Revisiting the evidence for precession in GW200129 with machine learning noise mitigation

Ronaldas Macas and Andrew Lundgren*
University of Portsmouth, Portsmouth, PO1 3FX, United Kingdom

Gregory Ashton
Department of Physics, Royal Holloway, University of London, Egham TW20 0EX, United Kingdom

(Received 23 November 2023; accepted 28 February 2024; published 18 March 2024)

GW200129 is claimed to be the first-ever observation of the spin-disk orbital precession detected with gravitational waves (GWs) from an individual binary system. However, this claim warrants a cautious evaluation because the GW event coincided with a broadband noise disturbance in LIGO Livingston caused by the 45 MHz electro-optic modulator system. In this paper, we present a state-of-the-art neural network that is able to model and mitigate the broadband noise from the LIGO Livingston interferometer. We also demonstrate that our neural network mitigates the noise better than the algorithm used by the LIGO-Virgo-KAGRA Collaboration. Finally, we reanalyze GW200129 with the improved data quality compared to the data used by the LIGO-Virgo-KAGRA Collaboration and show that the evidence for precession is still observed.

DOI: 10.1103/PhysRevD.109.062006

I. INTRODUCTION

The gravitational-wave (GW) era started in 2015 when a binary black hole (BBH) merger was detected [1]. Since then, the LIGO-Virgo-KAGRA (LVK) Collaboration observed binary neutron star and neutron star–black hole systems, as well as many other BBH mergers [2–4].

GW analyses often assume that the noise is Gaussian. While this is true in many cases, about 24% of the GW candidates from the third LVK observing run (O3) were flagged as having a non-Gaussian noise nearby [3,4]. If the non-Gaussian noise is not accounted for, GW searches and source parameter estimation can be affected, including the sky localization that is used for the electromagnetic follow-up efforts [5–9].

One of the GW candidate events from O3, GW200129, attracted interest because it is claimed to be the first-ever individual BBH that has strong evidence for spin-induced orbital precession [10]. Initial analysis performed by the LVK Collaboration with two waveform models found evidence for precession when using IMRPhenomXPHM but not with SEOBNRv4 [4]. Reference [4] gave equal weight to both analyses, thus leaving the properties of GW200129 unclear. In addition, Nitz *et al.* [11] found GW200129 to be precessing when analyzed with the IMRPhenomX waveform model [12]. Hannam *et al.* [10] put forward an analysis based on the NRSurrogate waveform model that is more faithful to numerical simulations [13] and found the orbital precession of GW200129 to be 10 orders of magnitude higher than any previous weak-field measurement.

The precession claim warrants a cautious evaluation because GW200129 coincided with excess noise at the Livingston interferometer caused by the 45 MHz electro-optic modulator system [14,15]. Luckily, this noise was also recorded by multiple witness channels, namely LSC-POP_A_RF45_I_ERR_DQ, LSC-POP_A_RF45_Q_ERR_DQ, and LSC-POP_A_RF9_I_ERR_DQ, which we refer to as RF45-I, RF45-Q, and RF9-I, respectively.

The LIGO-Virgo-KAGRA Collaboration used `gwsubtract` to model and mitigate the excess noise in the GW strain channel h(t) [4]. The `gwsubtract` algorithm was first developed to remove the "jitter" noise from LIGO Hanford interferometer during the second LIGO-Virgo observing run, which resulted in 30% improvement in LIGO Hanford sensitivity [16]. The algorithm was similarly applied for noise mitigation around GW200129 for the Livingston data: it estimated the coupling between the RF9-I and h(t) channels and removed the predicted excess noise from h(t). This cleaned version of the Livingston data was then used in the LVK and Ref. [10] analyses.

The implementation of `gwsubtract` for GW200129 and the algorithm itself has multiple shortcomings. Firstly, the derivation of `gwsubtract` relies on the assumptions that the data is Gaussian and stationary [16], while the excess power in the Livingston data is non-Gaussian by definition. Furthermore, the algorithm assumes that the noise coupling must be linear which is not necessarily the

*Andrew.Lundgren@port.ac.uk

case for GW200129. Finally, the LVK Collaboration used only a single channel, RF9-I, to estimate the noise coupling, even though RF45-I and RF45-Q also recorded the noise.

Additional studies suggest that the evidence for precession using the `gwsubtract` data is robust when using a NRSurogate waveform model [17,18]. However, it is unclear if the `gwsubtract` data are systematically biased, which is the reason why Payne *et al.* [18] also used an alternative method to clean the Livingston data. Reference [18] mitigates the excess noise with BayesWave, an algorithm that does not use witness channels and relies only on the strain data to model the noise [19,20]. Their analysis found that the evidence for precession is smaller than the statistical and systematic uncertainty of noise mitigation.

In this paper, we describe a novel method to mitigate the excess noise using machine learning, with a focus on removing the noise around GW200129. In Sec. II, we describe our data selection procedure, the machine learning model, and how the effectiveness of the model is tested, as well as the procedure to estimate the orbital precession for GW200129. In Sec. III, we show how well the machine learning model works, compare its effectiveness against `gwsubtract`, and present results of the spin-induced orbital precession estimate. We summarize our findings in Sec. IV.

II. METHODS

A. Data

The Livingston interferometer was locked and observing for more than five hours when GW200129 was detected and continued to observe for more than 38 hr. During this time, the Livingston interferometer had excess noise caused by the 45 MHz electro-optic modulator system; this noise is also known as the radio frequency (rf) noise [15]. Radio frequency noise coupled to the GW strain channel h(t), as well as the noise witness channels RF45-I, RF45-Q, and RF9-I (Fig. 1).

To learn the coupling between the witness channels and the GW strain channel h(t), we selected 27 hr of the total 43 hr when the detector was locked and observing. To be exact, we chose the interval from January 29, 2020 09:55:06 UTC to January 30, 2020 12:55:06 UTC.

We chose this interval for multiple reasons. First, the rf noise was exceptionally recurrent during this time. In addition, the noise coupling may change if the interferometer configuration changes; therefore, we selected only the data contained in this observational lock. Finally, we did not use the time around GW200129 to avoid any biases.

After data selection, we whitened the data using the inner product equation, where $\langle a|b\rangle$ is the inner product between the two time series $a(t)$ and $b(t)$ given by

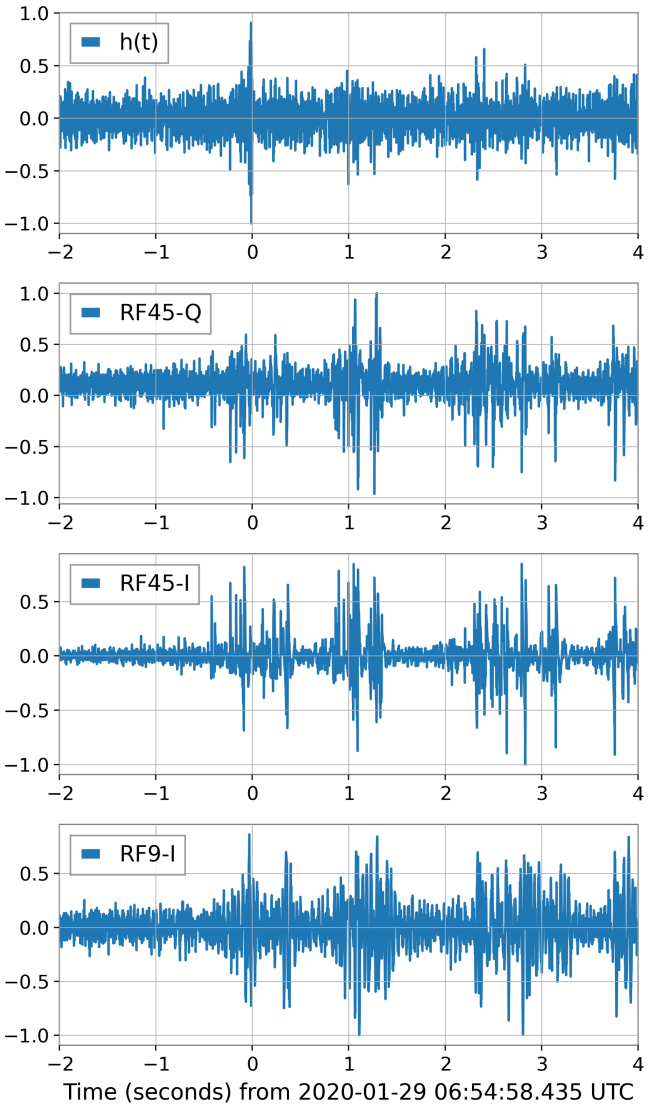


FIG. 1. Whitened time series data for the gravitational-wave strain channel h(t) and witness channels RF45-Q, RF45-I, and RF9-I. Gravitational-wave signal GW200129 is clearly seen in h(t) at 0 s. The non-Gaussian noise in h(t) caused by the electro-optic modulator system is also recorded by the witness channels.

$$\langle a|b\rangle = 4\int_0^\infty \frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}df, \tag{1}$$

where $S_n(f)$ denotes the power spectral density (PSD). To avoid using PSD biased by the rf noise, we inspected the data visually and selected 512 s of relatively quiet data for the PSD estimation.

Once the data were whitened, it was resampled to 512 Hz. Because our network memory requirement is proportional to f^3_{Nyquist}, data resampling from 4096 to 512 Hz significantly reduced the required memory to train the algorithm. Since the orbital precession for GW200129 is mostly observed below 50 Hz [18], reducing the sampling rate does not impact the precession measurement.

After resampling the data, we selected the time intervals when the rf noise was present in the witness channels. To identify these intervals, the Z score of 11.7 was used for the RF45-I channel (there was no particular reason why this specific channel was selected out of the three witness channels). If a data point passed this threshold, we would select 4 s around this data point.

We collated 477 noisy data intervals (about 32 min) after this procedure. Using ∼32 min of noise-only data instead of the full 27 hr of data allowed us to train the machine learning algorithm more efficiently.

B. Dense neural network

We use a dense neural network (DNN) to model the excess noise witnessed by the rf channels. DNNs are well known for their ability to generalize even extremely complicated relationships in data, including the nonlinear coupling that can happen among multiple noise witness channels. In addition, a DNN does not rely on the assumption that the data need to be Gaussian and stationary like the linear subtraction method `gwsubtract`.

We found that a relatively simple fully connected DNN with just two hidden layers and a dropout layer in between is enough to model the rf noise around GW200129 (see Table I for full network architecture). To predict a single noise point, i.e., $\frac{1}{512}$ s, in h(t), we use 2 s of data from each rf channel.

In order to reduce the overfitting and increase the learning efficiency, we use L2 kernel regularization of 10^{-4} and rectified linear unit activation function. The dataset from Sec. II A is split with 9∶1 ratio for testing and validating the algorithm. The network is trained for 500 epochs using Adam optimizer [21] with inverse time decay learning function, learning rate of 10^{-3}, and mean squared error loss function. Training the network written with TensorFlow [22] takes about 30 min on Nvidia A100 80 GB GPU.

After the training is performed, the model can be applied to mitigate the noise around GW200129. To do that, we perform the same steps as in Sec. II A for 4096 s around GW200129, i.e., the data are first whitened and resampled.

TABLE I. Dense neural network architecture. We use 2 s of data with $f_{\text{Nyquist}} = 256$ Hz from each rf channel to predict a single data point, i.e., $\frac{1}{512}$ s, in the gravitational-wave strain channel h(t).

Layer	Output shape (length, dimensions)
Input	(1024, 3)
Flatten	(3072, 1)
Dense	(1024, 1)
Dropout	(1024, 1)
Dense	(1024, 1)
Output	(1, 1)

After that, the three rf channels are used as input to get the rf noise contribution in h(t) estimated by the DNN. Once that is done, the data are up-sampled to 4096 Hz, dewhitened, and subtracted from the original 4096 s LIGO Livingston frame. This is the data frame that has rf noise mitigated using our machine learning algorithm. The data frame is publicly available on Zenodo [23].

C. Testing the cleaned data

We apply the method proposed by Macas and Lundgren [24] to estimate how much of the excess noise is removed using our algorithm and compare its effectiveness with `gwsubtract`. To do that, we first identify time intervals when the rf noise is present in the data around GW200129. Using the Z score of 6 for the RF45-I channel, we select 366 s out of 4096 s considered to contain the radio frequency noise.

We convert these data into the normalized time-frequency tiles with quality factor $Q = 8$, which allows us to estimate the average tile power. Then, using Bayesian statistical modeling, we calculate the amount of power in the non-Gaussian data versus the total power, defined as "fractional power." We repeat the same procedure for the original Livingston data, as well as the data that have the rf noise mitigated with `gwsubtract`.

D. Estimating the spin-induced orbital precession of GW200129

We finally apply the Bilby Bayesian inference package [25,26] to estimate the posterior probability density using our new cleaned Livingston data and the original data from Hanford and Virgo. We follow the configuration of the initial LVK gravitational-wave transient catalog 3 (GWTC-3) analysis [4], but use the NRSur7dq4 waveform approximant [13] with a suitably modified prior following Hannam *et al.* [10]. We also marginalize over the calibration uncertainty. However, the PSD used by the GWTC-3 analysis used an on-source estimate that depends on the analysis data. Since we have cleaned the Livingston data, we opt instead to use a Welch PSD estimate for all three detectors. Finally, we perform sampling using the Bilby Markov chain Monte Carlo (Bilby-MCMC) sampler [25].

III. RESULTS AND DISCUSSION

A. Data cleaning

Figure 2 shows the spectrograms of the Livingston data around GW200129 before and after cleaning and the difference between these two spectrograms. We note that the algorithm removed large portions of the excess noise after the GW signal, as well as some noise that happened during the GW signal and within the frequency range of the GW signal. Comparing with the `gwsubtract` results [14], it is clear that our algorithm removes noticeably more data at frequencies above 50 Hz.

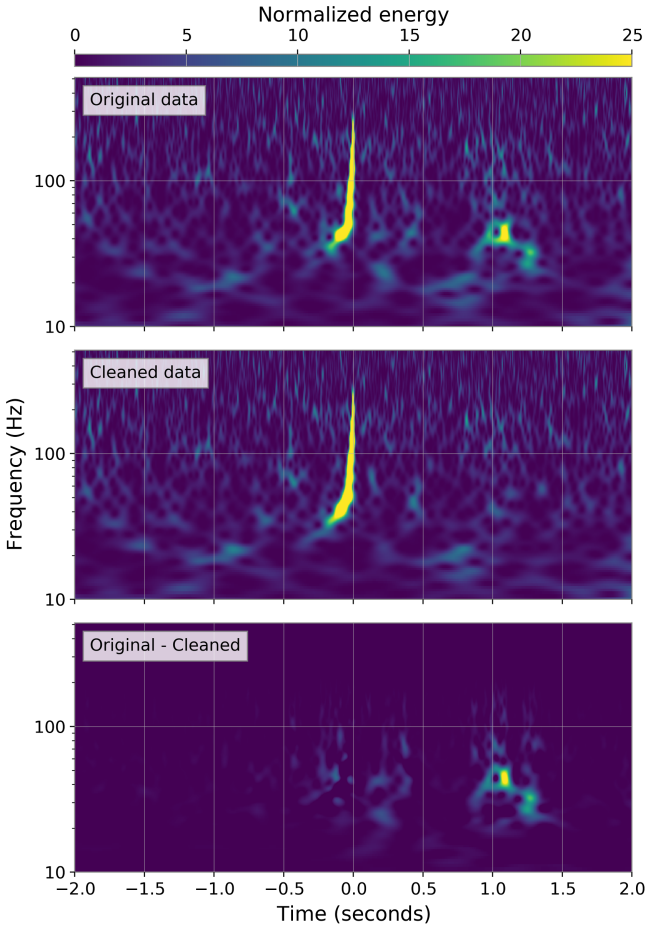


FIG. 2. Spectrograms of the original Livingston data around GW200129, cleaned Livingston data using our machine learning approach, and the difference between these two spectrograms. Our algorithm removes the non-Gaussian noise after the GW signal, as well as some noise that happened during the GW signal and within the frequency range of the GW signal.

As we can see from Fig. 3, our modeled radio frequency noise is closer to the `gwsubtract` noise prediction than the prediction made by BayesWave (glitch model A) in Payne *et al.* [18]. This makes sense since both the `gwsubtract` and machine learning (ML) approach use witness channel information. However, ML subtraction contains more high-frequency features than `gwsubtract`, which can be explained by the fact that ML uses more witness channels and/or the algorithm represents the nonlinearity better. Comparing BayesWave's noise model with `gwsubtract` and ML, it looks much simpler with a peak around −0.08 s before the merger.

Close to the merger time, we can see some differences between all three models. Around −0.025 s before the merger, BayesWave subtracts more data than `gwsubtract` and ML. Later, around −0.015 s before the merger, `gwsubtract` subtracts more data than BayesWave, but less data than ML.

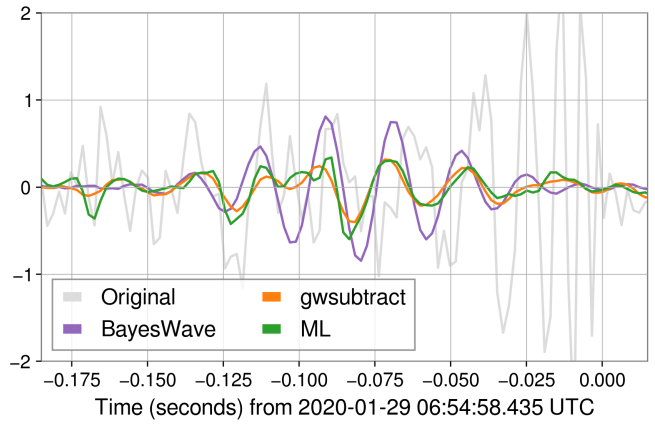


FIG. 3. Radio frequency noise modeled by BayesWave (purple), `gwsubtract` (orange), and our algorithm (green); the original unmitigated data are in gray. All data were whitened and bandpassed to 512 Hz.

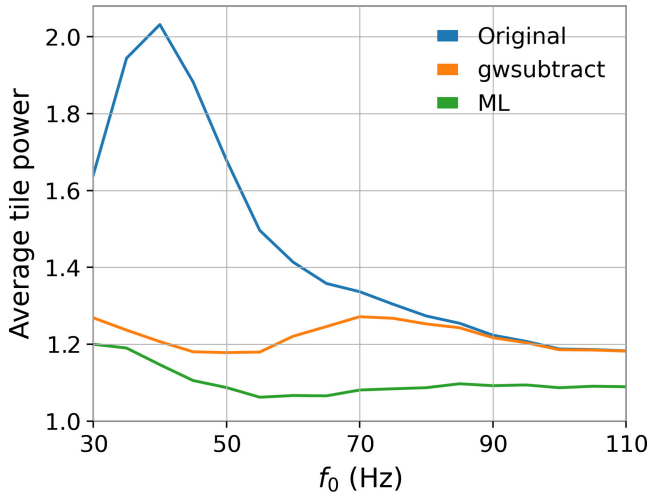
For a more quantitative comparison between the original, `gwsubtract`, and ML data frames, see Fig. 4. We omit results of Payne *et al.* [18] in this comparison because the Macas and Lundgren [24] comparison method requires significantly more data than the Ref. [18] cleaned.

The average tile power plot [Fig. 4(a)] shows that both `gwsubtract` and ML algorithms significantly reduce the excess power. For Gaussian data, the average tile power is 1, while the original data frame has as much as twice the amount of power than Gaussian data around 40 Hz.

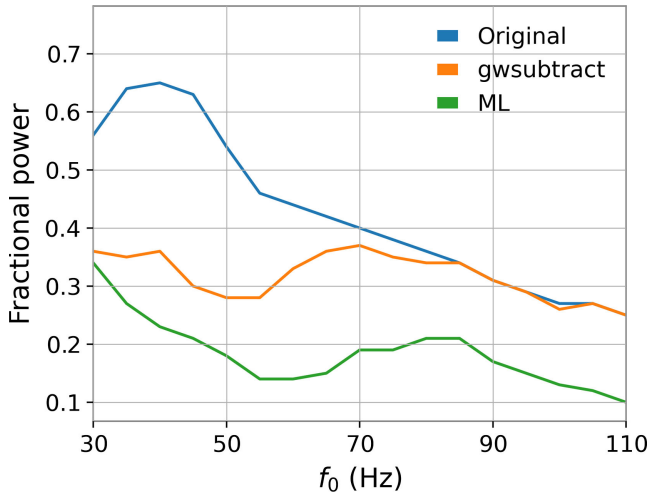
The ML algorithm removes more excess power than `gwsubtract` across all frequency ranges, and the difference becomes much higher at frequencies above ∼50 Hz. From ∼85 Hz, `gwsubtract` does not remove any excess noise anymore. This happens because `gwsubtract` uses only a single witness channel that does not contain any correlation with h(t) at those frequencies.

The fractional power plot [Fig. 4(b)] shows similar conclusions. The ML algorithm is better at removing the excess noise across all frequencies, reducing up to 10% more of the fractional power at frequencies below 50 Hz. At higher frequencies, the difference in fractional power between `gwsubtract` and ML can be as high as 20%.

Unfortunately, the average tile power and fractional power plots [Figs. 4(a) and 4(b)] indicate that neither the `gwsubtract` nor the ML algorithm remove the excess noise completely. This can happen due to a variety of reasons. First, it is possible that both algorithms cannot estimate the correlation function between the witness channels and the gravitational-wave strain ideally. Another possibility is that none of the witness channels are perfect, i.e., they do not record the full coherence between the rf noise in a witness channel and the rf noise in h(t).



(a)



(b)

FIG. 4. Average tile power (a) and fractional power (b) for data around GW200129 with quality factor $Q = 8$. Both tests indicate that our machine learning algorithm removes more non-Gaussian noise than `gwsubtract`, especially at frequencies above 50 Hz.

B. Spin-induced orbital precession of GW200129

In Fig. 5, we show the posterior distribution of the spin magnitude and tilt from our new analysis utilizing the cleaned data. Comparing with Fig. 2 of Hannam *et al.* [10], which used the `gwsubtract` cleaned frame, we see similar results within the expected sampling error (any subtle changes would most likely be caused by the difference in PSD construction). From this figure, we can confirm that the identification of evidence for precession, i.e., that the primary spin is constrained to be close to unity and highly misaligned, is robust to our new deglitching routine.

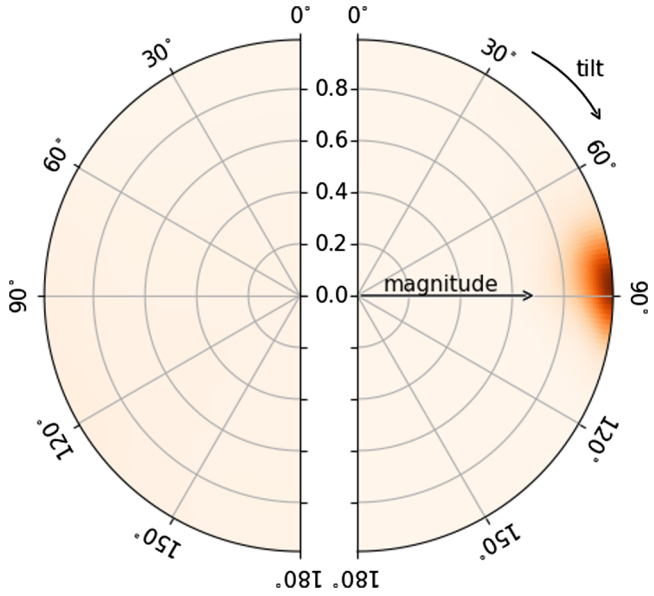


FIG. 5. A spin-disk plot showing the two-dimensional posterior distribution of the spin magnitude and tilt of the primary (right-hand side) and secondary (left-hand side) of GW200129.

In Fig. 6, we demonstrate the two-dimensional mass ratio q and precession parameter χ_p results using our noise mitigation routine. We see that the mass ratio peaks around 0.5 and that χ_p peaks around 0.95, indicating a highly precessing system. This agrees with results from Hannam *et al.* [10], but contrasts with the findings from Payne *et al.* [18] where BayesWave is used to model the noise. Payne *et al.* [18] find the mass ratio to be less

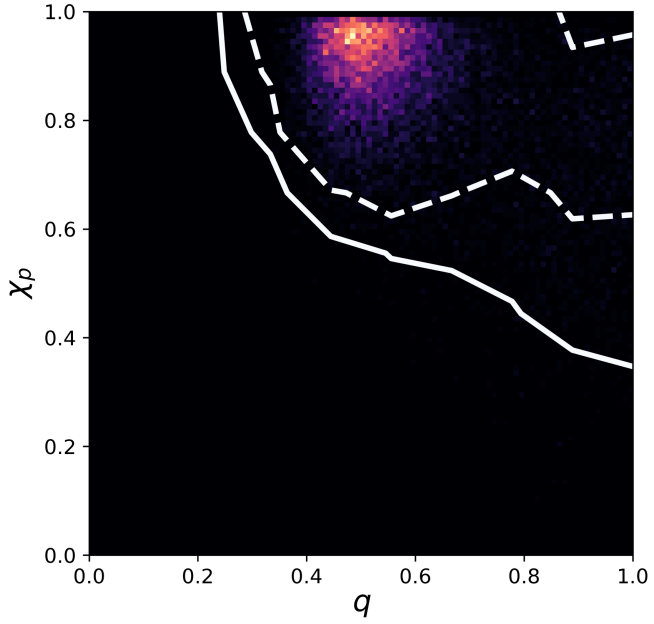


FIG. 6. Mass ratio q and the precession parameter χ_p plot for GW200129 using ML-mitigated data.

symmetric and the posterior on the precession parameter χ_p to be less informative (depending on the glitch model they use).

IV. CONCLUSIONS

In this paper, we demonstrated that machine learning can model the non-Gaussian noise recorded by witness channels. We use RF45-I, RF45-Q, and RF9-I channels as input to the fully connected dense neural network and gravitational-wave strain channel h(t) as output. Such an algorithm allowed us to mitigate the noise around the GW200129 signal.

We also found that our approach removes a significant part of the non-Gaussian noise in h(t), and that it actually removes more non-Gaussian noise than `gwsubtract`, the algorithm used by the LIGO-Virgo-KAGRA Collaboration.

At frequencies below 50 Hz, our algorithm removes up to 10% more fractional power compared with `gwsubtract`, while at higher frequencies the fractional power difference can be as high as 20%. However, neither `gwsubtract` nor our method is able to remove the non-Gaussian noise completely. This can happen due to these algorithms being imperfect, witness channels not containing the full information about the excess noise, or any other unknown reason.

Furthermore, we note that our method does not oversubtract the noise. The algorithm removes only the correlated noise between the gravitational-wave strain h(t) and the witness channels, as the algorithm does not have enough capacity to memorize all of the data it is trained on.

The reanalysis of GW200129 with the improved data quality finds a similar evidence of the spin-induced precession as reported by Hannam *et al.* [10]. However, we note that, while our method is able to remove more rf noise from the data than `gwsubtract`, we also show that there is still some noise left unmitigated.

The data frame used in this paper is publicly available on Zenodo [23]. This research has made use of data, software, and/or web tools obtained from the Gravitational Wave Open Science Center [27], a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. Various parts of the analysis used GWpy [28], PyCBC [29,30], Bilby [25,26], Bilby-MCMC [31], TensorFlow [22], Keras [32], Adam [21], PyMC [33], NumPy [34], SciPy [35], IPython [36], Jupyter NOTEBOOK [37], Pandas [38], Matplotlib [39], ARVIZ [40], and CORNER [41].

ACKNOWLEDGMENTS

We thank Adrian Helmling-Cornell, Charlie Hoy, Christopher Berry, Katerina Chatziioannou, Mark Hannam, Nikolaos Stergioulas, Thomas Dent, and members of the LIGO Detector Characterization group for their valuable input during the preparation of this manuscript. We also thank the anonymous referee for valuable comments, which improved the manuscript during the review. R. M. is supported by STFC Grants No. ST/S000550/1, No. ST/T000325/1, No. ST/V005715/1, and No. ST/X002225/1. The authors are grateful for computational resources provided by the LIGO Laboratory and supported by National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation. LIGO Laboratory and Advanced LIGO are funded by the U.S. National Science Foundation (NSF), as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN), and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA. is supported by Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Observation of gravitational waves from a binary black hole merger, Phys. Rev. Lett. **116**, 061102 (2016).

[2] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs, Phys. Rev. X **9**, 031040 (2019).

[3] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, Phys. Rev. X **11**, 021053 (2021).

[4] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, Phys. Rev. X **13**, 041039 (2023).
[5] T. D. Canton, S. Bhagwat, S. V. Dhurandhar, and A. Lundgren, Effect of sine-Gaussian glitches on searches for binary coalescence, Classical Quantum Gravity **31**, 015016 (2013).
[6] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Effects of data quality vetoes on a search for compact binary coalescences in Advanced LIGO's first observing run, Classical Quantum Gravity **35**, 065010 (2018).
[7] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals, Classical Quantum Gravity **37**, 055002 (2020).
[8] C. Pankow, K. Chatziioannou, E. A. Chase, T. B. Littenberg, M. Evans, J. McIver, N. J. Cornish, C.-J. Haster, J. Kanner, V. Raymond *et al.*, Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817, Phys. Rev. D **98**, 084016 (2018).
[9] R. Macas, J. Pooley, L. K. Nuttall, D. Davis, M. J. Dyer, Y. Lecoeuche, J. D. Lyman, J. McIver, and K. Rink, Impact of noise transients on low latency gravitational-wave event localization, Phys. Rev. D **105**, 103021 (2022).
[10] M. Hannam *et al.*, General-relativistic precession in a black-hole binary, Nature (London) **610**, 652 (2022).
[11] A. H. Nitz, S. Kumar, Y.-F. Wang, S. Kastha, S. Wu, M. Schäfer, R. Dhurkunde, and C. D. Capano, 4-OGC: Catalog of gravitational waves from compact binary mergers, Astrophys. J. **946**, 59 (2023).
[12] G. Pratten, S. Husa, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, and R. Jaume, Setting the cornerstone for a family of models for gravitational waves from compact binaries: The dominant harmonic for non-precessing quasicircular black holes, Phys. Rev. D **102**, 064001 (2020).
[13] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Surrogate models for precessing binary black hole simulations with unequal masses, Phys. Rev. Res. **1**, 033015 (2019).
[14] D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. D. Renzo, and G. Ashton, Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run, Classical Quantum Gravity **39**, 245013 (2022).
[15] LIGO Scientific and Virgo Collaborations, Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914, Classical Quantum Gravity **33**, 134001 (2016).
[16] D. Davis, T. Massinger, A. Lundgren, J. C. Driggers, A. L. Urban, and L. Nuttall, Improving the sensitivity of Advanced LIGO using noise subtraction, Classical Quantum Gravity **36**, 055011 (2019).
[17] T. Islam, A. Vajpeyi, F. H. Shaik, C.-J. Haster, V. Varma, S. E. Field, J. Lange, R. O'Shaughnessy, and R. Smith, Analysis of GWTC-3 with fully precessing numerical relativity surrogate models, arXiv:2309.14473.
[18] E. Payne, S. Hourihane, J. Golomb, R. Udall, D. Davis, and K. Chatziioannou, Curious case of GW200129: Interplay between spin-precession inference and data-quality issues, Phys. Rev. D **106**, 104017 (2022).
[19] N. J. Cornish and T. B. Littenberg, BayesWave: Bayesian inference for gravitational wave bursts and instrument glitches, Classical Quantum Gravity **32**, 135012 (2015).
[20] N. J. Cornish, T. B. Littenberg, B. Bé csy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, BayesWave analysis pipeline in the era of gravitational wave observations, Phys. Rev. D **103**, 044006 (2021).
[21] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.
[22] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, software available from https://www.tensorflow.org (2015).
[23] R. Macas, A. Lundgren, and G. Ashton, Data Release for "Revisiting the evidence for precession in GW200129 with machine learning noise mitigation", Zenodo, 10.5281/zenodo.10143337 (2023).
[24] R. Macas and A. Lundgren, A sensitive test of non-Gaussianity in gravitational-wave detector data, Phys. Rev. D **108**, 063016 (2023).
[25] G. Ashton and C. Talbot, Bilby-MCMC: An MCMC sampler for gravitational-wave inference, Mon. Not. R. Astron. Soc. **507**, 2037 (2021).
[26] I. M. Romero-Shaw *et al.*, Bayesian inference for compact binary coalescences with Bilby: Validation and application to the first LIGO-Virgo gravitational-wave transient catalogue, Mon. Not. R. Astron. Soc. **499**, 3295 (2020).
[27] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration and KAGRA Collaboration), Open data from the third observing run of LIGO, Virgo, KAGRA and GEO, Astrophys. J. Suppl. Ser. **267**, 29 (2023).
[28] D. Macleod *et al.*, GWpy/GWpy: 2.0.1 (2020).
[29] G. S. Davies, T. Dent, M. Tápai, I. Harry, C. McIsaac, and A. H. Nitz, Extending the PyCBC search for gravitational waves from compact binary mergers to a global network, Phys. Rev. D **102**, 022004 (2020).
[30] T. D. Canton, A. H. Nitz, B. Gadre, G. S. Davies, V. Villa-Ortega, T. Dent, I. Harry, and L. Xiao, Realtime search for compact binary mergers in Advanced LIGO and Virgo's third observing run using PyCBC live, Astrophys. J. **923**, 254 (2021).
[31] G. Ashton, M. Hübner, P. D. Lasky, C. Talbot, K. Ackley, S. Biscoveanu, Q. Chu, A. Divakarla, P. J. Easter, B. Goncharov, F. Hernandez 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, Astrophys. J. Suppl. Ser. **241**, 27 (2019).
[32] F. Chollet *et al.*, Keras (2015), https://github.com/fchollet/keras.
[33] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, Probabilistic programming in Python using PyMC3, PeerJ Comput. Sci. **2**, e55 (2016).
[34] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith *et al.*, Array programming with NumPy, Nature (London) **585**, 357 (2020).

[35] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).

[36] F. Perez and B. E. Granger, IPython: A system for interactive scientific computing, Comput. Sci. Eng. **9**, 21 (2007).

[37] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, and C. Willing, Jupyter NOTEBOOKS—A publishing format for reproducible computational workflows, in *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, edited by F. Loizides and B. Schmidt (IOS Press, Washington, DC, United States, 2016), pp. 87–90.

[38] W. McKinney *et al.*, Data structures for statistical computing in Python, in *Proceedings of the 9th Python in Science Conference Austin, TX* (IOS Press, Amsterdam, The Netherlands, 2010), Vol. 445, pp. 51–56.

[39] J. D. Hunter, Matplotlib: A 2d graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[40] R. Kumar, C. Carroll, A. Hartikainen, and O. Martin, ARVIZ a unified library for exploratory analysis of Bayesian models in Python, J. Open Source Software **4**, 1143 (2019).

[41] D. Foreman-Mackey, CORNER.PY: Scatterplot matrices in Python, J. Open Source Software **1**, 24 (2016).