

Population-level correlations in Bayesian statistics: An illustrative model for gravitational-wave astronomy

Caroline B. Owen¹,* Alexandre Toubiana, and Davide Gerosa²

Dipartimento di Fisica “G. Occhialini”, Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy and INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy



(Received 18 June 2026; accepted 14 September 2026; published 22 October 2026)

With increasingly large numbers of gravitational-wave events, population inference is now beginning to move beyond predictions of marginal distributions and to probe correlations between compact-binary parameters such as masses, spins, and redshifts. These correlations have strong constraining power for both astrophysics and tests of general relativity. In this paper, we present an idealized analytical model to study the interplay between single-event correlations, systematic biases, and population-level correlations. With this, we investigate the potential emergence of false-positive measurements of population-level correlations. We quantify how the presence of correlations at the single-event level between a pair of parameters increases the uncertainty of population-level correlations for those parameters, potentially obscuring the true underlying population correlation if present. We also find that if waveform systematics lead to biases that are correlated across the catalog (which is likely, because certain regions of the parameter space are more difficult to model), this can be effectively absorbed by a population analysis that targets correlations and can be misinterpreted as such. This simple Gaussian-based model may serve as a broad compass for future, more detailed explorations.

DOI: [10.1103/nhtw-rdx](https://doi.org/10.1103/nhtw-rdx)

I. INTRODUCTION

Population analyses of gravitational-wave (GW) sources, which combine multiple events within the hierarchical Bayesian framework [1,2], are becoming increasingly common tools to unveil the formation mechanisms of compact binaries, improve our ability to test general relativity, and measure cosmological parameters (e.g. [3–5]). As the size of GW catalog increases, we become able not only to infer the population-level marginal distributions of binary parameters, but also to characterize the correlations between them [6–17]. This, in turn, enhances our ability to distinguish between different formation scenarios [18,19] and to identify the origin of possible deviations from general relativity [20,21]. Given the central role that measurements of population-level correlations are likely to play in shaping our understanding of the Universe, it is essential to understand the factors that may affect them.

A negative correlation between the mean of the effective-spin distribution and the mass ratio was initially identified in the GWTC-2.1 catalog [22] and further confirmed with

subsequent data [23]. Interestingly, as more events have been added to the catalog, the evidence for this correlation weakened, while evidence for a correlation between the width of the effective-spin distribution and mass ratio has increased [3,24].

The initially inferred trend follows the same direction as the well-known degeneracy between effective spin and mass ratio at the single-event parameter-estimation level [25]. Reference [22] performed additional analyses to strengthen the interpretation that this single-event degeneracy was not responsible for the inferred population-level correlation. Indeed, hierarchical analyses are unbiased in the ideal limit: if the model used to analyze the data is correct, they should, on average, recover the correct population properties. However, whether correlations at the individual-event level can facilitate false measurements of population-level correlations, especially due to statistical fluctuations in the finite-number-of-events regime, remains an important question.

In addition to such statistical effects, systematic effects may also contribute to false measurements of population-level correlations. At the single-event level, the finite accuracy of the waveform models used in parameter estimation of individual events can lead to systematic biases in the inferred source properties [26–31]. Since these biases are source-dependent, they could in turn mimic spurious correlations at the population level. Additionally, population analyses require choosing a model for the

*Contact author: cbo2@williams.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

distribution of events. If this model is misspecified, the reconstruction of the population properties can be compromised [32–35], including the inferred correlations between binary parameters.

In this paper, we systematically investigate false-positive measurements of population-level correlations when no such correlations are present in the true population. We consider three possible causes: (i) statistical fluctuations, including the role of correlations between parameters in individual observations; (ii) systematic biases in single-event measurements; and (iii) systematic biases arising from an incorrect choice of population prior. We construct a simple toy model to explore how these factors affect the measurement of population-level correlations between pairs of binary parameters. While considerably simpler than the approaches commonly used to infer the astrophysical properties of compact-binary populations, our framework closely resembles the approach developed in Ref. [20] to constrain multidimensional deviations from general relativity. As a result, our findings are particularly relevant to and directly translatable to that science case, while also providing insight into qualitative behaviors in other inference setups.

The remainder of this paper is organized as follows. In Sec. II, we set up our toy model and demonstrate how it models statistical and systematic errors at the single-event and population levels. In Sec. III, we explore how statistical fluctuations in finite catalogs can impact the probability of obtaining a credible false-positive measurement of correlation. Finally, in Sec. IV we summarize the results and discuss their implications.

II. MODEL SETUP

We construct a simple toy model to examine correlations between two binary parameters $\theta = (\theta_1, \theta_2)$. We assume that we have observed a catalog $\{d_k\}$ of N signals with $k \in [1, \dots, N]$. Each signal $d_k = n_k + h^*(\theta_k^*)$ is a sum of detector noise n_k and the true¹ waveform h^* produced by true source parameters θ_k^* . The source parameters of each signal are drawn from some underlying population model $\pi^*(\theta|\Lambda^*)$ with population parameters Λ^* .

We assume that both the single-event and population likelihoods are well described by multivariate Gaussian distributions. Additionally, we employ a multivariate Gaussian model to recover the population properties. For completeness, we recall the expression of the multivariate Gaussian distribution. Consider a random vector $\mathbf{x}$ of dimension ℓ distributed according to a multivariate Gaussian distribution with mean vector μ and covariance matrix Σ . The probability distribution function (PDF) of $\mathbf{x}$ is

$$\mathcal{N}(\mathbf{x}|\mu, \Sigma) = \frac{\exp[-\frac{1}{2}(\mathbf{x} - \mu)^T \Sigma^{-1}(\mathbf{x} - \mu)]}{\sqrt{(2\pi)^\ell |\Sigma|}}. \quad (1)$$

The covariance matrix is a symmetric matrix where the diagonal components $\Sigma_{ii} = \sigma_i^2$ are the variances of the marginal distributions and the off-diagonal components $\Sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$ are proportional to Pearson's correlation coefficients $\rho_{ij} \in (-1, 1)$. These coefficients quantify linear correlations such that when ρ_{ij} is positive (negative), x_i and x_j are (anti)correlated, and when $\rho_{ij} = 0$, they are uncorrelated. The components of the inverse covariance matrix can be obtained by inverting the Fisher information matrix [36]

$$(\Sigma^{-1})_{ij} = -\frac{\partial^2}{\partial x^i \partial x^j} \log \mathcal{N}(\mathbf{x}|\mu, \Sigma)|_{\mathbf{x}=\mu}. \quad (2)$$

Typically, in GW astronomy, we quote 90% credible intervals on the marginal distributions of a given posterior. For a multivariate Gaussian, 90% of the probability density of the marginal distribution on x_i is contained within an interval of $2\sqrt{2}\sigma_i \text{Erf}^{-1}(0.9)$ centered about μ_i .

A. Single-event parameter estimation

Assuming the detector noise to be zero-mean and Gaussian distributed, the single-event likelihood is

$$\mathcal{L}(d|\theta) \propto \exp\left[-\frac{1}{2}(d - h(\theta)|d - h(\theta))\right], \quad (3)$$

where $(\cdot|\cdot)$ denotes the inner product weighted by the one-sided noise power spectral density and $h(\theta)$ is the chosen waveform approximant.

In the large signal-to-noise ratio (SNR) limit, this likelihood is well described by a multivariate Gaussian distribution on the single-event parameters

$$\mathcal{L}(d|\theta) \propto \mathcal{N}(\theta|\theta_{\text{ml}}, \Sigma_{\text{pe}}), \quad (4)$$

centered around the maximum likelihood values θ_{ml} and with a covariance matrix Σ_{pe} .

In this toy model, we prescribe Σ_{pe} via standard deviations $\sigma_{\text{pe},1}$ and $\sigma_{\text{pe},2}$ and correlation coefficient ρ_{pe} . The $\sigma_{\text{pe},i}$ are a surrogate for SNR, with which they scale inversely in the large-SNR limit for a given location in parameter space and model $h(\theta)$. The coefficient ρ_{pe} captures the correlation between θ_1 and θ_2 in the single-event likelihood that arises due to the functional form of $h(\theta)$. In this simplified approach, the maximum-likelihood point is a sufficient statistic for describing the data, so in the following we will refer to the data d_k through $\theta_{\text{ml},k}$.

The maximum likelihood point for a given event can be shifted from the true value for two reasons: (i) the presence of noise in the detector, which causes a statistical shift, and (ii) waveform mismodeling, which causes a systematic bias. Defining

¹Throughout this manuscript, we will use the symbol $\star$ to indicate true values of the single-event parameters θ^* , population parameters Λ^* , waveform h^* , and population distribution π^* .

$$\Delta\theta = \theta_{\text{ml}} - \theta^*, \quad (5)$$

we can write

$$\Delta\theta = \Delta\theta_{\text{stat}} + \Delta\theta_{\text{sys}}. \quad (6)$$

When the shift is small, and under the Gaussian likelihood approximation, these contributions can be approximated as [29,37]

$$\Delta\theta_{\text{stat}} = \Sigma_{\text{pe}} \cdot \left(\frac{\partial h}{\partial \theta} \right) \bigg|_{\theta=\theta^*}, \quad (7)$$

$$\Delta\theta_{\text{sys}} = \Sigma_{\text{pe}} \cdot \left(\frac{\partial h}{\partial \theta} \right) \bigg|_{\theta=\theta^*} (h - h^*). \quad (8)$$

With the likelihood model of Eq. (3), $\Delta\theta_{\text{stat}}$ is distributed according to a normal distribution with zero mean and covariance matrix Σ_{pe} . The difference between the waveform approximant and the true waveform $h - h^*$, and consequently $\Delta\theta_{\text{sys}}$, will have some particular dependence on θ^* and potentially additional binary parameters. In the following Section, we consider several illustrative examples.

Thus, our mock catalog $\{\theta_{\text{ml},k}\}$ is constructed as follows:

- (1) Draw N samples θ^* from the chosen population distribution $\pi^*(\theta|\Lambda^*)$.
- (2) For each θ^* , we draw $\Delta\theta_{\text{stat}}$ from the Gaussian distribution $\mathcal{N}(\mathbf{0}, \Sigma_{\text{pe}})$. If systematic effects are included, we additionally compute $\Delta\theta_{\text{sys}}$ according to the chosen prescription. Finally, $\theta_{\text{ml}} = \theta^* + \Delta\theta$.

B. Hierarchical inference

Once the mock catalog of observations is generated, the next step is hierarchical inference to recover the population parameters Λ . Regardless of the underlying distribution the events are drawn from, we model the population with the multivariate Gaussian distribution

$$\pi(\theta|\Lambda) = \mathcal{N}(\theta|\mu_{\text{pop}}, \Sigma_{\text{pop}}), \quad (9)$$

with population parameters

$$\Lambda = (\mu_{\text{pop},1}, \mu_{\text{pop},2}, \sigma_{\text{pop},1}, \sigma_{\text{pop},2}, \rho_{\text{pop}}). \quad (10)$$

The coefficient ρ_{pop} is the population parameter that quantifies the population-level correlation between θ_1 and θ_2 . It is the measurement of this parameter that interests us here.

We take the population likelihood of observing the catalog $\{\theta_{\text{ml},k}\}$ given Λ to be the product of the single-event likelihoods, marginalized over the population model

$$\mathcal{L}(\{\theta_{\text{ml},k}\}|\Lambda) = \prod_k \int d\theta_k \mathcal{L}(\theta_{\text{ml},k}|\theta_k) \pi(\theta_k|\Lambda). \quad (11)$$

This is a simplified likelihood that has been marginalized over the rate of binary mergers and that disregards selection effects. Note that selection effects are also not usually included in hierarchical tests of general relativity [5,38,39], although frameworks for doing so do exist [40].

In our setup, where both the population model $\pi(\theta|\Lambda)$ and parameter estimation likelihoods $\mathcal{L}(\theta_{\text{ml}}|\theta)$ are distributed according to multivariate Gaussian distributions, the above integrals can be done analytically

$$\mathcal{L}(\{\theta_{\text{ml},k}\}|\Lambda) = \prod_{k=1}^N \mathcal{N}(\theta_{\text{ml},k}|\mu_{\text{pop}}, \Sigma_{\text{pop}} + \Sigma_{\text{pe}}). \quad (12)$$

The population likelihood will be maximized by Λ_{ml} which is specified by $\mu_{\text{ml}} = \hat{\mu}$ and $\Sigma_{\text{ml}} = \hat{\Sigma} - \Sigma_{\text{pe}}$. The hatted quantities are the empirical mean and covariance matrix of the set of N observations $\{\theta_{\text{ml},k}\}$. In this model, both $\hat{\mu}$ and $\hat{\Sigma}$ can be computed algebraically

$$\begin{aligned} \hat{\mu} &= \frac{1}{N} \sum_{k=1}^N \theta_{\text{ml},k}^k, \\ \hat{\Sigma} &= \frac{1}{N} \sum_{k=1}^N (\theta_{\text{ml},k}^k - \hat{\mu})^T (\theta_{\text{ml},k}^k - \hat{\mu}), \end{aligned} \quad (13)$$

allowing us to determine $\rho_{\text{pop,ml}}$ directly. Note that the population likelihood is maximized by the biased estimator of the covariance matrix, rather than by the unbiased estimator, which uses a factor of $1/(N-1)$ instead of $1/N$. Nevertheless, this estimator is unbiased in the limit $N \rightarrow +\infty$.

In the absence of mismodeling in both waveform approximant and population model, the maximum likelihood values $\{\theta_{\text{ml},k}\}$ are generated by drawing the source parameters $\{\theta_k^*\}$ from the Gaussian distribution $\mathcal{N}(\theta|\mu_{\text{pop}}^*, \Sigma_{\text{pop}}^*)$ and shifting them by $\{\Delta\theta_{\text{stat},k}\}$ drawn from $\mathcal{N}(\theta|\mathbf{0}, \Sigma_{\text{pe}})$. In this case, $N\hat{\Sigma}$ is distributed according to a Wishart distribution with $N-1$ degrees of freedom and with scale matrix $\Sigma_{\text{pop}}^* + \Sigma_{\text{pe}}$. In general, the expectation value of a function of random variables is not equal to that function evaluated at the expectation values of those random variables. While, it is straightforward to show that $\langle \Sigma_{\text{ml}} \rangle = \Sigma_{\text{pop}}^*$, $\rho_{\text{pop,ml}}$ is a nontrivial function of the components of $\hat{\Sigma}$ and Σ_{pe} , and it is not necessarily the case that $\langle \rho_{\text{pop,ml}} \rangle = \rho_{\text{pop}}^*$. While we are not aware of an analytic expression for this expectation value, it will, in general, depend on N and the components of Σ_{pe} , including ρ_{pe} . However, as $N \rightarrow \infty$, the Wishart distribution becomes more strongly peaked at its expectation value and thus the peak of the distribution on $\rho_{\text{pop,ml}}$ is pushed toward ρ_{pop}^* .

If systematic error is present at either the single-event or population level, $N\hat{\Sigma}$ will no longer be governed by the Wishart distribution. The distribution that $N\hat{\Sigma}$, and

consequently $\rho_{\text{pop,ml}}$ takes depends on the specific form of the mismodeling at hand. For example, if $\{\Delta\theta_{\text{sys},k}\}$ depends on the source parameters such a way as to induce additional correlation in $\{\theta_{\text{ml},k}\}$, $\rho_{\text{pop,ml}}$ may be biased from ρ_{pop}^* .

In general, mismodeling at the population inference level can also lead to bias in the recovered population parameters. However, it is here that the simplicity of our toy model becomes a shortcoming. The empirical covariance matrix $\hat{\Sigma}$ as given in Eq. (13) will capture the correlation in $\{\theta_{\text{ml},k}\}$ regardless of the distribution from which $\{\theta_k^*\}$ is drawn. If the true population distribution of the single-event parameters is uncorrelated, then we expect $\rho_{\text{pop,ml}}$ to tend to 0 in the large N limit even if the Gaussian population model is not overall a good representation of the true distribution.

C. Quantifying false measurements

We restrict our consideration to underlying population models $\pi^*(\theta|\Lambda^*)$ in which θ_1 and θ_2 are uncorrelated at the population level, i.e. $\rho_{\text{pop}}^* = 0$, because we seek to understand the circumstances in which population-level correlation may be measured despite such correlation not being present in the true distribution.

As with the single-event parameters, $\rho_{\text{pop,ml}}$ can be shifted from ρ_{pop}^* due to both statistical and systematic effects. However, whether a given catalog yields a credible measurement of population-level correlation will depend not only on $\rho_{\text{pop,ml}}$ but also on the uncertainty of the measurement. To estimate the credible intervals, we make the additional assumption that the population likelihood is well described as a five-dimensional multivariate Gaussian distribution

$$\mathcal{L}(\{\theta_{\text{ml},k}\}|\Lambda) \propto \mathcal{N}(\Lambda|\Lambda_{\text{ml}}, \Sigma_{\text{hi}}) \quad (14)$$

centered about Λ_{ml} . Following Eq. (2), the inverse of the covariance matrix of the likelihood is

$$(\Sigma_{\text{hi}}^{-1})_{ij} = -\frac{\partial^2}{\partial \Lambda^i \partial \Lambda^j} \log \mathcal{L}(\{\theta_{\text{ml},k}\}|\Lambda) \Big|_{\Lambda=\Lambda_{\text{ml}}} \quad (15)$$

where $\mathcal{L}(\{\theta_{\text{ml},k}\}|\Lambda)$ is given by Eq. (12) [41,42].

We take the prior distribution on the population parameters to be uniform² and sufficiently wide such that the standard deviation $\sigma_{\rho_{\text{pop}}}$ of the marginalized posterior on ρ_{pop}

²The uniform prior on ρ_{pop} is an instance of a Lewandowski-Kuworicka-Joe (LKJ) distribution with shape parameter 1. The LKJ distribution is a probability distribution on correlation matrices and can be tuned to disfavor (favor) correlations by selecting a shape factor greater (less) than 1 [43]. While not investigated in this study, it is reasonable to suspect that increasing (decreasing) the shape factor would decrease (increase) the probability of a false-positive identification of correlation.

can then be extracted from Σ_{hi} . For several realizations of the data $\{\theta_{\text{ml},k}\}$, we employed an Markov Chain Monte Carlo algorithm to sample the true posterior and have verified that $\sigma_{\rho_{\text{pop}}}$ as computed above is a good approximation of that of the true posterior for the cases considered in this study, generally with a relative error of $\sim 1\%$. As with $\rho_{\text{pop,ml}}$, the value of $\sigma_{\rho_{\text{pop}}}$ for a given catalog will be impacted by statistical fluctuations and systematic effects.

We consider a measurement to be consistent with no correlation if $\rho_{\text{pop}} = 0$ falls within the 90% credible interval of the marginalized posterior. When this occurs, the ratio

$$r_\rho = \frac{|\rho_{\text{pop,ml}}|}{\sqrt{2}\sigma_{\rho_{\text{pop}}} \text{Erf}^{-1}(0.9)} \quad (16)$$

is less than 1. To understand the impact of statistical fluctuations on the measurement of population-level correlation, we set out to estimate probability distributions $p(\rho_{\text{pop,ml}})$, $p(\sigma_{\rho_{\text{pop}}})$, and $p(r_\rho)$ under various conditions.

The toy model we have laid out allows us to account for statistical uncertainty and both statistical and systematic shifts in measuring the single-event and population parameters. We investigate the effects of statistical uncertainty at the level of single-event parameter estimation by varying the standard deviations $\sigma_{\text{pe},1}$ and $\sigma_{\text{pe},2}$. Varying $\sigma_{\text{pe},1}$ and $\sigma_{\text{pe},2}$ will also impact the statistical shifts, the magnitude of which will tend to be larger when $\sigma_{\text{pe},1}$ and $\sigma_{\text{pe},2}$ are larger. The correlation coefficient ρ_{pe} allows us to model the effects of correlation in the measurement of binary parameters for individual events. We can explore the effects of various types of systematic error at the parameter estimation level by modeling $\Delta\theta_{\text{sys}}$ in different ways.

Varying the catalog size N allows us to explore the effects of statistical uncertainty and shifts at the population level. As the catalog size grows, the statistical uncertainty in measurements of population parameters decreases. Similarly, the effects of catalog variance, in which the properties of the finite catalog differ from those of the underlying distribution, will decrease, thereby reducing statistical shifts. We can explore the effects of population mismodeling by choosing an underlying population model $\pi^*(\theta|\Lambda^*)$ that does not match the recovery model.

Our procedure is as follows: We specify the underlying population distribution, $\pi^*(\theta|\Lambda^*)$, the catalog size N , and the single-event likelihood $\sigma_{\text{pe},1}$, $\sigma_{\text{pe},2}$, ρ_{pe} , $\Delta\theta_{\text{sys}}$. A single catalog consists of N sets of source parameters $\{\theta_k^*\}$ drawn from $\pi^*(\theta|\Lambda^*)$. We determine the maximum likelihood values of the single-event parameters $\{\theta_{\text{ml},k}\}$ based on the chosen single-event likelihood. Finally we compute $\rho_{\text{pop,ml}}$, $\sigma_{\rho_{\text{pop}}}$, and r_ρ for the catalog. For a given configuration we perform this calculation 2×10^5 times, allowing us to approximate the probability distributions $p(\rho_{\text{pop,ml}})$, $p(\sigma_{\rho_{\text{pop}}})$, and $p(r_\rho)$. By computing the fraction of catalogs for which $r_\rho > 1$, we can determine the probability that

hierarchical inference will imply a population-level correlation between θ_1 and θ_2 despite no correlation being present in the true underlying population.

III. RESULTS

A. No mismodeling

We begin by considering the case where there is no mismodeling, either at the single-event parameter estimation level, i.e., $\Delta\theta_{\text{sys}} = 0$, or at the population hierarchical inference level, i.e., $\pi^*(\theta|\Lambda^*) = \mathcal{N}(\theta|\mu_{\text{pop}}^*, \Sigma_{\text{pop}}^*)$. Specifically, we consider an uncorrelated population distribution with $\mu_{\text{pop},1}^* = \mu_{\text{pop},2}^* = 0$, $\sigma_{\text{pop},1}^* = \sigma_{\text{pop},2}^* = 1$ and $\rho_{\text{pop}}^* = 0$. With no mismodeling present, we investigate how statistical effects alone impact the measurement of ρ_{pop} .

We consider a grid of 15 single-event likelihoods with $\sigma_{\text{pe}} = \sigma_{\text{pe},1} = \sigma_{\text{pe},2} \in \{0.1, 0.5, 0.9\}$ and $\rho_{\text{pe}} \in \{-0.9, -0.5, 0.0, 0.5, 0.9\}$. For each configuration, we consider catalogs with $N \in \{50, 100\}$ events each, comparable to the current GW catalogs, and approximate the distributions $p(\rho_{\text{pop},\text{ml}})$ and $p(\sigma_{\rho_{\text{pop}}})$ in the manner described above.

The medians of these distributions are shown as functions of ρ_{pe} in the top four panels of Fig. 1. The 5% and 95% quantiles of the distributions enclose the shaded regions. The bottom panel plots the percentage of catalogs for which $r_\rho > 1$.

We first consider the cases where $\rho_{\text{pe}} = 0$. We see that even in the absence of single-event correlation, there is a nonzero probability of measuring $\rho_{\text{pop}} \neq 0$. However, the chance of a false-positive correlation measurement tends to decrease as both the precision of the single-event parameter estimation and the catalog size increase.

Now, turning to the full set of configurations, we find that the median of $p(\rho_{\text{pop},\text{ml}})$ is approximately zero for all configurations, as predicted for large catalogs in our above. Although the absolute value of the median does increase slightly with both $|\rho_{\text{pe}}|$ and σ_{pe} , and decrease with N . On the other hand, the spread of $p(\rho_{\text{pop},\text{ml}})$ varies much more apparently among configurations, in particular increasing with σ_{pe} and decreasing with N . Interestingly, while the shape of the distribution is much less impacted by ρ_{pe} , there is a slight negative correlation between ρ_{pe} and $\rho_{\text{pop},\text{ml}}$. When ρ_{pe} is positive (negative), there is a greater probability for values of $\rho_{\text{pop},\text{ml}}$ that are more negative (positive). This negative correlation results from our choice to quantify the correlation using the dimensionless coefficient ρ_{pop} rather than $\Sigma_{12,\text{pop}}$. This is a notable result because one might expect a false correlation identified in the population distribution to align with the correlation in the parameter estimation space. We see here that need not be the case.

The variance $\sigma_{\rho_{\text{pop}}}$ increases with both σ_{pe} and $|\rho_{\text{pe}}|$. This leads to a counterintuitive outcome: while wider or more strongly correlated single-event posteriors can lead to a larger shift in $\rho_{\text{pop},\text{ml}}$, these effects also tend to produce

wider credible intervals on the measurement, making it less likely that the $\rho_{\text{pop}} = 0$ case will be falsely ruled out, especially when σ_{pe} is large or N is small. This is demonstrated in the bottom panel of Fig. 1. We therefore infer that strong correlations at the single-event level could, in fact, obscure an existing population-level correlation simply by increasing measurement uncertainty.

On the other hand, increasing N causes the spread of $\rho_{\text{pop},\text{ml}}$ and the median and spread of $\sigma_{\rho_{\text{pop}}}$ to shrink. In other words, a larger catalog size means a measurement of ρ_{pop} that is both more accurate and more precise and, consequently, a decreased probability of measuring $\rho_{\text{pop}} \neq 0$. The one exception is when $|\rho_{\text{pe}}|$ and σ_{pe} are largest, which leads to such great measurement uncertainty in the $N = 50$ case that increasing the catalog size to $N = 100$ does not lower the probability of a false-positive measurement.

B. Mismodeling at the single-event level

We move on to the case where the waveform used in single-event parameter estimation does not match the true waveform of nature. However, we again consider a population distributed according to uncorrelated Gaussians with $\mu_{\text{pop},1}^* = \mu_{\text{pop},2}^* = 0$, $\sigma_{\text{pop},1}^* = \sigma_{\text{pop},2}^* = 1$, and $\rho_{\text{pop}}^* = 0$. To emphasize the effects of mismodeling, we consider the case where $\sigma_{\text{pe},1} = \sigma_{\text{pe},2} = 0.1$ and $\rho_{\text{pe}} = 0.9$, because we saw in the previous section that, in the absence of waveform mismodeling, this case differed very little from that in which $\rho_{\text{pe}} = 0$.

We consider several illustrative examples for $\Delta\theta_{\text{sys}}$. We begin by considering the case where the bias depends only on θ^* and no additional parameters. We select the specific realization

$$\Delta\theta_{\text{sys},i} = \varphi\theta_1^*\theta_2^* \quad (17)$$

for $i = 1, 2$. The constant φ parametrizes the size of the derivation and, for the time being, we take its value to be common among all events in a given catalog. We choose this realization for the bias because, while θ_1^* and θ_2^* are drawn from uncorrelated distributions, $\Delta\theta_{\text{sys},1}$ and $\Delta\theta_{\text{sys},2}$ will clearly be correlated across the catalog, inducing correlation in $\{\theta_{\text{ml}}^k\}$.

The statistical uncertainty in single-event parameter estimation scales inversely with the SNR, while the systematic bias remains constant. Therefore, as detectors improve and events are observed with greater sensitivity, the relative significance of systematic bias will grow. Significant systematic error is already present in some regions of parameter space [44–47].

We consider the systematic bias significant when the true source value of the parameter falls outside the 90% credible interval of the marginalized posterior. In other words

$$|\Delta\theta_{\text{sys},i}| < \sqrt{2}\text{Erf}^{-1}(0.9)\sigma_{\text{pe},i}. \quad (18)$$

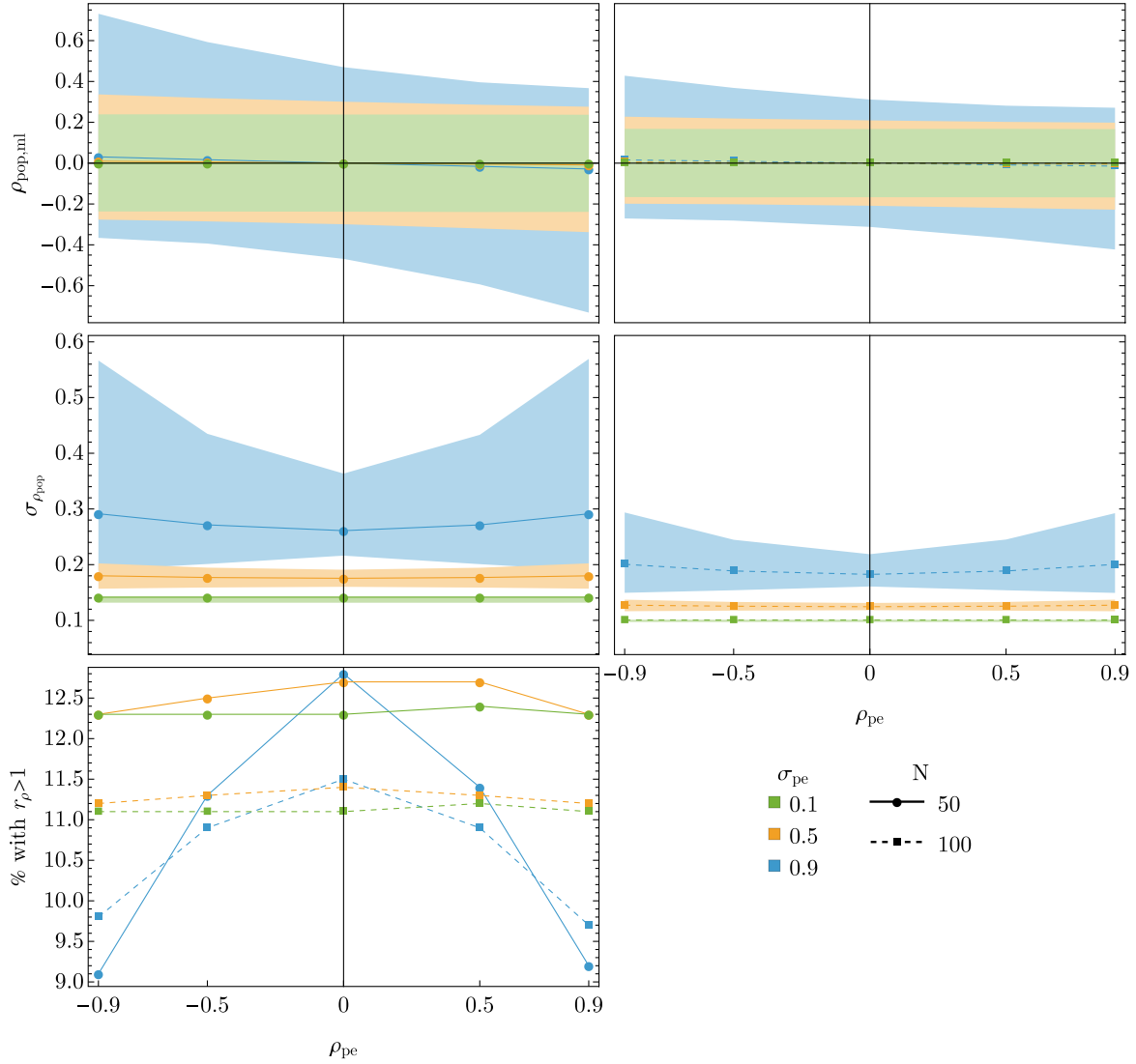


FIG. 1. Top four panels show the median of the distributions $p(\rho_{\text{pop,ml}})$ and $p(\sigma_{\rho_{\text{pop}}})$ as a function of ρ_{pe} for various values of σ_{pe} (colors) and N (left vs right; linestyles) in the absence of mismodeling at the single event or population levels. The 5% and 95% quantiles of the probability distributions enclose the shaded regions. The bottom panel shows the percentage of catalogs that produce a credible measurement of $\rho_{\text{pop}} \neq 0$, corresponding to $r_{\rho} > 1$. The true population correlation is $\rho_{\text{pop}}^* = 0$. We note two counterintuitive results: (1) There is a slight negative correlation between ρ_{pe} and $\rho_{\text{pop,ml}}$ such that when ρ_{pe} is positive (negative), there is a greater probability for values of $\rho_{\text{pop,ml}}$ that are more negative (positive), indicating that spurious population-level correlations are unlikely to be aligned with correlations in the single-event posteriors. (2) Because large values of ρ_{pe} tend to increase the uncertainty in the measurement of ρ_{pop} as quantified by $\sigma_{\rho_{\text{pop}}}$, larger single-event correlations can decrease the probability of a false positive measurement of $\rho_{\text{pop}} \neq 0$.

The single-event parameters θ_1 and θ_2 are unbounded, and the absolute value of the bias grows with distance from the $\theta_1 = \theta_2 = 0$ point. Therefore, for any choice of φ , there is a region at the edges of the parameter space where the inequality above is violated. Thus, our chosen bias in Eq. (17) is indicative of the current GW catalog.

We consider $\varphi = 5 \times 10^{\beta}$, for $\beta = \{-\infty, -1.75, -1.5, -1.25, -1.0, -0.75\}$. As a measure of the impact of the bias on the catalog, for each value of φ we compute the percent of the underlying population distribution for which the bias is significant according to the inequality in

Eq. (18). We find that the above values of φ correspond to systematic error that is significant for approximately $\{0, 10, 20, 40, 50, 70\}\%$ of the catalog, respectively. While the largest of these percentages are greater than we suspect for our current GW catalog, by considering large biases, we are able to qualitatively explore the relationship between systematic error and catalog size in producing false-positive measurements of correlation.

We carry out the procedure outlined in the previous section to approximate the distributions $p(\rho_{\text{pop,ml}})$, $p(\sigma_{\rho_{\text{pop}}})$, and to determine the percent of catalogs for which

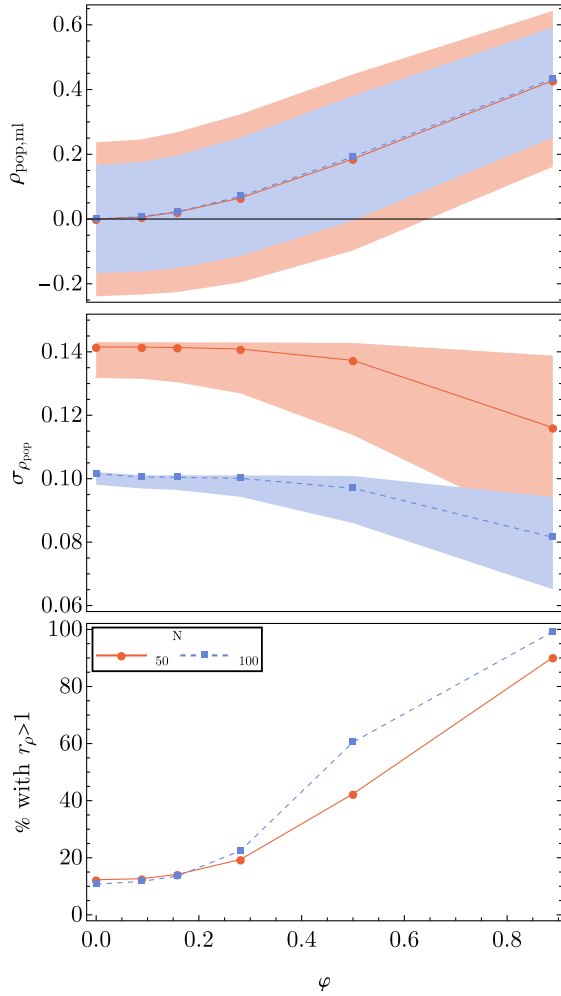


FIG. 2. Top two panels show the median of the distributions $p(\rho_{\text{pop, ml}})$ and $p(\sigma_{\rho_{\text{pop}}})$ as a function of φ for catalogs of size N with single-event likelihood specified by $\sigma_{\text{pe}} = 0.1$, $\rho = 0.9$ and systematic error of the form in Eq. (17). The 5% and 95% quantiles of the probability distributions enclose the shaded regions. The bottom panel shows the percentage of catalogs that produce a credible measurement of $\rho_{\text{pop}} \neq 0$, corresponding to $r_\rho > 1$. The true population correlation is $\rho_{\text{pop}}^* = 0$. The specific analyses considered here correspond to $\varphi = 5 \times 10^\beta$, for $\beta = \{-\infty, -1.75, -1.5, -1.25, -1.0, -0.75\}$ which constitute systematic error that is significant for $\{0, 10, 20, 40, 50, 70\}\%$ of the catalog. We note that when systematic error is small, or nonexistent, increasing the size of the catalog lowers the probability of a false-positive measurement of $\rho_{\text{pop}} \neq 0$. However, as the size of the systematic error grows, the resulting increase in precision means a greater probability of obtaining a credible measurement of $\rho_{\text{pop}} \neq 0$.

$r_\rho > 1$. We consider catalogs with $N \in \{50, 100\}$. Figure 2 shows the median of the distributions $p(\rho_{\text{pop, ml}})$ and $p(\sigma_{\rho_{\text{pop}}})$ as a function of φ . The 5% and 95% quantiles of the probability distributions enclose the shaded regions. The bottom panel shows the percentage of catalogs for which $r_\rho > 1$.

We see that as the significance of single-event bias grows, the median $\rho_{\text{pop, ml}}$ moves farther from zero, indicating a higher probability for bias in the measurement of ρ_{pop} . While catalog size has virtually no impact on the median $p(\rho_{\text{pop, ml}})$, increasing the number of events does tighten the spread of the distribution. On the other hand, as in the case with no mismodeling, catalog size is the dominant factor affecting the distribution $p(\sigma_{\rho_{\text{pop}}})$, with larger catalogs corresponding to more precise measurements of ρ_{pop} .

When waveform mismodeling is small or nonexistent, the bias in $\rho_{\text{pop, ml}}$ is negligible, and a larger catalog size means a lower probability of measuring $\rho_{\text{pop}} \neq 0$. However, as the waveform mismodeling and bias in $\rho_{\text{pop, ml}}$ grow, a larger catalog increases the probability of obtaining a credible measurement of $\rho_{\text{pop}} \neq 0$. The effects of statistical uncertainty in the measurement of population parameters decrease as catalog size grows, allowing systematic effects to dominate.

Thus far, we have considered the case where φ is common among all events. However, it is also possible that the bias in θ_1 and θ_2 depends on additional parameters that also vary across the catalog. These parameters may be accounted for in the waveform approximant, or they may be additional, unmodeled parameters such as black hole hair or environmental effects. We can represent such scenarios by allowing φ to vary among events.

We consider three specific scenarios: (1) φ is drawn from a Gaussian distribution $\mathcal{N}(\varphi|0, 0.5)$ centered at zero, (2) φ is the absolute value of a number drawn from a Gaussian distribution $\mathcal{N}(\varphi|0, 0.5)$ centered at zero, and (3) φ is drawn from a Gaussian $\mathcal{N}(\varphi|0.5, 0.5)$ with a nonzero mean. We compare these scenarios to the case with no bias ($\varphi = 0$) and to the case where φ is fixed ($\varphi = 0.5$, corresponding to significant bias for 50% of the population). The distributions resulting from these scenarios are illustrated in Fig. 3. We see that allowing φ to vary among events can either reduce or enhance the effects of systematic error with respect to the case of constant φ , depending on the distribution φ takes.

C. Mismodeling at the population level

Finally, we consider the case where the mismodeling occurs at the population level. That is, while we still employ the Gaussian model given in Eq. (9) in the hierarchical inference, we draw the catalogs from underlying distributions $\pi^*(\theta|\Lambda^*)$ that are non-Gaussian. We assume θ_1 and θ_2 are uncorrelated at the population level such that the joint probability distribution can be written as the product of the marginal distributions $\pi^*(\theta|\Lambda^*) = \pi_1^*(\theta_1|\Lambda_1^*)\pi_2^*(\theta_2|\Lambda_2^*)$. As with mismodeling at the single-event level, there are infinitely many ways in which the true population distribution might differ from the chosen model. We focus on several underlying distributions that

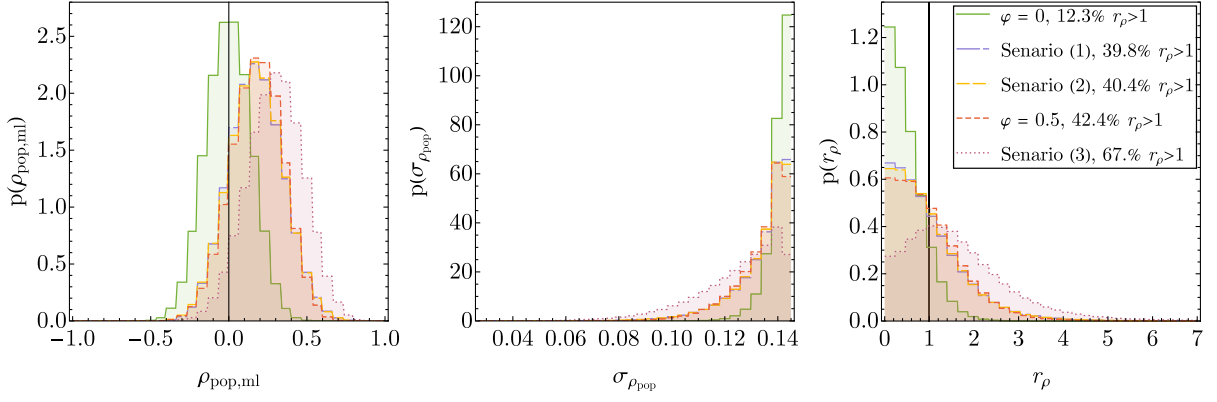


FIG. 3. Probability distributions $p(\rho_{\text{pop,ml}})$, $p(\sigma_{\rho_{\text{pop}}})$, and $p(r_\rho)$ for several scenarios in which systematic bias in single-event parameters takes the form given in Eq. (17). In scenario (1), φ is drawn from a Gaussian distribution $\mathcal{N}(\varphi|0, 0.5)$ centered at zero, in scenario (2), φ is the absolute value of a number drawn from a Gaussian distribution $\mathcal{N}(\varphi|0, 0.5)$ centered at zero, and in scenario (3), φ is drawn from a Gaussian $\mathcal{N}(\varphi|0.05, 0.5)$ with a nonzero mean. These scenarios are compared to the case with no systematic error, $\varphi = 0$, and the case where $\varphi = 0.5$ is constant across events. Each distribution corresponds to a catalog of $N = 50$ events, $\sigma_{\text{pe}} = 0.1$, and $\rho_{\text{pe}} = 0.9$. The black vertical line in the right-most panel indicates $r_\rho = 1$, above which we assume a credible false-positive measurement of correlation. The percentage of catalogs in each configuration that fall above this line is reported in the plot legend. We see that a systematic error that depends on a third parameter can either increase or decrease the probability of a false-positive measurement of population-level correlation, depending on how that additional parameter varies across the parameter space.

allow us to explore the effects of specific types of mismodeling.

We consider two families of underlying distributions that generalize the Gaussian distribution. The exponential power distribution,

$$\pi_i(\theta_i|\sigma_{\text{pop},i}, \kappa) \propto \exp\left(-\frac{1}{\kappa} \left|\frac{\theta_i}{\sigma_{\text{pop},i}}\right|^\kappa\right), \quad (19)$$

allows us to explore the effects of an underlying distribution with lighter or heavier tails than the model. When $\kappa = 2$, we recover a Gaussian distribution. When $\kappa > 2$ ($\kappa < 2$), the tails of the distribution are lighter (heavier) than those of a Gaussian distribution. The skew normal distribution

$$\begin{aligned} \pi_i(\theta_i|\sigma_{\text{pop},i}, \alpha) &\propto \left[1 + \text{Erf}\left(\frac{\alpha}{\sqrt{2}} \frac{\theta_i}{\sigma_{\text{pop},i}}\right)\right] \\ &\times \exp\left[-\frac{1}{2} \left(\frac{\theta_i}{\sigma_{\text{pop},i}}\right)^2\right] \end{aligned} \quad (20)$$

allows us to explore the effects of using a symmetric model to recover an asymmetric distribution. When $\alpha = 0$, we recover a Gaussian distribution. As the absolute value of α grows, the distribution becomes more asymmetric.

We consider four underlying population distributions, each with $\sigma_{\text{pop},1}^* = \sigma_{\text{pop},2}^* = 1$. Two populations are distributed according to joint exponential power distributions with $\kappa^* \in \{0.2, 200\}$. Two are distributed according to joint skew normal distributions with $\alpha^* \in \{1, 10\}$. For illustration, the marginal distributions are depicted in Fig. 4,

compared to a Gaussian distribution (corresponding to an exponential power distribution with $\kappa = 2$ or a skew normal distribution with $\alpha = 0$). We specify the single-event

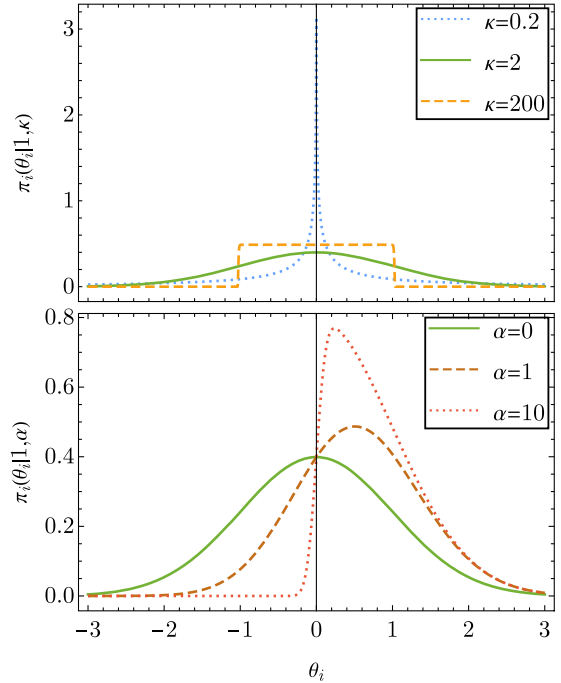


FIG. 4. Marginal underlying population distributions used to generate the distributions in Fig. 5. The top panel depicts the exponential power distribution of Eq. (19) and the bottom panel depicts the skew distribution of Eq. (20). For each plot, $\sigma_{\text{pop},i} = 1$. The green distributions ($\kappa = 2$ and $\alpha = 0$) are the same in both the upper and lower panels and correspond to a Gaussian underlying distribution.

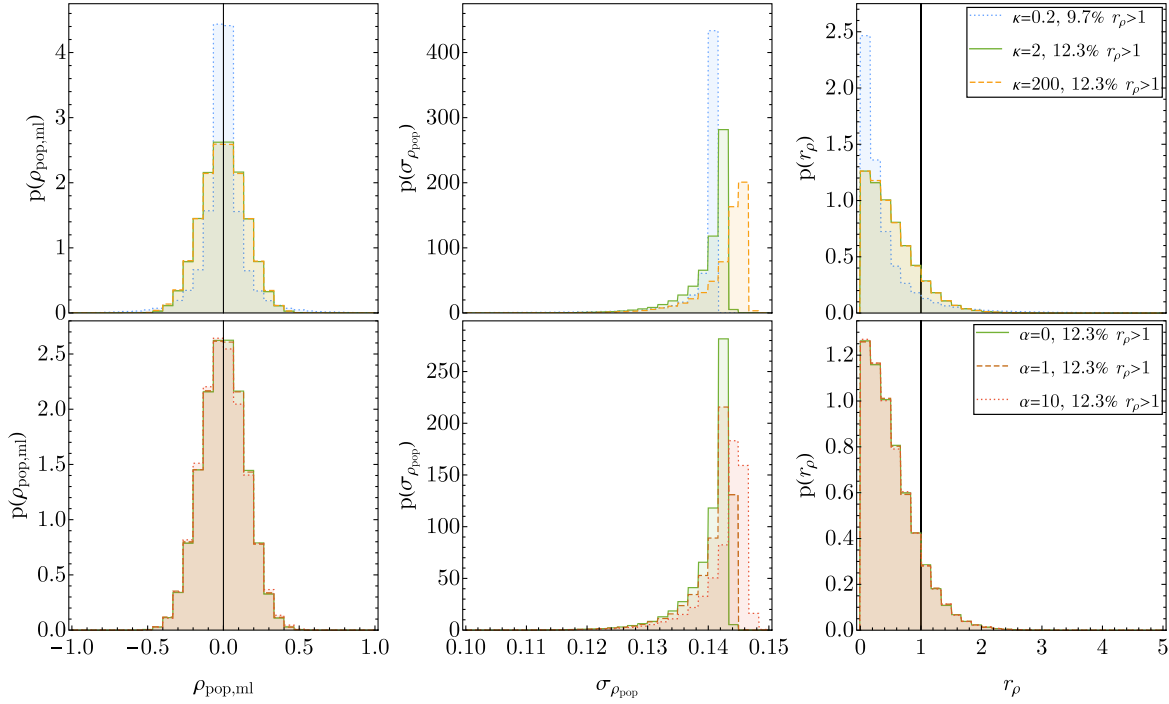


FIG. 5. Probability distributions $p(\rho_{\text{pop,ml}})$, $p(\sigma_{\rho_{\text{pop}}})$, and $p(r_\rho)$ corresponding to several different uncorrelated underlying population distributions. In the top panels, the true population distribution is the exponential power distribution of Eq. (19), and in the bottom panels, the true population distribution is the skew normal distribution of Eq. (20). The green distributions ($\kappa = 2$ and $\alpha = 0$) are the same in both the upper and lower panels and correspond to a Gaussian underlying distribution. Each distribution corresponds to a catalog of $N = 50$ events, a single-event likelihood specified by $\sigma_{\text{pe}} = 0.1$, and $\rho_{\text{pe}} = 0.9$, and the Gaussian population recovery model given in Eq. (9). The black vertical line in the right-most panel indicates $r_\rho = 1$, above which we assume a credible false-positive measurement of correlation. The percentage of catalogs in each configuration that fall above this line is reported in the plot legend. We see that for this simplified toy model, population mismodeling has a negligible effect on the probability of obtaining a credible measurement of the population-level correlation. The one exception is the exponential Gaussian population distribution with $\kappa = 0.2$, corresponding to a heavier-tailed distribution, which shows a reduced probability of a non-zero correlation measurement.

likelihood with $\sigma_{\text{pe}} = 0.1$ and $\rho_{\text{pe}} = 0.9$ and $\Delta\theta_{\text{sys}} = 0$. For each underlying population distribution, we consider catalogs with $N = 50$. The probability distributions $p(\rho_{\text{pop,ml}})$, $p(\sigma_{\rho_{\text{pop}}})$ and $p(r_\rho)$ are shown in Fig. 5 along with those generated in Sec. III A, from an underlying Gaussian distribution.

As discussed in Sec. II B, we do not expect $\rho_{\text{pop,ml}}$ to be biased from the true value $\rho_{\text{pop}}^* = 0$ in the large N limit, and this is indeed what we see, even for $N = 50$. On the other hand, the distribution $p(\sigma_{\rho_{\text{pop}}})$ varies across underlying distributions, indicating that using an inaccurate population model could affect the precision with which correlation is measured. However, for most of the population distributions considered, we observe no differences in the percentage of catalogs that ultimately yield a false-positive correlation measurement compared with the case of no population mismodeling. The one exception is when the population is distributed according to an exponential power distribution with a heavy tail. In this case, though we generally associate mismodeling with a higher probability of bias, $p(\rho_{\text{pop,ml}})$ has smaller variance about the true value

$\rho_{\text{pop}}^* = 0$, resulting in a reduced probability of making a false-positive measurement of population-level correlation.

We have discussed how the simplicity of our Gaussian toy model makes it ill-suited to fully exploring the effects of systematic bias arising from population-level mismodeling. To compound the issue, the population models used in the analysis of the actual GW transient catalog are more complex than the simple correlated Gaussians considered here, so mismodeling is likely to have effects that are not straightforward to predict. However, our results suggest that while population mismodeling can affect the probability of a false-positive measurement of population-level correlation, an otherwise poor model may still identify the presence or absence of correlation in a population.

IV. CONCLUSION AND DISCUSSION

We have used a toy model to explore how statistical and systematic effects at both the single-event and population levels may affect the measurement of population-level correlations between pairs of binary parameters in compact binaries observed by LVK. The simplified nature of the toy

model makes it difficult to draw quantitative comparisons with population analyses of the actual GW catalog. That being said, this study allows us to draw several qualitative conclusions that may inform our understanding of the robustness of the identified population-level correlation.

We tackled the effects of single-event correlations on the measurement of population-level correlations, and we found two notable and counterintuitive results:

- (i) While the presence of single-event correlation tends to increase the magnitude of the statistical shift in the maximum likelihood population correlation, this shift need not align with the correlation in parameter estimation. Rather, there seems to be a slight negative correlation between single-event correlation and inferred population correlation when quantifying correlation with the dimensionless coefficient ρ_{pop} .
- (ii) Because the presence of single-event correlation increases the uncertainty in the measurement of population correlations, it can actually reduce the probability of a credible false-positive correlation measurement at the population level. We conclude that, for pairs of parameters with strong correlation at the parameter estimation level, such as effective spin and mass ratio, this correlation could potentially obscure true underlying population correlations. This effect will diminish as the SNRs of detections increase and as more events are added to the catalog.

We next considered systematic effects resulting from mismodelling in both the waveform and population distribution. For the latter, we found evidence that an otherwise poor population model may still be useful for ruling out correlation in the underlying distribution. As for waveform systematics, we found that when waveform mismodeling introduces biases into recovered single-event parameters that are correlated across the catalog, these biases are likely to be interpreted as population-level correlations between those parameters. Unlike the effects of single-event correlation, this effect will tend to increase both as detections become louder and as more events are added to the catalog. Our results align with general intuition about systematic effects: increasing detector sensitivity and adding more events to the catalog will make our measurements of single-event and population parameters more precise, but can only make them more accurate to the extent that our waveforms and population models are good representations of reality.

The state-of-the-art approximants IMRPhenomXPHM-SpinTaylor [48,49], SEOBNRv5PHM [50,51], and NRSur7dq4 [52] used to analyze the current GWTC-5 events [47] are meticulously validated against numerical-relativity simulations. However, the former approaches recover numerical-relativity waveforms less faithfully in regions with more unequal mass ratios and higher spin

precession, while the latter can only capture sufficiently short signals. This mismodeling, which varies across the parameter space, has the potential to produce spurious population-level correlations. Several events in the current GW catalog exhibit discrepancies between the posteriors obtained using different waveforms, indicating the presence of systematic error. Unsurprisingly, this error is often most evident in the mass ratio and spin parameters [44–47]. Our results emphasize the importance of investigating how such biases could be affecting correlation measurements in the actual GW catalog.

Furthermore, we have focused here on systematic error arising from mismodeling in the waveform approximant and the population distribution model. However, systematic error is, more generally, any non-stochastic error arising from incorrect assumptions about the experiment. This could include, among other things, deviations from the assumption of stationary and Gaussian noise that inform the choice of single-event likelihood or mischaracterization of selection effects at the population level. While we did not consider such effects in our study, they could potentially contribute to biases that could be interpreted as population correlation.

As the number of detected GW events grows well into the hundreds, the marginal population distributions are becoming increasingly well understood, and identifying correlations and subpopulations is emerging as a prominent new frontier for characterizing the observed dataset and ultimately improving our (astro)physical understanding of compact binaries. Despite the simplicity of our model, we hope this exploration of the robustness of inferred correlations will serve as a useful compass for interpreting current and upcoming GW data releases.

ACKNOWLEDGMENTS

We thank Riccardo Sturani and Matthew Mould for discussions. C. B. O., A. T., and D. G. are supported by ERC Starting Grant No. 945155–GWmining, Cariplo Foundation Grant No. 2021-0555, Italian-French University (UIF/UFJ) Grant No. 2025-C3-386, MUR Grant “Progetto Dipartimenti di Eccellenza 2023-2027” (BiCoQ), and the INFN TEONGRAV initiative. A. T. and D. G. are supported by MUR Young Researchers Grant No. SOE2024-0000125. D. G. is supported by MSCA Fellowship No. 101149270–ProtoBH. Computational work was performed at CINECA with allocations through INFN and the University of Milano-Bicocca.

DATA AVAILABILITY

There are no publicly available research data or software supporting this manuscript. Requests for further information or data should be sent to the authors.

- [1] I. Mandel, W. M. Farr, and J. R. Gair, *Mon. Not. R. Astron. Soc.* **486**, 1086 (2019).
- [2] S. Vitale, D. Gerosa, W. M. Farr, and S. R. Taylor, in *Handbook of Gravitational Wave Astronomy* (Springer, Singapore, 2021).
- [3] A. G. Abac *et al.* (LIGO, Virgo, and KAGRA Collaborations), [arXiv:2605.27226](#).
- [4] A. G. Abac *et al.* (LIGO, Virgo, and KAGRA Collaborations), [arXiv:2605.27227](#).
- [5] A. G. Abac *et al.* (LIGO, Virgo, and KAGRA Collaborations), [arXiv:2603.19019](#).
- [6] T. A. Callister, in *Encyclopedia of Astrophysics*, 1st ed. (Elsevier, New York, 2026).
- [7] J. Heinzl, M. Mould, S. Álvarez-López, and S. Vitale, *Phys. Rev. D* **111**, 063043 (2025).
- [8] F. Antonini, I. Romero-Shaw, T. Callister, F. Dosopoulou, D. Chattopadhyay, Y. B. Ginat, M. Gieles, and M. Mapelli, *Nat. Astron.* **10**, 1049 (2026).
- [9] R. Tenorio, A. Toubiana, T. Bruel, D. Gerosa, and J. R. Gair, *Astrophys. J. Lett.* **994**, L52 (2025).
- [10] S. Banagiri, E. Thrane, and P. D. Lasky, *Phys. Rev. Lett.* **137**, 021403 (2026).
- [11] N. Guttman, E. Payne, P. D. Lasky, and E. Thrane, *Astrophys. J.* **996**, 144 (2026).
- [12] A. Ray and V. Kalogera, *Astrophys. J. Lett.* **998**, L20 (2026).
- [13] V. Tiwari, [arXiv:2510.25579](#).
- [14] H. Tong, T. A. Callister, M. Fishbach, E. Thrane, F. Antonini, S. Stevenson, I. M. Romero-Shaw, and F. Dosopoulou, [arXiv:2511.05316](#).
- [15] V. Gennari, S. Mastrogiovanni, N. Tamanini, S. Marsat, and G. Pierra, *Phys. Rev. D* **111**, 123046 (2025).
- [16] S. Alvarez-Lopez, J. Heinzl, M. Mould, and S. Vitale, *Astrophys. J. Lett.* **1006**, L5 (2026).
- [17] A. Q. Cheng, A. Toubiana, S. Biscoveanu, and J. Gair, [arXiv:2605.25980](#).
- [18] I. Mandel and A. Farmer, *Phys. Rep.* **955**, 1 (2022).
- [19] M. Mapelli, in *Handbook of Gravitational Wave Astronomy* (Springer, Singapore, 2021).
- [20] H. Zhong, M. Isi, K. Chatziioannou, and W. M. Farr, *Phys. Rev. D* **110**, 044053 (2024).
- [21] E. Payne, M. Isi, K. Chatziioannou, L. Lehner, Y. Chen, and W. M. Farr, *Phys. Rev. Lett.* **133**, 251401 (2024).
- [22] T. A. Callister, C.-J. Haster, K. K. Y. Ng, S. Vitale, and W. M. Farr, *Astrophys. J. Lett.* **922**, L5 (2021).
- [23] R. Abbott *et al.* (LIGO, Virgo, and KAGRA Collaborations), *Phys. Rev. X* **13**, 011048 (2023).
- [24] A. G. Abac *et al.* (LIGO, Virgo, and KAGRA Collaborations), *Astrophys. J. Lett.* **1005**, L51 (2026).
- [25] K. K. Y. Ng, S. Vitale, A. Zimmerman, K. Chatziioannou, D. Gerosa, and C.-J. Haster, *Phys. Rev. D* **98**, 083007 (2018).
- [26] C. B. Owen, C.-J. Haster, S. Perkins, N. J. Cornish, and N. Yunes, *Phys. Rev. D* **108**, 044018 (2023).
- [27] A. Dhani, S. H. Völkel, A. Buonanno, H. Estelles, J. Gair, H. P. Pfeiffer, L. Pompili, and A. Toubiana, *Phys. Rev. X* **15**, 031036 (2025).
- [28] V. Kapil, L. Reali, R. Cotesta, and E. Berti, *Phys. Rev. D* **109**, 104043 (2024).
- [29] R. S. Chandramouli, K. Prokup, E. Berti, and N. Yunes, *Phys. Rev. D* **111**, 044026 (2025).
- [30] A. G. Abac *et al.* (LIGO, Virgo, and KAGRA Collaborations), *Astrophys. J. Lett.* **993**, L25 (2025).
- [31] A. G. Abac *et al.* (LIGO, Virgo, and KAGRA Collaborations), *Astrophys. J. Lett.* **1004**, L23 (2026).
- [32] I. M. Romero-Shaw, E. Thrane, and P. D. Lasky, *Pub. Astron. Soc. Aust.* **39**, e025 (2022).
- [33] A. Q. Cheng, M. Zevin, and S. Vitale, *Astrophys. J.* **955**, 127 (2023).
- [34] S. J. Miller, S. Winney, K. Chatziioannou, and P. M. Meyers, [arXiv:2604.06090](#).
- [35] M. Mould, J. Heinzl, S. Alvarez-Lopez, C. Plunkett, N. E. Wolfe, and S. Vitale, *Phys. Rev. D* **113**, 103021 (2026).
- [36] M. Vallisneri, *Phys. Rev. D* **77**, 042001 (2008).
- [37] C. Cutler and M. Vallisneri, *Phys. Rev. D* **76**, 104018 (2007).
- [38] A. G. Abac *et al.* (LIGO, Virgo, and KAGRA Collaborations), [arXiv:2603.19020](#).
- [39] A. G. Abac *et al.* (LIGO, Virgo, and KAGRA Collaborations), [arXiv:2603.19021](#).
- [40] E. Payne, M. Isi, K. Chatziioannou, and W. M. Farr, *Phys. Rev. D* **108**, 124060 (2023).
- [41] J. R. Gair, A. Antonelli, and R. Barbieri, *Mon. Not. R. Astron. Soc.* **519**, 2736 (2022).
- [42] V. De Renzi, F. Iacovelli, D. Gerosa, M. Mancarella, and C. Pacilio, *Phys. Rev. D* **111**, 044048 (2025).
- [43] D. Lewandowski, D. Kurowicka, and H. Joe, *J. Multivariate Anal.* **100**, 1989 (2009).
- [44] R. Abbott *et al.* (LIGO and Virgo Collaborations), *Phys. Rev. D* **109**, 022001 (2024).
- [45] R. Abbott *et al.* (LIGO, Virgo, and KAGRA Collaborations), *Phys. Rev. X* **13**, 041039 (2023).
- [46] A. G. Abac *et al.* (LIGO, Virgo, and KAGRA Collaborations), *Astrophys. J. Lett.* **1004**, L22 (2026).
- [47] A. G. Abac *et al.* (LIGO, Virgo, and KAGRA Collaborations), [arXiv:2605.27225](#).
- [48] G. Pratten *et al.*, *Phys. Rev. D* **103**, 104056 (2021).
- [49] M. Colleoni, F. A. R. Vidal, C. García-Quirós, S. Akçay, and S. Bera, *Phys. Rev. D* **111**, 104019 (2025).
- [50] A. Ramos-Buades, A. Buonanno, H. Estellés, M. Khalil, D. P. Mihaylov, S. Ossokine, L. Pompili, and M. Shiferaw, *Phys. Rev. D* **108**, 124037 (2023).
- [51] L. Pompili *et al.*, *Phys. Rev. D* **108**, 124035 (2023).
- [52] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, *Phys. Rev. Res.* **1**, 033015 (2019).