Bayesian analysis of systematic errors in the determination of the constant of gravitation

Stefano Rinaldi[1,2*], Hannah Middleton[3,4], Walter Del Pozzo[1,2] and Jonathan Gair[5]

[1]Dipartimento di Fisica "E. Fermi", Università di Pisa, Pisa, I-56127, Italy.
[2]INFN, Sezione di Pisa, Pisa, I-56127, Italy.
[3]Institute for Gravitational Wave Astronomy & School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, United Kingdom.
[4]OzGrav-Melbourne & School of Physics, University of Melbourne, Parkville, Victoria 3010, Australia.
[5]Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam, 14476, Germany.

*Corresponding author(s). E-mail(s): stefano.rinaldi@phd.unipi.it;

Abstract

Measurements of the gravitational constant G are notoriously difficult. Individual state-of-the-art experiments have managed to determine the value of G with high precision: although, when considered collectively, the range in the measured values of G far exceeds individual uncertainties, suggesting the presence of unaccounted for systematic effects. Here, we propose a Bayesian framework to account for the presence of systematic errors in the various measurement of G while proposing a consensus value, following two paths: a parametric approach, based on the Maximum Entropy Principle, and a non-parametric one, the latter being a very flexible approach not committed to any specific functional form. With both our methods, we find that the uncertainty on this fundamental constant, once systematics are included, is significantly larger than what quoted in CODATA 2018. Moreover, the morphology of the non-parametric distribution hints towards the presence of several sources of unaccounted for systematics. In light of this, we recommend a consensus value for the gravitational constant $G = 6.6740^{+0.0015}_{-0.0015} \times 10^{-11}$ m^3 kg^{-1} s^{-2}.

1 Introduction

The Newtonian constant of gravitation, G, is one of the fundamental constants of modern physics. It was the first fundamental constant to be identified and yet it remains one of the least well known, with large disagreement between experimental measurements. Over several decades, huge experimental efforts have tried to determine the value of G.

Individually, these experiments report relative uncertainties that can be as low as 1.2×10^{-5}: however different experiments find values of G that can be several standard deviations away from each other. With such a range in measurement, combining results into a single best estimate of G is understandably challenging [1, 2].

1

The current accepted value of G comes from the Committee on Data for Science and Technology (CODATA). CODATA periodically provides a set of self-consistent values of the fundamental constants for use by the scientific and technological communities. The recommended value of G from the CODATA 2010 results is $6.67384(80) \times 10^{-11}$ m^3 kg^{-1} s^{-2} [3]. After the addition of three more experimental results, the CODATA 2014 recommended value is $6.67408(31) \times 10^{-11}$ m^3 kg^{-1} s^{-2} [4]. In 2017, a CODATA Special Adjustment [5] was released with the purpose of obtaining the best numerical values of the Planck constant h, the electron mass e, Boltzmann's constant k, and Avogadro's number N_A: however, the value of G was not updated.

The most recent CODATA 2018 recommendation for G comes from [6], where two new experimental results are included and a correction is made to a previously reported value. The current recommended value for G is

$$G = 6.67430(15) \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}\,.$$

Any experiment is affected by noise; the effect of the noise is to induce uncertainty on the quantity of interest. The uncertainty can be *statistical* – the statistical error is the difference between a value measured in a single experiment and the value averaged over many experiments – or *systematic*, which is the difference between the averaged value and the true value of the parameter(s) of interest. The main difference among the two classes of uncertainties is that while the statistical error causes a random shift with zero mean of the measured quantity (thus, in principle, the statistical error can be averaged out simply by repeating the experiment a very large number of times), the systematic error for a specific experiment will shift the expected value away from the *true* value of the measured quantity.

The considerable disagreements among different experiments aimed at determining the gravitational constant suggests the presence of an overarching unidentified source of uncontrolled systematic effects leading to such disparate results but, to the best of our knowledge, no other work addresses the presence of systematic errors in a statistical way.

In this paper, we model systematics within the context of Bayesian probability theory. In particular, we will introduce a so-called *hierarchical* model to infer a probability distribution for the unknown systematic errors. In doing so, we will explore several different assumptions, each reflecting a particular choice regarding the nature and magnitude of the errors. The measurements included in this work are the ones listed in [6].

The rest of the paper is organised as follows: in Sec. 2 we briefly review the measurements of G used in this work and two existing statistical methods to propose a consensus value. In Sec. 3 we describe the Bayesian hierarchical framework used to estimate G. In Sec. 4 we present our results and finally, in Sec. 5, we discuss our findings and conclude with a recommendation on the value of G.

2 Measurements of the gravitational constant

The analysis presented in this paper makes use of 16 experimental results dating from 1982 to 2018. Here we briefly review the methods used in each of the 16 experiments. For a comprehensive and detailed review of the measurements of G, we refer the interested reader to [1] or [2].

Following the approach taken by Cavendish in 1797–1798, the majority of experiments listed in [6] involve precision measurements of a torsion balance. Free deflection was used in BIPM-01 and BIPM-14 [7, 8] as well as electrostatic compensation (see below). Time-of-swing experiments (NIST-82, TR&D-97, LANL-97, HUST-05, HUST-09, UCI-14 and HUST$_T$-18) instead measure the change in oscillation period of the torsion balance with different source mass positioning [9–16]. A third variation on the torsion balance uses electrostatic compensation (BIPM-01, MSL-03 and BIPM-14) [7, 8, 17]. The gravitational torque on the test masses is balanced by an electrostatic torque so that they do not rotate. For UWash-00 and HUST$_A$-18, the torsion balance is rotated on a turntable and feedback is used to change the rotation rate so that the fibre twist is minimised and the angular acceleration of the turntable is equal to the gravitational angular acceleration of the balance [16, 18].
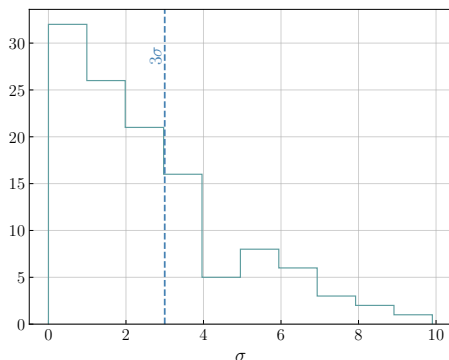
Figure 1 Histogram of σ-levels for all the possible pairs of experiments included in [6]. Assuming a threshold of 3σ for a pair of measurements to be in agreement with each other, 41 pairs out of 120 fall beyond this threshold.

Four experiments listed in [6] do not use a torsion balance method. UWup-02 uses a microwave Fabry-Perot interferometer whose resonance frequency is influenced by the placement of source masses behind each of the reflectors [19]. Similarly, JILA-18 uses a laser Fabry-Perot interferometer to measure the spacing between the test masses of a double pendulum as the positions of source masses around it are changed [20, 21]. UZur-06 uses a beam balance to weigh test masses in the presence of movable source masses [22]. LENS-14 uses atom interferometry to measure how a source mass influences the atom's acceleration [23, 24].

[6] reports the value of G and one-sigma uncertainties σ for each experiment. The range of the measured values is $\approx 0.0037 \times 10^{-11}$ m^3 kg^{-1} s^{-2}: however, the largest individual one-sigma uncertainty is 0.00099×10^{-11} m^3 kg^{-1} s^{-2}, from LENS-14. On the other hand, other measurements report uncertainties as small as $\approx 0.00008 \times 10^{-11}$ m^3 kg^{-1} s^{-2} (HUST-18).

Most importantly, individual observations are often inconsistent with others within their stated measurement uncertainties. If we assume that measurements are inconsistent above the 3σ level, we find that 34% of all the possible pairs of experiments are inconsistent among themselves (see Figure 1).

Recommending a single value from this variety of measurements is understandably difficult.

2.1 Existing statistical frameworks

This is not the first work that tries to reconcile the plethora of different values for G into a single, recommended value. Most of the previous effort, however, is devoted to the identification of potential sources of discrepancy among different experiments, ranging from systematic errors in the measurement apparatuses to the potential presence of an unknown oscillatory factor affecting the measurement process over time or accounting for inaccuracies in Newtonian theory [25, 26].

Among the few works that tries to address the discrepancy among different experiments from a statistical point of view, we outline here two existing frameworks to propose a consensus value for G. The first is the one used by [6], whereas the second is introduced in [27]. Neither of them, however, proposes a framework to account for systematic errors. Here, we briefly review these techniques and in Section 3 we present a new statistical method to account for the presence of systematics in the consensus value.

2.1.1 Tiesinga *et al.*: least-squares procedure

The value proposed in [6], as well as the ones from previous CODATA recommendations, is obtained using a least-squares procedure. In particular, according to [2], having n different measurements $\mathbf{y} = \{y_1, \ldots, y_n\}$ of an unknown quantity $\bar{y}$ with covariance matrix $\mathbf{C}$, minimising the quantity

$$\chi^2 = (\mathbf{y} - \bar{y})\mathbf{C}^{-1}(\mathbf{y} - \bar{y})^{\mathrm{T}} \qquad (1)$$

with respect to $\bar{y}$ leads to the variance-weighted mean of the measurement and its uncertainty.

This method, however, relies on a fundamental assumption: the provided uncertainties have to be statistical in nature rather than systematic. This procedure weights the different measurements according to their precision, trusting more the measurements with the smallest associated error, which is reasonable under the assumption that all the measurements are in agreement with the same *true* value.

In presence of systematic errors, however, there is no reason to believe that the most *precise* measurement is also the most *accurate*. As stated above, the presence of systematic errors shifts the expected value of the affected measurement: thus,

the line of reasoning in which we favour the measurement obtained with a very precise experiment might end up being biased.

2.1.2 Merkatas *et al.*: shades of dark uncertainty

This work [27] suggests that the uncertainty associated with each of the different measurements of the gravitational constant does not account for *all* the statistical uncertainty and that there could be several different latent sources of statistical uncertainty.

The idea is that different experiments are affected by these *shades of dark uncertainty* and that the same source can be shared among different experiments, providing random amounts of additional statistical uncertainty. They propose a Bayesian framework to infer the magnitude of these additional uncertainties required to reach consensus among measurements and then propose a value for G. Their work, however, does not consider the possibility of systematic errors being present.

This approach, in our view, makes sense while grouping experiments that are correlated in some sense (e.g. using the same methodology or being performed by the same people): this way, there is room to believe that uncertainties in similar experiments could have been similarly underestimated.

3 Bayesian hierarchical analysis

In this section we propose a method to combine different measurements of G by employing a hierarchical framework based on Bayesian inference as a way of marginalising over the unknown systematic effects.

Let us begin by defining the value of the gravitational constant as G; we wish to determine G given the ensemble of N experiments $\mathbf{D} = \{D_1, \ldots, D_N\}$ and a model H.

The whole idea of *setting up an experiment*, given the presence of a different systematic error in every measurement, can be represented as follows: every experiment, in presence of systematics, will measure an *experiment value* G_i, which is not the true value of the constant of gravitation G. This value is a realisation of a stochastic process governing the source of systematic errors, being drawn from $p(G|\theta)$:

$$G_i \sim p(G|\theta)\,. \tag{2}$$

θ, here, represents the parameters of our model for the distribution of systematic errors.

Each individual experiment i, in turn, will result in a probability distribution for its own G_i. Hierarchically combining different measurements, therefore, allows us to characterise the probability distribution of systematic errors. This distribution, at the same time, acts as a probability distribution for G, characterising the probability of deviating from the unknown true value: therefore, at the end of this work, we will recommend a value for G based on this probability density.

The application of this population study-like approach is possible thanks to the independence of all the systematic errors at play[1] under the assumption that, although two different experiments may share the same source of systematics (e.g. using the same experimental setup), the magnitude of the systematic is different for each of them.

The probability distribution $p(G|\theta)$ is determined by its functional form and by a set of parameters θ. If one wants to reconstruct the systematic error distribution – which means assuming a functional form for $p(G|\theta)$ and inferring its parameters θ – and therefore give a probability distribution for G, the data to use in such inference are the experiment values $\mathbf{G} = \{G_1, \ldots G_N\}$. Unfortunately, due to the presence of statistical uncertainty, these values are unknown, since the G_is are the values that would be measured by experiments in absence of statistical error: every experiment, implicitly, gives a probability distribution for this quantity while reporting a value with an associated error. Having at hand only the N posterior distributions $\mathbf{D}$ provided by our experiments, we need to combine the experiment outcomes $\mathbf{D}$ in a hierarchical fashion.

Within the Bayesian framework and under the assumption of a functional form for the systematic error distribution[2], the inference is completely

---

[1]For most of the measurements included in this work, this is a safe assumption. Three pairs of these experiments, however, are correlated (see caption of Table XXIX in [6]). Given the fact that these correlations are low, with the highest correlation coefficient being r(NIST-82,NANL-97) = 0.351, we opted to neglect these correlations.

[2]This assumption is included in the hypothesis H.

described by the posterior distribution for θ:

$$p(\theta|\mathbf{D}, H) = \frac{p(\mathbf{D}|\theta, H)p(\theta|H)}{p(\mathbf{D}|H)}, \qquad (3)$$

where $p(\theta|H)$ is the prior probability distribution for θ, describing our *a priori* expectation for its value. The likelihood function $p(\mathbf{D}|\theta, H)$ is known only conditioned on the knowledge of $\mathbf{G}$. Marginalising over this quantity, we get

$$p(\theta|\mathbf{D}, H) = \frac{p(\theta|H)\int p(\mathbf{D}|\theta, \mathbf{G}, H)p(\mathbf{G}|\theta, H)\mathrm{d}\mathbf{G}}{p(\mathbf{D}|H)}. \qquad (4)$$

Here $p(\mathbf{G}|\theta, H)$ represents the systematic error distribution. Under the assumption of statistical independence of each experiment, can be factorised into the product of probabilities:

$$p(\mathbf{G}|\theta, H) = \prod_i^N p(G_i|\theta, H). \qquad (5)$$

The denominator $p(\mathbf{D}|H)$ is the so-called *evidence*, which is given by the integral over all the parameters characterising the statistical model induced by the hypothesis H.

3.1 Likelihood function

The likelihood function $p(\mathbf{D}|\theta, \mathbf{G}, H)$ describes, in fact, the likelihood of observing the available data given a specific value for the parameters that we want to infer. Making use, once again, of the assumption of statistical independence of each experiment and of the fact that each D_i is independent of G_j for $j \neq i$, the likelihood factorises into the product of individual likelihoods:

$$p(\mathbf{D}|\theta, \mathbf{G}, H) = \prod_i^N p(D_i|G_i, H). \qquad (6)$$

Once the *experiment value* of G, G_i, is known, the posterior distribution for each experiment does not depend on the values of the parameters θ, since these describes only the systematic error distribution: this is a consequence of the fact that systematic errors cannot be removed or accounted for *a posteriori*.

Every experiment implicitly gives a posterior distribution for G_i, hence $p(G_i|D_i, H)$: making use

of the Bayes' theorem, we get

$$p(\mathbf{D}|\theta, \mathbf{G}, H) = \prod_i^N \frac{p(G_i|D_i, H)p(D_i|H)}{p(G_i|H)}. \qquad (7)$$

$p(G_i|H)$ is the prior on each G_i, which we take uniform between $G_{\min} = 6.668 \times 10^{-11}$ m^3 kg^{-1} s^{-2} and $G_{\max} = 6.678 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, and $p(D_i|H)$ is the evidence for the single experiment outcome.

Our framework needs to include a functional form for these N posterior distributions. Given that the only information we have available are the central value $\hat{G}_i$ and the uncertainty σ_i around it – therefore $D_i = \{\hat{G}_i, \sigma_i\}$ – following the Maximum Entropy Principle (MEP) [28] we assume a Gaussian distribution[3]. Under this assumption, the likelihood for each measurement reads

$$p(G_i|D_i, H) \propto \exp\left[-\frac{1}{2}\left(\frac{G_i - \hat{G}_i}{\sigma_i}\right)^2\right]. \qquad (8)$$

3.2 Systematic effects modelling

In order to reconstruct the probability distribution $p(G|\theta)$, we need to assume a model for this distribution. Here, we propose two different models, based on different assumptions.

3.2.1 Maximum Entropy Principle: Gaussian distribution

We model the effect of the unknown systematic errors as follows: since we consider only the dispersion of systematics, we once again appeal to the MEP to choose the probability distribution $p(G|\theta)$. Given that we want to give an expected value and an uncertainty for G, the distribution is taken to be a Gaussian distribution with mean $\hat{G}$ and unknown standard deviation Σ, therefore $\theta = \{\hat{G}, \Sigma\}$.

Under this assumption, we can write:

$$p(G_i|\theta, H) = \exp\left[-\frac{1}{2}\left(\frac{G_i - \hat{G}}{\Sigma}\right)^2\right]. \qquad (9)$$

---

[3]The MEP states that the probability distribution that maximise the information entropy making use of the least amount of information or, in some late sense, the most conservative choice knowing only the expected value and the variance is the Gaussian distribution.

The assumption of a Gaussian distribution both for $p(G_i|D_i)$ and $p(G_i|\theta)$ is particularly useful, since it is possible to marginalise over G_i analytically. In fact, making use of the fact that the integrals are independent,

$$\prod_i^N \int \mathcal{N}(G_i|\hat{G}, \sigma_i)\mathcal{N}(G_i|\hat{G}, \Sigma)\mathrm{d}G_i$$
$$= \prod_i^N \mathcal{N}\left(\hat{G}_i \middle| \hat{G}, \sqrt{\sigma_i^2 + \Sigma^2}\right), \quad (10)$$

where we denoted with $\mathcal{N}(\cdot|\mu, \sigma)$ the Gaussian distribution.

The prior $p(\theta|H)$ is composed by the prior on $\hat{G}$, which we take uniform between $G_{\min}$ and $G_{\max}$, and the prior on Σ. We will consider several possible choices for this distribution, following some of the prescriptions discussed in [29]:

- UN: a uniform distribution for Σ. Using this prior probability distribution means assuming that we have no information at all regarding the value of the systematic error;
- JF: a uniform distribution over $\log \Sigma$. This is the so- called Jeffreys' prior, corresponding to the assumption that we have no information about the value of the order of magnitude of Σ. A change of variable shows that the probability density function for Σ is proportional to $1/\Sigma$, hence reflecting the expectation that the systematic errors are small;
- IG: an Inverse Gamma distribution

$$p(\Sigma|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\Sigma^{-(\alpha+1)}\exp\left[-\frac{\beta}{\Sigma^2}\right], \quad (11)$$

where $\alpha > 0$ and $\beta > 0$ are called the *shape* and *scale* parameters that determine the morphology of the Inverse Gamma distribution, and $\Gamma(\alpha)$ is the complete Gamma function. We infer α and β from the experimental values, assigning uniform priors between 0 and 100. The Inverse Gamma distribution is *conjugate* to the Gaussian distribution. This guarantees that the posterior on Σ is still an Inverse gamma distribution.

3.2.2 Non-parametric reconstruction: (H)DPGMM

The second model we use is (H)DPGMM, a non-parametric model introduced in [30]. In what follows, we will give a brief overview of the model, referring the interested reader to the relevant papers for more details. Bayesian non-parametric methods are powerful tools that allow us to perform an inference without committing to any specific model prescription. This results in an extreme flexibility when it comes to modelling unknown distributions: all the information that is encoded in the inferred distribution is extracted from the data themselves.

In particular, this model relies on the Dirichlet Process Gaussian Mixture Model [31] or DPGMM, an infinite weighted sum of Gaussian distributions with a Dirichlet Process [32] as prior distribution on weights, to approximate the unknown probability distribution:

$$p(x) \approx \sum_i^\infty w_i \mathcal{N}(x|\mu_i, \sigma_i). \quad (12)$$

The *standard* DPGMM is used to reconstruct an *outer* probability distribution when samples $\mathbf{x} = \{x_1, \ldots, x_N\}$ from the unknown distribution $p(x)$ are available. This is not always the case: there are situations, like the mass function inference described in [30], in which we do not have direct access to samples, but rather we have N sets of *inner* samples drawn from the N posterior distributions (*inner distributions*) for each sample x_i.

To infer the outer distribution having at hand only the N sets of inner posterior samples, one needs to specify a model for both the outer distribution and for the N inner posterior distributions: (H)DPGMM models both the inner and the outer distributions as DPGMM, linking them in a hierarchical fashion.

This is a very similar situation to the one we are addressing in this paper. $p(G)$ is the outer distribution and $p(G_i|D_i)$ are the N posterior distributions we want to use to infer $p(G)$. In general, in order to apply (H)DPGMM, we would need to approximate $p(G_i|D_i)$ with a weighted sum of Gaussian distributions: however, we can interpret the likelihood (8) as a DPGMM with a single component with $w_i = 1$, whereas every other Gaussian

component has $w_j = 0$, and use it as a very simple non-parametric reconstruction.

In this case, the parameter vector $\theta = \{\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\sigma}\}$ is composed of a vector of relative weights, a vector of means and a vector of standard deviations. The length of these vectors is, *a priori*, not limited.

The outcome of such a model, applied to the problem we are dealing with, is a phenomenological distribution for the gravitational constant G.

3.3 Inference

We proceed now to specify how the parameters θ for each hypothesis are inferred.

The expression for the posterior distribution, under the MEP hypothesis, becomes

$$p(\theta|\mathbf{D}, H) \propto p(\theta|H) \prod_i^N \mathcal{N}\left(\hat{G}_i \middle| \hat{G}, \sqrt{\sigma_i^2 + \Sigma^2}\right). \tag{13}$$

For each of the three different prior prescriptions for Σ (UN, JF and IG), we generate samples from Eq. (13) using a nested sampling algorithm [33], CPNEST [34].

For the non-parametric hypothesis, (H)DPGMM, we explore the posterior distribution drawing different realisations for $p(G)$ using FIGARO, a Gibbs sampler presented in [35].

4 Results

We summarise here our findings for each of the systematic error models considered in this work.

Posterior distributions for G under the UN, JF, IG and (H)DPGMM hypothesis are shown in Figure 2. The shape of the non-parametric distribution for G is very different from the shape of the simple Gaussian distribution assumed under the MEP hypothesis: in fact, three different modes are clearly distinguishable. We retain this fact as qualitative evidence of the presence of uncontrolled systematics.

In the following, we summarise the main findings for the proposed models by discussing the inferred posterior distribution, reporting median and 68% credible interval from the median posterior:

- UN: $G = 6.6739(10) \times 10^{-11}$ m^3 kg^{-1} s^{-2} ;
- JF: $G = 6.6739(9) \times 10^{-11}$ m^3 kg^{-1} s^{-2} ;

- IG: $G = 6.6739(10) \times 10^{-11}$ m^3 kg^{-1} s^{-2} .

Concerning (H)DPGMM, the median posterior distribution for G under this hypothesis gives median and 68% credible interval

$$G = 6.6740^{+0.0007}_{-0.0009} \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2} ,$$

and median and 90% credible interval

$$G = 6.6740^{+0.0015}_{-0.0015} \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2} .$$

The presence of three different modes in the non-parametric reconstruction could hint towards the interpretation of *families* of systematic effects, similar to the founding idea of [27]: before commenting on this, however, we want to make clear that an extensive discussion of the systematics that might affect the individual experiments is well beyond our area of expertise. Therefore, the following discussion must be taken as heuristic and driven by statistical considerations only: before claiming that two or more experiments are affected by the same, or at least similar, systematics, it is necessary a dedicated study on the potential sources of such systematic errors.

These three modes might suggest that at least three (or two, if we assume that one of these modes is free of systematics) different effects are at play. While it is not possible to tell which (if any) of the three modes is unaffected by systematics, we note that the rightmost mode contains two measurements only, BIPM-01 and BIPM-14. These two experiments share both the same methodology and the same group, making plausible (not *likely*) for their results to be affected by the same source of systematics[4].

An alternative – but incorrect – interpretation of these results might be to use the mean parameter of the Gaussian distribution $\hat{G}$ from the MEP model as the true value of the gravitational constant G. The inferred value is $\hat{G} = 6.6739 \pm 0.0003 \times 10^{-11}$ m^3 kg^{-1} s^{-2}, estimated via Monte Carlo sampling. This value is similar to the method and result presented in [27], in which the authors address the same issue by proposing an additional, latent source of uncertainty.

---

[4]For intellectual honesty, we note that it might also be possible to suggest, making use of the same line of reasoning, that these two are the only experiments unaffected by systematics.
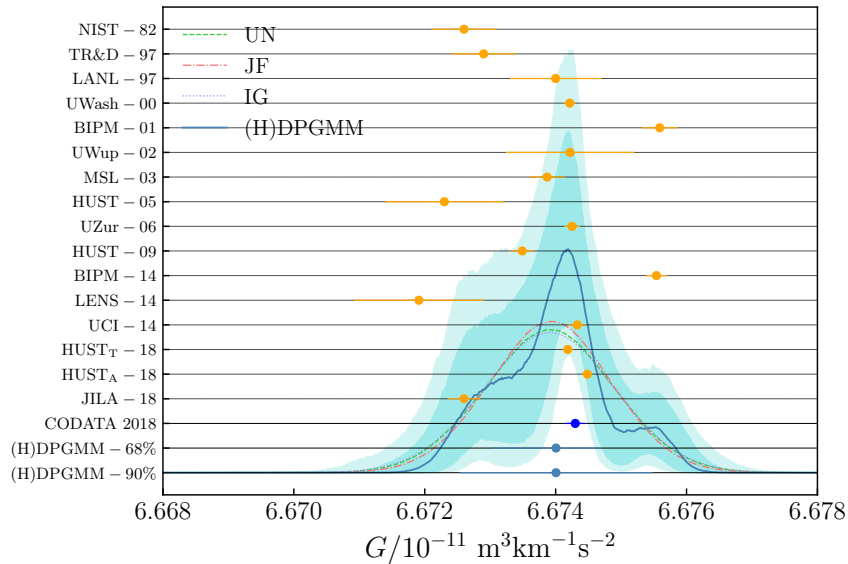
Figure 2 Median posterior distribution for G under the models presented in this paper. For the (H)DPGMM reconstruction, the solid blue line represents the median and the shaded regions indicates the 68% (dark turquoise) and 90% (light turquoise) credible regions for the posterior on G. As a comparison, we report the experimental values and their standard error (orange symbols), as well as the [6] (CODATA 2018) recommended value (blue symbol) and our most probable value for G and 68% and 90% (conservative) credible intervals for (H)DPGMM.

Interpreting this quantity, which we report for completeness, as the true value for the gravitational constant, in this framework, is conceptually incorrect: $\hat{G}$ is a parameter of the posterior distribution for G. For a Gaussian distribution, the median coincides with the mean parameter: the uncertainty on the inferred mean parameter is, however, in general much smaller than the variance of the distribution, leading to an underestimation of the uncertainty on G.

5 Conclusions

In this paper, we proposed two different models to reconstruct the probability distribution $p(G)$ assuming the presence of systematic effects. We found that, although the numerical values for the gravitational constant are very similar among the two models, the functional form reconstructed by the non-parametric one is morphologically different from the Gaussian distribution that arises from the MEP hypothesis.

This suggests that the systematic effects at work behind the experiments we considered are not under control: although some of these measurements are extremely *precise*, they are not very *accurate*. Therefore, further studies are required

both to understand the systematics that affect these experiments and to pinpoint the value of the gravitational constant. Such studies are already taking place, as described in [36] and references therein or in [37].

In light of our investigations, we find that the latest CODATA recommended value is heavily underestimating the actual uncertainty on G. Hence, although this is not the purpose of this paper, we think that the best value to adopt for G is the most conservative we find under the most general assumptions, the one from the (H)DPGMM model:

$$G = 6.6740^{+0.0015}_{-0.0015} \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}.$$

Acknowledgements

Statements and declarations

Funding

Code availability

All the results and figures presented in this paper can be reproduced using the code available at https://github.com/sterinaldi/fund_const. FIGARO is publicly available at https://github.com/sterinaldi/figaro.

Data availability

This manuscript has no associated data.

Competing interests

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

[1] C. Speake, T. Quinn, The search for Newton's constant. Physics Today **67**(7), 27–33 (2014). https://doi.org/10.1063/PT.3.2447

[2] B.M. Wood, Recommending a value for the Newtonian gravitational constant. Philosophical Transactions of the Royal Society of London Series A **372**(2026), 20140,029–20140,029 (2014). https://doi.org/10.1098/rsta.2014.0029

[3] P.J. Mohr, B.N. Taylor, D.B. Newell, CODATA recommended values of the fundamental physical constants: 2010. Reviews of Modern Physics **84**(4), 1527–1605 (2012). https://doi.org/10.1103/RevModPhys.84.1527. https://arxiv.org/abs/1203.5425 [physics.atom-ph]

[4] P.J. Mohr, D.B. Newell, B.N. Taylor, CODATA recommended values of the fundamental physical constants: 2014. Reviews of Modern Physics **88**(3), 035009 (2016). https://doi.org/10.1103/RevModPhys.88.035009. https://arxiv.org/abs/1507.07956 [physics.atom-ph]

[5] P. Mohr, D.B. Newell, B.N. Taylor, E. Tiesinga, Data and analysis for the CODATA 2017 special fundamental constants adjustment. Metrologia **55**(1), 125 (2018). https://doi.org/10.1088/1681-7575/aa99bc

[6] E. Tiesinga, P.J. Mohr, D.B. Newell, B.N. Taylor, CODATA recommended values of the fundamental physical constants: 2018. Reviews of Modern Physics **93**(2), 025010 (2021). https://doi.org/10.1103/RevModPhys.93.025010

[7] T.J. Quinn, C.C. Speake, S.J. Richman, R.S. Davis, A. Picard, A New Determination of *G* Using Two Methods. Phys. Rev. Lett. **87**, 111,101 (2001). https://doi.org/10.1103/PhysRevLett.87.111101. URL https://link.aps.org/doi/10.1103/PhysRevLett.87.111101

[8] T. Quinn, C. Speake, H. Parks, R. Davis, Erratum: Improved Determination of *G* Using Two Methods [Phys. Rev. Lett. 111, 101102 (2013)]. Phys. Rev. Lett. **113**, 039,901 (2014). https://doi.org/10.1103/PhysRevLett.113.039901. URL https://link.aps.org/doi/10.1103/PhysRevLett.113.039901

[9] G.G. Luther, W.R. Towler, Redetermination of the Newtonian gravitational constant G. Phys. Rev. Lett.**48**(3), 121–123 (1982). https://doi.org/10.1103/PhysRevLett.48.121

[10] O.V. Karagioz, V.P. Izmailov, Measurement of the gravitational constant with a torsion balance. Measurement Techniques **39**(10), 979–987 (1996). https://doi.org/10.1007/BF02377461. URL https://doi.org/10.1007/BF02377461

[11] C.H. Bagley, G.G. Luther, Preliminary Results of a Determination of the Newtonian Constant of Gravitation: A Test of the Kuroda Hypothesis. Phys. Rev. Lett. **78**, 3047–3050 (1997). https://doi.org/10.1103/PhysRevLett.78.3047. URL https://link.aps.org/doi/10.

1103/PhysRevLett.78.3047

[12] Z.K. Hu, J.Q. Guo, J. Luo, Correction of source mass effects in the HUST-99 measurement of *G*. Phys. Rev. D **71**, 127,505 (2005). https://doi.org/10.1103/PhysRevD.71.127505. URL https://link.aps.org/doi/10.1103/PhysRevD.71.127505

[13] J. Luo, Q. Liu, L.C. Tu, C.G. Shao, L.X. Liu, S.Q. Yang, Q. Li, Y.T. Zhang, Determination of the Newtonian Gravitational Constant *G* with Time-of-Swing Method. Phys. Rev. Lett. **102**, 240,801 (2009). https://doi.org/10.1103/PhysRevLett.102.240801. URL https://link.aps.org/doi/10.1103/PhysRevLett.102.240801

[14] L.C. Tu, Q. Li, Q.L. Wang, C.G. Shao, S.Q. Yang, L.X. Liu, Q. Liu, J. Luo, New determination of the gravitational constant *G* with time-of-swing method. Phys. Rev. D **82**, 022,001 (2010). https://doi.org/10.1103/PhysRevD.82.022001. URL https://link.aps.org/doi/10.1103/PhysRevD.82.022001

[15] R. Newman, M. Bantel, E.C. Berg, W.D. Cross, A measurement of *G* with a cryogenic torsion pendulum. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **372**(2026), 20140,025 (2014). https://doi.org/10.1098/rsta.2014.0025. URL https://doi.org/10.1098/rsta.2014.0032

[16] Q. Li, C. Xue, J.P. Liu, J.F. Wu, S.Q. Yang, C.G. Shao, L.D. Quan, W.H. Tan, L.C. Tu, Q. Liu, H. Xu, L.X. Liu, Q.L. Wang, Z.K. Hu, Z.B. Zhou, P.S. Luo, S.C. Wu, V. Milyukov, J. Luo, Measurements of the gravitational constant using two independent methods. Nature **560**(7720), 582–588 (2018). https://doi.org/10.1038/s41586-018-0431-5. URL https://doi.org/10.1038/s41586-018-0431-5

[17] T.R. Armstrong, M.P. Fitzgerald, New Measurements of *G* Using the Measurement Standards Laboratory Torsion Balance. Phys. Rev. Lett. **91**, 201,101 (2003). https://doi.org/10.1103/PhysRevLett.91.201101. URL https://link.aps.org/doi/10.1103/PhysRevLett.91.201101

[18] J.H. Gundlach, S.M. Merkowitz, Measurement of Newton's Constant Using a Torsion Balance with Angular Acceleration Feedback. Phys. Rev. Lett.**85**(14), 2869–2872 (2000). https://doi.org/10.1103/PhysRevLett.85.2869. https://arxiv.org/abs/gr-qc/0006043 [gr-qc]

[19] U. Kleinevoß, Bestimmung der Newtonschen Gravitationskonstanten *G*. Ph.D. thesis, University of Wuppertal, Wuppertal, Germany (2002)

[20] H.V. Parks, J.E. Faller, Simple Pendulum Determination of the Gravitational Constant. Phys. Rev. Lett. **105**, 110,801 (2010). https://doi.org/10.1103/PhysRevLett.105.110801. URL https://link.aps.org/doi/10.1103/PhysRevLett.105.110801

[21] H.V. Parks, J.E. Faller, Erratum: Simple Pendulum Determination of the Gravitational Constant [Phys. Rev. Lett. 105, 110801 (2010)]. Phys. Rev. Lett. **122**, 199,901 (2019). https://doi.org/10.1103/PhysRevLett.122.199901. URL https://link.aps.org/doi/10.1103/PhysRevLett.122.199901

[22] S. Schlamminger, E. Holzschuh, W. Kündig, F. Nolting, R.E. Pixley, J. Schurr, U. Straumann, Measurement of Newton's gravitational constant. Phys. Rev. D**74**(8), 082001 (2006). https://doi.org/10.1103/PhysRevD.74.082001. https://arxiv.org/abs/gr-qc/0609027 [gr-qc]

[23] M. Prevedelli, L. Cacciapuoti, G. Rosi, F. Sorrentino, G. Tino, Measuring the Newtonian constant of gravitation *G* with an atomic interferometer. Philosophical transactions. Series A, Mathematical, physical, and engineering sciences **372** (2014). https://doi.org/10.1098/rsta.2014.0030

[24] G. Rosi, F. Sorrentino, L. Cacciapuoti, M. Prevedelli, G.M. Tino, Precision measurement of the Newtonian gravitational constant using cold atoms. Nature**510**(7506), 518–521 (2014). https://doi.org/10.1038/nature13433. https://arxiv.org/abs/1412.7954 [physics.atom-ph]

[25] M. Pitkin, Comment on "Measurements of Newton's gravitational constant and the length of day" by Anderson J. D. et al. EPL (Europhysics Letters) **111**(3), 30,002 (2015). https://doi.org/10.1209/0295-5075/111/30002. URL https://doi.org/10.1209/0295-5075/111/30002

[26] J.D. Anderson, G. Schubert, V. Trimble, M.R. Feldman, Measurements of Newton's gravitational constant and the length of day. EPL (Europhysics Letters) **110**(1), 10002 (2015). https://doi.org/10.1209/0295-5075/110/10002. https://arxiv.org/abs/1504.06604 [gr-qc]

[27] C. Merkatas, B. Toman, A. Possolo, S. Schlamminger, Shades of dark uncertainty and consensus value for the Newtonian constant of gravitation. Metrologia **56**(5), 054,001 (2019). https://doi.org/10.1088/1681-7575/ab3365. URL https://dx.doi.org/10.1088/1681-7575/ab3365

[28] E.T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003). https://doi.org/10.1017/CBO9780511790423

[29] A. Gelman, Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). Bayesian Analysis **1**(3), 515 – 534 (2006). https://doi.org/10.1214/06-BA117A. URL https://doi.org/10.1214/06-BA117A

[30] S. Rinaldi, W. Del Pozzo, (H)DPGMM: a hierarchy of Dirichlet process Gaussian mixture models for the inference of the black hole mass function. MNRAS **509**(4), 5454–5466 (2021). https://doi.org/10.1093/mnras/stab3224. URL https://doi.org/10.1093/mnras/stab3224. https://arxiv.org/abs/2109.05960 [astro-ph]

[31] M.D. Escobar, M. West, Bayesian Density Estimation and Inference Using Mixtures. Journal of the American Statistical Association **90**(430), 577–588 (1995). URL http://www.jstor.org/stable/2291069

[32] T.S. Ferguson, A Bayesian Analysis of Some Nonparametric Problems. The Annals of Statistics **1**(2), 209 – 230 (1973). https://doi.org/10.1214/aos/1176342360. URL https://doi.org/10.1214/aos/1176342360

[33] J. Skilling, Nested sampling for general Bayesian computation. Bayesian Analysis **1**(4), 833 – 859 (2006). https://doi.org/10.1214/06-BA127. URL https://doi.org/10.1214/06-BA127

[34] J. Veitch, W. Del Pozzo, A. Lyttle, M.J. Williams, C. Talbot, M. Pitkin, G. Ashton, M. Hübner, A. Nitz, D. Mihaylov, D. Macleod, G. Carullo, G. Davies. CPNest (2022). URL https://github.com/johnveitch/cpnest

[35] S. Rinaldi, W. Del Pozzo, Rapid localization of gravitational wave hosts with FIGARO. MNRAS: Letters **517**(1), L5–L10 (2022). https://doi.org/10.1093/mnrasl/slac101. URL https://doi.org/10.1093/mnrasl/slac101. https://arxiv.org/abs/2205.07252 [astro-ph]

[36] C. Rothleitner, S. Schlamminger, Invited review article: Measurements of the newtonian constant of gravitation, g. Review of Scientific Instruments **88**(11), 111,101 (2017). https://doi.org/10.1063/1.4994619. URL https://doi.org/10.1063/1.4994619. https://arxiv.org/abs/https://doi.org/10.1063/1.4994619

[37] S. Schlamminger, L. Chao, V. Lee, D.B. Newell, C.C. Speake, The crane operator's trick and other shenanigans with a pendulum. American Journal of Physics **90**(3), 169–176 (2022). https://doi.org/10.1119/10.0006965. URL https://doi.org/10.1119/10.0006965. https://arxiv.org/abs/https://doi.org/10.1119/10.0006965