Bayesian test of the mass of the graviton with planetary ephemerides

Vincenzo Mariani[*]

Géoazur, Université Côte d'Azur, Observatoire de la Côte d'Azur,
CNRS, 250 avenue A. Einstein, 06250 Valbonne, France

Agnès Fienga[†]

Géoazur, Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS,
250 avenue A. Einstein, 06250 Valbonne, France
and IMCCE, Observatoire de Paris, PSL University, CNRS,
77 avenue Denfert-Rochereau, 75014 Paris, France

Olivier Minazzoli[‡]

Artemis, Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS,
BP4229, 06304, Nice Cedex 4, France
and Bureau des Affaires Spatiales, 2 rue du Gabian, 98000 Monaco

Mickaël Gastineau[§] and Jacques Laskar[‖]

IMCCE, Observatoire de Paris, PSL University, CNRS,
77 avenue Denfert-Rochereau, 75014 Paris, France

In this work, we investigated Bayesian methodologies for constraining in the Solar System a Yukawa suppression of the Newtonian potential—which we interpret as the effect of a non-null graviton mass—by considering its impact on planetary orbits. Complementary to the previous results obtained with INPOP (Intégrateur Numérique Planétaire de l'Observatoire de Paris) planetary ephemerides, we consider here a Markov chain Monte Carlo approach associated with a Gaussian process regression for improving the resolution of the constraints driven by planetary ephemerides on the graviton mass in the Solar System. At the end of the procedure, a posterior for the mass of the graviton is presented, providing an upper bound at 1.01×10^{-24} eV c^{-2} (respectively, $\lambda_g \geq 122.48 \times 10^{13}$ km) with a 99.7% confidence level. The threshold value represents an improvement of 1 order of magnitude relative to the previous estimations. This updated determination of the upper bound is mainly due to the Bayesian methodology, although the use of new planetary ephemerides (INPOP21a used here versus INPOP19a used previously) already induces a gain of a factor of 3 with respect to the previous limit. The INPOP21a ephemerides are characterized by the addition of new Juno and Mars orbiter data but also by a better Solar System modeling, with notably a more realistic model of the Kuiper belt. Finally, by testing the sensitivity of our results to the choice of the *a priori* distribution of the graviton mass, it turns out that the selection of a prior more favorable to zero-mass graviton (that is, here, general relativity) seems to be more supported by the observations than nonzero mass graviton, leading to a possible conclusion that planetary ephemerides are more likely to favor general relativity.

I. INTRODUCTION

Planetary ephemerides have evolved with the observational accuracy obtained for the astrometry of planets and natural satellites thanks to the navigation tracking of spacecraft orbiting these systems. Since the late 19th century the astrometry of planets has known a significant improvement leading to an increased accuracy of the dynamical theories describing their motions. The motion of the planets and asteroids in our Solar System can be solved directly by the numerical integration of their equations of motion. The improved (present and future) accuracy in the measurements of observables from space missions like Cassini-Huygens, Mars Express, Venus Express, BepiColombo etc. makes the Solar System a suitable arena to test general relativity theory (GRT) as well

[*]vmariani@geoazur.unice.fr
[†]agnes.fienga@oca.eu
[‡]ominazzoli@gmail.com
[§]mickael.gastineau@obspm.fr
[‖]jacques.laskar@obspm.fr

as alternative theories of gravity by mean of Solar System ephemerides. INPOP (Intégrateur Numérique Planétaire de l'Observatoire de Paris) planetary ephemerides are developed since 2003, integrating numerically the Einstein-Infeld-Hoffmann equations of motion proposed by [1,2], and fitting the parameters of the dynamical model to the most accurate planetary observations following [3]. Testing alternative theories of gravity with planetary ephemerides consist in changing the metric of GRT into alternative frameworks and consequently modifying the equations of motion, the light time computation and the definition of timescales used for the construction of planetary ephemerides (see Ref. [4] for a review). In principle, such modifications of GRT can be summarized as considering additional terms to GRT fundamental equations, such as a Yukawa suppression of the Newtonian potential. At the phenomenological level, a mass of the gravitational interaction[1] is often assumed to either lead to a modification of the dispersion relation of gravitational waves [5] or to lead to a Yukawa suppression of the Newtonian potential[2] [8]. Recently[3] in [9], Will argued that Solar System observations and planetary dynamics could be used to improve the constraints on the mass of the graviton m_g. However Will used results based on statistics of postfit residuals of the Solar System ephemerides that are performed without including the effect of a massive graviton inside the equations of motion. In order to overcome the consistency issues that are raised by such type of analyses —that is, which are based on postfit residuals—we investigate an original approach that is based on a statistical inference of the mass of the graviton m_g within the framework of INPOP as presented in [10,11].

In this work, the approach we decided to use is partially Bayesian and several tools like Markov chain Monte Carlo (MCMC), Gaussian process and Bayes factor are employed, in order to exploit at the best the values of INPOP χ^2. Contrarily to [11], a full posterior distribution of the graviton mass m_g is deduced, using a Metropolis-Hastings algorithm which optimizes the information contained in such a distribution, and a Gaussian process regression which improves the resolution of the search process. Moreover, by considering two different assumptions for the *a priori* distribution of the graviton mass m_g, we are able (i) with an uniform distribution, to give a new upper limit for m_g, improving the limit by a factor of 30 relative to [11] (see Secs. III B and III B 2) and (ii) with a Laplace distribution, to demonstrate that, with the present accuracy of the planetary ephemerides, GRT is sufficient for explaining the observations (see Sec. III C). In other words we can infer that the nowadays planetary ephemerides

tend to prefer a non-massive graviton. In the Appendix, we give more details on the methods applied.

II. THEORETICAL FRAMEWORK

A. Massive graviton phenomenology

There is not an unique definition of what a massive gravity may mean [7]. Massive interactions usually lead to a Yukawa suppression of those interactions on the scale of the Compton wavelength, λ_g. But gravity is different, and therefore it may not be the case for a fully consistent theory of massive gravity [7]. Nevertheless, from a phenomenological point of view [8], one can test whether or not there is a Yukawa suppression of the gravitational potentials at the level of the Solar System—see, e.g., [6,9] and references therein. Let us note that, while one often talks about a "graviton mass" in the literature [6,9], the word "graviton" is mostly used for convenience since everything is considered at the classical level only. (The same way that some may use the word "photon" for classical electrodynamics). Formally, this would lead to the following modification of the Newtonian potential w_{Newton} [6,9]

$$w = w_{\text{Newton}} \exp(-r/\lambda_g), \tag{1}$$

which can be developed as [10]

$$w = w_{\text{Newton}} \left(1 + \frac{1}{2} \frac{r^2}{\lambda_g^2} \right) + \mathcal{O}(\lambda_g^{-3}), \tag{2}$$

after a convenient change of coordinate system that absorbs the constant term in the gravitational potential, and which has no impact on the observables. By analogy with standard quantum physics, the Compton length can also be interpreted in terms of a mass of the graviton m_g following the relation

$$\lambda_g = \frac{\hbar}{c m_g}, \tag{3}$$

with $\hbar$ the Planck constant, and c the speed of light. (In some sense this is also one of the reasons why, beyond simple convenience, one often talks about a "graviton mass" despite working at the classical level, since the translation of the Yukawa suppression length in terms of a mass involves the quantum of action $\hbar$.) In that situation, the equation of motion has only one extra term with respect to the usual Einstein-Infeld-Hoffmann-Droste-Lorentz equations that reads

$$\delta a^i = \frac{1}{2} \sum_P \mu_P \frac{c^2 m_g^2}{\hbar^2} \frac{x^i - x_P^i}{r} + \mathcal{O}(m_g^3), \tag{4}$$

where μ_P is the gravitational parameter $\mu_P = G M_P$. Further assuming that light still propagates along null geodesics, the Shapiro delay reads

---

[1]Reported as *the mass of the graviton* in the rest of the paper, for convenience.

[2]For more information on the status of current theoretical models of massive gravity, we refer the reader to [6,7].

[3]Following his more-than-20-years-old seminal work [8].

$$c(t_r - t_e) = c(t_r - t_e)_{GRT}$$
$$+ \sum_A \frac{\mu_A}{c^2} \frac{c^2 m_g^2}{2\hbar^2} \ln \left[b^2 \frac{\boldsymbol{n} \cdot \boldsymbol{r}_{rA} + r_{rA}}{\boldsymbol{n} \cdot \boldsymbol{r}_{eA} + r_{eA}} \right.$$
$$\left. + \boldsymbol{n} \cdot (r_{rA}\boldsymbol{r}_{rA} - r_{eA}\boldsymbol{r}_{eA}) \right] + \mathcal{O}(m_g^3). \quad (5)$$

The term $c(t_r - t_e)_{GRT}$ corresponds to the GRT light time, b is the minimal distance between the light path and the central body (here the Sun). This expression is an approximation at the c^{-2} level. Let us note, however, that the correction to the Shapiro delay due to the graviton mass has been found to be negligible in practice as it was discussed in [10] using the Compton wavelength formalism.

In [10,11], planetary ephemerides have been fully developed in the massive gravity framework of Eqs. (4) and (5) using the equivalent Compton wavelength formalism and fitted over the data sample for INPOP17a and INPOP19a, respectively. The results of these investigations are gathered in [4].

It is somewhat interesting to compare these constraints to the ones deduced from the observation of gravitational waves [8,9]. Indeed, one can assume that a massive gravitational field that leads to Eq. (1) might also modify the dispersion relation of gravitational waves as follows [8,12]:

$$E^2 = p^2 c^2 + m_g c^2, \quad (6)$$

where E and p are the energy and momentum of the wave. Such a modified dispersion relation causes gravitational wave frequency modes to propagate at different speeds, leading to an overall modification of the phase morphology of gravitational waves with respect to the GRT predictions. Since the morphology of gravitational wave phase has been consistent with general relativity so far, it led to severe constraints on the value of the graviton mass: $m_g \leq 1.27 \times 10^{-23}$ eV c^{-2} at 90% confidence level [5], whereas previous results with ephemerides were at the few 10^{-23} eV c^{-2} level [11].

Each type of constraints is relevant in its own right given that they test different phenomenologies, which may (or may not) be related, depending on the underlying massive gravity theory that one is considering. For instance, screening mechanisms may suppress one effect but not the other [7].

B. Planetary ephemerides construction

INPOP is a planetary ephemerides that is built by integrating numerically the equations of motion of the Solar System objects following the formulation of Moyer [2], and by adjusting to Solar System observations such as space mission navigation and radio science data, ground-based optical observations, or lunar laser ranging [13,14].

In addition to adjusting the astronomical intrinsic parameters, it can be used to constrain parameters that encode deviations from GRT [15–18], such as the Compton wavelength λ_g as defined in Eq. (3). As long as m_g is small enough, the gravitational phenomenology in the Newtonian regime recovers the one of GRT.

But, as discussed in [9], a graviton mass would indeed lead to a modification of the perihelion advance of Solar System bodies. Based on published constraints on the perihelion advance of Mars—or on the post-Newtonian parameters γ and β—derived from Mars Reconnaissance Orbiter data, Will had estimated that the graviton mass should be smaller or equal to $(4 - 8) \times 10^{-24}$ eV c^{-2} depending on the specific analysis. But, as an input for his analysis, Will uses results based on interpreting statistics of postfit residuals of the Solar System ephemerides obtained in various frameworks (post-Newtonian and GRT) as a possible outcome of the graviton influence. However, first of all—unlike the historical occurrence of the substantial error in the perihelion advance of Mercury computed in Newton's theory—a lot of different contributions from the details of the Solar System model being used could explain the rather small postfit differences between computed and observed positions [4]. Furthermore various parameters of the ephemerides (e.g., masses, semimajor axes, etc.) are more or less correlated to m_g as it is shown in [10]. Therefore any kind of signal introduced by $m_g > 0$ can, in part, be reabsorbed during the fit of other correlated parameters. In order to overcome the correlation issues described previously, we investigate a new approach based on a statistical inference on the mass of the graviton m_g within the full framework of INPOP. Planetary ephemerides are developed in the framework described in Sec. II A, where planetary equations of motion but also Shapiro delays are modified according to Eqs. (4) and (5).

In this work, we will use the INPOP21a planetary ephemerides [14] that benefits from the latest Juno and Mars orbiter tracking data up to 2020 as well as a fit of the Moon-Earth system to LLR observations also up to 2020. For a more detailed review about this specific version, the reader can refer to [14] whereas [19] gives more descriptions regarding recent GRT tests obtained with INPOP21a and INPOP19a. INPOP21a is more accurate than INPOP19a especially for Jupiter and Saturn orbits as additional Juno observations of Jupiter were used covering a 4-year period when only 2.5 years were considered in INPOP19a. Consequently, a more realistic model of the Kuiper belt was implemented in INPOP21a, leading to an improvement of about 1.4 on the Jupiter orbit accuracy. Additionally, two years of Mars Express navigation data have been added to the 13 years already implemented in INPOP19a. This increase of Mars orbiter data improves mainly the stability of the ephemerides and its extrapolation capabilities [20]. In terms of adjustment, in addition to the initial conditions of the planetary orbit, the gravitational

mass of the Sun, its oblateness and the ratio between the mass of the Earth and the one of the Moon, 343 asteroid masses are fitted in INPOP21a following the procedure described, for example, in [20]. A mass representing the average effect of 500 trans-neptunian objects has also been added as described in [14]. A total of 401 parameters are accounted for the INPOP21a construction. They constitute the list of astronomical parameters we will refer to in the following.

III. METHODOLOGY AND RESULTS

Starting from [17], as well in [21–23], several tools to assess the goodness of the INPOP fit with respect to modifications in the equations of motion or in the global framework of the ephemeris or observations have been used. In particular the computation of the INPOP χ^2 plays an essential role to determine which data or which model improve significantly the ephemeris computation. In our work, as in [11], the computation of $\chi^2(m_g)$ is the output, for a given value of m_g, of the INPOP iterative fit, after the adjustment of all its astronomical parameters (see Sec. II B). The $\chi^2(m_g)$ is computed following

$$\chi^2(m_g, \mathbf{k}) \equiv \frac{1}{N_{\text{obs}}} \sum_{i=1}^{N_{\text{obs}}} \left(\frac{g^i(m_g, \mathbf{k}) - d_{\text{obs}}^i}{\sigma_i} \right)^2, \quad (7)$$

where m_g is a fixed value, $\mathbf{k}$ are the astronomical parameters fitted with INPOP (see step 1 in Fig. 1 and Secs. III A and II B), N_{obs} is the number of observations, the function g^i represents the computation of observables, the vector $\mathbf{d}_{\text{obs}} = (d_{\text{obs}}^i)_i$ is the vector of observations and σ_i are the observational uncertainties.

A. Methodology in a nutshell

As said in Sec. I, we want to obtain a posterior for the mass of the graviton m_g. The general pipeline used to obtain such posteriors is summarized in Fig. 1 and works as follows. First (step 1 in Fig. 1), we compute the value of $\chi^2(m_g)$ [see Eq. (7)] for several different values of m_g spreading over the domain of our interest. For a given m_g the value $\chi^2(m_g)$ is obtained as the outcome of the full INPOP iterative adjustment letting fixed m_g. In such a fit, the astronomical parameters $\mathbf{k}$ of Eq. (7) are adjusted with the least squares procedure, whereas m_g is a fixed value. This χ^2 estimation is necessary since it ensures the m_g
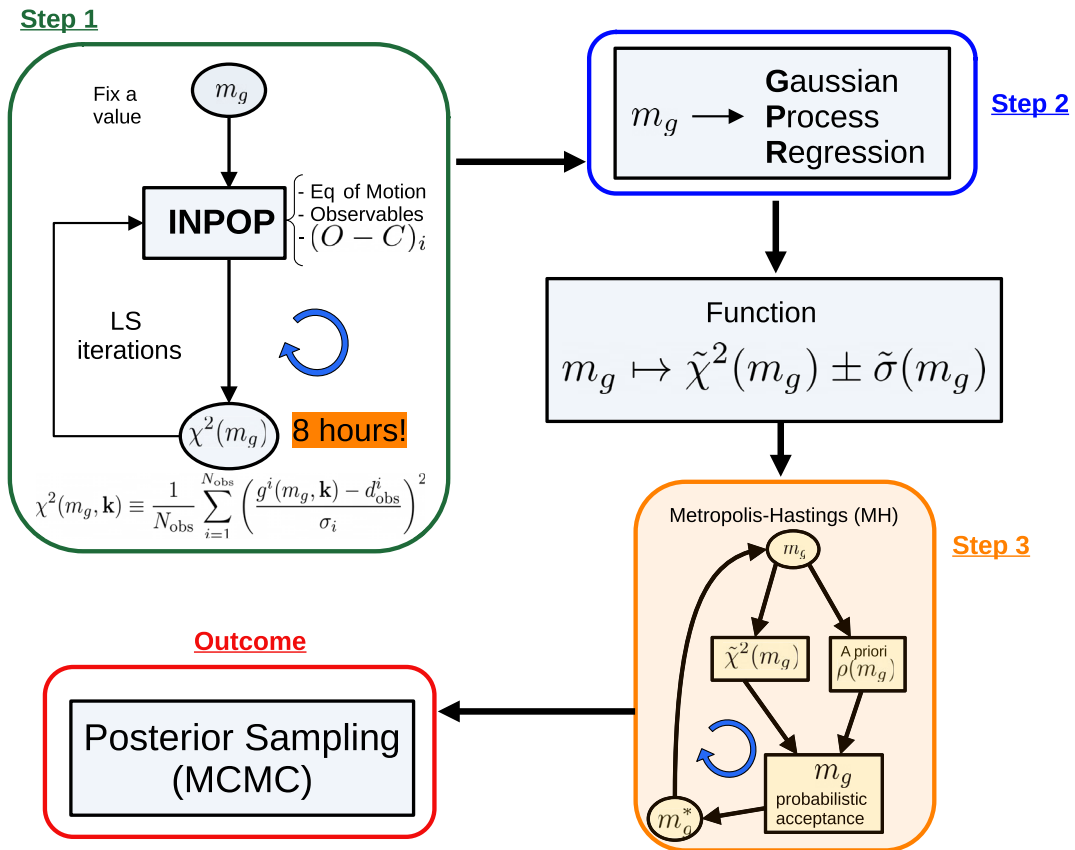

FIG. 1. General pipeline used to obtain the posterior of m_g. The first step is the computation of $\chi^2(m_g)$ for m_g spread over the domain of our interest. Whereupon a GPR is computed, obtaining the interpolation $m_g \longmapsto \tilde{\chi}^2(m_g)$ with the corresponding uncertainty. The outcome is the posterior for the mass of the graviton m_g. The fully detailed description of method and pipeline can be found in [24].

contribution to the dynamics considered, avoiding high correlations between m_g and the other astronomical parameters **k** (see Ref. [10]). The iterative fit can take up to eight hours of computation for one given m_g value. But, in order to use the MCMC algorithm, it is necessary to evaluate *sequentially* the likelihood, and so the χ^2, for thousands of different values of m_g (see the Appendix). In this context, the direct χ^2 computation becomes difficult. In order to overcome the problem of the computation time, the second step of the method (step 2 in Fig. 1) is the use of a Gaussian process regression (GPR) to interpolate among the values $(m_g, \chi^2(m_g))$, already computed during step 1. Starting from this set of points we obtain with the GPR a function $m_g \mapsto \tilde{\chi}^2(m_g)$ which is the interpolation of that set of points, along with an uncertainty $\tilde{\sigma}(m_g)$ relative to the possible error of interpolation in $\tilde{\chi}^2(m_g)$. The interpolation $m_g \mapsto \tilde{\chi}^2(m_g)$ is necessary to run the Metropolis-Hasting (MH) algorithm (step 3 in Fig. 1), which has, as an outcome, the posterior density for m_g in the form of a MCMC. Step 3 of the method is then the MH algorithm for selecting possible values of m_g. The algorithm is described in Appendix A 1 and more details can be found in [24]. In the MH algorithm, a prior density distribution has to be settled for the parameter to sample with the MCMC. For this work, we used two different prior density distributions of the graviton mass as inputs for the MH algorithm in order to test the sensitivity of the posterior to the choice of the prior and if there is any gain in considering a nonzero mass graviton (with an uniform prior distribution, see Secs. III B and III B 2) versus a massless graviton (with a half-Laplace prior distribution, see Sec. III C).

B. Results with the uniform prior distribution and upper bound

In terms of prior for the mass of the graviton, we firstly chose an uniform distribution with large intervals of values encompassing the latest results from [11] but without giving preference to any possible values. Following the step 2 of Fig. 1, for each value of the graviton mass proposed by the MH algorithm, the value $\tilde{\chi}^2(m_g)$ is computed instead of $\chi^2(m_g)$. Based on a first grid of fully converged runs, we use the Gaussian process regression as a form of interpolation to increase the resolution of the MH algorithm, along with GPR error estimations to assess the uncertainties of such regression. Details about GPR and GPR error estimations are given in [24].

1. MH algorithm and GPR

In Fig. 2 is plotted the posterior density of probability obtained with MH algorithm associated with GPR, supposing an uniform prior. Contrarily to a detection curve—which in our case would look close to a Gaussian-like posterior centered on a positive value (see i.e. [24])—from Fig. 2, we do not have a single figure that
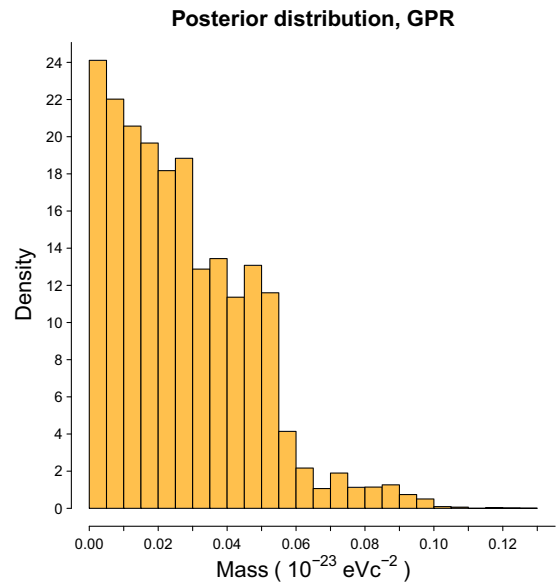

FIG. 2.   Density for the posterior probability distribution used as target probability. The prior was a uniform prior between 0 and 3.62×10^{-23} eV c^{-2}.

we could choose as value for the mass of the graviton. Indeed the shape is not a *bell*, nor does it show an individual peak. The quantile at 97% is 0.0985×10^{-23} eV c^{-2}.

The posterior plotted in Fig. 2 also tends to concentrate close to $m_g = 0$ with decreasing steps for larger m_g up to $m_g < 0.15 \times 10^{-23}$ eV c^{-2}. The conclusion drawn is that, within the GPR approximation of χ^2 we do not have any detection for $m_g \neq 0$. Thus we cannot provide an estimated value for the mass of the graviton, but we can give a 99.7% upper limit as quantile of the deduced mass posterior, that is $m_g \leq 0.98 \times 10^{-24}$ eV c^{-2}.

2. Uncertainty assessment

In order to explore the uncertainties induced by the GPR interpolation on $\tilde{\chi}^2(m_g)$ values, we ran 300 MCMC simulations using the MH algorithm. Each MH is performed on a different Gaussian process uncertainty estimation, GPUE (see the black line in Fig. 10). One GPUE is a perturbation of the GPR within the uncertainty provided by the GPR itself. The GPUE interpolates possible χ^2 deviations from $\tilde{\chi}^2$ (see Ref. [24] for a detailed description on how to encompass uncertainty) exploring the space of uncertainties in the GP regression. Running the MCMC simulation on the GPUE is a way to propagate the uncertainty induced by the GPR in the posterior sampling. We, therefore, built a posterior distribution presented in blue in Fig. 3 accounting for the GPR uncertainty, and we call such a posterior Gaussian process uncertainty realization (GPUR).

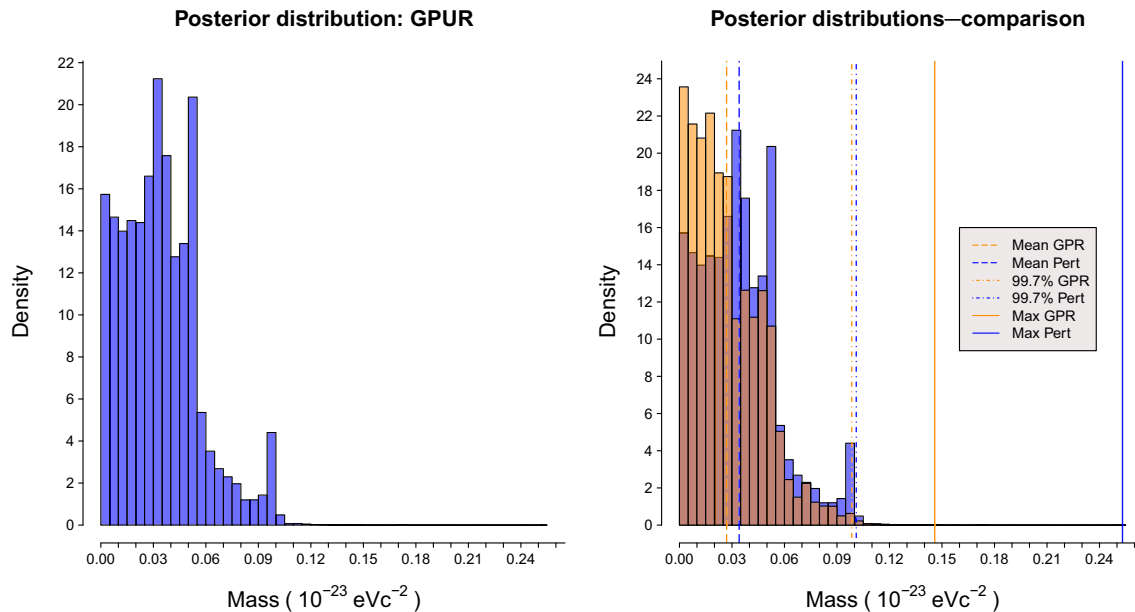The GPUR result presented in Fig. 3 is consistent with the nominal case (see Sec. III B 1), and it shows the

FIG. 3.   Posterior probability distributions obtained from GPR (in orange) and the posterior with GPR error assessment (GPUR left-hand side and in blue). The dashed lines represent the averages of the posterior with GPR and GPUR (respectively, orange and blue). The dot-and-dashed lines represent the 99.7% quantiles of the two densities. The solid lines are instead in place of the maximum m_g for each one of the two densities. The brown area of the histogram represents the overlaid zone between the two posteriors presented.

posterior with a slightly larger interval of masses. This is actually what it is expected, since the uncertainty of GPR in the present case is small, but not absent. In Fig. 3 it is easy to see how much the maximum value of m_g is shifted towards larger m_g, passing from the MCMC with GPR to the MCMC on GPUEs. In particular the average going from 0.26×10^{-24} eV c^{-2} to 0.34×10^{-24} eV c^{-2} whereas the maximum mass in the posterior going from 1.45×10^{-24} eV c^{-2} to 2.53×10^{-24} eV c^{-2}. The strategy we propose in the Appendix relies on the assumption that if we compute the real values of $\chi^2(m_g)$, then we can estimate χ^2 values in the zones of domain for which $\chi^2(m_g)$ is unknown, with an uncertainty based on χ^2 values already computed. In our specific case, the strategy looks consistent. The outcome of the 300 MCMC runs on different GPUEs is similar with respect to the nominal GPR case with slight differences (see, e.g., Table I and Fig. 3). As previously indicated (see Sec. III B 1), again the GPUR posterior is not similar to the one obtained with a positive detection. We are however able to provide limits for the mass. The upper bound for GPUR we would provide at

99.7% C.L. is $m_g \leq 1.01 \times 10^{-24}$ eV c^{-2} (at 90% C.L. is $m_g \leq 0.59 \times 10^{-24} eVc^{-2}$). This represents an improvement of about 1 order of magnitude from the previous estimations in terms of upper limit provided for m_g at 99.7% C.L. Such a limit is taken from the GPUR posterior (see Fig. 3) since it takes into consideration also the GPR uncertainty.

Finally, as stressed in Table I, from a very large uniform prior between 0 and 3.62×10^{-23} eV c^{-2}, the MCMC algorithm indicates a posterior between 0 and a maximum of 2.53×10^{-24} eV c^{-2}, inducing a significant improvement also on the possible maximum value for the mass of the graviton.

C. Results with the half-Laplace prior distribution

The absence of positive detection can be interpreted in two ways: either because the data employed are not sensitive enough, or because GRT is sufficient to explain the data. The lack of detection could also depend on the hypothesis made on our *a priori* knowledge of the mass of the graviton. In order to discriminate between the former and the latter and to test the sensitivity of our results to the prior, we made an additional experiment, changing the prior density for m_g in the MH algorithm from an uniform prior to a half-Laplace distribution, which gives a prior preference to small graviton masses. We ran the MH algorithm again, with the same GPR, by using a half-Laplace prior (red line in Fig. 4). The half-Laplace prior is chosen such that the zone of higher probability of the half-Laplace density shares the same domain of the posterior

TABLE I.   Summary of the outcome for MCMC run based on GPR and the posterior GPUR. These values are plotted as vertical lines in Fig. 3. The unit is 10^{-24} eV c^{-2}. The prior was a uniform prior between 0 and 3.62×10^{-23} eV c^{-2}.

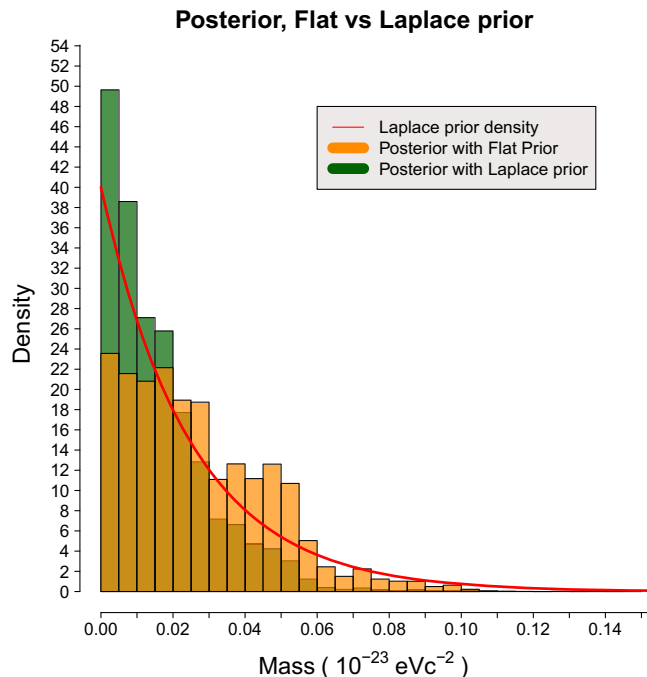|	$\langle m_g \rangle$	99.7% quantile	$\max\{m_g\}$
GPR	0.26	0.98	1.45
GPUR	0.34	1.01	2.53

FIG. 4. Densities for the posterior probability distributions obtained from GPR with an uniform (flat) prior (in orange) and with a half-Laplace prior (dark green). In red, the shape of the half-Laplace prior used. The brown area is the overlaid zone of the two posteriors.

obtained with the uniform prior. The underlying idea of half-Laplace prior is to give preference to the mass value $m_g = 0$, being in our case representative of GRT. Doing so we can discern whether the data have enough information to flatten this new prior or not. As one can see in Fig. 4, the MH algorithm does not provide the same outcome with the two different priors, and in particular the new posterior (green in Fig. 4) turns out to be piled up around $m_g = 0$. For a validation of the MCMC convergence in the case of half-Laplace prior we refer to [24].

IV. DISCUSSION

A. Towards nondetection

1. Nondetection with half-Laplace prior

Comparing the outcome obtained with the two different priors (uniform and half-Laplace), we may see whether the new posterior (green in Fig. 4) resembles the old one (orange in Fig. 4) or not: we found that it is not the case. In particular, in Fig. 4, we can see that the peak of the Laplace posterior is almost at 50 (in terms of value assumed by the density), whereas the maximum value of the half-Laplace prior is 40. So 25% smaller than the posterior. Moreover, even if the green posterior is close to a half-Laplace density, its tail is shorter. We can thus conclude that the information contained in the dataset (and by using this methodology)

slightly prefers GRT since the posterior is even more peaked toward $m_g = 0$.

2. The Bayes factor

In Secs. III A and III B and Appendix A 1 we explain that, because of the time of computation, it has been chosen to not use the direct evaluation of the χ^2 function but an approximation based on the Gaussian process. We also used the Gaussian process as a way to assess the uncertainties of this interpolation and show in Sec. III C that the results were only marginally impacted by the interpolation uncertainties. However, we will see that the result of the nondetection presented in Sec. III C can also be sustained by fully integrated and adjusted ephemerides without any interpolation. In order to assess this point, we computed an estimation of the Bayes factor using the same set of masses that have been used for the GPR interpolation and for which fully integrated and adjusted INPOP ephemerides have been built, and a χ^2 computed (see Appendix A 1). The Bayes factor is a tool used in the context of model selection (see, e.g., [25]). The central notion is that prior and posterior information should be combined in a ratio that provides evidence of one model specification M_1 over an other M_2. The Bayes factor can be interpreted as a quantity saying which model between M_1 and M_2 represents at the best the observed data set $\mathbf{d}_{\text{obs}}$. In our case selecting a specific model means to select a specific value of m_g. The Bayesian setup requires a prior distribution for the parameter m_g we are dealing with, that, for us, is $\rho(m_g)$. The quantity of interest is then the ratio:

$$BF = \frac{\pi(M_1|\mathbf{d}_{\text{obs}})}{\pi(M_2|\mathbf{d}_{\text{obs}})} \times \frac{\rho(M_2)}{\rho(M_1)}. \qquad (8)$$

In the case of equal priors for M_1 and M_2 the Bayes factor equals to the ratio of the likelihoods. For major details about the Bayes factor and a more extensive explanation the reader is addressed to [25]. The interpretation of the Bayes factor (BF) is not always easy. We will take as reference the so-called Jeffreys' scale which is a qualitative judgment on the evidence. Generally speaking, if $BF \gg 1$, then it indicates that the weight of model 1 is greater than model 2, and on the other hand if $BF \ll 1$ then model 2 has more weight. In the zones of the domain for which $BF \simeq 1$ neither model 1 nor 2 is predominant. The Bayes factor is computed with real unapproximated χ^2 and not with interpolated values. As model 1, we use an extremely small value for the mass of graviton i.e. $m_g \simeq 10^{-40} \times 10^{-23}$ eV c^{-2} mimicking GRT; whereas with model 2, we employ a massive graviton.

In Fig. 5 the Bayes factor is presented relative to the $\log_{10}$ of the graviton mass. The BF values tend to be above 1 for almost all the values of m_g, and it is close to 1 for values of m_g roughly smaller than 10^{-25} eV c^{-2}.
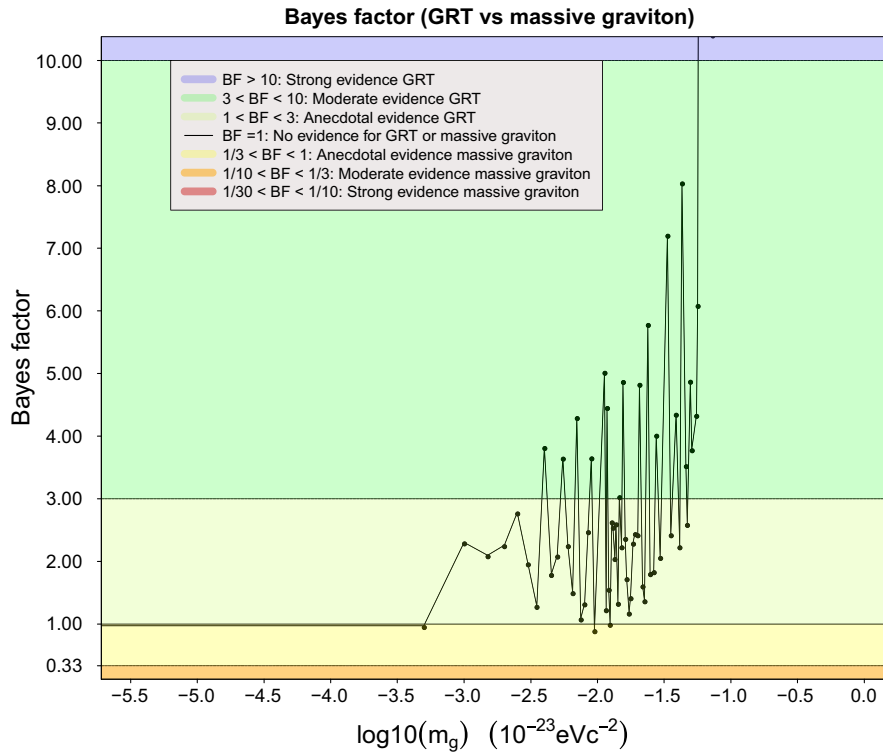
FIG. 5. Value of the Bayes factor as function of m_g. Each point corresponds to one INPOP run. On the x axis we put the masses in terms of $\log_{10}(m_g)$.

The conclusion is also that GRT remains the most likely model ($BF > 10$) up to $m_g = 0.06 \times 10^{-23}$ eV c^{-2}. For $m_g \leq 0.06 \times 10^{-23}$ eV c^{-2}, $BF > 1$ almost everywhere, except in some points for which we have $BF \simeq 1$. Below 0.06×10^{-23} eV c^{-2} the general trend is in any case with anecdotal or moderate evidence towards GRT. In Fig. 5 one can see that we do not obtain any evidence for a massive graviton favored over GRT, for any possible ephemerides with $m_g \neq 0$. This conclusion is consistent with what has been discussed in Secs. III C and IV A 1 with less resolution as with the GPR and MH algorithms but in using full estimated χ^2 function without approximations.

B. Comparisons with previous estimates

1. Comparison with previous INPOP estimations

References [10,11] have given upper bounds for the mass of the graviton obtained with INPOP17b and INPOP19a. In our work are presented a generalization and an improvement of the results given in [10,11], using INPOP21a. By using a more general semi-Bayesian approach, we are showing that GRT is enough for explaining the data, and the massive graviton is not inducing any improvement in the planetary model. In order to compare with previous works, we can give an upper limit of the posterior. In particular in [10,11] the upper bound for the mass with a 99.7% confidence level is $m_g \leq 3.62 \times 10^{-23}$ eV c^{-2}. In order to proceed with a comparison, we take the 99.7% confidence level on GPUR (see Sec. III B 2 and Table I) which corresponds to a mass of $m_g \leq 1.01 \times 10^{-24}$ eV c^{-2}. The result is an improvement by 1 order of magnitude in comparison with Bernus *et al.* [11]. It is however interesting to understand if this result is due to the INPOP improvement, or due to the change of methodology. Hence, we used the formalism proposed in [11] with INPOP19a, but using INPOP21a χ^2 values instead. Reference [11] computed a "likelihood" interpreted as the probability of a tested theory to be likely. We refer to such a likelihood as L_B. The results obtained using INPOP21a are shown in Fig. 6. In this figure, one can see that the INPOP21a L_B value is improved by a factor of 3, going from 3.62×10^{-23} to 1.18×10^{-23} eV c^{-2} for a C.L. at 99.7% (and going from 3.16×10^{-23} to 1.03×10^{-23}eVc^{-2} for a C.L. at 90%). A spike is present in Fig. 6 for $m_g = 1.03 \times 10^{-23}$ eV c^{-2}: this is due to a local minimum in the χ^2 function. But let us stress that, even at this local minimum one has the likelihood L_B (at $m_g = 1.03 \times 10^{-23}$ eV c^{-2}) that is still smaller than the likelihood L_B obtained in GRT. Thus, GRT is still the most favourite guess between the two possibilities. The local minimum is also present in the function $m_g \longmapsto \tilde{\chi}^2(m_g)$ used in the MH algorithm: the MCMC process takes into account this local minimum and overcomes it going towards the global minimum at about $m_g \simeq 0$. For further details on the computation of L_B, see Ref. [11]. We can conclude that both the new
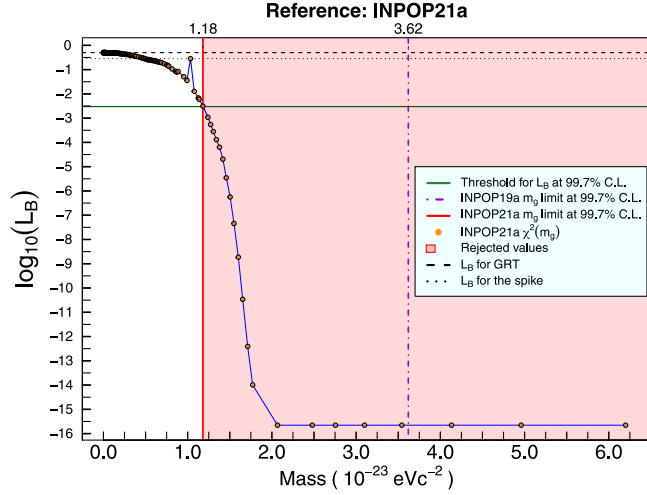
FIG. 6.   Values of the likelihood L_B as computed in [11]. The orange dots represent the $\log_{10}(L_B)$ values computed with IN-POP21a. The green horizontal line is a threshold value for L_B at 99.7% with the same criteria used in [11]. The vertical red line is the m_g limit at 99.7% C.L. with the same criteria adopted in [11] but using INPOP21a. The violet vertical line is the m_g limit at 99.7% C.L. proposed in [11] using INPOP19a. The shaded red zone is the zone of rejected value we would obtain with INPOP21a at 99.7% C.L. The horizontal dashed line represent the L_B value for GRT, whereas the horizontal dotted line represents the L_B value for the spike around $m_g = 1.03 \times 10^{-23}$ eV c^{-2}.

model and the new observations introduced with INPOP21a induce a factor of 3 improvement relative to [11] determination on m_g. The MH algorithm implemented here goes even further and explores a zone close to $m_g = 0$.

2. Comparison with LIGO-Virgo-KAGRA estimations

The LIGO-Virgo-KAGRA collaboration presents an updated bound of the mass of the graviton m_g at 90% confidence level, that is $m_g \leq 1.27 \times 10^{-23}$ eV c^{-2} in [5]. The posterior then obtained from the MCMC in our work seems to give more than 1 order of magnitude better constraint. However, it is important to stress that the two studies are perfectly complementary because they focus on different aspects of the massive gravity phenomenology (radiative versus orbital), and also use totally different observations (gravitational waves versus astrometry in the Solar System). In particular, decoupling mechanisms (in some specific theories of massive gravity) could in principle suppress one aspect of the massive gravity phenomenology and not the other, such that, from a phenomenological point of view, it will always be important to probe both aspect of the massive gravity phenomenology.

C. Improvement with the BepiColombo mission

The BepiColombo mission will arrive at Mercury in 2026 for at least two years of close circular orbit about the planet. Thanks to the MORE radio science experiment,

measurements of the Mercury to Earth distances will be obtained with an accuracy of about 1 cm in KaKa band [26]. Based on such an accuracy, we suppose for our simulations a daily acquisition of range tracking data during a period of 2.5 years, from 2026 to 2028.5 [27] and in using INPOP21a as reference solution in the GRT framework. In order to assess if the MORE range data will be sensitive to the mass of the graviton and up to which level these measurements can help to improve the results of this work, we computed the differences for the Earth-Mercury distances during the two years of the MORE observation period between INPOP21a and ephemerides built and fitted with different values of the graviton mass m_g. As it has been already discussed, a given m_g value fixed in the computation of the ephemerides yields to a perturbation of the orbits, and then on the residuals, with respect to ephemerides in which GRT is assumed. Quantifying such a perturbation provides a way to see if data from future interplanetary missions may play a role in the m_g limit determination, keeping in mind that after fit of the perturbed ephemeris, most of the perturbation will be absorbed by other parameters (see discussion about correlations in i.e. [10] or [4]). As the expected BepiColombo accuracy for the range is at 1 cm level, one can expect that for perturbations below this threshold of 10^{-2} m would not be detectable by the MORE measurements. In Fig. 7, the standard deviations $\Delta\sigma$ of the Earth-Mercury distance perturbations are presented as function of the massive graviton m_g considered. We can see that the smallest m_g that might produce a significant perturbation is roughly $m_g = 0.087 \times 10^{-23}$ eV c^{-2}, which is only 13.8% below
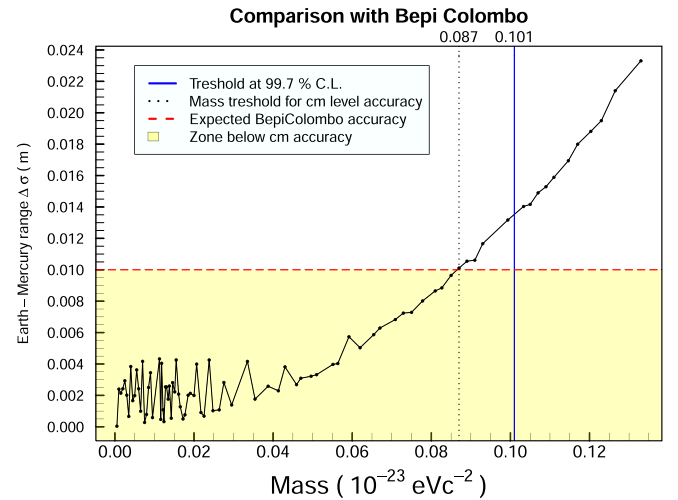

FIG. 7.   Standard deviations of the differences on Earth-Mercury distances ($\Delta\sigma$) between INPOP21a and massive graviton ephemerides as function of the massive graviton m_g. The red dashed horizontal line represents the BepiColombo measurement uncertainty whereas the yellow area represents the zone below the 1 cm accuracy. The black dotted vertical line gives the smallest detectable m_g value.

the bound provided in Sec. III B 2. Because of correlations between parameters, we are not considering this limit as the one that would minimize the future INPOP χ^2 including the MORE data, but more as a minimum threshold below which the mass of the graviton will not be detectable by the BepiColombo radio science experiment. As a consequence, we expect a marginal improvement on the graviton mass from the BepiColombo mission as the threshold of detection with MORE will be only 13.8% smaller than the limit given in Sec. III B 2

V. CONCLUSION

We have presented our work on the use of MCMC algorithm and INPOP in order to get an improvement of the detection limit of the mass of the graviton m_g using the Solar System dynamics as arena. A key strength of the present study is to include the m_g contribution in terms of accelerations and light times computation within the full dynamics of the Solar System. Moreover, by considering a semi-Bayesian approach, and using MCMC and MH algorithms, we avoid correlations between m_g and other INPOP astronomical parameters. We used the GPR to obtain an approximation of the χ^2 ready to use within the MH algorithm and to asses the uncertainty of approximation afterwards. From the posterior obtained, we can give an upper bound of $m_g \leq 1.01 \times 10^{-24}$ eV c^{-2} at 99.7% C.L. (respectively, $\lambda_g \geq 122.48 \times 10^{13}$ km) and $m_g \leq 0.59 \times 10^{-24}$ eV c^{-2} at 90% C.L. (respectively, $\lambda_g \geq 209.67 \times 10^{13}$ km), including approximation uncertainty, and we had shown with a change of the prior (from flat to half-Laplace) that no significant information is detectable in planetary ephemerides for masses smaller than this limit. Observations from the BepiColombo mission will provide new data that, according to our analysis, should not lead to a major improvement on the mass of the graviton.

APPENDIX: METHOD

In this section are given the most important aspects of the method for the MH algorithm associated with an interpolation of the χ^2 function and the uncertainty assessments based on the Gaussian process (GP). Complementary information and fully detailed implementation, as well as validation tests, can be found in [24].

1. Metropolis-Hastings algorithm and Gaussian process regression

In the past years there have been already some attempts to deal with the problem of high correlations among parameters inside the INPOP planetary ephemerides fit, as in the case of the determination of asteroid masses (see Ref. [20]). For testing alternative theories and thus assessing threshold values for the violation of GRT, [17] had tested genetic algorithm approaches for identifying intervals of values for parameters such as parameterized post-Newtonian, β, γ, the Sun oblateness J_2 and secular variations of the gravitationnal mass of the Sun $\frac{\dot{\mu}}{\mu}$, with which planetary ephemerides can be computed and fitted to the observations with a comparable accuracy than the ephemerides built in GRT.

Keeping in mind the problem of correlation between planetary ephemerides and GRT parameters, we propose a new procedure with a semi-Bayesian approach to test a possible deviation from GRT in a particular case: we investigate the posterior probability distribution of a possible nonzero mass of the graviton m_g employing MCMC techniques.

Our approach is semi-Bayesian in the sense that only the mass of the graviton is actually sampled with the MH algorithm procedure, the INPOP astronomical parameters being fitted with a least square procedure. We follow here the algorithm already used by [10,11,19]: for a fixed value of m_g, we integrate the motion of the planets with INPOP and we fit to planetary observations, the astronomical parameters listed in Sec. II B in using the least square iterative procedure described in [20]. We then obtain a fully fitted ephemeris built for a fixed value of m_g. The fitted ephemerides are necessary in order to compute the likelihood, essential part of any MCMC procedure. Generally speaking, the Metropolis-Hastings algorithm is one of the first Markov chain Monte Carlo methods developed, providing a sequence of random samples drawing them from a given probability distribution. The probability distribution from which we draw our samples is the posterior, taking into account the Solar System dynamics and all the observations used in INPOP. The MH algorithm is based on the idea that the drawings are done sequentially and according to a random acceptance process. The outcome of the algorithm is a sequence of random samples from the posterior. Indeed, this sequence is a Markov chain, and its equilibrium distribution is the posterior itself. At each step of the algorithm, one new element of the sequence is computed and proposed, being accepted or rejected. The acceptance/rejection is random, according to a certain probability (changing at each step) based on the last accepted element of the sequence and the likelihood of the last element and of the new candidate element. We will

not provide a proof of convergence of the method since it is out of the goal of our work and it can be found easily in [28,29]. For a detailed overview on the MCMC method, see Refs. [28,29]. The details about the specific MH implementation adopted for this work can be found in [24]. The MH algorithm will play a role only on the selection of the parameter of interest (i.e. the mass of the graviton m_g) whereas the 402 other INPOP parameters are obtained by regular least squares. This hybrid approach has been chosen because, from one side, it would be very costly in term of time of computation to explore the full parameter space with a MH algorithm, and, on the other side, this is a way to avoid correlation of m_g with the other INPOP parameters. Similarly to [11], the $\chi^2(m_g)$ computed will be estimated setting fixed the value of m_g and fitting the remaining **k** parameters [see Eq. (7)] with the full INPOP adjustment. Differently from [11], we provide a posterior probability distribution for m_g, and it is done with the MH algorithm. Finally, we approximate the χ^2 function with $m_g \longmapsto \tilde{\chi}^2(m_g)$ considering a set of values $S = \{(\vec{m}_g, \chi^2(\vec{m}_g))\}$ and interpolating among the points of the set S. $\vec{m}_g$ gathers values of the masses for which we are going to compute the actual corresponding χ^2 unapproximated. The interpolation $m_g \longmapsto \tilde{\chi}^2(m_g)$ has been built in exploiting a GP among the points of the set S. The GPR is a method to predict a continuous variable as a function of one or more dependent variables, where the prediction takes the form of a probability distribution (see, e.g., [30,31]). A full description of the method can be found in [24] and the reader is referred to [31] for a more detailed discussion on the GPs and to [32] for the documentation of the package GPfit that we used to produce our GPR. The notation used for the general GPR description is similar to what is found in [30,31]. To compute a Gaussian process regression among a given set of points, an function m (in GPR jargon the *mean function*) and a function K (in GPR jargon the *kernel function*) are necessary. In our case the m function is the best linear unbiased predictor (BLUP), as described in [32,33]. As kernel we used a standard exponential kernel such as Eq. (A1)

$$K(x, x') = \exp\left(-\frac{|x - x'|^\alpha}{\ell^2}\right) \qquad (A1)$$

for suitable ℓ and α parameters to tune. The points we are interpolating with the GPR (in GPR jargon the *observations*, see Ref. [31]) are considered as noise-free since they are computer simulations. We refer the reader to [34] for further details about this assumption. However, as we are using the GPR for interpolating χ^2 values that will be seen as forward model outcome by the MCMC runs, it is interesting to consider the uncertainty of this interpolation (see Refs. [34,35]) for the interpretation of the MCMC results, propagating this uncertainty in the posterior sampled by the MH algorithm.
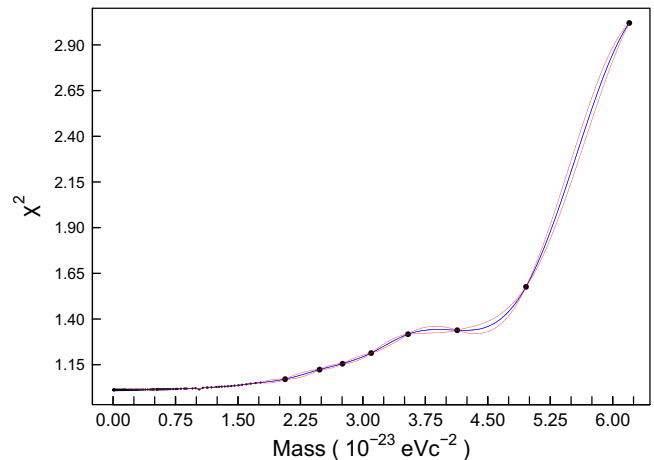

FIG. 8. Plot of the function $m_g \longmapsto \tilde{\chi}^2(m_g)$. The dots correspond to χ^2 value estimated with INPOP and not extrapolated. The red lines correspond to the 2-σ uncertainties provided by the GPR.

In Fig. 8 we see the evolution of the χ^2 as a function of m_g as well as the GPR outcome. The black dots indicated the χ^2 obtained with fully integrated and converged ephemerides when the blue line represents the mean value obtained from GPR (see details in [24]). Moreover the red lines indicate the estimates of uncertainty provided by the GPR at 2σ level. In particular the graphs of the two red lines are $m_g \longmapsto \tilde{\chi}^2(m_g) \pm 2\tilde{\sigma}(m_g)$. We see that $m_g \longmapsto \tilde{\sigma}(m_g)$ is zero or close to zero when m_g is used to compute $\chi^2(m_g)$ from INPOP. This is due to the assumption that the $\chi^2(\vec{m}_g)$ are computed with zero noise as these values are directly obtained from INPOP construction. In Fig. 9 an enlargement of Fig. 8 is provided on the final interval of interest of the MH algorithm.
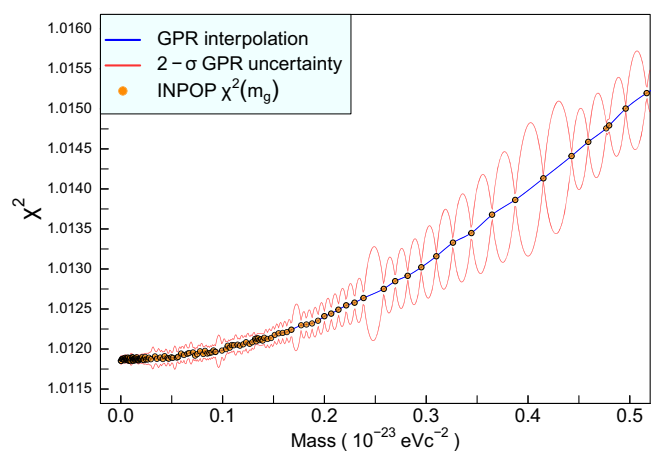

FIG. 9. Plot of the function $m_g \longmapsto \tilde{\chi}^2(m_g)$ made up with INPOP χ^2. The GPR interpolation is plotted in blue among the points $\chi^2(m_g)$ computed with the full INPOP (orange dots). The red lines represent the 2-σ uncertainty (on the interpolation) provided by the GPR.

2. Gaussian process and uncertainty assessments

One of the advantages about GPR, is that, in principle, we can consider the value of $\tilde{\chi}^2(m_g)$ as a normal random variable with given average and standard deviation provided as outcome of the GPR. As we said previously the estimator in our case is the BLUP, and we call it $m_g \longmapsto \tilde{\chi}^2(m_g)$.

We would like also to exploit these confidence intervals and the fact that, due to the hypothesis of Gaussian process and the assumptions we did, Eq. (A2) holds

$$\forall\, m_g, \quad \chi^2(m_g) \sim \mathcal{N}(\tilde{\chi}^2(m_g), \quad \tilde{\sigma}^2(m_g)). \quad (A2)$$

Based on Eq. (A2), we can produce a *perturbation* of the interpolation $m_g \mapsto \tilde{\chi}^2(m_g)$, that, for sake of clarity, we are going to call GPUE. The underlying idea of GPUE being that, for each m_g of the domain, you can consider the corresponding $\chi^2(m_g)$ as outcome of a random draw of a normal random variable according to the uncertainty of the GPR, i.e. following Eq. (A2).

In Eq. (A2) $\tilde{\sigma}(m_g)$ is the value provided by the GPR to use as standard deviation when we consider $\chi^2(m_g)$ as a normal random variable. We indicate as

$$m_g \longmapsto \check{\chi}^2(m_g), \quad (A3)$$

one *estimation* of the GPR with its own uncertainty. In Fig. 10 the black line represents one GPUE. For a detailed description of how to build such a GPUE, the reader is addressed to [24].

By definition $m_g \longmapsto \check{\chi}^2(m_g)$ is locally defined as a realization of a normal random variable. In order to propagate the GPR uncertainty in the MCMC runs, we are going to use the GPUE. We compute several GPUEs, all different,
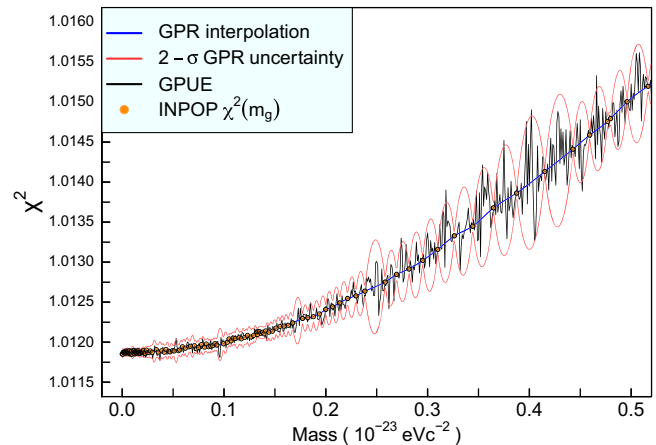

FIG. 10. Plot of the function $m_g \longmapsto \tilde{\chi}^2(m_g)$ made up with INPOP χ^2 (orange dots). The GPR interpolation is plotted in blue among the points $\chi^2(m_g)$ computed with the full INPOP (orange dots). The red lines represent the 2-σ uncertainty (on the interpolation) provided by the GPR. The black line is one possible GPUE (in other words, the GPR interpolation perturbed at each point of the domain). The plot is enlarged on a portion of the domain.

whereupon for each of the maps $m_g \longmapsto \check{\chi}^2(m_g)$ produced, we run separately a Markov chain using the MH algorithm. Doing so we are running MH algorithm to produce a Markov chain on a noisy version of the nominal interpolation obtained with GPR. All the Markov chains obtained are joined together, producing the (blue) posterior presented in Fig. 3. This method is an attempt to assess the *uncertainty of interpolation* that we have. We call such a final posterior a GPUR.

[1] T. D. Moyer, DPODP manual, JPL-TR-32-1527, JPL, https://ntrs.nasa.gov/citations/19710017134.

[2] T. D. Moyer, *Formulation for Observed and Computed Values of Deep Space Network Data Types for Navigation (JPL Deep-Space Communications and Navigation Series)*, 1st ed. (John Wiley & Sons, Inc., Hoboken, NJ, 2003).

[3] M. Soffel, S. A. Klioner, G. Petit, P. Wolf, S. M. Kopeikin, P. Bretagnon, V. A. Brumberg, N. Capitaine, T. Damour, T. Fukushima, B. Guinot, T.-Y. Huang, L. Lindegren, C. Ma, K. Nordtvedt, J. C. Ries, P. K. Seidelmann, D. Vokrouhlick, C. M. Will, and C. Xu, The IAU 2000 resolutions for astrometry, celestial mechanics, and metrology in the relativistic framework: Explanatory supplement, Astron. J. **126**, 2687 (2003).

[4] A. Fienga and O. Minazzoli, Testing GR and alternative theories with planetary ephemerides, arXiv:2303.01821.

[5] The LIGO Scientific, The Virgo, and The KAGRA Collaborations, Tests of general relativity with GWTC-3, arXiv:2112.06861.

[6] Claudia de Rham, J. Tate Deskins, Andrew J. Tolley, and Shuang-Yong Zhou, Graviton mass bounds, Rev. Mod. Phys. **89**, 025004 (2017).

[7] C. de Rham, Massive gravity, Living Rev. Relativity **17**, 7 (2014).

[8] C. M. Will, Bounding the mass of the graviton using gravitational-wave observations of inspiralling compact binaries, Phys. Rev. D **57**, 2061 (1998).

[9] C. M. Will, Solar system versus gravitational-wave bounds on the graviton mass, Classical Quantum Gravity **35**, 17LT01 (2018).

[10] L. Bernus, O. Minazzoli, A. Fienga, M. Gastineau, J. Laskar, and P. Deram, Constraining the Mass of the Graviton with the Planetary Ephemeris INPOP, Phys. Rev. Lett. **123**, 161103 (2019).

[11] L. Bernus, O. Minazzoli, A. Fienga, M. Gastineau, J. Laskar, P. Deram, and A. Di Ruscio, Constraint on the Yukawa suppression of the Newtonian potential from the

planetary ephemeris INPOP19a, Phys. Rev. D **102**, 021501 (2020).

[12] R. Abbott *et al.* (The LIGO Scientific, The Virgo, and The KAGRA Collaborations), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, arXiv:2111.03606.

[13] A. Fienga, H. Manche, J. Laskar, and M. Gastineau, INPOP06: A new numerical planetary ephemeris, Astron. Astrophys. **477**, 315 (2008).

[14] A. Fienga, P. Deram, A. Di Ruscio, V. Viswanathan, J. I. B. Camargo, L. Bernus, M. Gastineau, and J. Laskar, INPOP21a planetary ephemerides, Notes Scientifiques et Techniques de l'Institut de Mecanique Celeste, 110, 2021.

[15] A. Fienga, J. Laskar, P. Kuchynka, H. Manche, G. Desvignes, M. Gastineau, I. Cognard, and G. Theureau, The INPOP10a planetary ephemeris and its applications in fundamental physics, Celestial Mech. **111**, 363 (2011).

[16] A. K. Verma, A. Fienga, J. Laskar, H. Manche, and M. Gastineau, Use of MESSENGER radioscience data to improve planetary ephemeris and to test general relativity, Astron. Astrophys. **561**, A115 (2014).

[17] A. Fienga, J. Laskar, P. Exertier, H. Manche, and M. Gastineau, Numerical estimation of the sensitivity of INPOP planetary ephemerides to general relativity parameters, Celest. Mech. Dyn. Astron. **123**, 325 (2015).

[18] V. Viswanathan, A. Fienga, O. Minazzoli, L. Bernus, J. Laskar, and M. Gastineau, The new lunar ephemeris INPOP17a and its application to fundamental physics, Mon. Not. R. Astron. Soc. **476**, 1877 (2018).

[19] A. Fienga, L. Bigot, D. Mary, P. Deram, A. Di Ruscio, L. Bernus, M. Gastineau, and J. Laskar, Evolution of INPOP planetary ephemerides and Bepi-Colombo simulations, IAU Symp. **364**, 31 (2022).

[20] A. Fienga, C. Avdellidou, and J. Hanuš, Asteroid masses obtained with INPOP planetary ephemerides, Mon. Not. R. Astron. Soc. **492**, 589 (2019).

[21] A. Di Ruscio, A. Fienga, D. Durante, L. Iess, J. Laskar, and M. Gastineau, Analysis of Cassini radio tracking data for the construction of INPOP19a: A new estimate of the Kuiper belt mass, Astron. Astrophys. **640**, A7 (2020).

[22] A. Fienga, A. Di Ruscio, L. Bernus, P. Deram, D. Durante, J. Laskar, and L. Iess, New constraints on the location of P9 obtained with the INPOP19a planetary ephemeris, Astron. Astrophys. **640**, A6 (2020).

[23] A. Fienga, J. Laskar, H. Manche, and M. Gastineau, Constraints on the location of a possible 9th planet derived from the Cassini data, Astron. Astrophys. **587**, L8 (2016).

[24] V. Mariani, A. Fienga, and O. Minazzoli, Testing the mass of the graviton with Bayesian planetary numerical ephemerides B-INPOP, arXiv:2303.05298.

[25] J. Gill, *Bayesian Methods: A Social and Behavioral Sciences Approach*, 2nd ed., Statistics in the Social and Behavioral Sciences Series (CRC Press, Boca Raton, FL, 2008).

[26] L. Iess *et al.*, Gravity, geodesy and fundamental physics with BepiColombo's MORE investigation, Space Sci. Rev. **217**, 21 (2021).

[27] R. N. Thor, R. Kallenbach, U. R. Christensen, A. Stark, G. Steinbrügge, A. Di Ruscio, P. Cappuccio, L. Iess, H. Hussmann, and J. Oberst, Prospects for measuring Mercury's tidal love number h_2 with the BepiColombo laser altimeter, Astron. Astrophys. **633**, A85 (2020).

[28] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed., Springer Texts in Statistics (Springer, New York, 2004), 10.1007/978-1-4757-4145-2.

[29] S. Brooks, A. Gelman, G. Jones, and X. L. Meng, *Handbook of Markov Chain Monte Carlo*, Chapman Hall/CRC Handbooks of Modern Statistical Methods (CRC Press, Boca Raton, FL, 2011), https://www.mcmchandbook.net/.

[30] S. H. Strub, L. Ferraioli, C. Schmelzbach, S. C. Stähler, and D. Giardini, Bayesian parameter estimation of Galactic binaries in LISA data with Gaussian process regression, Phys. Rev. D **106**, 062003 2022.

[31] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning Series (MIT Press, Cambridge, MA, 2005).

[32] B. MacDonald, P. Ranjan, and H. Chipman, GPfit: An R package for fitting a Gaussian process model to deterministic simulator outputs, J. Stat. Softw. **64**, 1 (2015).

[33] G. K. Robinson, That BLUP is a good thing: The estimation of random effects, Stat. Sci. **6**, 15 (1991).

[34] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, Design and analysis of computer experiments, Stat. Sci. **4**, 409 (1989).

[35] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*, 1st ed., Breakthroughs in Statistics (Springer, New York, 2003).