Assessing and marginalizing over compact binary coalescence waveform systematics with RIFT

A. Z. Jan,[1] A. B. Yelikar,[1] J. Lange,[2,1] and R. O'Shaughnessy[1]

[1]*Center for Computational Relativity and Gravitation, Rochester Institute of Technology, Rochester, New York 14623, USA*

[2]*Institute of Computational and Experimental Research in Mathematics, Brown University, Rhode Island 02903, USA*

(Received 6 November 2020; accepted 10 December 2020; published 29 December 2020)

As Einstein's equations for binary compact object inspiral have only been approximately or intermittently solved by analytic or numerical methods, the models used to infer parameters of gravitational wave (GW) sources are subject to waveform modeling uncertainty. Using a simple scenario, we illustrate these differences then introduce a very efficient technique to marginalize over waveform uncertainties, relative to a prespecified sequence of waveform models. Being based on RIFT, a very efficient parameter inference engine, our technique can directly account for any available models, including very accurate but computationally costly waveforms. Our evidence- and likelihood-based method works robustly on a point-by-point basis, enabling accurate marginalization for models with strongly disjoint posteriors while simultaneously increasing the reusability and efficiency of our intermediate calculations.

DOI: 10.1103/PhysRevD.102.124069

I. INTRODUCTION

Since the first gravitational wave detection GW150914 [1], the Advanced Laser Interferometer Gravitational Wave Observatory (LIGO) [2] and Virgo [3,4] detectors have continued to discover gravitational waves (GW) from coalescing binary black holes (BBHs) and neutron stars. The properties of each source are inferred by comparing each observation to some estimate(s) for the GW emitted when a BBH merge, commonly called an approximant. As illustrated most recently by GW190521 [5,6], GW190814 [7], GW190412 [8], and the discussion in GWTC-2 [9], these approximations disagree more than enough to produce noticable differences, consistent with prior work [10–12]. Despite the ongoing generation of new waveforms with increased accuracy [13–18], these previous investigations suggest that waveform model systematics can remain a limiting factor in inferences about individual events [10] and populations [12,19].

Recently, Ashton and Khan [20] described and illustrated marginalizing between a discrete set of waveform models in a fully Bayesian way. In this procedure, the waveform-marginalized posterior is the weighted average of the posteriors $p_k(\theta)$ derived from each waveform model k alone, weighted by the evidence Z_k for (and prior p_k for) each model k: $p(\theta) = [\sum_k p_k(\theta) p_k Z_k] / \sum_q p_q Z_q$. This extremely simple procedure faces one obvious limitation: analysis must be performed for every waveform model of interest. Unfortunately, as many of the most accurate time-domain waveform models incur exceptionally high evaluation costs, and as most conventional parameter estimation (PE) engines like LALInference [21] or BILBY [22] are limited by this cost, the universe of possible waveforms must often omit the most expensive and accurate waveform models. As the RIFT parameter inference engine circumvents several issues associated with waveform evaluation cost [23,24], despite retaining the original waveform implementation (i.e., no surrogate generation), in this work we examine novel extensions of this waveform-marginalization technique that are uniquely adapted to RIFT's algorithm. Using a simple toy model, we demonstrate the pernicious effects of model systematics, then show how our technique efficiently mitigates them.

This paper is organized as follows. In Sec. II, we review the use of RIFT for parameter inference, the two waveform models used in this work, the use of probability-probability (PP) plots to diagnose systematic error with noise, the use of zero-noise PE to isolate the systematic uncertainty between waveforms, and our waveform marginalization technique. In Sec. III, we use two well-studied waveform models to demonstrate the impact of contemporary model systematics, then marginalize over them. We emphasize that all calculations in this section adopt signal amplitudes and masses consistent with current observations. In Sec. V, we summarize our results and discuss their potential applications to future GW source and population inference.

II. METHODS

A. RIFT review

A coalescing compact binary in a quasicircular orbit can be completely characterized by its intrinsic and extrinsic parameters. By intrinsic parameters we refer to the binary's masses m_i, spins, and any quantities characterizing matter in the system. For simplicity and reduced computational overhead, in this work we assume all compact object spins are aligned with the orbital angular momentum. By extrinsic parameters we refer to the seven numbers needed to characterize its spacetime location and orientation. We will express masses in solar mass units and dimensionless nonprecessing spins in terms of Cartesian components aligned with the orbital angular momentum $\chi_{i,z}$. We will use λ, θ to refer to intrinsic and extrinsic parameters, respectively.

RIFT [23] consists of a two-stage iterative process to interpret gravitational wave data d via comparison to predicted gravitational wave signals $h(\lambda,\theta)$. In one stage, for each λ_α from some proposed "grid" $\alpha = 1, 2, \ldots N$ of candidate parameters, RIFT computes a marginal likelihood

$$\mathcal{L}_{\rm marg} \equiv \int \mathcal{L}(\lambda,\theta) p(\theta) d\theta \tag{1}$$

from the likelihood $\mathcal{L}(\lambda,\theta)$ of the gravitational wave signal in the multidetector network, accounting for detector response; see the RIFT paper for a more detailed specification. In the second stage, RIFT performs two tasks. First, it generates an approximation to $\mathcal{L}(\lambda)$ based on its accumulated archived knowledge of marginal likelihood evaluations $(\lambda_\alpha, \mathcal{L}_\alpha)$. This approximation can be generated by Gaussian processes, random forests, or other suitable approximation techniques. Second, using this approximation, it generates the (detector-frame) posterior distribution

$$p_{\rm post} = \frac{\mathcal{L}_{\rm marg}(\boldsymbol{\lambda}) p(\boldsymbol{\lambda})}{\int d\boldsymbol{\lambda} \mathcal{L}_{\rm marg}(\boldsymbol{\lambda}) p(\boldsymbol{\lambda})}, \tag{2}$$

where $p(\boldsymbol{\lambda})$ is the prior on intrinsic parameters like mass and spin. The posterior is produced by performing a Monte Carlo integral: the evaluation points and weights in that integral are weighted posterior samples, which are fairly resampled to generate conventional independent, identically distributed "posterior samples." For further details on RIFT's technical underpinnings and performance, see [23–25].

B. Waveform models

In this work, we employ two well-studied models for nonprecessing binaries, whose differences are known to be significant. We use SEOBNRV4 [15], an effective-one-body model for quasicircular inspiral, and IMRPHENOMD [26,27], a phenomenological frequency-domain inspiral-merger-ringdown model.

The effective-one-body (EOB) approach models the inspiral and spin dynamics of coalescing binaries via an ansatz for the two-body Hamiltonian [28], whose corresponding equations of motion are numerically solved in the time domain. For nonprecessing binaries, outgoing gravitational radiation during the inspiral phase is generated using an ansatz for resumming the post-Newtonian expressions for outgoing radiation including nonquasicircular corrections, for the leading-order $\ell = 2$ subspace. For the merger phase of nonprecessing binaries, the gravitational radiation is generated via a resummation of many quasinormal modes, with coefficients chosen to ensure smoothness. The final BH's mass and spin, as well as some parameters in the nonprecessing inspiral model, are generated via calibration to numerical relativity simulations of BBH mergers.

The IMRPHENOMD model is a part of an approach that attempts to approximate the leading-order ($\ell = 2$) gravitational wave radiation using phenomenological fits to the Fourier transform of the gravitational wave strain, computed from numerical relativity simulations, effective-one-body waveforms and post-Newtonian calculation [13,29,30]. Also using information about the final BH state, this phenomenological frequency-domain approach matches standard approximations for the post-Newtonian gravitational wave phase to an approximate, theoretically motivated spectrum characterizing merger and ringdown.

C. Fiducial synthetic sources and PP tests

We will only explore the impact of systematics over a limited fiducial population. Specifically, we consider a universe of synthetic signals for three-detector networks, with masses drawn uniformly in m_i in the region bounded by $\mathcal{M}/M_\odot \in [30, 60]$ and $\eta \in [0.2, 0.25]$ and with extrinsic parameters drawn uniformly in sky position and isotropically in Euler angles, with source luminosity distances drawn proportional to d_L^2 between 1.5 Gpc and 4 Gpc. These bounds are expressed in terms of $\mathcal{M} = (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$ and $\eta = m_1 m_2/(m_1 + m_2)^2$ and encompass the detector-frame parameters of many massive binary black holes seen in GWTC-1 [31] and GWTC-2 [9]. All our sources have nonprecessing spins, with each component assumed to be uniform between $[-1, 1]$. For complete reproducibility, we use SEOBNRV4 and IMRPHENOMD, starting the signal evolution at 18 Hz but the likelihood integration at 20 Hz, performing all analysis with 4096 Hz time series in Gaussian noise with known advanced LIGO design power spectral densities [32]. For each synthetic event and for each interferometer, the same noise realization is used for both waveform approximations. Ensuring convergence of the analyses, the differences between them therefore arise solely due to waveform systematics. For context, Fig. 1 shows the cumulative signal to noise ratio

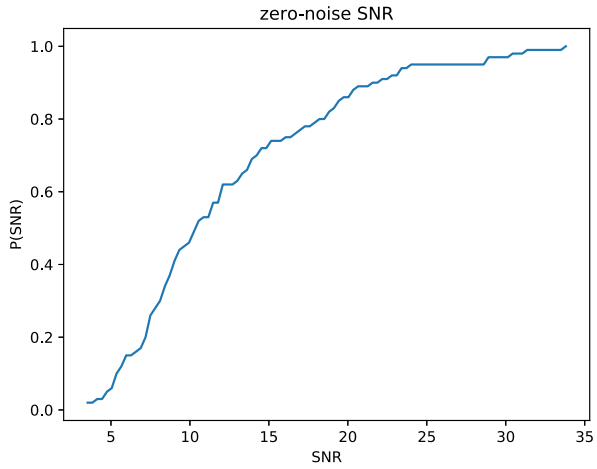

FIG. 1. Cumulative SNR distribution for a synthetic population of 100 events drawn from the fiducial BBH population described in Sec. II C. To avoid ambiguity, this figure shows the expected SNR (i.e., the SNR evaluated using a zero-noise realization).

(SNR) distribution of one specific synthetic population generated from this distribution. Though a small fraction have substantial signal amplitudes, most events are near or below the level of typical detection candidates. By using a very modest-amplitude population to assess the impact of waveform systematics, we demonstrate their immediate impact on the kinds of analyses currently being performed on real observations, let alone future studies.

One way to assess the performance of parameter inference is a probability-probability plots (usually denoted PP plot) [33]. Using RIFT on each source k, with true parameters λ_k, we estimate the fraction of the posterior distributions that is below the true source value $\lambda_{k,\alpha}$ $[\hat{P}_{k,\alpha}(<\lambda_{k,\alpha})]$ for each intrinsic parameter α, again assuming all sources have zero spin. After reindexing the sources so $\hat{P}_{k,\alpha}(\lambda_{k,\alpha})$ increases with k for some fixed α, the top panel of Fig. 2 shows a plot of k/N versus $\hat{P}_k(\lambda_{k,\alpha})$ for all binary parameters. For the top panel, both injections and inference are performed with the same model, and the recovered probability distribution is consistent with $P(<p) = p$, as expected.

D. Zero noise runs to assess systematic biases

Our synthetic data consists of expected detector responses $h(t)$ superimposed on detector noise realization $n(t)$. The recovered posterior distribution's properties and in particular maximum-likelihood parameters depend on the specific noise realization used. To disentangle the deterministic effects of waveform systematics from the stochastic impact of different noise realizations, we also repeat our analyses with the "zero noise" realization: $n(t) = 0$.

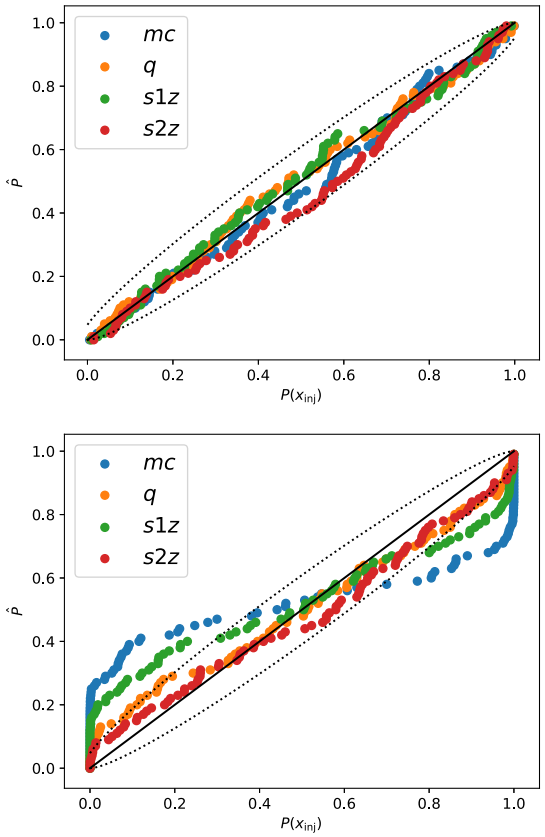

FIG. 2. PP plot of events injected with SEOBNRV4 and recovered with SEOBNRV4 (top panel) and IMRPHENOMD (bottom panel) waveform. The dashed line indicates the 90% credible interval expected for a cumulative distribution drawn from 100 uniformly distributed samples.

E. Model-model mismatch

Several previous investigations (e.g., [34–40] and references therein) have phenomenologically argued that the magnitude of systematic biases are related to the model-model *mismatch*, a simple inner-product-based estimate of waveform similarity between two model predictions $h_1(\lambda)$ and $h_2(\lambda)$ at identical model parameters λ:

$$\mathcal{M}(\lambda) = 1 - \max_{t_c,\phi_c} \frac{|\langle h_1 | e^{i(2\pi f t_c + \phi_c)} h_2 \rangle|}{|h_1||h_2|}. \tag{3}$$

In this expression, the inner product $\langle a|b\rangle_k \equiv \int_{-\infty}^{\infty} 2df \tilde{a}(f)^* \tilde{b}(f)/S_{h,k}(|f|)$ is implied by the kth detector's noise power spectrum $S_{h,k}(f)$, which for the purposes of waveform similarity is assumed to be the advanced LIGO instrument, H1. In practice we adopt a low-frequency cutoff $f_{\rm min}$ so all inner products are modified to

$$\langle a|b\rangle_k \equiv 2 \int_{|f|>f_{\rm min}} df \frac{[\tilde{a}(f)]^* \tilde{b}(f)}{S_{h,k}(|f|)}. \tag{4}$$

Figure 3 shows the distribution of mismatches for our synthetic population, where h_1 is generated using

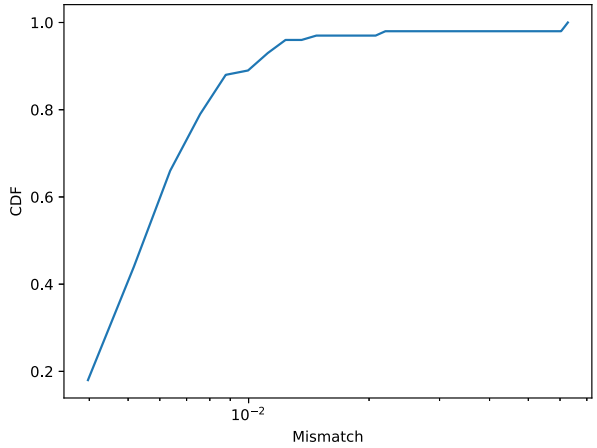

FIG. 3. Cumulative mismatch distribution for all the synthetic sources in our population. We evaluate the GW strain along the z axis using SEOBNRV4 and IMRPHENOMD and then compute the mismatch between them. This figure shows the cumulative distribution of these mismatches, most of which are slightly less than 10^{-2}.

SEOBNRV4 and h_2 with IMRPHENOMD. For simplicity, we regenerate all signals at zero inclination, to avoid polarization-related effects associated with the precise emission direction. For our fiducial compact binary population, the mismatches between these two models are typically below 10^{-2}, consistent with previous reports on systematic differences between these two waveforms and with their similarity to even more accurate models and simulations [12,15,41].

F. Marginalizing over waveform systematics

Suppose we have two models A and B for GW strain, and use them to interpret a particular GW source. We have prior probabilities $p(A|\lambda)$ and $p(B|\lambda)$, characterizing our relative confidence in these two models for a source with parameters λ.[1] Suppose we have produced a RIFT analysis with each model for this event, and have marginal likelihood functions $\mathcal{L}_A(\lambda)$ and $\mathcal{L}_B(\lambda)$ evaluated at a *single* point λ. We can therefore construct the marginal likelihood for λ by averaging over both models:

$$\mathcal{L}_{av}(\lambda) = p(A|\lambda)\mathcal{L}_A(\lambda) + p(B|\lambda)\mathcal{L}_B(\lambda). \tag{5}$$

For simplicity the calculations in this work always adopt $p(A|\lambda) = p(B|\lambda) = 1/2$. We can therefore transparently integrate multimodel inference into RIFT as follows. We assume we have a single grid of points λ_k such that both $(\lambda_k, \mathcal{L}_A(\lambda_k))$ and $(\lambda_k, \mathcal{L}_B(\lambda_k))$ can be interpolated to produce reliable likelihoods and thus posterior distributions $p_A(\lambda)$ and $p_B(\lambda)$, respectively. At each point λ_k we therefore construct $\mathcal{L}_{av}(\lambda_k)$ by the above procedure. We then interpolate to approximate $\hat{\mathcal{L}}(\lambda)$ versus the continuous parameters λ.

Operationally speaking, we construct model-averaged marginal likelihoods by the following procedure. First, we construct a fiducial grid for models A and B, for example by joining the grids used to independently analyze A and B. We use an algorithm to integrate the extrinsic likelihood, a process where each candidate GW signal is compared to a regular grid of candidate source parameters to produce an array of candidate likelihood values, to evaluate $\mathcal{L}_A(\lambda_k)$ and $\mathcal{L}_B(\lambda_k)$ on this grid [23,42]. We construct $\mathcal{L}_{av}(\lambda_k)$ as above. We use the combinations $(\lambda_k, \mathcal{L}_{av})$ with an algorithm to construct the intrinsic posterior from this sampled data to construct a model-averaged posterior distribution [23].

Our procedure bears considerable resemblance to the approach suggested by Ashton and Khan, but we have organized the calculation differently. In that approach, Ashton and Khan used the evidences $Z_A = \int \mathcal{L}_A p(\lambda) d\lambda$ and Z_B for the two waveform models. While we can compute both quantities with very high accuracy, we prefer to directly average between waveform models at the same choice of intrinsic parameters [i.e., via Eq. (5)] to ensure that marginalization over waveform models is completely decoupled from the interpolation techniques used to construct $\hat{\mathcal{L}}$ from the sampled data.

III. RESULTS

Using our fiducial BBH population, we generated 100 synthetic signals using IMRPHENOMD, and another 100 synthetic signals with SEOBNRV4. For each signal, we performed parameter inference with *both* IMRPHENOMD and SEOBNRV4. These inferences allow us both to assess the impact of waveform systematics in our fiducial population and mitigate them.

A. Demonstrating and quantifying waveform systematics

The PP plot provides the most compelling demonstration of waveform systematics' pernicious impact. Ideally, when recovering a known model and a known population, we expect to recover the injected values as often as they occur, producing a diagonal PP plot. The top panel of Fig. 2 shows precisely what we expect when we inject and recover with the same model (here, SEOBNRV4). By contrast, the bottom panel shows a PP plot generated using inference from IMRPHENOMD on the same SEOBNRV4 injections. The PP plot is considerably nondiagonal, reflecting frequent and substantial parameter biases in our fiducial population.

Parameter biases introduced by waveform systematics vary in magnitude and direction over the parameter space. To illustrate these offsets for the parameters $x = \mathcal{M}, q, \chi_{\rm eff}$, we have evaluated the parameter shift Δx between the mean

[1]For simplicity I will assume there are no internal model hyperparameters, but the method is easily generalized to include them.

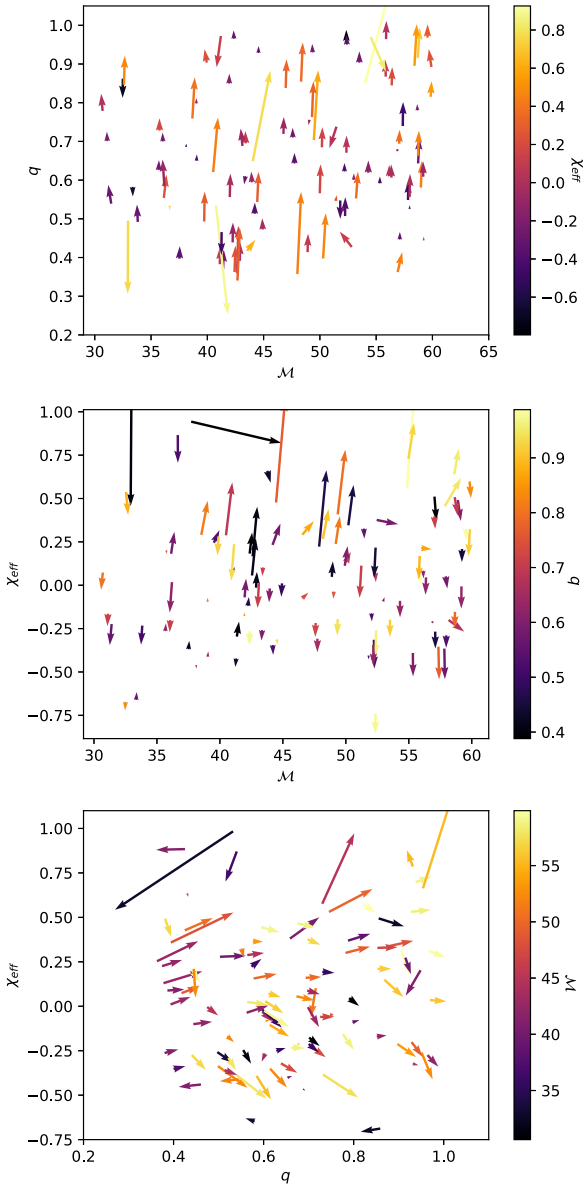

FIG. 4. Vector plot showing amplitude-scaled offsets between SEOBNRV4 and IMRPHENOMD for parameters $\mathcal{M}$ and q (top panel), $\mathcal{M}$ and $\chi_{\rm eff}$ (middle panel), and q and $\chi_{\rm eff}$ (bottom panel) as a function of the respective parameters with color map being the value of the parameter mentioned on the color scale.

inferred with IMRPHENOMD and the mean inferred with SEOBNRV4, *relative* to $\sigma\rho$, which is a product of ρ (the signal-to-noise ratio, a measure of the signal amplitude) and the statistical error (as measured by the standard deviation σ of the posterior of the parameter x in question). (The combination $\sigma\rho$ is approximately independent of signal amplitude, allowing us to measure the effect of waveform systematics for a fiducial amplitude.) Figure 4 shows a vector plot of these scaled offsets, $\Delta x/\rho\sigma$, as a function of two of the parameters at a time. The length of the arrow corresponds to the scaled shifts in the parameters $\mathcal{M}$, q, and $\chi_{\rm eff}$, plotted against the injected parameter values. The color scale shows the remaining parameter. The top two panels show that shifts in $q = m_2/m_1$, $\chi_{\rm eff} = (m_1\chi_{1,z} + m_2\chi_{2,z})/(m_1 + m_2)$ are substantial. Parameter shifts for q generally increase with $\chi_{\rm eff}$. Shifts in $\chi_{\rm eff}$ are generally positive for positive $\chi_{\rm eff}$, negative for negative $\chi_{\rm eff}$, and strongly dependent on mass ratio, with more substantial shifts at either comparable mass or at very high mass ratio, respectively. In both cases, chirp mass $\mathcal{M}$ has modest impact, with somewhat larger shifts occurring at somewhat larger values of chirp mass. Most extreme waveform systematics seem to be associated with large mass ratio.

Relative differences in mean value only imperfectly captures the differences between the two posteriors. As a sharper diagnostic that includes parameter correlations, we use the mean and covariance of each distribution in $\mathcal{M}$, q, $\chi_{\rm eff}$ to generate a local Gaussian approximation for each posterior, and then compute the KullbackLeibler (KL) divergence between these two Gaussian approximations [23]. We expect more substantial differences and thus larger KL divergence for stronger signals, whose posteriors are more sharply constrained. To corroborate our intuition, Fig. 5 shows a scatter plot with these KL divergences on the horizontal axis and the largest value of $\ln\mathcal{L}$ on the vertical axis. As expected, for the strongest signals, differences between the two waveform models are the most pronounced.

One might expect that large parameter offsets are more likely to occur when the data favors one model or another. While conceivably true asymptotically, for our specific synthetic population, we do not find a strong correlation between the Bayes factor (BF) ($\mathcal{Z}_{\rm SEOBNRV4}/\mathcal{Z}_{\rm IMRPHENOMD}$) and any parameter offsets. Figure 6 shows this BF plotted versus the scaled parameter offsets in $\mathcal{M}$, q, $\chi_{\rm eff}$. Large offsets can occur without the data more strongly favoring one model or the other and vice versa.

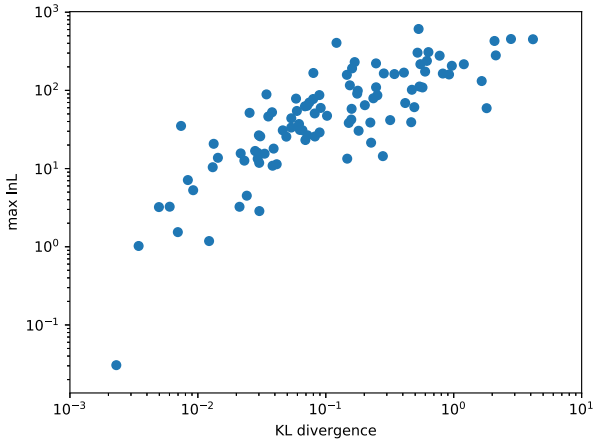

FIG. 5. Figure showing KL divergences between the two waveform models versus the log of the maximum likelihood for the combined posteriors of $\mathcal{M}$, q and $\chi_{\rm eff}$.

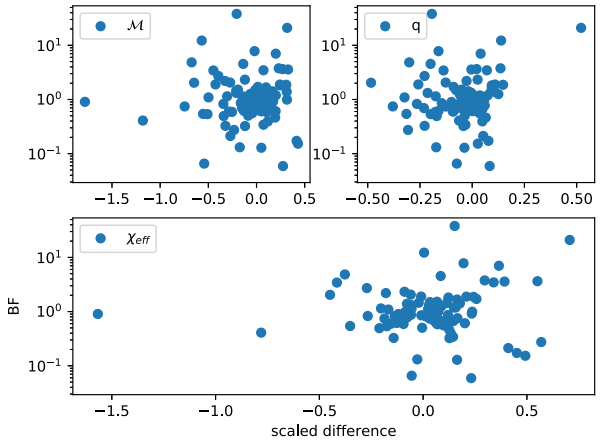

FIG. 6. Figure showing BF for SEOBNRV4 versus IMRPHENOMD plotted against differences between the SEOBNRV4 and IMRPHENOMD waveforms for parameters $\mathcal{M}$, q and $\chi_{\rm eff}$.

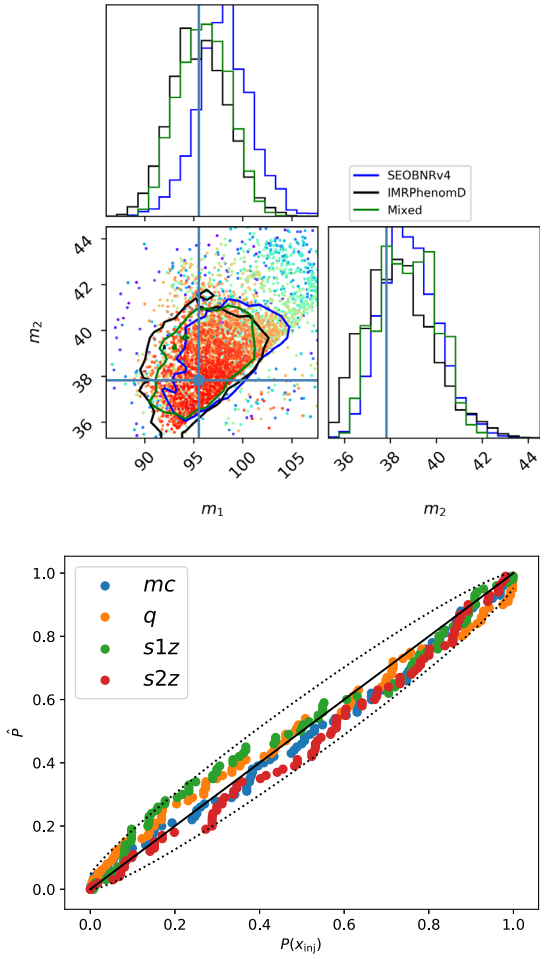

FIG. 7. Top panel: example of a model averaged result. The blue and black curves show the 1D marginal distributions and 2D 90% credibles for SEOBNRV4 and IMRPHENOMD inferences, respectively. The green curves show the corresponding model-averaged result. Bottom panel: PP plot test for our model-marginalized procedure.

B. A PP plot test for marginalizing over waveform errors

We test our model-averaged waveform procedure using a full synthetic PP plot procedure. Specifically, we use the $n_s = 100$ synthetic source parameters. For each source, we pick one waveform model A, B with probabilities $p(A)$, $p(B)$, and use it to generate the signal. We then analyze the signal using the model-averaged procedure described above.

As a concrete example, the top panel of Fig. 7 shows our analysis of one fiducial event in our synthetic sample. The colored points show likelihood evaluations, with color scale corresponding to the marginalized likelihood evaluated with IMRPHENOMD. The blue and black contours show the 90% credible intervals for SEOBNRV4 and IMRPHENOMD, respectively; the two posteriors differ substantially (i.e., the shift in mean in m_1 is of order one standard deviation), illustrating the impact of model systematics on parameter inference. The green contour shows our model-marginalized posterior. For comparison, the cross shows the injected source parameters, and the model was IMRPHENOMD.

The bottom panel of Fig. 7 shows one PP plot corresponding to applying our model-marginalized procedure to a population where each source is randomly selected from either IMRPHENOMD or SEOBNRV4. The dotted line shows a 90% frequentist interval for the largest of four random cumulative distributions. This figure shows our PP plots are consistent with the diagonal, as desired.

IV. DISCUSSION

In this work, we performed simple tests which reproduce significant differences between the models SEOBNRV4 and IMRPHENOMD, and can be extended to other available waveforms easily using RIFT, an efficient parameter estimation engine. The PP plot test, a commonly used statistical test, can be used to confirm differences between waveform models and, as shown in Fig. 2, parameter estimation performed using a model different from the injected model gives a nondiagonal PP plot for most parameters. We calculated the magnitude and direction of the offsets introduced due to using a waveform model different to the injected model, and these differences are higher for extreme case scenarios, as expected. A linear correlation between the KL divergence computed for the two models and the log of the maximum likelihood of the injected model, shows that high-SNR signal will have larger differences in the inferred parameter from various models. Because the most informative signals exhibit the largest parameter biases, waveform systematics have the potential to strongly contaminate population inference. Most importantly, we also demonstrated a method to mitigate these waveform systematics by marginalizing over the models used for parameter estimation analyses.

Our method requires as input some prior probabilities $p(X_k|\lambda)$ for different waveform models X_k. One way these prior probabilities could be selected is by waveform faithfulness studies between models and numerical relativity simulations. These fidelity studies inevitably suggest waveform models vary in reliability over their parameter space (e.g., [43,44]), suggesting $p(X_k|\lambda)$ will depend nontrivially on λ. Operationally, these model priors propagate into each model's posterior inferences as if parameter inferences for model X are performed using a model-dependent prior $\propto p_{\rm prior}(\lambda)p(X_k|\lambda)$, instead of a common prior for all models. RIFT can seamlessly perform these calculations at minimal added computational expense, while simultaneously returning results for each model derived from the conventional prior alone.

V. CONCLUSIONS

Many waveform models exist currently that describe compact binary coalescences. Even though these are derived by solving Einstein's equations, the various analytical or numerical approximation considered bring in differences and affect the parameter estimation process leading to biased interpretation of results. Averaging over the waveform models can mitigate these biases. Building on prior directly comparable work [20], we have demonstrated an efficient method to perform such model marginalization.

Other techniques have been proposed to marginalize over waveform model systematics. Notably, several groups have proposed using the error estimates provided by their model regressions (e.g., the Gaussian process error) [45]. Relative to regression-based methods, our method has two notable advantages. Our method can be immediately generalized to include multiple waveform models. Critically, we plan to introduce parameter-dependent weighting of the likelihood from a waveform, since different waveforms are accurate in different regimes. No other model-marginalization technique can presently provide this level of control.

ACKNOWLEDGMENTS

R. O. S. gratefully acknowledges support from NSF Grants No. PHY-1707965, No. PHY-2012057, and No. AST-1909534. A. Y. acknowledges support from NSF Grant No. PHY-2012057. The authors are grateful for computational resources provided by the LIGO Laboratories at CIT and LHO supported by National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459.

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 061102 (2016).
[2] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32**, 074001 (2015).
[3] T. Accadia *et al.*, J. Instrum. **7**, P03012 (2012).
[4] F. Acernese *et al.* (Virgo Collaboration), Classical Quantum Gravity **32**, 024001 (2015).
[5] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **125**, 101102 (2020).
[6] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya *et al.* (LIGO Scientific and Virgo Collaborations), Astrophys. J. Lett. **900**, L13 (2020).
[7] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya *et al.* (LIGO Scientific and Virgo Collaborations), Astrophys. J. Lett. **896**, L44 (2020).
[8] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. D **102**, 043015 (2020).
[9] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya *et al.* (LIGO Scientific and Virgo Collaborations), Report No. LIGO-P2000061, 2020, https://dcc.ligo.org/LIGO-P2000061.
[10] F. H. Shaik, J. Lange, S. E. Field, R. O'Shaughnessy, V. Varma, L. E. Kidder, H. P. Pfeiffer, and D. Wysocki, Phys. Rev. D **101**, 124054 (2020).
[11] A. Williamson, J. Lange, R. O'Shaughnessy, J. Clark, P. Kumar, J. Bustillo, and J. Veitch, Phys. Rev. D **96**, 124041 (2017).
[12] M. Prrer and C.-J. Haster, Phys. Rev. Research **2**, 023151 (2020).
[13] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113**, 151101 (2014).
[14] S. Khan, K. Chatziioannou, M. Hannam, and F. Ohme, Phys. Rev. D **100**, 024059 (2019).
[15] A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Pürrer, V. Raymond *et al.*, Phys. Rev. D **95**, 044028 (2017).
[16] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Phys. Rev. Research **1**, 033015 (2019).
[17] G. Pratten, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson *et al.*, arXiv:2004.06503.
[18] S. Ossokine, A. Buonanno, S. Marsat, R. Cotesta, S. Babak, T. Dietrich, R. Haas, I. Hinder, H. P. Pfeiffer, M. Pürrer *et al.*, Phys. Rev. D **102**, 044055 (2020).

[19] D. Wysocki, J. Lange, and R. O'Shaughnessy, Phys. Rev. D **100**, 043012 (2019).
[20] G. Ashton and S. Khan, Phys. Rev. D **101**, 064037 (2020).
[21] J. Veitch *et al.*, Phys. Rev. D **91**, 042003 (2015).
[22] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov *et al.*, Astrophys. J. Suppl. Ser. **241**, 27 (2019).
[23] J. Lange, R. O'Shaughnessy, and M. Rizzo, arXiv:1805.10457.
[24] D. Wysocki, R. O'Shaughnessy, J. Lange, and Y.-L. L. Fang, Phys. Rev. D **99**, 084026 (2019).
[25] J. Lange, RIFT'ing the Wave: Developing and Applying an Algorithm to Infer Properties Gravitational Wave Sources (2020), https://dcc.ligo.org/LIGO-P2000268.
[26] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044006 (2016).
[27] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044007 (2016).
[28] A. Taracchini, Y. Pan, A. Buonanno, E. Barausse, M. Boyle, T. Chu, G. Lovelace, H. P. Pfeiffer, and M. A. Scheel, Phys. Rev. D **86**, 024011 (2012).
[29] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, J. T. Whelan, B. Brügmann, P. Diener, J. Gonzalez, M. Hannam *et al.*, Classical Quantum Gravity **24**, S689 (2007).
[30] L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, N. Dorband, M. Hannam, S. Husa, P. Mösta, D. Pollney, C. Reisswig *et al.*, Phys. Rev. D **82**, 064016 (2010).
[31] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **9**, 031040 (2019).
[32] LIGO Scientific Collaboration, https://dcc.ligo.org/LIGO-T1800044 (2018).
[33] S. Cook, A. Gelman, and D. Rubin, J. Comput. Graph. Stat. **15**, 675 (2006).
[34] L. Lindblom, B. J. Owen, and D. A. Brown, Phys. Rev. D **78**, 124020 (2008).
[35] J. S. Read, C. Markakis, M. Shibata, K. Uryú, J. D. E. Creighton, and J. L. Friedman, Phys. Rev. D **79**, 124033 (2009).
[36] L. Lindblom, J. G. Baker, and B. J. Owen, Phys. Rev. D **82**, 084020 (2010).
[37] H. Cho, E. Ochsner, R. O'Shaughnessy, C. Kim, and C. Lee, Phys. Rev. D **87**, 02400 (2013).
[38] M. Hannam, S. Husa, F. Ohme, and P. Ajith, Phys. Rev. D **82**, 124052 (2010).
[39] P. Kumar, T. Chu, H. Fong, H. P. Pfeiffer, M. Boyle, D. A. Hemberger, L. E. Kidder, M. A. Scheel, and B. Szilagyi, Phys. Rev. D **93**, 104050 (2016).
[40] M. Pürrer and C.-J. Haster, Phys. Rev. Research **2**, 023151 (2020).
[41] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044007 (2016).
[42] C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy, Phys. Rev. D **92**, 023002 (2015).
[43] P. Kumar, I. MacDonald, D. A. Brown, H. P. Pfeiffer, K. Cannon, M. Boyle, L. E. Kidder, A. H. Mroué, M. A. Scheel, B. Szilágyi *et al.*, Phys. Rev. D **89**, 042002 (2014).
[44] P. Kumar, K. Barkett, S. Bhagwat, N. Afshari, D. A. Brown, G. Lovelace, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **92**, 102001 (2015).
[45] A. J. K. Chua, N. Korsakova, C. J. Moore, J. R. Gair, and S. Babak, Phys. Rev. D **101**, 044027 (2020).