we only consider a subset of these parameters. We choose two vacuum-GR parameters, $\vec{v} := \{\ln M, z\}$, two local effect parameters, $\vec{\ell} := \{A_\ell, n_\ell\}$, and one global effect parameter $\vec{g} := \{A_g\}$. Hence we can define the source parameter set $\vec{\theta} := \{\vec{v}, \vec{\ell}, \vec{g}\}$ for data analysis.

B. Population model

Given the source model, the EMRI population can be described by N_{pop} draws of the source parameters $\vec{\theta} \sim \pi(\vec{\theta}|\vec{\lambda}, \mathcal{H})$. We decompose the hyperparameter set similarly to the source parameters, $\vec{\lambda} := \{\vec{\lambda}_v, \vec{\lambda}_\ell, \vec{\lambda}_g\}$, and assume independent priors on $\vec{v}$, $\vec{\ell}$, and $\vec{g}$, such that the joint prior $\pi(\vec{\theta}|\vec{\lambda}, \mathcal{H})$ can be written piecewise,

$$\pi(\vec{\theta}|\vec{\lambda}, \mathcal{H}) = \pi(\vec{v}|\vec{\lambda}_v) \times \begin{cases} 1 & \mathcal{H} = \mathcal{H}_v \\ \pi(\vec{\ell}|\vec{\lambda}_\ell) & \mathcal{H} = \mathcal{H}_\ell \\ \pi(\vec{g}|\vec{\lambda}_g) & \mathcal{H} = \mathcal{H}_g \end{cases}. \quad (10)$$

For the vacuum-GR parameters $\vec{v} = \{\ln M, z\}$, we choose a prior pdf $p(\vec{v}|\vec{\lambda}_v)$ that scales with the MBH mass M and the source redshift z , and also accounts for the size of the comoving volume. Explicitly, we choose (see e.g. [20])

$$\pi(\vec{v}|\vec{\lambda}_v) \propto d_c^2(z) \frac{dn}{d \ln M dz}, \quad (11)$$

$$\frac{dn}{d \ln M dz} := K \left(\frac{M}{M_*} \right)^\alpha (1+z)^\beta. \quad (12)$$

Here, the comoving distance $d_c(z)$ is related to the redshift z given H_0 , $\Omega_{m,0}$, and $\Omega_{\Lambda,0}$ [50]. Additionally, when calculating the number density of sources, $dn/d \ln M dz$, K determines the size of the population, and M_* sets the scale for the MBH masses. The corresponding hyperparameter set is $\vec{\lambda}_v := \{K, \alpha, \beta\}$, and we fix $M_* = 3 \times 10^6 M_\odot$, following Babak et al. [19].

The local effect parameters $\vec{\ell} = \{A_\ell, n_\ell\}$ must be drawn from a distribution such that only a fraction $:= f$ of the sources in the population have a non-null value of $\vec{\ell}$, while the remaining $(1-f)$ sources follow a vacuum-GR evolution. To incorporate this dependence on f , we choose a bimodal distribution for the prior,

$$\pi(\vec{\ell}|\vec{\lambda}_\ell) := (1-f)\delta^2(\vec{\ell}) + f\mathcal{N}(\vec{\mu}_\ell|\vec{\ell}, \Sigma_\ell). \quad (13)$$

With this choice, a fraction f of the sources is chosen from a two-dimensional normal distribution $\mathcal{N}(\vec{\mu}_\ell|\vec{\ell}, \Sigma_\ell)$

with a mean $\vec{\mu}_\ell := (\mu_{A_\ell}, \mu_{n_\ell})$ and covariance $\Sigma_\ell := \text{diag}(\sigma_{A_\ell}^2, \sigma_{n_\ell}^2)$. The remaining $(1-f)$ draws are identically zero, i.e., they assume the null value of $A_\ell = n_\ell = 0$, enforced by the Dirac-delta $\delta^2(\vec{\ell})$. Correspondingly, the local effect hyperparameter set is $\vec{\lambda}_\ell := \{f, (\mu_{A_\ell}, \mu_{n_\ell}), (\sigma_{A_\ell}, \sigma_{n_\ell})\}$.

Finally, for the global effect, we only model the intrinsic strength of time-varying G with $\vec{g} = \{A_g\}$. Since this strength is modeled to be constant in all sources, we choose a Dirac-delta prior,

$$\pi(\vec{g}|\vec{\lambda}_g) := \delta(A_g - \dot{G}) \quad (14)$$

where $\dot{G} [\text{yr}^{-1}]$ is the value of time-varying G at the population level. Hence, the global effect hyperparameter $\vec{\lambda}_g = \{\dot{G}\}$. We stress that this distinction between the inferred source parameter A_g and the hyperparameter $\dot{G}$ is important in the data analysis context, since the individual source likelihood only estimates the MLE for A_g , the value of which may be biased about the true value of $\dot{G}$ in the population depending on the assumed hypothesis, the detector-noise realization, etc. A constraint on $\dot{G}$ is only obtained hierarchically from these “noisy” estimates of A_g .

IV. DATA ANALYSIS TECHNIQUES

Assuming a fiducial true hypothesis $\mathcal{H}^*$, and a corresponding set of true hyperparameters $\vec{\lambda}^*$, we can now construct a population of N_{pop} sources with inferred parameters $\{\vec{\theta}^*\}$ drawn from the prior $\pi(\vec{\theta}|\vec{\lambda}^*, \mathcal{H}^*)$. Here, we formulate an approximate analysis model to infer these parameters, which can then be used to perform hypothesis testing. For simplicity, we assume the zero-detector-noise realization in this paper, such that our results may be interpreted as averages over many noise realizations drawn from a zero-mean Gaussian process.

A. Obtaining the MLE estimate of observed sources

To claim an observation, we define a detectability criterion based on the optimal signal-to-noise ratio (SNR) of the signal. Explicitly, we set the criterion $\rho_{\text{opt}}^2 := \langle \vec{h}(\vec{\theta}^*) | \vec{h}(\vec{\theta}^*) \rangle \geq \rho_{\text{thresh}}^2$. Here $\langle \cdot | \cdot \rangle$ is the detector-noise-weighted inner-product [51, 52], and $\vec{h}$ is the GW template evaluated at $\vec{\theta}^*$. ρ_{thresh} is the fiducial choice of the SNR threshold, which we will quantify below.

The value of the MLE $\hat{\theta}_i$ for the i^{th} observed source depends on the hypothesis assumed in the analysis. If the correct hypothesis $\mathcal{H}^*$ is assumed, the MLE is simply $\hat{\theta}_i = \vec{\theta}_i^*$. Otherwise, the MLE will be systematically biased away from the truth [52–54]. Assuming the linearized-signal approximation (LSA) [55], we can approximate these biases by the linear-bias approximation for nested

parameter	lower bound	upper bound
$\ln M$	$\log(10^{5.5})$	$\log(10^{6.5})$
z	0.01	1.0
A_l	0.0	10^{-5}
n_l	-20.0	20.0
A_g	-5×10^{-12}	5×10^{-12}

TABLE I: Inference bounds on the source parameters. Given a hypothesis, any source in the true population with an MLE that lies beyond these bounds effectively remains undetected.

models [32]. We provide its explicit form in Eq. B1 of the Appendix B1.

Since the LSA (and hence the linear-bias approximation) holds for “large” SNRs, we can impose that it is valid for the observed sources by setting a similarly large ρ_{thresh} . We choose $\rho_{\text{thresh}} = 20$, which is commonly chosen in the literature (see, e.g., [51, 52, 55]).

In realistic inference setups, the employed sampling scheme may be limited to a small region of the parameter space for efficiency. Thus, given a hypothesis, any source with an MLE (as estimated by the linear-bias approximation) that lies beyond this region will effectively remain undetected. To include this effect in our analysis, we set fiducial lower and upper bounds $[a_k, b_k]$ on all $k = 1, \dots, D$ inferred model parameters, and omit any sources in the observed population for which the MLE is outside these bounds.² We describe the chosen bounds for our analysis in Table I. In practice, we found that reasonably changing these bounds does not significantly affect our results.

B. Approximate hyperlikelihoods

To evaluate the Bayes factor in Eq. (8), we must calculate the full-space integral over the source parameters $\vec{\theta}$ (7) to obtain the single source hyperlikelihood (3) in the alternative hypothesis. In general, evaluating this integral can be computationally expensive or practically infeasible. This is especially true in realistic inference setups in GW data analysis, which involve a high-dimensional parameter space.

² When the local effect amplitude A_ℓ is small, the linear-bias approximation may predict the slope n_ℓ ’s MLE estimate to be significantly beyond any realistic bounds, due to its near-zero information content. Thus, for small A_ℓ , we will exempt n_ℓ from the inference bounds rule (in such cases, n_ℓ is uninformative anyway and hence does not contribute to the hyperlikelihood’s value).

With the framework constructed above, we show in Appendices B2, B3, and B4 that Eq. (7) can be analytically approximated under each of the three hypotheses $\mathcal{H}_v$, $\mathcal{H}_\ell$, and $\mathcal{H}_g$, respectively. Our results make use of the LSA to approximate the individual source likelihood in a given hypothesis $\mathcal{H}$ by a normal distribution,

$$p(\hat{\theta}_i | \vec{\theta}, \mathcal{H}) \approx \mathcal{N}(\vec{\theta} | \hat{\theta}_i, \Gamma_i^{-1}). \quad (15)$$

Here, Γ_i is the Fisher information matrix (FIM), with elements

$$(\Gamma_i)_{km} := \langle \partial_k h | \partial_m h \rangle_{\hat{\theta}_i} \approx \langle \partial_k h | \partial_m h \rangle_{\vec{\theta}_i^*} \quad (16)$$

and $\partial_k h := \partial h / \partial \vec{\theta}^k$.

We also validate the constructed approximate hyperlikelihoods against the direct Monte Carlo integration of Eq. (7). We describe the validation scheme in detail in Appendix B5. As shown in Fig. 6 of the Appendix, the results are consistent up to stochastic errors in the Monte Carlo integration.

V. RESULTS

With the hierarchical framework developed in Sec. II, the description of the EMRI model in Sec. III, and the data analysis setup in Sec. IV, we now perform hypothesis tests of beyond-vacuum-GR effects in various example cases. In the first subsection, we study how well our framework can recover the true hypothesis under which a given EMRI population was generated, considering three different population sizes. Then, in the second subsection, we show the dependence of the Bayes factors on the true local and global effect hyperparameters f^* and $\dot{G}^*$ in the population. This will allow us to comment on the robustness of our framework, particularly in disentangling these effects. All results presented below are produced using the `Hierarchical` package, publicly available on Github (<https://github.com/perturber/Hierarchical/>).

A. Fixed population studies

1. Setup

Population and source parameters — We construct three EMRI populations P_v , P_ℓ , and P_g , generated assuming $\mathcal{H}_v$, $\mathcal{H}_\ell$, and $\mathcal{H}_g$, as the true hypothesis, respectively. In addition, we consider a fourth population P_{mix} , in which the (global) time-varying G effect is present in all sources and a fraction f of the population is simultaneously affected by (local) migration torques. The true hyperparameters in all four populations are given in Table II in the Appendix.

In P_ℓ and P_{mix} , we choose the mean of the local effect amplitude and slope as $\mu_{A_\ell}, \mu_{n_\ell} = (10^{-6}, 8)$. μ_{n_ℓ} is chosen as the value of the migration torque slope for an α -disk model and μ_{A_ℓ} is chosen to be an order-of-magnitude smaller than typical values of the effective migration torque amplitudes $\sim 10^{-5}$ [24]. We choose a smaller value to ensure that the linear-bias approximation, valid for small corrections, conservatively holds for all generated sources. The fractional size of the population with the local effect is fixed to $f = 0.5$. In P_g and P_{mix} , the global effect hyperparameter $\dot{G}^*$ is chosen to be $10^{-12} [\text{yr}^{-1}]$. This value is consistent with current bounds on $\dot{G} \lesssim 10^{-12} [\text{yr}^{-1}]$ [41].

Given the true hyperparameters, we consider EMRI populations of three sizes, with $N_{\text{pop}} = 100, 500$, and 1000 sources. The inferred source parameters $\vec{v} = \{\ln M, z\}$, $\vec{\ell} = \{A_\ell, n_\ell\}$, and $\vec{g} = \{A_g\}$ are drawn from their respective prior pdfs, i.e. Eqs. (11), (13), and (14). Since the priors on $\ln M, z$ are improper, we additionally set minimum and maximum cutoffs on their ranges. We choose $M \in [10^{5.5}, 10^{6.5}]$, and $z \in [0.01, 1.0]$. The luminosity distance, $d_L(z)$, of the source is computed from the drawn redshifts z for the assumed values of H_0 , $\Omega_{m,0}$, and $\Omega_{\Lambda,0}$.

The remaining (uninferred) source parameters, $\mu, a, p_0, \theta_S, \phi_S, \theta_K, \phi_K, \Phi_0, n_g$, are drawn from some fiducial priors, as summarized in Table III in the Appendix. All except p_0 and n_g , i.e., the initial semi-latus rectum of the CO's orbit and the global effect slope, respectively, are drawn uniformly. p_0 is set to $p_0(T_{\text{plunge}})$ where the CO with an initial semi-latus rectum $p_0(T_{\text{plunge}})$ plunges into the MBH after a randomly chosen time $T_{\text{plunge}} \sim \mathcal{U}[0.5, 2]$ years; $n_g = 4$ is fixed in all sources because the chosen global effect, i.e., time-varying G is a -4 PN order effect [29, 41]; We thus obtain the parameter set $\{\vec{\theta}^*\}$ of all N_{pop} sources in the population.

Detector response and GW observables — For a given source with true parameters $\vec{\theta}^*$, we generate a time-domain GW waveform $\vec{h}(\vec{\theta}^*) := \vec{h}_+(\vec{\theta}^*) - i\vec{h}_\times(\vec{\theta}^*)$ in the solar system barycenter (SSB) frame. We use the adiabatic Kerr equatorial inspiral model **FastKerrEccentricEquatorial** of the **FASTEMRI-WAVEFORMS** (FEW) package [56–59]. Additionally, we fix the detector's observation window to 1 year, such that sources that plunge earlier are observed until merger. Otherwise, their waveforms are cut off at the 1-year mark.

This SSB frame waveform is then modulated by the LISA response function. Working in the 1st-generation time-delay interferometry (TDI) configuration with equal armlengths, we can express the GW signal as three TDI observables, A , E , and T [60, 61]. However, the T channel can be dropped because its GW strain content is negligible compared to the A and E channels in this configuration.³ We use the **FastLISAResponse** module of the **LISA-ON-GPU** package [62] to obtain the A , E observables

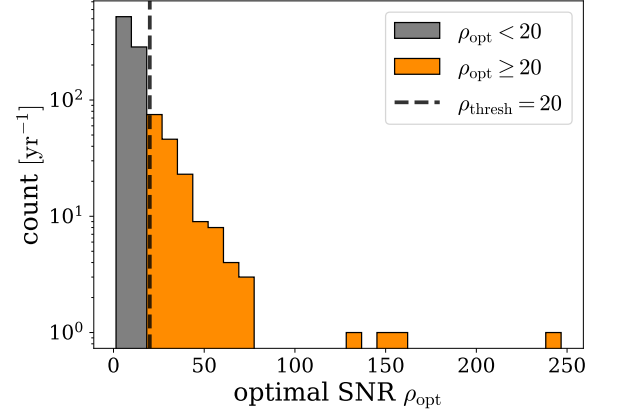


FIG. 1: The distribution of optimal SNR ρ_{opt} in the vacuum-GR population P_v for the $N_{\text{pop}} = 1000$ case. SNRs $\rho_{\text{opt}} \geq 20$ are highlighted in orange, representing $\approx 17\%$ of all sources.

$:= \vec{h}_{A,E}(\vec{\theta}^*)$.

Detected sources and MLE estimates — Of the N_{pop} GW signals generated above, only $N_{\text{obs}} < N_{\text{pop}}$ will be observed. They satisfy the detectability criterion (Sec. IV)

$$\rho_{\text{opt}}^2 = \sum_{C=A,E} \langle \vec{h}_C(\vec{\theta}^*) | \vec{h}_C(\vec{\theta}^*) \rangle \geq \rho_{\text{thresh}}^2. \quad (17)$$

This constitutes the set of $i = 1, \dots, N_{\text{obs}}$ observed sources in the population. For the i^{th} source, we calculate the FIM in the full parameter space, Γ_i , using the STABLEM-RIFISHER (SEF) package [63]. The corresponding MLE estimate, $\hat{\theta}_i$, is obtained assuming the linear-bias approximation, Eq. (B1).

With the set of MLEs $\{\hat{\theta}_i\}$ in hand, we can now calculate the Bayes factor from Eq. (8). Explicitly, the Bayes factors comparing the null vacuum-GR hypothesis against the alternative local effect and global effect hypotheses are, respectively,

$$\mathcal{B}_\ell^v = \frac{1}{\pi(f=0|\mathcal{H}_\ell)} \frac{\prod_i p(\hat{\theta}_i|f=0, \mathcal{H}_\ell)}{\int df \prod_i p(\hat{\theta}_i|f, \mathcal{H}_\ell)}, \quad (18)$$

$$\mathcal{B}_g^v = \frac{1}{\pi(\dot{G}=0|\mathcal{H}_g)} \frac{\prod_i p(\hat{\theta}_i|\dot{G}=0, \mathcal{H}_g)}{\int d\dot{G} \prod_i p(\hat{\theta}_i|\dot{G}, \mathcal{H}_g)}. \quad (19)$$

In practice, the integral in the denominator of these expressions can be approximated by a Monte Carlo integral with N random draws of $f, \dot{G}$ from their respective hyperpriors $\pi(\vec{\lambda}|\mathcal{H})$. Our choice of hyperpriors for all modeled hyperparameters is summarized in Table IV. The Bayes factor for the global effect hypothesis over the local one is given simply by the ratio $\mathcal{B}_g^v = \mathcal{B}_\ell^v / \mathcal{B}_g^v$.

³ Remember that we assumed a zero detector-noise realization in

Sec. III A such that the noise information in the T channel is also zero.

2. Results

With an SNR threshold of $\rho_{\text{thresh}} = 20.0$, we find that only about 17% of the sources are observable in all four populations, P_v , P_ℓ , P_g , and P_{mix} . We present the SNR distribution for the vacuum-GR population P_v with $N_{\text{pop}} = 1000$ sources in Fig. 1.

To infer the hyperposteriors under the local effect hypothesis, we draw $N = 5000$ samples of the vacuum-GR and local effect hyperparameters $\vec{\lambda}_v = \{K, \alpha, \beta\}$ and $\vec{\lambda}_\ell = \{f, (\mu_{A_\ell}, \mu_{n_\ell}), (\sigma_{A_\ell}, \sigma_{n_\ell})\}$ from narrow uniform priors around the true value, and weight them by the approximate hyperlikelihood in the local effect hypothesis, given by Eqs. (B10), (B18), and (B25). The marginalized hyperposteriors of f for all four populations P_v , P_ℓ , P_g , and P_{mix} in the $N_{\text{pop}} = 1000$ case are plotted in the top four panels of Fig. 2. In P_v and P_ℓ , we find that the expectation $\langle f \rangle$ of f (dashes vertical line) lies within $\sim 1\sigma$ of the true value f^* (solid vertical line), indicating a robust recovery of the fractional population with migration torques, i.e. the local effect, in these populations. However, in P_g and P_{mix} , we recover $|\langle f \rangle - f^*| \approx 0.06$ representing biases $\sim 3.5\sigma$ and $\sim 1.5\sigma$ for the two populations, respectively. These systematic biases are a direct consequence of correlations between the migration torque parameters A_ℓ, n_ℓ with the intrinsic strength of the global effect $\dot{G}$ [yr $^{-1}$]. These correlations induce biases in the MLE estimates $\hat{A}_\ell, \hat{n}_\ell$ in all sources when analyzed under the local effect hypothesis. The extent to which these biases are measurable is reflected in the recovered value of $\langle f \rangle$. Similar results are obtained in the $N_{\text{pop}} = 100, 500$ cases.

Similarly, for the global effect hypothesis, we obtain 5000 samples from narrow uniform priors of $\vec{\lambda}_g$ and $\vec{\lambda}_g = \{\dot{G}\}$ around the true value and weight them by the corresponding hyperlikelihood (B31). The marginalized hyperposteriors of $\dot{G}$ in the four populations for $N_{\text{pop}} = 1000$ are plotted in the bottom panel of Fig. 2. We find that the expectation $\langle \dot{G} \rangle$ is consistent at $\lesssim 0.01\sigma$ with the true value $\dot{G}^*$ in all four populations in the $N_{\text{pop}} = 1000$ case. Notably, the width of the hyperlikelihood on $\dot{G}$ is broader by a factor of 2 in P_ℓ and P_{mix} . This broadening is induced by the systematic biases in the MLE estimate of the source parameter $A_{g,i}$ due to non-null migration torques in these populations. Since these biases are independent, they average out over all observed sources, but still contribute to the measurement uncertainty of $\dot{G}$. This is also because the local effect is present in only $f = 0.5$ fraction of the sources, while all other sources have the correct MLE estimate, $A_{g,i} = \dot{G}^*$, such that the average over all observed sources is closer to the truth. Similar results are obtained in the $N_{\text{pop}} = 100, 500$ cases.

We now calculate the Bayes factors $\mathcal{B}_\ell^v$ and $\mathcal{B}_g^v$ according to Eqs. (18) and (19) (and $\mathcal{B}_\ell^g = \mathcal{B}_\ell^v/\mathcal{B}_g^v$). We again draw $N = 5000$ samples of the vacuum-GR, local effect, and global effect hyperparameters, $\vec{\lambda}_v, \vec{\lambda}_\ell, \vec{\lambda}_g$, respectively, this

time from their full priors (Table IV), and weight them by the approximate hyperlikelihoods calculated in the appendices B2, B3, and B4 for the vacuum-GR, local effect, and global effect hypotheses, respectively. We plot the log Bayes factors for all four populations and the different population sizes in Fig. 3. For brevity, we notate, e.g., $\log_{10} \mathcal{B}_\ell^v$ as v/ℓ in this plot. We find that the correct hypothesis is always preferred in all of the populations. Notably, in P_g , the local effect hypothesis is also preferred over vacuum-GR even though the true population does not have the local effect. This is again explained by the systematic biases induced on the MLE estimates of A_ℓ, n_ℓ due to a non-null $\dot{G}$ in this population. The two hypotheses $\mathcal{H}_\ell, \mathcal{H}_g$ can still be disentangled by calculating $\mathcal{B}_\ell^g$, found in this case to be extremely favorable for the (correct) global effect hypothesis. This is represented by the “ g/ℓ ” bars in the third panel. The effect of the size of the EMRI population is most evident in the calculation of $\mathcal{B}_\ell^g$ in the population P_{mix} , where the preference for the global effect hypothesis over the local one flips when N_{pop} is small. This is because with fewer sources in the population, the measurement uncertainty on $\dot{G}$ is larger, and hence increasingly consistent with the null value.

B. Robustness check

To more broadly comment on the robustness of the hierarchical modeling framework for decoupling the considered beyond-vacuum-GR effects, i.e., migration torques and time-varying G , we now study the variability of the Bayes factor $\mathcal{B}_\ell^g$ as a function of the population’s hyperparameters.

We consider 11 populations of fixed size $N_{\text{pop}} = 500$. In the first check, we fix $\dot{G}^* = 10^{-12}$ [yr $^{-1}$] and vary f^* linearly in the range $[0.0, 1.0]$. The setup otherwise is equivalent to that in the previous subsection. As shown in the top panel of Fig. 4, $\log_{10}(\mathcal{B}_\ell^g)$, which quantifies the data’s preference for the global effect hypothesis over the local, scales inversely with f^* . This behavior is expected: as f^* increases in the population, the MLE estimate of the global effect’s source parameter $\hat{A}_g$ is more biased among sources, and the inferred uncertainty in the hyperparameter $\dot{G}$ grows. This makes the hyperposterior on $\dot{G}$ more consistent with the null value, and hence the preference of the data for the global effect drops. We also calculate the relative error $r_{\dot{G}} := |\dot{G}^* - \langle \dot{G} \rangle|/\dot{G}^*$ in all 11 populations. For $f \geq 0.9$, we find that $r_{\dot{G}} \gtrsim 1.0$, i.e., the expectation value $\langle \dot{G} \rangle$ incurs significant biases with respect to the truth $\dot{G}^*$. This is due to the biases induced by the local effect across a larger fraction of the population.

In the second check, we fix $f^* = 0.5$ and vary $\dot{G}^*$ linearly in the range $[10^{-13}, 10^{-12}]$ across the 11 populations. As shown in the bottom panel of Fig. 4, $\log_{10} \mathcal{B}_\ell^g$ scales proportionately with the global effect hyperparameter $\dot{G}^*$. At smaller values, $\dot{G}^* \lesssim 4 \times 10^{-13}$ [yr $^{-1}$], $\log_{10} \mathcal{B}_\ell^g \lesssim 0.0$, i.e., the global effect hypothesis is no longer favored. This

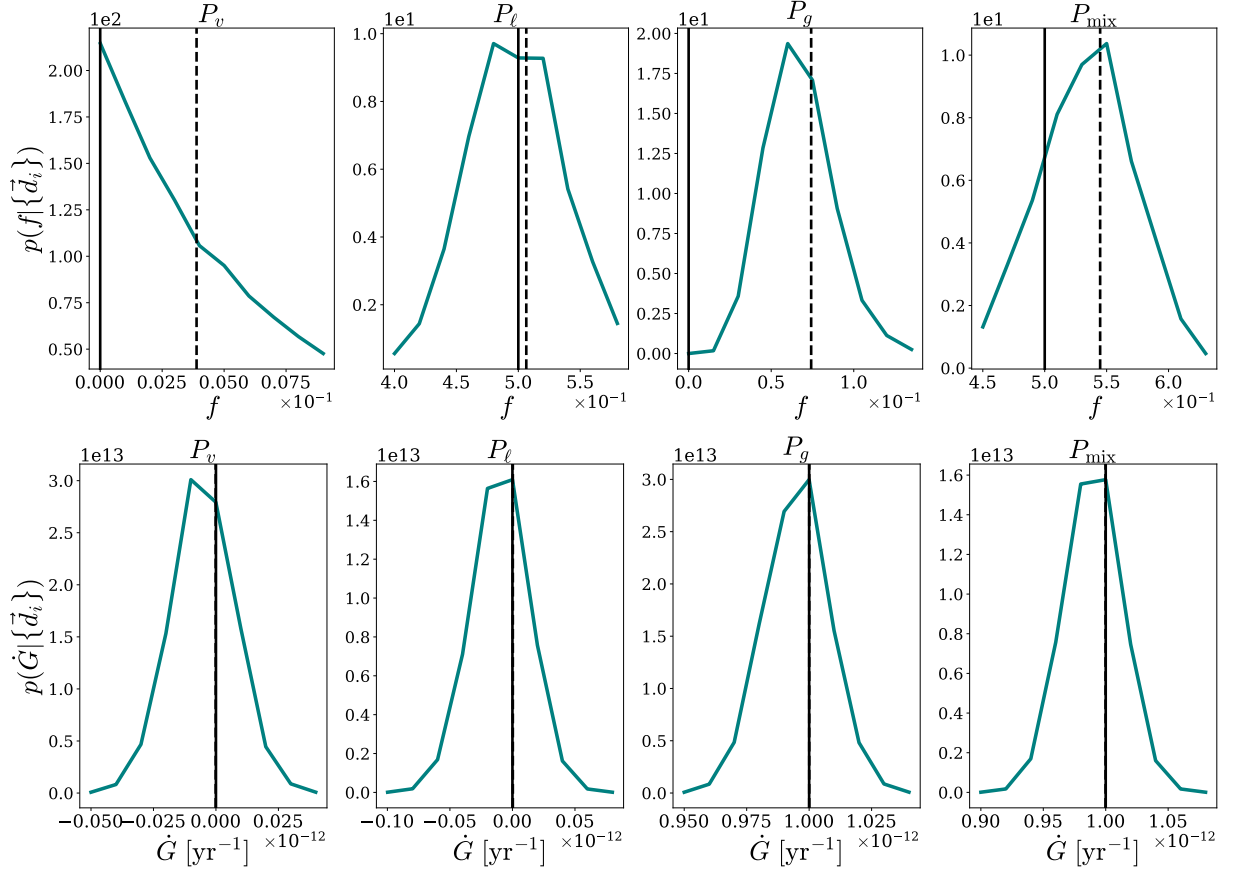


FIG. 2: Marginalized hyperposteriors on f (top panels) and $\dot{G}$ (bottom panels) constructed from $N = 5000$ samples from the analysis priors weighted by the hyperposterior densities. We plot four populations, namely, P_v , P_ℓ , P_g , and P_{mix} (left to right) with population size $N_{\text{pop}} = 1000$. The solid vertical line in each plot represents the true value of f or $\dot{G}$ while the dashed vertical line represents the sample expectation. f is found to be systematically biased by $\sim 3.5\sigma$ in P_g and $\sim 1.5\sigma$ in P_{mix} , while $\dot{G}$ is recovered consistently with the truth at $\lesssim 0.01\sigma$ in all cases.

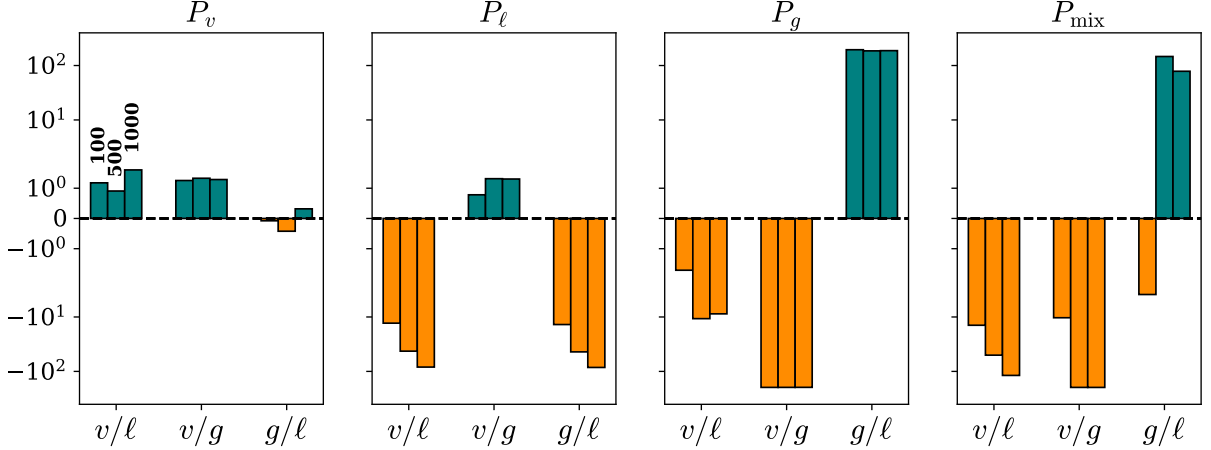


FIG. 3: Bayes factors $\mathcal{B}_\ell^v$, $\mathcal{B}_g^v$, and $\mathcal{B}_\ell^g$ in the considered populations P_v (left panel), P_ℓ (left-center panel), P_g (right-center panel), and P_{mix} (right panel). For brevity, we notate, e.g., $\log_{10} \mathcal{B}_\ell^v$ as v/ℓ , which is to be understood as the preference for the hypothesis $\mathcal{H}_v$ over $\mathcal{H}_\ell$. The left, center, and right vertical bars in each panel represent the Bayes factors for the three different population sizes, $N_{\text{pop}} = 100, 500$, and 1000 , respectively. The horizontal dashed line in each panel represents a log Bayes factor of zero.

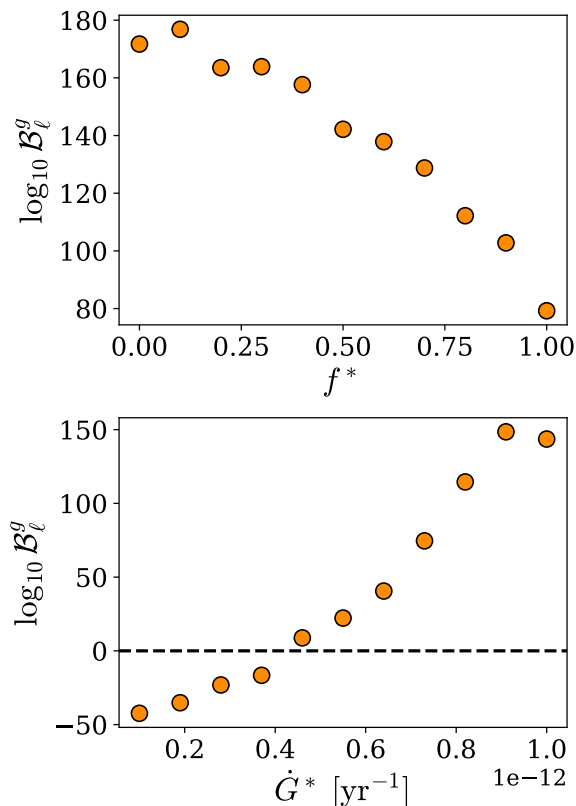


FIG. 4: The log-Bayes factor $\log_{10} \mathcal{B}_\ell^g$ as a function of f^* (top panel) and $\dot{G}^*$ (bottom panel) in 11 different populations of size $N_{\text{pop}} = 500$. In the top panel, $\dot{G}^* = 10^{-12} [\text{yr}^{-1}]$ is fixed. In the bottom panel, we fix $f^* = 0.5$ instead. In the bottom panel, the dashed horizontal line represents $\log_{10} \mathcal{B}_\ell^g = 0.0$, i.e., for which the data does not favour either of the hypotheses.

is because the data is less sensitive to the global effect if it is too weak, such that the measurement uncertainty in $\dot{G}$ grows. Effectively, this analysis suggests that only population realizations with intrinsic strength of the global effect $\dot{G}^* \gtrsim 4 \times 10^{-13} [\text{yr}^{-1}]$ will be discernible in this framework. We highlight that, at the zero crossing point of $\log_{10} \mathcal{B}_\ell^g$, even though both migration torques and time-varying G are present in the population, they effectively “cancel” each other out. In such realizations of the population, the hierarchical framework will struggle to disentangle the two effects without supplementary checks.

VI. DISCUSSION

A. Summary

In this paper, we developed a hierarchical framework for GW populations with beyond-vacuum-GR effects, with the motivation to study how well local effects like astrophysical environments, which may be present in only a

fraction of the population, can be decoupled from global ones like modified-GR contributions, which are expected to be present in all sources simultaneously. We consider specific models of local and global beyond-vacuum-GR effects: for the local effect, we choose migration torques induced by an accretion disk around the MBH. For the global effect, we model time-varying G , a modified gravity effect present universally with a constant amplitude $\dot{G} [\text{yr}^{-1}]$.

Considering four simulated EMRI populations, P_v , P_ℓ , P_g , and P_{mix} and three different population sizes $N_{\text{pop}} = 100, 500, 1000$, we find that the framework consistently recovers the fiducial true hypothesis that was used in constructing these populations, and even indicates whether both local and global effects are present in the population simultaneously. The recovered value of the global effect hyperparameter $\dot{G}$, i.e., the intrinsic strength of the time-varying G effect, is consistent with the truth $\dot{G}^*$ at $\lesssim 0.01\sigma$ in all populations. However, the recovered value of the local effect hyperparameter f , i.e., the fractional population with migration torques, incurs systematic biases $\sim 3\sigma$ and $\sim 1.5\sigma$ compared to the true value f^* in P_g and P_{mix} , respectively, for all three population sizes $N_{\text{pop}} = 100, 500, 1000$.

We check that the Bayes factor $\mathcal{B}_\ell^g$ scales proportionally to the intrinsic strength $\dot{G}^* [\text{yr}^{-1}]$ of the global effect in the population, and inversely with the fraction of sources affected by the local effect, f^* . We also find that the expectation of the global effect hyperparameter $\langle \dot{G} \rangle$ may significantly deviate from the ground truth if $f^* \geq 0.9$. We further highlighted a scenario where the Bayes factor $\log_{10} \mathcal{B}_\ell^g \approx 0.0$ even when both effects are present in the population, such that the ability to tell them apart with the data is lost. Future studies should study such situations in more detail to further assess the robustness of the hierarchical framework.

B. Future directions and outlook

While we only consider circular-equatorial Kerr EMRIs in this paper, future studies should study the robustness of the hierarchical framework for beyond-vacuum-GR effects in more generic orbits, i.e., including the effects of eccentricity and inclination in the data. Such extensions may reduce parameter space degeneracies and, therefore, the biases incurred in recovering the local and global effect hyperparameters in mixed populations. More accurate models for EMRI waveforms and beyond-vacuum-GR effects are becoming increasingly available [28, 59, 64–66].

When constructing the populations in Sec. III, we used independent prior pdfs on different classes of effects, and employed the same priors in the analysis setup. This simplification may not represent reality since the true distribution of the source parameters in the universe will not be known a priori. Additionally, both the local and global effect’s strength may depend on the binary’s vacuum-GR properties, such that the distributions are no longer

independent. For more extensive applications of the hierarchical framework, future studies should investigate cases where all population parameters are realistically correlated and are drawn from distributions that differ from the priors used in the analysis.

A broader challenge for the hierarchical modeling framework lies in incorporating it within the proposed global fit pipeline for LISA data analysis [67–69]. The global fit aims to infer all source types in the data stream simultaneously, and will be expensive to evaluate even in the simplest vacuum-GR hypothesis $\mathcal{H}_v$. Redoing the global fit for the alternative hypotheses $\mathcal{H}_\ell$ and $\mathcal{H}_g$ (which also include additional parameters) may not be practically feasible. Yet, it may be possible to perform hypothesis tests in a post-processing style analysis by bias-correcting and importance-weighting the samples obtained in the $\mathcal{H}_v$ analysis [70]. This procedure effectively provides a new set of samples from the hyperposterior in the alternative hypothesis and is particularly well-suited for beyond-vacuum-GR type effects (see [70] for more details). Recently, simulation-based inference (SBI) techniques have been proposed for feasible hierarchical modeling of GW sources [71], and may enable efficient modeling and inference of beyond-vacuum-GR effects using source populations.

Finally, we caution that some systematic biases may persist in the hierarchical framework, e.g., due to inaccuracies in waveform modeling which can mimic global

effects, as discussed in [72], and should be addressed in future work. In our framework, these systematic biases can be understood to be induced by “local” waveform modeling inaccuracies, which may strongly depend on the source parameters [73], biasing the inference of the intrinsic strength $\dot{G}$ of the global effect. Our setup can be readily employed to study such systematics in the future. Extensions of the current framework, shaped by these careful considerations, will pave the way for robust and realistic tests of GR in the presence of astrophysical environments with the next generation of GW detectors.

ACKNOWLEDGMENTS

SK thanks Patrick Meyers and Michele Vallisneri for useful discussions during a research visit at ETH Zurich. SK also acknowledges the computing resources provided by the NUS IT Research Computing group and the support of the NUS Research Scholarship. EB acknowledges support from the European Union’s Horizon ERC Synergy Grant “Making Sense of the Unexpected in the Gravitational-Wave Sky” (Grant No. GWSky-101167314) and the PRIN 2022 grant “GUVIRP - Gravity tests in the UltraViolet and InfraRed with Pulsar timing”. This work has been supported by the Agenzia Spaziale Italiana (ASI), Project n. 2024-36-HH.0, “Attività per la fase B2/C della missione LISA”.

-
- [1] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].
 - [2] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. X* **9**, 031040 (2019), arXiv:1811.12907 [astro-ph.HE].
 - [3] R. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. X* **11**, 021053 (2021), arXiv:2010.14527 [gr-qc].
 - [4] R. Abbott *et al.* (LIGO Scientific, VIRGO), *Phys. Rev. D* **109**, 022001 (2024), arXiv:2108.01045 [gr-qc].
 - [5] R. Abbott *et al.* (KAGRA, VIRGO, LIGO Scientific), *Phys. Rev. X* **13**, 041039 (2023), arXiv:2111.03606 [gr-qc].
 - [6] (2025), arXiv:2508.18082 [gr-qc].
 - [7] (2025), arXiv:2508.18083 [astro-ph.HE].
 - [8] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. D* **100**, 104036 (2019), arXiv:1903.04467 [gr-qc].
 - [9] R. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. D* **103**, 122002 (2021), arXiv:2010.14529 [gr-qc].
 - [10] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2112.06861 [gr-qc].
 - [11] S. Hild *et al.*, *Class. Quant. Grav.* **28**, 094013 (2011), arXiv:1012.0908 [gr-qc].
 - [12] B. P. Abbott *et al.* (LIGO Scientific), *Class. Quant. Grav.* **34**, 044001 (2017), arXiv:1607.08697 [astro-ph.IM].
 - [13] D. Reitze *et al.*, *Bull. Am. Astron. Soc.* **51**, 035 (2019), arXiv:1907.04833 [astro-ph.IM].
 - [14] N. Singh, T. Bulik, K. Belczynski, and A. Askar, *Astron. Astrophys.* **667**, A2 (2022), arXiv:2112.04058 [astro-ph.HE].
 - [15] K. K. Y. Ng, S. Vitale, W. M. Farr, and C. L. Rodriguez, *Astrophys. J. Lett.* **913**, L5 (2021), arXiv:2012.09876 [astro-ph.CO].
 - [16] J. R. Gair, I. Mandel, M. C. Miller, and M. Volonteri, *Gen. Rel. Grav.* **43**, 485 (2011), arXiv:0907.5450 [astro-ph.CO].
 - [17] M. Colpi *et al.* (LISA), (2024), arXiv:2402.07571 [astro-ph.CO].
 - [18] J. R. Gair, A. Sesana, E. Berti, and M. Volonteri, *Class. Quant. Grav.* **28**, 094018 (2011), arXiv:1009.6172 [gr-qc].
 - [19] S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. L. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein, *Phys. Rev. D* **95**, 103012 (2017), arXiv:1703.09722 [gr-qc].
 - [20] J. R. Gair, C. Tang, and M. Volonteri, *Phys. Rev. D* **81**, 104014 (2010), arXiv:1004.1921 [astro-ph.GA].
 - [21] E. Barausse, L. Rezzolla, D. Petroff, and M. Ansorg, *Phys. Rev. D* **75**, 064026 (2007), arXiv:gr-qc/0612123.
 - [22] E. Barausse and L. Rezzolla, *Phys. Rev. D* **77**, 104027 (2008), arXiv:0711.4558 [gr-qc].
 - [23] B. Kocsis, N. Yunes, and A. Loeb, *Phys. Rev. D* **84**, 024032 (2011), arXiv:1104.2322 [astro-ph.GA].
 - [24] E. Barausse, V. Cardoso, and P. Pani, *Phys. Rev. D* **89**, 104059 (2014), arXiv:1404.7149 [gr-qc].
 - [25] E. Barausse, N. Yunes, and K. Chamberlain, *Phys. Rev. Lett.* **116**, 241104 (2016), arXiv:1603.04075 [gr-qc].
 - [26] K. G. Arun *et al.* (LISA), *Living Rev. Rel.* **25**, 4 (2022), arXiv:2205.01597 [gr-qc].
 - [27] P. S. Cole, G. Bertone, A. Coogan, D. Gaggero, T. Karydas, B. J. Kavanagh, T. F. M. Spieksma, and G. M.

- Tomaselli, *Nature Astron.* **7**, 943 (2023), arXiv:2211.01362 [gr-qc].
- [28] F. Duque, S. Kejriwal, L. Sberna, L. Speri, and J. Gair, *Phys. Rev. D* **111**, 084006 (2025), arXiv:2411.03436 [gr-qc].
- [29] L. Speri, A. Antonelli, L. Sberna, S. Babak, E. Barausse, J. R. Gair, and M. L. Katz, *Phys. Rev. X* **13**, 021035 (2023), arXiv:2207.10086 [gr-qc].
- [30] L. Speri, S. Barsanti, A. Maselli, T. P. Sotiriou, N. Warburton, M. van de Meent, A. J. K. Chua, O. Burke, and J. Gair, (2024), arXiv:2406.07607 [gr-qc].
- [31] L. Copparoni, E. Barausse, L. Speri, L. Sberna, and A. Derdzinski, *Phys. Rev. D* **111**, 104079 (2025), arXiv:2502.10087 [gr-qc].
- [32] S. Kejriwal, L. Speri, and A. J. K. Chua, (2023), arXiv:2312.13028 [gr-qc].
- [33] A. Cárdenas-Avedaño and C. F. Sopuerta, “Testing Gravity with Extreme-Mass-Ratio Inspirals,” (2024) arXiv:2401.08085 [gr-qc].
- [34] T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, and A. Vecchio, *Phys. Rev. D* **85**, 082003 (2012), arXiv:1110.0530 [gr-qc].
- [35] M. Isi, K. Chatziioannou, and W. M. Farr, *Phys. Rev. Lett.* **123**, 121101 (2019), arXiv:1904.08011 [gr-qc].
- [36] C. Liu, D. Laghi, and N. Tamanini, *Phys. Rev. D* **109**, 063521 (2024), arXiv:2310.12813 [astro-ph.CO].
- [37] X. Yuan, J.-d. Zhang, and J. Mei, *Phys. Rev. D* **111**, 104050 (2025), arXiv:2412.00915 [gr-qc].
- [38] X. Yuan, (2025), arXiv:2506.22921 [gr-qc].
- [39] B. Bonga, H. Yang, and S. A. Hughes, *Phys. Rev. Lett.* **123**, 101103 (2019), arXiv:1905.00030 [gr-qc].
- [40] N. Yunes, F. Pretorius, and D. Spergel, *Phys. Rev. D* **81**, 064018 (2010), arXiv:0912.2724 [gr-qc].
- [41] K. Chamberlain and N. Yunes, *Phys. Rev. D* **96**, 084039 (2017), arXiv:1704.08268 [gr-qc].
- [42] C. Brans and R. H. Dicke, *Phys. Rev.* **124**, 925 (1961).
- [43] T. P. Sotiriou and S.-Y. Zhou, *Phys. Rev. D* **90**, 124063 (2014), arXiv:1408.1698 [gr-qc].
- [44] A. Dima, E. Barausse, N. Franchini, and T. P. Sotiriou, *Phys. Rev. Lett.* **125**, 231101 (2020), arXiv:2006.03095 [gr-qc].
- [45] A. Maselli, N. Franchini, L. Gualtieri, and T. P. Sotiriou, *Phys. Rev. Lett.* **125**, 141101 (2020), arXiv:2004.11895 [gr-qc].
- [46] I. Mandel, W. M. Farr, and J. R. Gair, *Monthly Notices of the Royal Astronomical Society* **486**, 1086–1093 (2019).
- [47] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science (Taylor & Francis, 2013).
- [48] J. M. Dickey, *The Annals of Mathematical Statistics* **42**, 204 (1971).
- [49] C. E. A. Chapman-Bird, C. P. L. Berry, and G. Woan, *Mon. Not. Roy. Astron. Soc.* **522**, 6043 (2023), arXiv:2212.06166 [astro-ph.HE].
- [50] D. Baumann, *Cosmology* (Cambridge University Press, 2022).
- [51] L. S. Finn, *Phys. Rev. D* **46**, 5236 (1992), arXiv:gr-qc/9209010.
- [52] C. Cutler and E. E. Flanagan, *Phys. Rev. D* **49**, 2658 (1994), arXiv:gr-qc/9402014.
- [53] C. Cutler and M. Vallisneri, *Phys. Rev. D* **76**, 104018 (2007), arXiv:0707.2982 [gr-qc].
- [54] E. E. Flanagan and S. A. Hughes, *Phys. Rev. D* **57**, 4566 (1998), arXiv:gr-qc/9710129.
- [55] M. Vallisneri, *Phys. Rev. D* **77**, 042001 (2008), arXiv:gr-qc/0703086.
- [56] A. J. K. Chua, M. L. Katz, N. Warburton, and S. A. Hughes, *Phys. Rev. Lett.* **126**, 051102 (2021), arXiv:2008.06071 [gr-qc].
- [57] M. L. Katz, A. J. K. Chua, L. Speri, N. Warburton, and S. A. Hughes, *Phys. Rev. D* **104**, 064047 (2021), arXiv:2104.04582 [gr-qc].
- [58] L. Speri, M. L. Katz, A. J. K. Chua, S. A. Hughes, N. Warburton, J. E. Thompson, C. E. A. Chapman-Bird, and J. R. Gair, *Front. Appl. Math. Stat.* **9** (2024), 10.3389/fams.2023.1266739, arXiv:2307.12585 [gr-qc].
- [59] C. E. A. Chapman-Bird *et al.*, (2025), arXiv:2506.09470 [gr-qc].
- [60] M. Vallisneri, *Phys. Rev. D* **71**, 022001 (2005), arXiv:gr-qc/0407102.
- [61] S. V. Dhurandhar, K. Rajesh Nayak, and J. Y. Vinet, *Phys. Rev. D* **65**, 102002 (2002), arXiv:gr-qc/0112059.
- [62] M. L. Katz, J.-B. Bayle, A. J. K. Chua, and M. Vallisneri, *Phys. Rev. D* **106**, 103001 (2022), arXiv:2204.06633 [gr-qc].
- [63] S. Kejriwal, O. Burke, C. Chapman-Bird, and A. J. K. Chua, “Stableemriferfisher (sef),” (manuscript in preparation).
- [64] V. Cardoso, K. Destounis, F. Duque, R. P. Macedo, and A. Maselli, *Phys. Rev. Lett.* **129**, 241103 (2022).
- [65] T. Zi and S. Kumar, (2025), arXiv:2508.00516 [gr-qc].
- [66] F. Duque, L. Sberna, A. Spiers, and R. Vicente, (2025), arXiv:2510.02433 [gr-qc].
- [67] M. Vallisneri, *Class. Quant. Grav.* **26**, 094024 (2009), arXiv:0812.0751 [gr-qc].
- [68] T. B. Littenberg and N. J. Cornish, *Phys. Rev. D* **107**, 063004 (2023), arXiv:2301.03673 [gr-qc].
- [69] M. L. Katz, N. Karnesis, N. Korsakova, J. R. Gair, and N. Stergioulas, (2024), arXiv:2405.04690 [gr-qc].
- [70] S. Kejriwal, F. Duque, A. J. K. Chua, and J. Gair, *Phys. Rev. D* **112**, 024005 (2025), arXiv:2503.01120 [gr-qc].
- [71] R. Srinivasan, E. Barausse, N. Korsakova, and R. Trotta, (2025), arXiv:2506.22543 [astro-ph.GA].
- [72] C. J. Moore, E. Finch, R. Busicchio, and D. Gerosa, *iScience* **24**, 102577 (2021), arXiv:2103.16486 [gr-qc].
- [73] H. Khalvati, P. Lynch, O. Burke, L. Speri, M. van de Meent, and Z. Nasipak, (2025), arXiv:2509.08875 [gr-qc].
- [74] C. Bishop, *Pattern recognition and machine learning*, Vol. 4 (Springer New York, 2006).
- [75] D. W. Hogg, A. M. Price-Whelan, and B. Leistedt, (2020), arXiv:2005.14199 [stat.CO].
- [76] C. Robert and G. Casella, *Monte Carlo statistical methods* (Springer Verlag, 2004).

Appendix A: Selection effects for beyond-vacuum-GR hyperparameters

We denote the null vacuum-GR hypothesis as $\mathcal{H}_v$ and the alternative local and global effect hypotheses jointly as $\mathcal{H}_a = \mathcal{H}_\ell, \mathcal{H}_g$. We decompose the hyperparameter set as $\vec{\lambda} = \{\vec{\lambda}_v, \vec{\lambda}_a\}$, and the source parameters $\vec{\theta} = \{\vec{\theta}_v, \vec{\theta}_a\}$, where a subscript v denotes vacuum-GR and a denotes

beyond-vacuum-GR effect parameters. In this notation, the selection function can be written as [46]

$$\alpha(\vec{\lambda}|\mathcal{H}_a) = \int d\vec{\theta} p_{\text{det}}(\vec{\theta}) \pi(\vec{\theta}|\vec{\lambda}, \mathcal{H}_a) \quad (\text{A1})$$

where $p_{\text{det}}(\vec{\theta})$ is the probability of detecting a source with parameters $\vec{\theta}$, and $\pi(\vec{\theta}|\vec{\lambda}, \mathcal{H}_a)$ is the prior pdf of $\vec{\theta}$ for the given hyperparameter set $\vec{\lambda}$ in the alternative hypothesis.

Since $\mathcal{H}_a$ describes a set of (perturbative) beyond-vacuum-GR effects, we argue that the detection probability is approximately independent of $\vec{\theta}_a$, such that

$$p_{\text{det}}(\vec{\theta}) \approx p_{\text{det}}(\vec{\theta}_v). \quad (\text{A2})$$

Additionally, if we assume that the priors of $\vec{\theta}_v$ and $\vec{\theta}_a$ are independent, i.e., $\pi(\vec{\theta}|\vec{\lambda}, \mathcal{H}_a) = \pi(\vec{\theta}_v|\vec{\lambda}_v)\pi(\vec{\theta}_a|\vec{\lambda}_a)$, we get

$$\alpha(\vec{\lambda}_a|\mathcal{H}_a) \approx \mathcal{C}_v \int d\vec{\theta}_a p(\vec{\theta}_a|\vec{\lambda}_a). \quad (\text{A3})$$

Here, $\mathcal{C}_v := \int d\vec{\lambda}_v d\vec{\theta}_v p_{\text{det}}(\vec{\theta}_v) \pi(\vec{\theta}_v|\vec{\lambda}_v)$ is a constant. Since the remaining integral in Eq. (A3) is independent of the detection probability, it is effectively calculated over the full space of $\vec{\theta}_a$. In this paper, we only consider proper priors for the beyond-vacuum-GR effects, and hence $\int d\vec{\theta}_a \pi(\vec{\theta}_a|\vec{\lambda}_a) = 1$ for arbitrary $\vec{\lambda}_a$. Therefore, the selection function $\alpha(\vec{\lambda}_a|\mathcal{H}_a) \approx \mathcal{C}_v$ is constant.

Appendix B: Derivations of the approximate hyperlikelihoods

In this appendix, we explicitly derive the analytical approximations for the hyperlikelihoods $p(\hat{\theta}_i|\vec{\lambda}, \mathcal{H})$ (7) in the vacuum-GR, local effect, and global effect hypotheses. We validate our construction against the direct Monte Carlo integration of Eq. (7) in the final subsection.

1. Linear-bias approximation

For notational convenience, we introduce an alternate decomposition of the inferred source parameter set $\vec{\theta} := \{\vec{\psi}, \vec{\varphi}\}$. Here, $\vec{\psi}$ is the set of parameters that are inferred in the analysis, and $\vec{\varphi}$ are the remaining model parameters fixed to some null value $\vec{\varphi}_0$ in the analysis. We re-emphasize that $\vec{\varphi}_0$ is different from the true value of the uninferred parameters $\vec{\varphi}^*$ in general. Then, from the linear-bias approximation, the MLE estimate $\hat{\psi}_i$ of the inferred parameters is [32]

$$\hat{\psi}_i \approx \vec{\psi}_i^* + (\Gamma_i^{\psi\psi})^{-1} \cdot \Gamma_i^{\psi\varphi} \cdot (\vec{\varphi}_i^* - \vec{\varphi}_{i,0}) \quad (\text{B1})$$

where Γ_i is the Fisher information matrix (FIM) of the i^{th} observed source. It is calculated at $\vec{\theta}_i^*$. $\Gamma_i^{\psi\psi}$ and $\Gamma_i^{\psi\varphi}$ are submatrices of Γ_i corresponding to the $\vec{\psi}$, $\vec{\psi}$ and $\vec{\psi}$, $\vec{\varphi}$ elements, respectively.

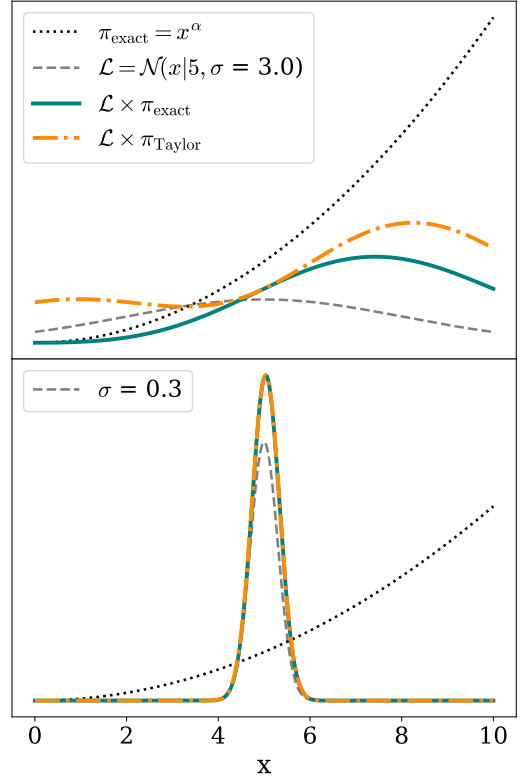


FIG. 5: 1-dimensional visualization of the product of a normal likelihood $\mathcal{L}(x) = \mathcal{N}(x|5, \sigma)$ with a power-law prior on x , $p(x) = x^\alpha$ where $\alpha = 2$ is the slope hyperparameter. In both panels, the green solid curve represents this product in its exact form, while the orange dash-dotted curve represents the approximate product, where $p(x)$ is Taylor-expanded to the second order in $x - 5$ about 5. In the top panel, the normal likelihood has a large standard deviation $\sigma = 3.0$, such that its rate of change is similar to that of the power-law prior, and the Taylor approximation fails. When we set $\sigma = 0.3$ as in the bottom panel, the normal likelihood varies at much smaller scales (as expected for GW data analysis), and the Taylor approximation holds well.

2. Vacuum-GR

In the vacuum-GR hypothesis, $\mathcal{H}_v$, the inferred set of parameters is $\vec{\psi} \equiv \{\vec{v}\}$, while the fixed set is $\vec{\varphi} \equiv \{\vec{\ell}, \vec{g}\}$, which assume the null value $\vec{\varphi}_{i,0} = \{\vec{0}, \vec{0}\}$ in the analysis. Correspondingly, the MLE $\hat{\psi}_i \equiv \hat{v}_i$ for the i^{th} source is obtained from Eq. (B1) and the source likelihood (15) can be written as

$$p(\vec{d}_i|\vec{v}, \mathcal{H}_v) = \mathcal{N}(\vec{v}|\hat{v}_i, (\Gamma_i^{vv})^{-1}). \quad (\text{B2})$$

The prior pdf in the vacuum-GR hypothesis, $\pi(\vec{v}|\vec{\lambda}_v)$ is given in Eq. (11). Substituting in Eq. (7),

$$p(\vec{d}_i|\vec{\lambda}_v, \mathcal{H}_v) = \int d\vec{v} \mathcal{N}(\vec{v}|\hat{v}_i, (\Gamma_i^{vv})^{-1}) \pi(\vec{v}|\vec{\lambda}_v). \quad (\text{B3})$$

While this integral is over the full space of $\vec{v}$, the integrand is a product of a normal distribution centered at $\hat{v}_i$ and a power-law in $\vec{v}$. Thus, at large distances $\Delta\vec{v}_i := \|\vec{v} - \hat{v}_i\| \gg \hat{v}_i$ away from $\hat{v}_i$, the integrand is exponentially suppressed, such that

$$\mathcal{N}(\vec{v}|\hat{v}_i, (\Gamma_i^{vv})^{-1})\pi(\vec{v}|\vec{\lambda}_v) \approx \mathcal{N}(\vec{v}|\hat{v}_i, (\Gamma_i^{vv})^{-1})\pi(\vec{v}|\vec{\lambda}_v)_{\text{Taylor}} \quad (\text{B4})$$

where

$$\begin{aligned} \pi(\vec{v}|\vec{\lambda}_v)_{\text{Taylor}} &:= \pi(\hat{v}_i|\vec{\lambda}_v) + \Delta\vec{v}_i^k \partial_k \pi(\hat{v}_i|\vec{\lambda}_v) \\ &\quad + \Delta\vec{v}_i^k \Delta\vec{v}_i^m \frac{\partial_k \partial_m \pi(\hat{v}_i|\vec{\lambda}_v)}{2} \end{aligned} \quad (\text{B5})$$

is the Taylor expansion of the power-law prior pdf to quadratic order in $\Delta\vec{v}_i$. Here, $\partial_k \pi(\hat{v}_i|\vec{\lambda}_v) := \partial \pi(\vec{v}|\vec{\lambda}_v) / \partial \vec{v}^k|_{\hat{v}_i}$ and we assume the Einstein summation convention over the k and m indices.

We remark that the order in $\Delta\vec{v}_i$ to which the prior pdf should be Taylor expanded, such that Eq. (B4) holds, is determined by the scale at which the exponential suppression dominates the integrand around $\hat{v}_i$. This subsequently relies on the relative rate of change of the power-law and the normal distributions. We demonstrate this scale-dependence with a simple one-dimensional example in Fig. 5. Since we are working in the LSA, the normal approximated likelihoods are tightly centered around $\hat{v}_i$ and the power-law prior evolves at much slower rates, such that a quadratic-order Taylor expansion suffices. We also verified that Eq. (B4) holds for typical sources in the results presented in Sec. V.

Substituting Eq. (B4) in (B3), we get

$$p(\hat{v}_i|\vec{\lambda}_v, \mathcal{H}_v) \approx \pi(\hat{v}_i|\vec{\lambda}_v) + \left((\Gamma_i^{vv})^{-1} \right)^{km} \frac{\partial_k \partial_m \pi(\hat{v}_i|\vec{\lambda}_v)}{2} \quad (\text{B6})$$

using standard results of multinormal integration over the full space,

$$\int d\vec{v} \mathcal{N}(\vec{v}|\hat{v}_i, (\Gamma_i^{vv})^{-1}) = 1, \quad (\text{B7})$$

$$\int d\vec{v} \mathcal{N}(\vec{v}|\hat{v}_i, (\Gamma_i^{vv})^{-1}) \Delta\vec{v}_i^k = \mathbb{E}[\Delta\vec{v}_i^k] = 0, \quad (\text{B8})$$

$$\begin{aligned} \int d\vec{v} \mathcal{N}(\vec{v}|\hat{v}_i, (\Gamma_i^{vv})^{-1}) \Delta\vec{v}_i^k \Delta\vec{v}_i^m &= \mathbb{E}[\Delta\vec{v}_i^k \Delta\vec{v}_i^m] \\ &= \left((\Gamma_i^{vv})^{-1} \right)^{km} \end{aligned} \quad (\text{B9})$$

for all k, m . Here the expectation $\mathbb{E}[\cdot]$ is calculated over the LSA likelihood $\mathcal{N}(\vec{v}|\hat{v}_i, (\Gamma_i^{vv})^{-1})$.

3. Local effect

In the local effect hypothesis, $\mathcal{H}_\ell$, the inferred set of parameters is $\vec{\psi} \equiv \{\vec{v}, \vec{\ell}\}$, while the uninferred set is $\vec{\varphi} \equiv \{\vec{g}\}$. Hence, with the MLE $\hat{\psi}_i$ given by Eq. (B1) and the prior $\pi(\vec{\theta}|\vec{\lambda}, \mathcal{H}_\ell) \equiv \pi(\vec{v}|\vec{\lambda}_v)\pi(\vec{\ell}|\vec{\lambda}_\ell)$ as described in Eqs. (11), (13), the hyperlikelihood can be written in two pieces

$$p(\hat{\psi}_i|\vec{\lambda}, \mathcal{H}_\ell) = p^{(1)}(\hat{\psi}_i|\vec{\lambda}, \mathcal{H}_\ell) + p^{(2)}(\hat{\psi}_i|\vec{\lambda}, \mathcal{H}_\ell) \quad (\text{B10})$$

where

$$p^{(1)}(\hat{\psi}_i|\vec{\lambda}, \mathcal{H}_\ell) := (1 - f) \int d\vec{\psi} \mathcal{N}(\vec{\psi}|\hat{\psi}_i, (\Gamma_i^{\psi\psi})^{-1}) \times \delta^2(\vec{\ell}) \pi(\vec{v}|\vec{\lambda}_v), \quad (\text{B11})$$

and

$$p^{(2)}(\hat{\psi}_i|\vec{\lambda}, \mathcal{H}_\ell) := f \int d\vec{\psi} \mathcal{N}(\vec{\psi}|\hat{\psi}_i, (\Gamma_i^{\psi\psi})^{-1}) \times \mathcal{N}(\vec{\ell}|\vec{\mu}_\ell, \Sigma_\ell) \pi(\vec{v}|\vec{\lambda}_v) \quad (\text{B12})$$

can be evaluated separately. In the following, we notate the dimension of a generic vector $\vec{a}$ as D_a .

a. Approximating the first piece

In approximating (B11), we first evaluate the integral over $\vec{\ell}$, such that the product of the normal distribution and the Dirac-delta prior gives us a joint normal pdf with $\vec{\ell} = \vec{0}$ fixed,

$$p^{(1)}(\hat{\psi}_i|\vec{\lambda}, \mathcal{H}_\ell) = (1 - f) \int d\vec{v} \mathcal{N}(\{\vec{v}, \vec{0}\}|\hat{\psi}_i, (\Gamma_i^{\psi\psi})^{-1}) \pi(\vec{v}|\vec{\lambda}_v). \quad (\text{B13})$$

The joint normal pdf can be decomposed into a conditional and a marginal (see, e.g., section 2.3 of [74]),

$$\mathcal{N}(\{\vec{v}, \vec{0}\}|\hat{\psi}_i, (\Gamma_i^{\psi\psi})^{-1}) = \mathcal{N}(\vec{v}|\hat{v}_i^\dagger, (\Gamma_i^{vv})^{-1}) \times \mathcal{S}_i^{(1)} \quad (\text{B14})$$

where,

$$\hat{v}_i^\dagger := \hat{v}_i + (\Gamma_i^{vv})^{-1} \cdot \Gamma_i^{v\ell} \cdot \hat{\ell}_i. \quad (\text{B15})$$

is the mean of the normal pdf.⁴ $\mathcal{N}(\vec{v}|\hat{v}_i^\dagger, (\Gamma_i^{vv})^{-1})$ conditioned at $\vec{\ell} = \vec{0}$, and

⁴ Note that the form of $\hat{v}_i^\dagger$ is objectively equivalent to the linear-bias formula for nested models (B1) with the “true” source parameter

$$\mathcal{S}_i^{(1)} := \frac{(2\pi)^{D_v/2}}{|\Gamma_i^{vv}|^{1/2}} \frac{|\Gamma_i^{\psi\psi}|^{1/2}}{(2\pi)^{D_\psi/2}} \exp \left[-\frac{1}{2} \hat{\ell}_i^T \cdot (\Gamma_i^{\ell\ell} - \Gamma_i^{\ell v} \cdot (\Gamma_i^{vv})^{-1} \cdot \Gamma_i^{v\ell}) \cdot \hat{\ell}_i \right], \quad (\text{B16})$$

is the normalized marginal probability of $\vec{\ell} = 0$ given the data realization $\hat{\ell}_i$. Substituting in (B13),

$$p^{(1)}(\hat{\psi}_i | \vec{\lambda}, \mathcal{H}_\ell) = (1 - f) \mathcal{S}_i^{(1)} \times \int d\vec{v} \mathcal{N}(\vec{v} | v_i^\dagger, (\Gamma_i^{vv})^{-1}) \pi(\vec{v} | \vec{\lambda}_v). \quad (\text{B17})$$

Thus, we obtain an integral similar in form to Eq. (B3), where the integrand is the product of a normal distribution centered at $v_i^\dagger$ and a power-law that evolves at a much slower rate. Following a similar argument as for the vacuum-GR integral (Sec. B2), the power-law prior can be Taylor expanded to second order in $\|\vec{v} - v_i^\dagger\|$ about $v_i^\dagger$ to obtain the final result for the first piece of the local effect hyperlikelihood,

$$p^{(1)}(\hat{\psi}_i | \vec{\lambda}, \mathcal{H}_\ell) \approx (1 - f) \mathcal{S}_i^{(1)} \times \left[\pi(v_i^\dagger | \vec{\lambda}_v) + ((\Gamma_i^{vv})^{-1})^{km} \frac{\partial_k \partial_m \pi(v_i^\dagger | \vec{\lambda}_v)}{2} \right]. \quad (\text{B18})$$

b. Approximating the second piece

To approximate the second piece (B12), we first note that the integrand has a product of (different dimensional) Gaussian pdfs. We evaluate this product by defining a projection matrix P of dimensionality $D_\ell \times D_\psi$ such that the local effect parameter vector can be written as $\vec{\ell} = P\vec{\psi}$. Then [75],

$$\mathcal{N}(\vec{\psi} | \hat{\psi}_i, (\Gamma_i^{\psi\psi})^{-1}) \mathcal{N}(\vec{\ell} | \vec{\mu}_\ell, \Sigma_\ell) = \mathcal{S}_i^{(2)} \mathcal{N}(\vec{\psi} | \hat{\psi}_i, \tilde{\Gamma}_i^{-1}) \quad (\text{B19})$$

where

$$\tilde{\Gamma}_i := \Gamma_i^{\psi\psi} + P^T \Sigma_\ell^{-1} P, \quad (\text{B20})$$

$$\tilde{\psi}_i \equiv \{\tilde{v}_i, \tilde{l}_i\} := \tilde{\Gamma}_i^{-1} (\Gamma_i^{\psi\psi} \hat{\psi}_i + P^T \Sigma_\ell^{-1} \vec{\mu}_\ell), \quad (\text{B21})$$

$$\mathcal{S}_i^{(2)} := \mathcal{N}(\vec{\mu}_\ell | \hat{\ell}_i, P(\Gamma_i^{\psi\psi})^{-1} P^T + \Sigma_\ell). \quad (\text{B22})$$

set $\vec{\psi}_i^* := \{\hat{v}_i, \hat{\ell}_i\}$ and where $\vec{\varphi} := \vec{\ell} = \vec{0}$ in the “null” hypothesis. This is directly related to the assumption of the LSA in obtaining (B1), such that the MLE in the null hypothesis is precisely the expectation of the (normal) LSA likelihood conditioned at $\vec{\ell} = \vec{0}$.

Substituting in (B12),

$$p^{(2)}(\hat{\psi}_i | \vec{\lambda}, \mathcal{H}_\ell) = f \mathcal{S}_i^{(2)} \int d\vec{\psi} \mathcal{N}(\vec{\psi} | \hat{\psi}_i, \tilde{\Gamma}_i^{-1}) \pi(\vec{v} | \vec{\lambda}_v) \quad (\text{B23})$$

$$= f \mathcal{S}_i^{(2)} \int d\vec{v} \mathcal{N}(\vec{v} | \tilde{v}_i, (\tilde{\Gamma}_i^{-1})^{vv}) \pi(\vec{v} | \vec{\lambda}_v). \quad (\text{B24})$$

In the second equality, we evaluate the integral over $\vec{\ell}$ to obtain the marginal likelihood of $\vec{v}$. Taylor-expanding $\pi(\vec{v} | \vec{\lambda}_v)$ about $\tilde{v}_i$ to second order in $\|\vec{v} - \tilde{v}_i\|$, we obtain the final form of the second integral piece,

$$p^{(2)}(\hat{\psi}_i | \vec{\lambda}, \mathcal{H}_\ell) \approx f \mathcal{S}_i^{(2)} \times \left[\pi(\tilde{v}_i | \vec{\lambda}_v) + (\tilde{\Gamma}_i^{-1})^{km} \frac{\partial_k \partial_m \pi(\tilde{v}_i | \vec{\lambda}_v)}{2} \right], \quad (\text{B25})$$

where we use for brevity the equivalence $(\tilde{\Gamma}_i^{-1})^{km} \equiv ((\tilde{\Gamma}_i^{-1})^{vv})^{km}$ for indices k, m that cycle only through the vacuum-GR parameters.

4. Global effect

Finally, in the global effect hypothesis, $\mathcal{H}_g$, the inferred set of parameters is $\vec{\psi} = \{\vec{v}, \vec{g}\} = \{\vec{v}, A_g\}$ and $\vec{\varphi} = \{\vec{\ell}\}$ is the uninferred set. Given the MLE $\hat{\psi}_i = \{\hat{v}_i, \hat{g}_i\} = \{\hat{v}_i, \hat{A}_{g,i}\}$ from Eq. (B1), and the prior $\pi(\vec{\theta} | \vec{\lambda}, \mathcal{H}_g) \equiv \pi(\vec{v} | \vec{\lambda}_v) \pi(\vec{g} | \vec{\lambda}_g)$ from Eqs. (11), (14), the hyperlikelihood can be written as

$$p(\hat{\psi}_i | \vec{\lambda}, \mathcal{H}_g) = \int d\vec{\psi} \mathcal{N}(\vec{\psi} | \hat{\psi}_i, (\Gamma_i^{\psi\psi})^{-1}) \times \delta(A_g - \dot{G}) \pi(\vec{v} | \vec{\lambda}_v). \quad (\text{B26})$$

It is notably similar in form to the first piece of the hyperlikelihood in the local effect case (B11), and can be solved following a similar argument.

We first evaluate the integral over $\vec{g} \equiv \{A_g\}$. We have

$$p(\vec{\psi}_i | \vec{\lambda}, \mathcal{H}_g) = \int d\vec{v} \mathcal{N}(\{\vec{v}, \dot{G}\} | \hat{\psi}_i, (\Gamma_i^{\psi\psi})^{-1}) \pi(\vec{v} | \vec{\lambda}_v). \quad (\text{B27})$$

Then, we decompose the joint normal pdf into a conditional and a marginal [74],

$$\mathcal{N}(\{\vec{v}, \dot{G}\} | \hat{\psi}_i, (\Gamma_i^{\psi\psi})^{-1}) = \mathcal{N}(\vec{v} | v_i^\dagger, (\Gamma_i^{vv})^{-1}) \times \mathcal{S}_i \quad (\text{B28})$$

where,

$$v_i^\dagger := \hat{v}_i + (\Gamma_i^{vv})^{-1} \cdot \Gamma_i^{vg} \cdot (\hat{A}_{g,i} - \dot{G}) \quad (\text{B29})$$

and

$$\mathcal{S}_i := \frac{(2\pi)^{D_v/2}}{|\Gamma_i^{vv}|^{1/2}} \frac{|\Gamma_i^{\psi\psi}|^{1/2}}{(2\pi)^{D_\psi/2}} \exp \left[-\frac{1}{2} (\Gamma_i^{gg} - \Gamma_i^{gv} \cdot (\Gamma_i^{vv})^{-1} \cdot \Gamma_i^{vg}) (\hat{A}_{g,i} - \dot{G})^2 \right]. \quad (\text{B30})$$

Substituting Eq. (B28) into (B27), and following arguments similar to Appendix B 3b, the analytical approximation of the global effect hyperlikelihood is given as

$$p(\hat{\psi}_i | \vec{\lambda}, \mathcal{H}_g) \approx \mathcal{S}_i \left[\pi(v_i^\dagger | \vec{\lambda}_v) + \frac{((\Gamma_i^{vv})^{-1})^{km} \partial_k \partial_m \pi(v_i^\dagger | \vec{\lambda}_v)}{2} \right]. \quad (\text{B31})$$

5. Validating the approximate hyperlikelihoods

To summarize, in this section, we have analytically approximated the hyperlikelihood (7) by assuming the validity of the LSA and utilizing various standard properties of normal distributions. Our approximate results are given in the vacuum-GR hypothesis, $\mathcal{H}_v$, by Eq. (B6); for the local-effect hypothesis, $\mathcal{H}_\ell$, by Eqs. (B10), (B18), and (B25); and in the global-effect hypothesis, $\mathcal{H}_g$, by Eq. (B31).

We will now validate the approximate results in the EMRI context by comparing them with the direct Monte Carlo integral of Eq. (7). In each hypothesis, we initialize an EMRI source on circular and equatorial orbits evolving according to Eq. (9), with true source parameters $\vec{\theta}^*$ and population hyperparameters $\vec{\lambda}^*$ satisfying the respective hypothesis: the vacuum-GR parameters are fixed across all three sources and are chosen to be representative of typical EMRIs, while the local and global effect parameters are non-null only in the local effect and global effect hypothesis, respectively. By construction, since each EMRI is initialized in its true hypothesis, the MLE $\hat{\psi}$ of the inferred parameters is not biased, such that we can bypass the bias-correction step (which is anyway irrelevant for integral validation). The optimal SNR of

each source is $\rho_{\text{opt}} \approx 123$, such that the LSA holds.

With the LSA likelihood (15), the Monte Carlo integral approximation of Eq. (7) is [76]

$$p(\hat{\psi}_i | \vec{\lambda}, \mathcal{H}) \approx \frac{V}{N} \sum_{n=1}^N \mathcal{N}(\vec{\psi}_n | \hat{\psi}, (\Gamma^{\psi\psi})^{-1}) \pi(\vec{\psi}_n | \vec{\lambda}, \mathcal{H}) \quad (\text{B32})$$

where $\{\vec{\psi}_n\}$ is a set of N random samples drawn uniformly from a suitably large hypercube of volume V around the inferred parameters $\hat{\psi}$. The corresponding approximate hyperlikelihoods for the three hypotheses were derived above. To gauge the validity of our results, we define the ratio $r_{\text{approx}}^{\text{MC}}$ of the Monte Carlo integral to the approximate hyperlikelihood in each hypothesis, such that $r_{\text{approx}}^{\text{MC}} = 1$ implies an exact match between the two approaches. As we show in Fig. 6, $r_{\text{approx}}^{\text{MC}}$ is consistent with unity for large N in all three hypotheses up to stochastic errors in the Monte Carlo integral, validating the constructed approximate hyperlikelihoods. We emphasize that, given the FIM Γ at $\vec{\theta}^*$ and the MLE $\hat{\psi}$, the analytical hyperlikelihood evaluation is $\mathcal{O}(N)$ more efficient than the Monte Carlo approach for *each* source in the population, underscoring the relevance of such approximations for the feasibility of our analysis.

Appendix C: Description of parameters used in the results

The hyperparameters used to generate the four populations P_v , P_ℓ , P_g , and P_{mix} in Sec. V A are presented in Table II. The corresponding priors on all source parameters are summarized in Table III, and the priors on the population hyperparameters used in the inference are summarized in Table IV.

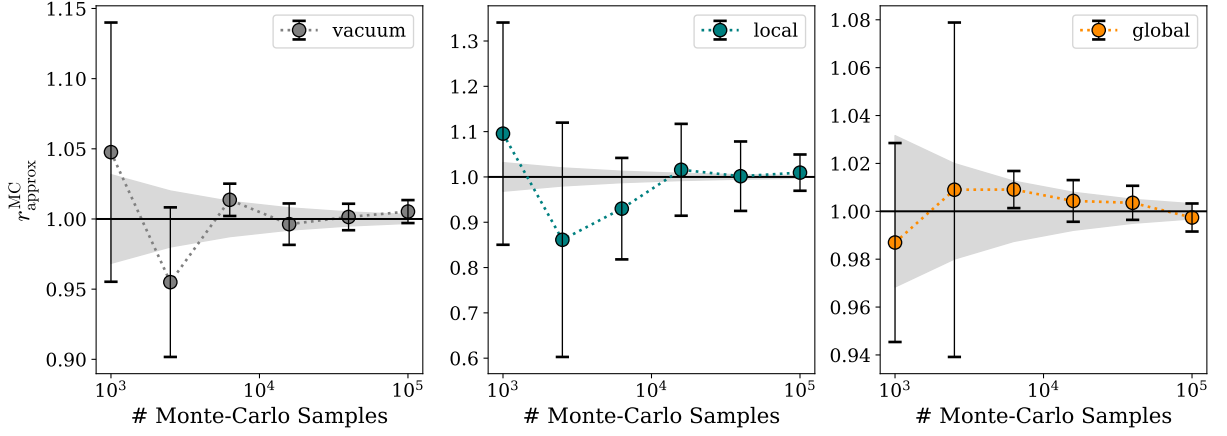


FIG. 6: Validating the approximate hyperlikelihoods calculated in the vacuum-GR hypothesis (Eq. (B6), left panel), the local effect hypothesis (Eqs. (B10), (B18), and (B25), middle panel), and the global effect hypothesis (Eq. (B31), right panel) against Monte-Carlo integral of the exact expression (Eq. (7)) for EMRIs. The vertical axis plots the ratio of the Monte Carlo integral to the approximate hyperlikelihood, $r_{\text{approx}}^{\text{MC}}$. The horizontal black solid line in each panel represents $r_{\text{approx}}^{\text{MC}} = 1.0$, i.e., where the two approaches match exactly. The vertical bars represent 1σ errors on the Monte Carlo integral for a given sample size, estimated empirically from five independent trials, while the shaded region represents the “theoretical” 1σ error in the Monte Carlo integral approximation with N samples, $1/\sqrt{N}$.

hyper-parameter	P_v	P_ℓ	P_g	P_{mix}
K^*	0.005	0.005	0.005	0.005
α^*	0.0	0.0	0.0	0.0
β^*	0.0	0.0	0.0	0.0
f^*	0.0	0.5	0.0	0.5
$\mu_{A_\ell}^*$	-	10^{-6}	-	10^{-6}
$\mu_{n_\ell}^*$	-	8.0	-	8.0
$\sigma_{A_\ell}^*$	-	10^{-7}	-	10^{-7}
$\sigma_{n_\ell}^*$	-	1.0	-	1.0
$\dot{G}^* [\text{yr}^{-1}]$	0.0	0.0	10^{-12}	10^{-12}

TABLE II: True hyperparameters in the constructed populations P_v , P_ℓ , P_g , and P_{mix} . P_v , P_ℓ , and P_g are constructed assuming $\mathcal{H}_v$, $\mathcal{H}_\ell$, and $\mathcal{H}_g$ as the true hypothesis, respectively, while P_{mix} is such that the full population is affected by the global time-varying G effect while a fraction f of the population simultaneously experiences the local migration torque effect.

parameter	prior pdf	description
$\ln M, z$	Eq. (11)	detector-frame log-MBH mass, source redshift
$\log_{10} q$	$\mathcal{U}[-5.5, -4.5]$	log of the small-mass-ratio $q := m_2/m_1$
μ	Mq	detector-frame CO mass
a	$\mathcal{U}[0.5, 0.99]$	MBH spin
θ_S	$\mathcal{U}[0, \pi]$	polar sky location
θ_K	$\mathcal{U}[0, \pi]$	polar spin orientation
ϕ_S	$\mathcal{U}[0, 2\pi]$	azimuthal sky location
ϕ_K	$\mathcal{U}[0, 2\pi]$	azimuthal spin orientation
Φ_0	$\mathcal{U}[0, 2\pi]$	initial orbital phase
T_{plunge}	$\mathcal{U}[0.5, 2]$ years	time to plunge at the start of observation
p_0	$p_0(T_{\text{plunge}})$ (fixed)	initial semi-latus rectum
A_ℓ, n_ℓ	Eq. (13)	amplitude & slope of the (local) migration torque
A_g	Eq. (14)	intrinsic strength of the (global) time-varying G effect
n_g	4 (fixed)	PN-order of the (global) time-varying G effect

TABLE III: Priors on the source parameters used to generate the true population. $p_0 = p_0(T_{\text{plunge}})$ where a CO with semi-latus rectum $p_0(T_{\text{plunge}})$ plunges into the MBH at the end of T_{plunge} for a given set of $\{M, \mu, a\}$.

hyperparameter	hyperprior pdf	description
K	$\mathcal{U}[0.9K^*, 1.1K^*]$	MBH population size scaling factor
α	$\mathcal{U}[-0.1, 0.1]$	MBH mass slope
β	$\mathcal{U}[-0.1, 0.1]$	MBH redshift slope
f	$\mathcal{U}[0, 1]$	fractional population with the (local) migration torques
μ_{A_ℓ}	$\mathcal{U}[0.9\mu_{A_\ell}^*, 1.1\mu_{A_\ell}^*]$	mean of the (local) migration torque amplitudes
μ_{n_ℓ}	$\mathcal{U}[0.9\mu_{n_\ell}^*, 1.1\mu_{n_\ell}^*]$	mean of the (local) migration torque slopes
σ_{A_ℓ}	$\mathcal{U}[0.9\sigma_{A_\ell}^*, 1.1\sigma_{A_\ell}^*]$	standard deviation of the (local) migration torque amplitudes
σ_{n_ℓ}	$\mathcal{U}[0.9\sigma_{n_\ell}^*, 1.1\sigma_{n_\ell}^*]$	standard deviation of the (local) migration torque slopes
$\dot{G}$	$\mathcal{U}[-5, 5] \times 10^{-12}$	intrinsic strength of the (global) time-varying G effect (per year)

TABLE IV: Hyperpriors on the population hyperparameters. * denotes the true value of the hyperparameter in the chosen population (Table II).