Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy

Anonymous Author(s)
Affiliation
Address
email

Abstract

Gravitational wave (GW) detection is now commonplace [3, 4] and as the sensitivity of the global network of GW detectors improves, we will observe $\mathcal{O}(100)$s of transient GW events per year [5]. The current methods used to estimate their source parameters employ optimally sensitive [21] but computationally costly Bayesian inference approaches [28] where typical analyses have taken between 6 hours and 5 days [2]. For binary neutron star (BNS) and neutron star black hole (NSBH) systems prompt counterpart electromagnetic (EM) signatures are expected on timescales of 1 second – 1 minute and the current fastest method for alerting EM follow-up observers [22], can provide estimates in $\mathcal{O}(1)$ minute, on a limited range of key source parameters. Here we show that a conditional variational autoencoder (CVAE) [26, 20] pre-trained on binary black hole (BBH) signals can return Bayesian posterior probability estimates. The training procedure need only be performed once for a given prior parameter space and the resulting trained machine can then generate samples describing the posterior distribution ~ 6 orders of magnitude faster than existing techniques.

1 Introduction

The problem of detecting GWs has largely been solved through the use of template based matched-filtering, a process recently replicated using machine learning techniques [14, 12, 13]. Once a GW has been identified through this process, Bayesian inference, known to be the optimal approach [21], is used to extract information about the source parameters of the detected GW signal.

In the standard Bayesian GW inference approach, we assume a signal and noise model and both may have unknown parameters that we are either interested in inferring or prefer to marginalise away. Each parameter is given a prior astrophysically motivated probability distribution and in the GW case, we typically assume a Gaussian additive noise model (in reality, the data is not truly Gaussian). Given a noisy GW waveform, we would like to find an optimal procedure for inferring some set of the unknown GW parameters. Such a procedure should be able to give us an accurate estimate of the parameters of our observed signal, whilst accounting for the uncertainty arising from the noise in the data.

According to Bayes' Theorem, a posterior probability distribution on a set of parameters, conditional on the measured data, can be represented as

$$p(x|y) \propto p(y|x)p(x), \tag{1}$$

where x are the parameters, y is the observed data, $p(x|y)$ is the posterior, $p(y|x)$ is the likelihood, and $p(x)$ is the prior on the parameters. The constant of proportionality, which we omit here, is $p(y)$,

Submitted to 34th Conference on Neural Information Processing Systems (NeurIPS 2020). Do not distribute.

the probability of our data, known as the Bayesian evidence or the marginal likelihood. We typically ignore $p(y)$ since it is a constant and for parameter estimation purposes we are only interested in the shape of the posterior.

Due to the size of the parameter space typically encountered in GW parameter estimation and the volume of data analysed, we must stochastically sample the parameter space in order to estimate the posterior. Sampling is done using a variety of techniques including Nested Sampling [23, 27, 25] and Markov chain Monte Carlo methods [11, 29]. The primary software tools used by the advanced Laser Interferometer Gravitational wave Observatory (LIGO) parameter estimation analysis are `LALInference` and `Bilby` [28, 8], which offer multiple sampling methods.

Machine learning has featured prominently in many areas of GW research over the last few years. These techniques have shown to be particularly promising in signal detection [14, 12, 13], glitch classification [31], earthquake prediction [9], and to augment existing Bayesian sampling methods [15]. These methods, including the one presented in this paper, are known as "likelihood-free" approaches in which there is no requirement for explicit likelihood evaluation [10], only the need to sample from the likelihood. Nor is it the case that pre-computed posterior distributions are required in the training procedure.

Recently, a type of neural network known as CVAE was shown to perform exceptionally well when applied towards computational imaging inference [26, 24], text to image inference [30], high-resolution synthetic image generation [19] and the fitting of incomplete heterogeneous data [18]. CVAEs, as part of the variational family of inference techniques are ideally suited to the problem of function approximation and have the potential to be significantly faster than existing approaches. It is therefore this type of machine learning network that we apply in the GW case to accurately approximate the Bayesian posterior $p(x|y)$, where x represents the physical parameters that govern the GW signal, and are the quantities we are interested in inferring. The data y represents the noisy measurement containing the GW signal and obtained from a network of GW detectors.

2 Methods

The construction of a CVAE begins with the definition of a quantity to be minimised (referred to as a cost function). In our case we use the cross entropy, defined as

$$H(p, r) = -\int dx\, p(x|y) \log r_\theta(x|y) \tag{2}$$

between the true posterior $p(x|y)$ and $r_\theta(x|y)$, the parametric distribution that we will use neural networks to model and which we aim to be equal to the true posterior. The parametric model is constructed from a combination of 2 (encoder and decoder) neural networks $r_{\theta_1}(z|y)$ and $r_{\theta_2}(x|y, z)$ where

$$r_\theta(x|y) = \int dz\, r_{\theta_1}(z|y) r_{\theta_2}(x|y, z). \tag{3}$$

In this case the θ subscripts represent sets of trainable neural network parameters and the variable z represents locations within a *latent space*. This latter object is typically a lower dimensional space within which an encoder can represent the input data, and via marginalisation allows the construction of a rich family of possible probability densities.

Starting from Eq. 2 it is possible to derive a computable bound for the cross-entropy that is reliant on the r_{θ_1} and r_{θ_2} networks and a third "recognition" encoder network $q_\phi(z|x, y)$ governed by the trainable parameter-set ϕ. The details of the derivation are described in the methods section and in [26] but equate to an optimisation of the evidence lower bound (ELBO). The final form of the cross-entropy cost function is given by the bound

$$H \lesssim \frac{1}{N} \sum_{n=1}^{N_b} \Big[\overbrace{-\log r_{\theta_2}(x_n|z_n, y_n)}^{L} + \overbrace{\mathrm{KL}\left[q_\phi(z|x_n, y_n)||r_{\theta_1}(z|y_n)\right]}^{\mathrm{KL}} \Big]. \tag{4}$$

The cost function is composed of 2 terms, the "reconstruction" cost L which is a measure of how well the decoder network r_{θ_2} predicts the true signal parameters x, and the Kullback–Leibler (KL)-divergence cost that measures the similarity between the distributions modelled by the r_{θ_1} and q_ϕ encoder networks. In practice, for each iteration of the training procedure, the integrations over x, y and z are approximated by a sum over a batch of N_b draws from the user defined prior $p(x)$, the known likelihood $p(y|x)$, and the recognition function $q_\phi(z|, x, y)$. Details of the training procedure are further explained in [7].

The implementation of the CVAE that we employ has a number of specific features that were included in order to tailor the analysis to GW signals. The details of these enhancements are described in [7] but in summary, the primary modifications are as follows, 1) Physically appropriate output decoder distributions are used for each output parameter: von Mises-Fisher distribution on the sky location parameters, von Mises distributions on periodic parameters, conditional truncated Gaussians for the component masses, and truncated Gaussians for parameters with defined prior bounds. 2) Each of the functions $r_{\theta_1}, r_{\theta_2}$, and q_ϕ are modelled using deep convolutional neural networks with multi-detector time-series represented as independent input channels. 3) The r_{θ_1} encoder models an $M = 16$ component Gaussian mixture model within the $n_z = 10$ dimensional latent space in order to capture the corresponding typical multi-modal nature of GW posterior distributions.

3 Results

We present results on 256 multi-detector GW test BBH waveforms in simulated advanced detector noise [1] from the LIGO Hanford, Livingston and Virgo detectors. We compare between variants of the existing Bayesian approaches and our CVAE implementation which we call `VItamin`. Posteriors produced by the `Bilby` inference library [8] are used as a benchmark in order to assess the efficiency and quality of our machine learning approach with the existing methods for posterior sampling.

For the benchmark analysis we assume that 9 parameters are unknown[1]: the component masses m_1, m_2, the luminosity distance d_L, the sky position α, δ, the binary inclination Θ_{jn}, the GW polarisation angle ψ, the time of coalescence t_0, and the phase at coalescence ϕ_0. For each parameter we use a uniform prior with the exception of the declination and inclination parameters for which we use priors uniform in $\cos\delta$ and $\sin\Theta_{jn}$ respectively. We use a sampling frequency of 256 Hz, a time-series duration of 1 second, and the waveform model used is `IMRPhenomPv2` [16] with a minimum cutoff frequency of 20Hz. For each input test waveform we run the benchmark analysis using multiple sampling algorithms available within `Bilby`. For each run and sampler we extract $\mathcal{O}(10^4)$ samples from the posterior on the 9 physical parameters.

The `VItamin` training process uses as input 10^7 whitened waveforms corresponding to parameters drawn from the same priors as assumed for the benchmark analysis. The waveforms are also of identical duration, sampling frequency, and use the same waveform model as in the benchmark analysis. The signals are whitened[2] using the same advanced detector PSDs [1] as assumed in the benchmark analysis. When each whitened waveform is placed within a training batch it is given a unique detector Gaussian noise realisation (after signal whitening this is simply zero mean, unit variance Gaussian noise). The `VItamin` posterior results are produced by passing each of our 256 whitened noisy testing set of GW waveforms as input into the testing path of the pre-trained CVAE. For each input waveform we sample until we have generated 10^4 posterior samples on 7 physical parameters $x = (m_1, m_2, d_\mathrm{L}, t_0, \Theta_{jn}, \alpha, \delta)$. We choose to output a subset of the full 9-dimensional space to demonstrate that parameters (such as ϕ_0 and ψ in this case) can (if desired) be marginalised out within the CVAE procedure itself, rather than after training.

We can immediately illustrate the accuracy of our machine learning predictions by directly plotting 2 and one-dimensional marginalised posteriors generated using the output samples from our `VItamin` and `Bilby` approaches superimposed on each other. We show this for one example test dataset in Fig. 1 where strong agreement between 2 `Bilby` samplers (`Dynesty` in blue, and `ptemcee` in green) and the CVAE (red) is clear. It is also evident that whilst we refer to the `Bilby` sampler results as benchmark cases, different existing samplers do not perfectly agree with each other. For each of

[1] Our analysis omits the 6 additional parameters required to model the spin of each BBH component mass.

[2] The whitening is used primarily to scale the input to a magnitude range more suitable to neural networks. The *true* power spectral density (PSD) does not have to be used for whitening, but training data and test data must be contain signals that share the same PSD.

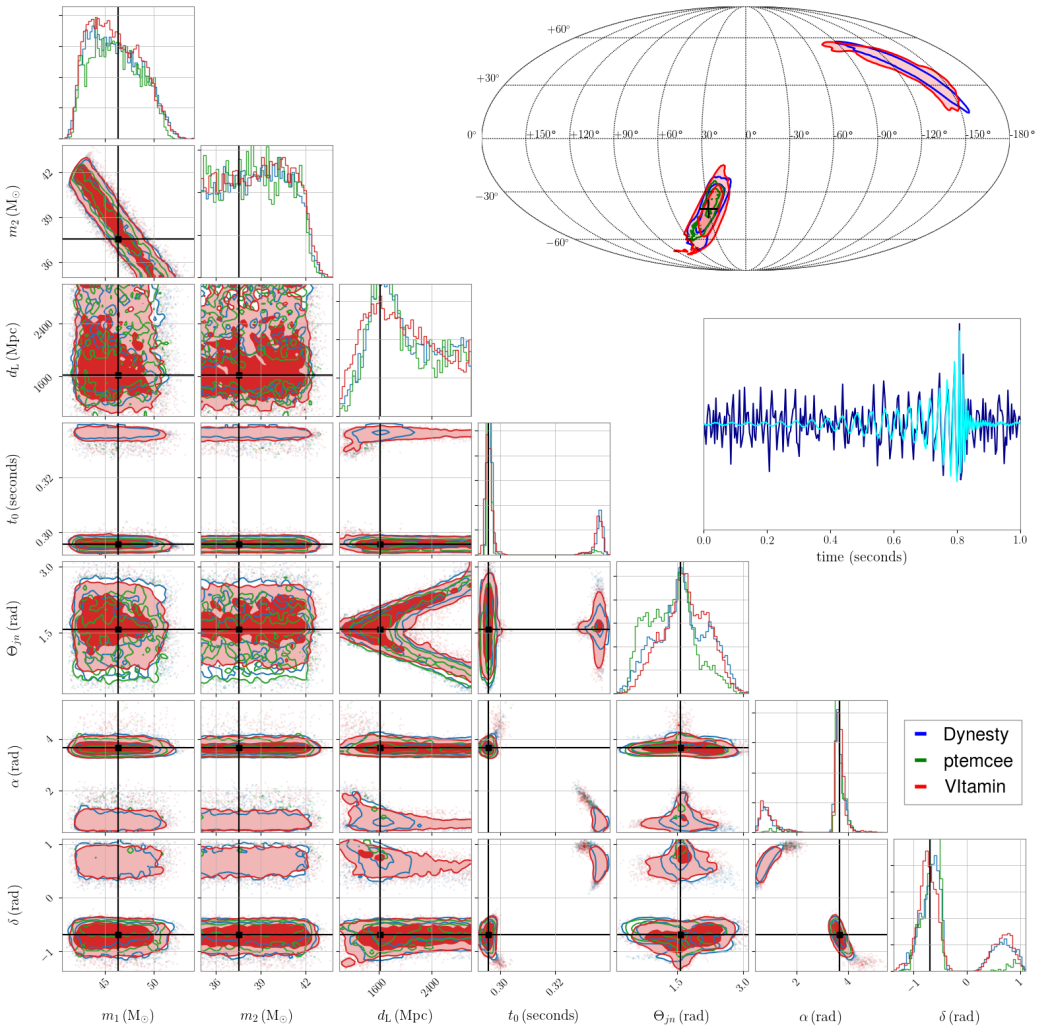

Figure 1: Corner plot showing one and two-dimensional marginalised posterior distributions on the GW parameters for one example test dataset. Filled red contours represent the two-dimensional joint posteriors obtained from `VItamin` and solid blue and green contours are the corresponding posteriors output from our benchmark analyses (using the `Dynesty` and `ptemcee` samplers within `Bilby`). In each case, the contour boundaries enclose $68, 90$ and 95% probability. One dimensional histograms of the posterior distribution for each parameter from both methods are plotted along the diagonal. Black vertical and horizontal lines denote the true parameter values of the simulated signal. At the top of the figure we include a Mollweide projection of the sky location posteriors from all three analyses. All results presented in this letter correspond to a three-detector configuration but for clarity we only plot the H1 whitened noisy time-series y and the noise-free whitened signal (in blue and cyan respectively) to the right of the figure. The test signal was simulated with an optimal multi-detector signal-to-noise ratio of 17.2.

our 256 test cases we see equivalent levels of disparity between pairs of benchmark samplers *and* between any benchmark sampler and our CVAE results.

In [7] we also show the results of 2 statistical tests (the probability-probability (p-p) plot test and KL-divergence tests) performed on the entire test dataset and between all samplers (`Dynesty`, `ptemcee`, `CPNest`, `emcee`, and `VItamin`). In both tests the quality of the `VItamin` results are indistinguishable from the benchmark samplers. The p-p plot results specifically indicate that the Bayesian one-dimensional marginalised posteriors from each approach are self-consistent from a frequentist perspective (e.g., the true values lie within the $X\%$ confidence interval for $X\%$ of the test cases). The second test computes the distribution of KL-divergences between posteriors conditioned on the same test data y from pairs of samplers. In all cases this measure of "distribution similarity" between `VItamin` and any particular benchmark sampler is entirely consistent with the distribution between that benchmark sampler and any other.

The dominating computational cost of running `VItamin` lies in the training time, which takes $\mathcal{O}(1)$ day to complete. We stress that once trained, there is no need to retrain the network unless the user wishes to use different priors $p(x)$ or assume different noise characteristics. Run-time for the benchmark samplers is defined as the time to complete their analyses when configured using the parameter choices. For `VItamin`, this time is defined as the total time to produce 10^4 samples. For our test case of BBH signals `VItamin` produces samples from the posterior at a rate which is ~ 6 orders of magnitude faster than our benchmark analyses using current inference techniques.

4 Conclusions

In this letter we have demonstrated that we are able to reproduce, to a high degree of accuracy, Bayesian posterior probability distributions generated through machine learning. This is accomplished using a CVAE trained on simulated GW signals and does not require the input of precomputed posterior estimates. We have demonstrated that our neural network model, which when trained, can reproduce complete and accurate posterior estimates in a fraction of a second, achieves the same quality of results as the trusted benchmark analyses used within the LIGO-Virgo Collaboration.

The significance of our results is most evident in the orders of magnitude increase in speed over existing algorithms. We have demonstrated the approach using BBH signals but with additional work to increase sample rate and signal duration, the method can also be extended for application to signals from BNS mergers (e.g., GW170817 [4], and GW190425 [6]) and NSBH systems where improved low-latency alerts will be especially pertinent. By using our approach, parameter estimation speed will no longer be limiting factor[3] in observing the prompt EM emission expected on shorter time scales than is achievable with existing LIGO-Virgo Collaboration (LVC) analysis tools such as Bayestar [22].

The predicted number of future detections of BNS mergers (~ 180 [5]) will severely strain the GW community's current computational resources using existing Bayesian methods. We anticipate that future iterations of our approach will provide full-parameter estimation on all classes of compact binary coalescence (CBC) signals in $\mathcal{O}(1)$ second on single graphics processing units (GPUs). Our trained network is also modular, and can be shared and used easily by any user to produce results. The specific analysis described in this letter assumes a uniform prior on the signal parameters. However, this is a choice and the network can be trained with any prior the user demands, or users can cheaply resample accordingly from the output of the network trained on the uniform prior. We also note that our method will be invaluable for population studies since populations may now be generated and analysed in a fully-Bayesian manner on a vastly reduced time scale. Our work can naturally be extended to include the full range of CBC signal types but also to any and all other parameterised GW signals and to analyses of GW data beyond that of ground based experiments. Given the abundant benefits of this method, we hope that a variant of this of approach will form the basis for future GW parameter estimation.

[3] A complete low-latency pipeline includes a number of steps. The process of GW data acquisition is followed by the transfer of data. There is then the corresponding candidate event identification, parameter estimation analysis, and the subsequent communication of results to the EM astronomy community after which there are physical aspects such as slewing observing instruments to the correct pointing.

Broader Impact

References

[1] Advanced LIGO sensitivity design curve. https://dcc.ligo.org/LIGO-T1800044/public. Accessed: 2019-06-01.

[2] Gracedb — gravitational-wave candidate event database (ligo/virgo o3 public alerts). https://gracedb.ligo.org/superevents/public/O3/. Accessed: 2019-09-16.

[3] B. P. Abbott et al. Binary black hole mergers in the first advanced LIGO observing run. *Phys. Rev. X*, 6:041015, Oct 2016.

[4] B. P. Abbott et al. Gw170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017.

[5] B. P. Abbott et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, 21(1):3, Apr 2018.

[6] B. P. Abbott et al. GW190425: Observation of a Compact Binary Coalescence with Total Mass 3.4 M. *Astrophysical Journal Letters*, 892(1):L3, Mar. 2020.

[7] Anonymous et al. Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy, 2019.

[8] G. Ashton, M. Huebner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divarkala, P. J. Easter, B. Goncharov, F. H. Vivanco, J. Harms, M. E. Lower, G. D. Meadors, D. Melchor, E. Payne, M. D. Pitkin, J. Powell, N. Sarin, R. J. E. Smith, and E. Thrane. Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *Astrophysical Journal Supplement Series*, 2018.

[9] M. Coughlin, P. Earle, J. Harms, S. Biscans, C. Buchanan, E. Coughlin, F. Donovan, J. Fee, H. Gabbard, M. Guy, N. Mukund, and M. Perry. Limiting the effects of earthquakes on gravitational-wave interferometers. *Classical and Quantum Gravity*, 34(4):044004, feb 2017.

[10] K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 2020.

[11] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The mcmc hammer. *PASP*, 125:306–312, 2013.

[12] H. Gabbard, M. Williams, F. Hayes, and C. Messenger. Matching matched filtering with deep networks for gravitational-wave astronomy. *Phys. Rev. Lett.*, 120:141103, Apr 2018.

[13] T. Gebhard, N. Kilbertus, G. Parascandolo, I. Harry, and B. Schölkopf. Convwave: Searching for gravitational waves with fully convolutional neural nets. In *Workshop on Deep Learning for Physical Sciences (DLPS) at the 31st Conference on Neural Information Processing Systems (NIPS)*, 2017.

[14] D. George and E. Huerta. Deep learning for real-time gravitational wave detection and parameter estimation: Results with advanced ligo data. *Physics Letters B*, 778:64 – 70, 2018.

[15] P. Graff, F. Feroz, M. P. Hobson, and A. Lasenby. BAMBI: blind accelerated multimodal Bayesian inference. *Monthly Notices of the Royal Astronomical Society*, 421(1):169–180, Mar. 2012.

[16] S. Khan, K. Chatziioannou, M. Hannam, and F. Ohme. Phenomenological model for the gravitational-wave signal from precessing binary black holes with two-spin effects, 2018.

[17] T. B. Littenberg and N. J. Cornish. Bayesian inference for spectral estimation of gravitational wave detector noise. *Physical Review D*, 91(8):084034, Apr 2015.

[18] A. Nazabal, P. M. Olmos, Z. Ghahramani, and I. Valera. Handling incomplete heterogeneous data using VAEs, 2018.

[19] A. Nguyen, J. Clune, Y. Bengio, A. Dosovitskiy, and J. Yosinski. Plug and play generative networks: Conditional iterative generation of images in latent space, 2016.

[20] A. Pagnoni, K. Liu, and S. Li. Conditional variational autoencoder for neural machine translation, 2018.

[21] A. C. Searle, P. J. Sutton, and M. Tinto. Bayesian detection of unmodeled bursts of gravitational waves. *Classical and Quantum Gravity*, 26(15):155017, Aug 2009.

[22] L. P. Singer and L. R. Price. Rapid Bayesian position reconstruction for gravitational-wave transients. *Physical Review D*, 93(2):024013, Jan 2016.

[23] J. Skilling. Nested sampling for general bayesian computation. *Bayesian Anal.*, 1(4):833–859, 12 2006.

[24] K. Sohn, H. Lee, and X. Yan. Learning structured output representation using deep conditional generative models. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 3483–3491. Curran Associates, Inc., 2015.

[25] J. S. Speagle. dynesty: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, 2019.

[26] F. Tonolini, A. Lyons, P. Caramazza, D. Faccio, and R. Murray-Smith. Variational inference for computational imaging inverse problems, 2019. To appear in JMLR.

[27] J. Veitch, W. D. Pozzo, C. Messick, and M. Pitkin. Cpnest. Jul 2017.

[28] J. Veitch, V. Raymond, B. Farr, W. M. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. D. Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. V. D. Sluys, A. Vecchio, W. Vousden, and L. Wade. Robust parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library. *Physical Review D*, 2014.

[29] W. Vousden, W. M. Farr, and I. Mandel. Dynamic temperature selection for parallel-tempering in Markov chain Monte Carlo simulations. 2015.

[30] X. Yan, J. Yang, K. Sohn, and H. Lee. Attribute2image: Conditional image generation from visual attributes, 2015.

[31] M. Zevin, S. Coughlin, S. Bahaadini, E. Besler, N. Rohani, S. Allen, M. Cabero, K. Crowston, A. K. Katsaggelos, S. L. Larson, T. K. Lee, C. Lintott, T. B. Littenberg, A. Lundgren, C. Østerlund, J. R. Smith, L. Trouille, and V. Kalogera. Gravity spy: integrating advanced ligo detector characterization, machine learning, and citizen science. *Classical and Quantum Gravity*, 34(6):064003, 2017.