Deep learning waveform anomaly detector for numerical relativity catalogs

Tibério Pereira*
Departamento de Física
Universidade Federal do Rio Grande do Norte,
Natal 59078-970, RN, Brazil

Riccardo Sturani†
Instituto de Física Teórica, UNESP-Universidade Estadual Paulista & ICTP South American Institute for Fundamental Research, Sao Paulo 01140-070, SP, Brazil

(Dated: October 17, 2022)

Numerical Relativity has been of fundamental importance for studying compact binary coalescence dynamics, waveform modelling, and eventually for gravitational waves observations. As the sensitivity of the detector network improves, more precise template modelling will be necessary to guarantee a more accurate estimation of astrophysical parameters. To help improving the accuracy of numerical relativity catalogs, we developed a deep learning model capable of detecting anomalous waveforms. We analyzed 1341 binary black hole simulations from the SXS catalog with various mass-ratios and spins, considering waveform dominant and higher modes. In the set of waveform analyzed, we found and categorised seven types of anomalies appearing close to the merger phase.

I. INTRODUCTION

The network of gravitational wave (GW) detectors composed by the two LIGO [1] and Virgo [2] observatories has already completed three successful observation runs [3], with the detection of over 90 coalescences of compact binary systems. To maximize the possibility of detections and their (astro-)physics output, collected data are analyzed via matched-filtering techniques [4, 5] by correlating them with pre-computed waveform templates, whose development is the object of intense investigation [6–12]. Improving the accuracy of GW templates straightforwardly enhances the quality of astrophysical information obtained from these sources.

Waveforms can be expressed as time series resulting from the spin-weighted spherical harmonic decomposition of the gravitational wave strain. While the dominant, quadrupolar mode of GW templates has been enough to analyze the vast majority of signals, in a few cases the imprints of sub-dominant (or higher modes, HM henceforth) have been detected [13, 14]. With increasing sensitivity and widening the network of observatories in future observation runs with KAGRA [15], it is expected that HM will have a larger impact on the detected signals [16]. GW waveforms generated via Numerical Relativity (NR) simulations [17] have been widely used for construction of semi-analytical and phenomenological templates, and have produced accurate HM waveforms over a vast parameter space.

The goal of the present paper is to provide a new tool to assess the data quality of the various simulations presented in a NR catalog. The numerical codes' complexity and the simulations' long runtime can generate a systematic accumulation of numerical residuals, leading to defects in the morphology of the waveforms. Furthermore, there are cases in which a catalog presents simulations with different numerical resolutions due to code updating.

In this work, we developed the deep learning model *Waveform AnomaLy DetectOr* (WALDO), capable of signaling possible anomalous waveforms in a NR catalog [18]. In our searches within binary black hole (BBH) simulations, we categorized seven different types of anomalies during the stages of coalescence. Identifying and excluding such waveforms is critical to the quality of research in GW analysis and surrogate modeling [19].

Applications of deep learning models to gravitational wave data is not new [20–34], but to the best of our knowledge this is the first work using deep learning to check the consistency of numerical simulations.

The paper is structured as follows. Section II is intended to help the reader providing a reference to our notations, Section III describes the dataset we used for our analysis, and Section IV describes our machine learning-based process to identify anomalous waveforms, whose results are presented in Section V. Finally we summarize our conclusions in Section VI.

II. DEFINITIONS

We adopt geometric units $G = c = 1$, and we denote by u the dimension-less time obtained by dividing physical time by the total mass M of the binary system, whose zero is set by the epoch of the peak of the dominant mode amplitude. The BBH mass-ratio q is taken to be larger than 1, the dimensionless spins $\vec{\chi}_i \equiv \vec{S}_i/M^2$, $i = \{1, 2\}$, being $\vec{S}_i$ the standard spin and the orbital eccentricity is denoted by e. From the GW *strain*, i.e. the GW polarization complex combination $h_+ - ih_\times$, we extract (and rescale as usual by distance r and mass M) spherical har-

* tiberio@fisica.ufrn.br
† riccardo.sturani@unesp.br

monic modes $h_{lm}(t)$, indexed by integers $l \geq 2$, $|m| \leq l$, resulting from the decomposition on the spin-weighted spherical harmonics base ${}_{-2}Y_{lm}$ of the strain,

$$h_{lm}(t) = \int d\Omega \; (h_+ - i h_\times)\,(\theta,\phi)\; {}_{-2}Y_{lm}(\theta,\,\phi)\,, \quad (1)$$

where Ω is the solid angle parametrized by θ and ϕ, which are respectively the angle between the radiation direction and the normal to the orbital plane, and a phase corresponding to a rotation in the orbital plane.

III. THE DATASET

We create a dataset using 1341 BBH simulations from the *Simulating eXtreme Spacetimes* (SXS) catalog [17], whose parameters are in the region $q = [1.0 : 4.0]$, $|\vec{\chi}_i| = [-0.9 : 0.9]$, and $e \simeq [0.0 : 8.0 \times 10^{-4}]$ [35, 36]. All simulation names are listed in the WALDO's repository [18]. Considering the modes ($l \leq 4$, $l-1 \leq |m| \leq l$), in total our dataset is composed by $N_d = 8046$ waveforms. Figure 1 shows the parameter space distribution of eccentricity, spin-aligned parameter χ_{eff}, spin-precession parameter χ_p [37, 38] and the mass-ratio q.

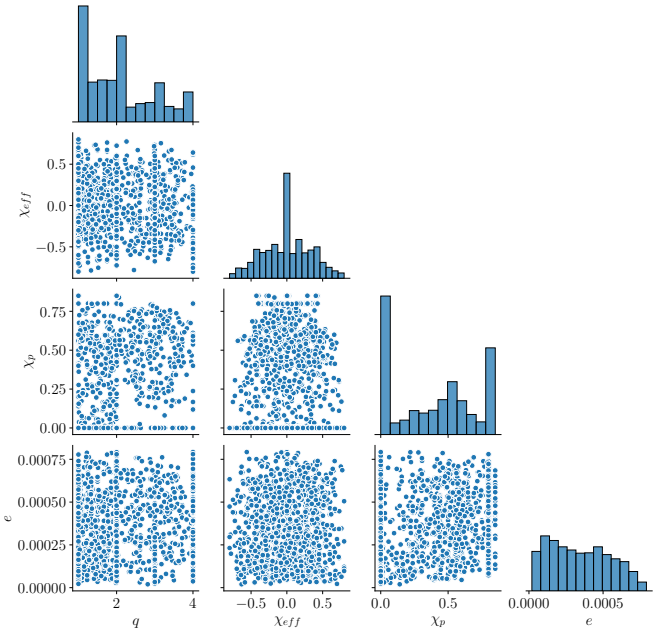

FIG. 1. Representation of the parameter space of the 1341 simulations: corner plot of the distribution of eccentricity e, mass-ratio q, spin-aligned parameter χ_{eff} and spin-precession χ_p.

To facilitate a unified treatment of all waveforms, we cut the waveforms to the highest initial time value of the entire dataset, *i.e.*, the inspiral starts at $u = -1700$, with corresponding instantaneous frequency for the dominant mode $f_{min} \simeq 14 - 19Hz\,(M/100M_\odot)^{-1}$, depending on the mass ratio. This conditioning is necessary to guarantee the same resolution of the waveforms – with the same time intervals – during the neural network (NN) training.

Also, we find it convenient to re-sample all modes via the time-reparametrization

$$\begin{aligned} u'(u) &= c_1\, r(u) + c_2\,, \\ r(u) &= \tanh\left[a\,(u - u_0 + b)\right]\,, \\ c_1 &= (u_F - u_0)\,/\,[r(u_F) - r(u_0)]\,, \\ c_2 &= u_0 - c_1\, r(u_0)\,, \end{aligned} \quad (2)$$

where u_0 and u_F are the initial and final value of the dimension-less time $u = [-1700 : 100]$; $a = 5 \times 10^{-4}$ and $b = 5. \times 10^2$ are constants. The rationale for this parameterization is to make smoother the transition from the wider spacing during the inspiral to a smaller one in the merger-ringdown phase, while keeping the number of samples equal to $N_s = 2048$ for all waveforms, without degrading the sampling rate in the merger-ringdown phase.

For deep learning *feature engineering* – the preprocessing procedures for improving NN computations – we normalize the dataset with the highest waveform amplitude value,

$$h^{(k)}_{lm} \to h^{(k)}_{lm} / \max\left(|h^{(k)}_{22}|\right)\,, \quad (3)$$

where the index (k) denotes the simulation number, $k = [1 : 8046]$. We define the dataset as the numerical three-dimensional array,

$$\mathbf{X} \equiv \left(Re\left\{h_{lm}\right\}, Im\left\{h_{lm}\right\}\right)^{(k)}\,, \quad (4)$$

whose dimensionality is $(N_d,\, N_s, 2)$.

IV. WALDO

The Waveform AnomaLy DetectOr (WALDO) holds a U-Net architecture [39], where the waveform input $X^{(k)}$ is reproduced as the output $\bar{X}^{(k)}$. During the training, the model learns all possible waveform features related to the parameter space. Evaluating its performance with the *mismatch* $\mathcal{M}$ between $h^{(k)}_{lm}$ and its prediction $\bar{h}^{(k)}_{lm}$, the measurement of high mismatch values can flag the presence of waveforms whose morphology do not match the dataset one, *i.e.*, we can find *anomalous waveforms*. The mismatch is defined as $\mathcal{M} \equiv 1 - \mathcal{O}$, where $\mathcal{O}$ denotes the *match* between two time series $h_{1,2}$, defined in terms of the scalar product

$$\langle h_1 | h_2 \rangle = 2 \int_0^\infty \left(\tilde{h}_1(f)\tilde{h}^*_2(f) + \tilde{h}_2(f)\tilde{h}^*_1(f)\right) df\,, \quad (5)$$

where f is variable conjugate to time under Fourier transform $h \to \tilde{h}$, the match being defined by maximization over initial phase and time of the scalar product of normalized waveforms

$$\mathcal{O} \equiv \overset{Max}{t, \phi_0} \frac{\langle h_1 | h_2 \rangle}{\left(\langle h_1 | h_1 \rangle \langle h_2 | h_2 \rangle\right)^{1/2}}\,. \quad (6)$$

The WALDO's encoder part presents convolutional layers of *kernel-size* = 3 intermediated by max-pooling layers with *pool-size* = 2. The input dimensionality reduces as the number of channels C increases for C = [32, 64, 128, 256, 512]. The decoder part duplicates the layers' dimension with up-sampling layers, followed by convolutional, concatenation, and another convolutional layer. We compose the architecture with ReLU activation functions, except for the hyperbolic tangent in the output layer.

A. Training and validation

To examine the training performance, we split the dataset into 70% for training data, which is computed in parallel with 20% for validation data, and 10% for testing data. Since 8046 waveforms forms a small dataset, we use the *K-fold validation* method for $K = 3$, and *batch-size* = 1. We minimize the model parameters using the mean squared error loss function and Adagrad optimizer [40]. During the validation, the NN can be trained through 200 epochs without over-fitting. We retrain the model using 90% of the dataset and obtain the mean square error average 2.8×10^{-7} over the testing data.

V. RESULTS

After the training, WALDO evaluates the mismatch and packs the values with (l, m) mode labels, together with the identification simulation number (ID) – that comes from the SXS simulation names as SXS:BBH:ID – the parameter space $(q, \vec{\chi}_1, \vec{\chi}_2, e)^{(k)}$, the waveforms $h_{lm}^{(k)}$ and their predictions $\bar{h}_{lm}^{(k)}$. Creating a histogram for each mode (l, m), we isolated 1% of the highest mismatch waveforms to verify any possible morphological discrepancy in the predictions. Figure 2 shows the (3, 2) waveform mismatch distribution of average 1.26×10^{-5} represented by the vertical green line; the *quantile* = 0.99, marked by the vertical pink line, separates 14 simulations that call for examination. The lowest mismatch value is on the order of 5.0×10^{-6} due to the low number of simulations $N_d = 1341$ and varied waveform morphology. We reinforce that the NN training is usually done with hundreds of thousands of data, however our small dataset does not interfere with the quality of waveform reproduction.

In this case, we found high mismatches due to noise accumulation in the predicted waveforms, even if qualitatively they follow the NR morphological patterns, as shown in Fig. 3 – where the blue line is h_{32}, the dashed orange line is $\bar{h}_{32}$, and the green one is $Re\left\{h_{32} - \bar{h}_{32}\right\}$ amplified 10 times. On the other hand, we also found morphological discrepancies between NN predicted and NR waveform modes, confined in specific sectors of the coalescence, causing high mismatches. An irregularity fairly common for the higher modes, is that the mode

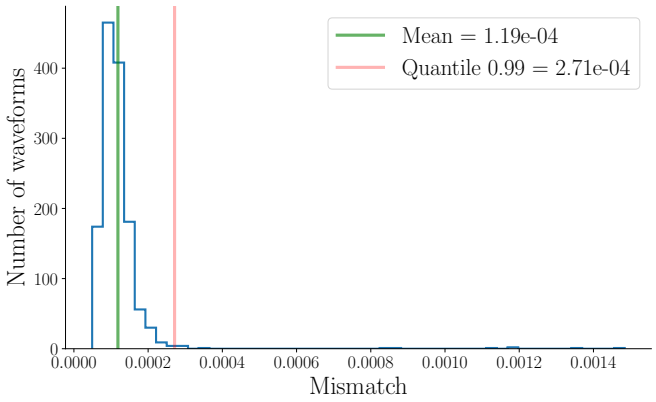

FIG. 2. Mismatch histogram of 1341 waveforms $h_{32}^{(k)}$ and the predictions $\bar{h}_{32}^{(k)}$. The vertical green line shows the average value 1.26×10^{-5}, and the pink one marks the *quantile* = 0.99 with 14 simulations on its right side.

amplitude around the merger peak has a greater magnitude than expected – what we call the *merger-peak* (MP) anomaly, as seen in Fig. 4.

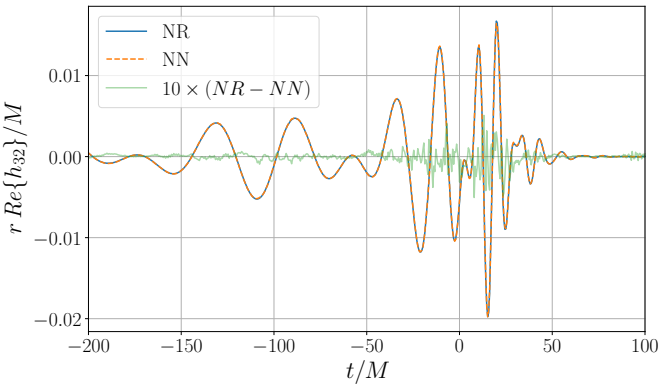

FIG. 3. NR vs NN comparison of the (3, 2)-waveforms from the silmulation ID = 1018. $Mismatch = 3.0 \times 10^{-4}$; $q = 1.2$, $\vec{\chi}_1 = (-0.59, 0.15, -0.39)$, $\vec{\chi}_2 = (-0.73, 0.19, 0.22)$, and $e = 4.2 \times 10^{-4}$.

For even m modes, we found MP anomalies in $\mathrm{ID}_{(3,2)} = [1863, 1991 - 1993, 1995]$ and $\mathrm{ID}_{(4,4)} = [1982, 1983, 1991, 1992, 1995]$. In the search for odd m modes, we restrict the mass ratio to $q > 1.0$, where we found $\mathrm{ID}_{(2,1)} = [1982, 1993]$, $\mathrm{ID}_{(3,3)} = [1982, 1983, 1989, 1991 - 1993, 1995]$, and $\mathrm{ID}_{(4,3)} = [0601, 1982, 1983, 1989, 1991 - 1993, 1995]$.

In some waveform (4, 4) modes the ringdown decay begins a little later in NR simulations than in our NN predictions. Figure 5 shows an example of the *lazy-ringdown* (LR) anomaly, also found in $\mathrm{ID}_{(4,4)} = [0230, 1477, 1481, 2104]$. In the simulation $\mathrm{ID}_{(4,4)} = 0155$, on the other hand, the ringdown amplitude does

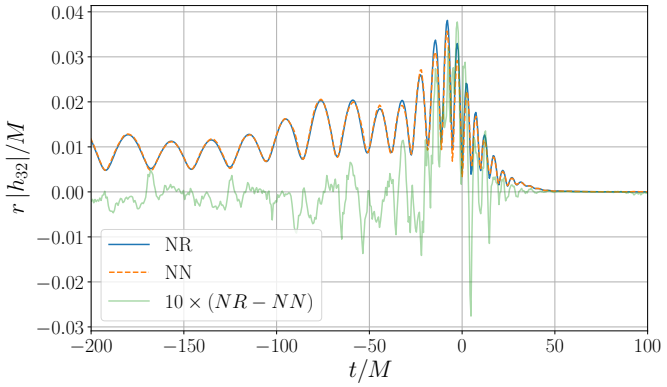

FIG. 4. NR vs NN comparison of the (3, 2)-waveforms from the silmulation ID = 1995, displayng a MP anomaly. $Mismatch = 1.48 \times 10^{-3}$; $q = 4.0$, $\vec{\chi}_1 = (0.51, -0.29, -0.54)$, $\vec{\chi}_2 = (0.07, 0.07, -0.79)$, and $e = 1.2 \times 10^{-4}$.

not exhibit appropriate asymptotic behavior. The *asymptotic-ringdown* (AR) anomaly is shown in Fig. 6.

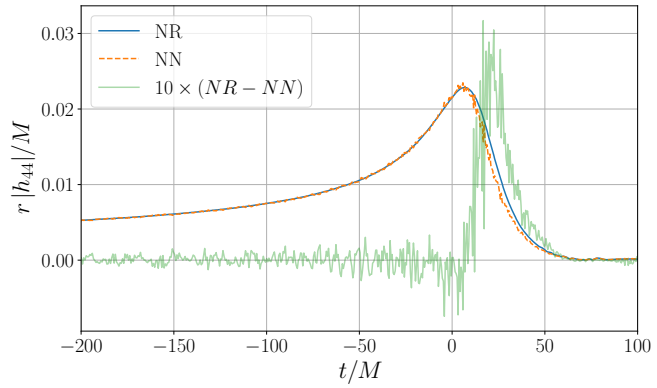

FIG. 5. NR vs. NN comparison of the (4, 4)-waveforms from the simulation ID = 0328 displaying a LR anomaly. $Mismatch = 1.44 \times 10^{-3}$; $q = 1.0$, $\vec{\chi}_1 = (0.0, 0.0, 0.8)$, $\vec{\chi}_2 = (0.0, 0.0, 0.8)$, and $e = 1.1 \times 10^{-4}$.

A. Constraint dataset search

The homogeneity of the parameter space distribution is important to avoid WALDO's prediction bias. For instance, a dataset containing 500 simulations of spin-aligned BBH and 20 precessing binaries can lead to high mismatch values for waveforms whose features indicate precession.

Thereby, we focus our search for anomalies on $|\vec{\chi}_{1,2}| \leq 0.4$, where we have a higher simulation density. We choose *quantile* = 0.90 to isolate 15 waveforms. This constraint leads us to find more AR anomalies in $\mathrm{ID}_{(3,3)}$ = [0116, 0115, 0119, 0129] and $\mathrm{ID}_{(4,4)}$ =

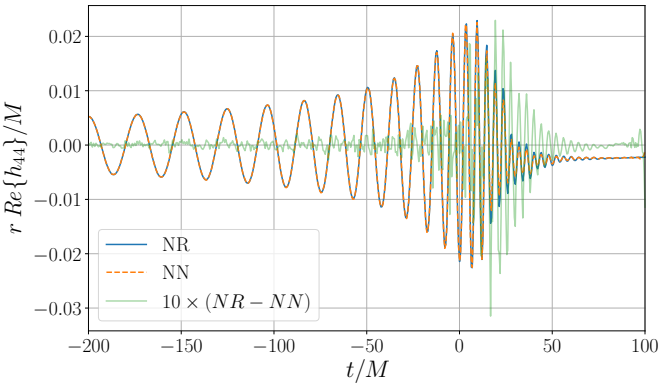

FIG. 6. NR vs. NN comparison of the (4, 4)-waveforms from the silmulation ID = 0155 exemplifying a AR anomaly. $Mismatch = 1.21 \times 10^{-3}$; $q = 1.0$, $\vec{\chi}_1 = (0.0, 0.0, 0.8)$, $\vec{\chi}_2 = (0.0, 0.0, 0.8)$, and $e = 5.1 \times 10^{-4}$.

[0070, 0115, 0124, 0135, 0150]. In addition, we found waveforms with similar decay as in Fig. 6, but with non-oscilatory patterns. In this case, those ringdowns were affected by the time interpolation of Eq. 2 because their final time is smaller than $u = 70$, giving rise to what we dubbed *short-ringdown* (SR) anomaly, found in $\mathrm{ID}_{(l,m)}$ = [1112, 1114, 1133], for all (l, m). Fig. 7 shows the h_{22} modes from these simulations.

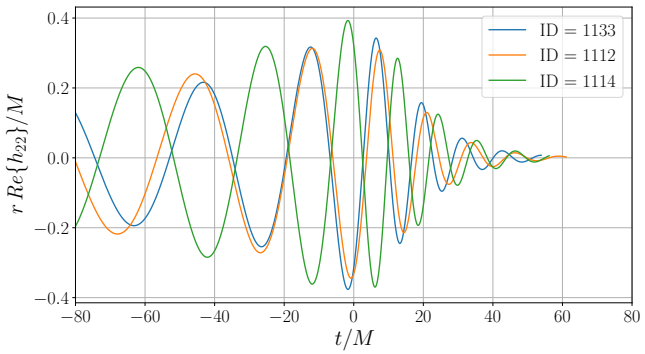

FIG. 7. Short-ringdown anomaly in the simulations ID = [1112, 1114, 1133].

The ringdown amplitude of the dominant mode is expected to show a smooth, quasi-exponential decay, however, in the simulations $\mathrm{ID}_{(2,2)}$ = [0066, 0067, 0070, 0072, 0126, 0136] appear small ripples up to $u = 40$ as in Fig. 8, which we call the *rippled-ringdown* (RR) anomaly.

Note that these small ripples are present in several of the original NR simulations, and they are reproduced by the NN predictions, however with high enough mismatch to be uncovered.

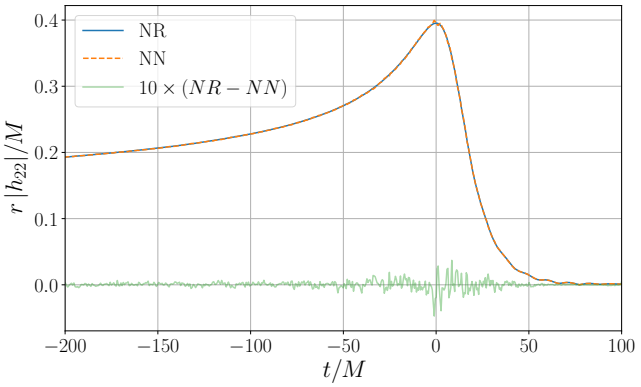

FIG. 8. NR vs. NN comparison of the (2, 2)-waveforms from the simulation ID = 0072, showing a RR anomaly. $Mismatch = 1.33 \times 10^{-5}$; $q = 1.0$, $\vec{\chi}_1 = (0.0, 0.0, 0.0)$, $\vec{\chi}_2 = (0.0, 0.0, 0.0)$, and $e = 1.5 \times 10^{-4}$.

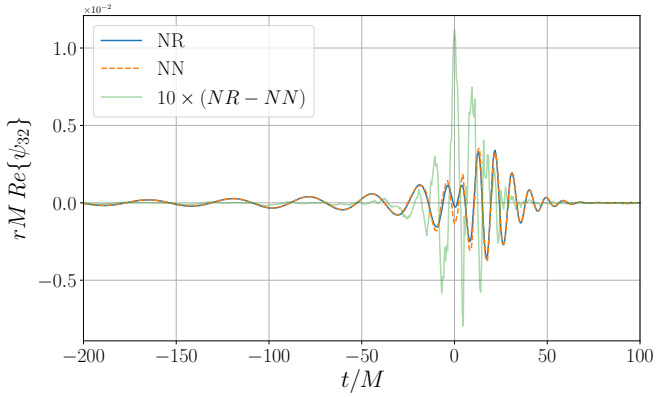

FIG. 9. NR vs NN comparison of the ψ_{32} from the simulation ID = 1989. $Mismatch = 2.7 \times 10^{-3}$; $q = 4.0$, $\vec{\chi}_1 = (-0.50, -0.29, -0.55)$, $\vec{\chi}_2 = (0.02, -0.10, -0.79)$, and $e = 6.1 \times 10^{-4}$.

B. Radiation field search

In NR the strain modes are computed from the integration of the radiation field Ψ_4. This operation is not trivial, and it can lead to accumulated numerical noise [41]. For this reason, we investigate the $\psi_{lm} = \ddot{h}_{lm}$ modes within the whole dataset to ensure whether those anomalies can be ascribed to integration residues.

We retrained the NN and evaluate the mismatch between NR ψ_{lm} waveforms and their reproductions $\bar{\psi}_{lm}$. From previous analyses, we found exclusively MP anomalies in h_{lm} modes with $\mathrm{ID}_{(2,2)} = [1993, 1995]$, $\mathrm{ID}_{(3,2)} = [1863, 1878, 1983, 1991 - 1993, 1995]$, $\mathrm{ID}_{(3,3)} = [1982, 1989, 1991-1993]$, $\mathrm{ID}_{(4,3)} = [1103, 1982, 1983, 1989, 1991 - 1993]$, and $\mathrm{ID}_{(4,4)} = [0024, 0126, 1982, 1989, 1993, 1995]$.

Still, the ψ_{lm} modes revealed smaller magnitudes than the predictions in the initial region of the merger as seen in Fig. 9. The *initial-merger* (IM) anomalies are present in ψ_{lm} with $\mathrm{ID}_{(2,2)} = [1982, 1989]$, $\mathrm{ID}_{(3,2)} = [1982, 1989]$, $\mathrm{ID}_{[(3,3),\,(4,3),\,(4,4)]} = 1995$.

In the ψ_{lm} mode (2, 1), we did not find MP or IM anomalies but a qualitatively different discrepancy in $\mathrm{ID}_{(2,1)} = [0013, 0252, 0292, 0388, 0513, 0575, 0464, 1508, 2126]$. The *dephased-inspiral* (DI) anomaly appears as if the NN inspiral was generated by a higher eccentric orbit than the NR one, see e.g. Fig. 10. In fact, simulations with eccentricity $e > 1.0 \times 10^{-4}$ have to be considered non-negligibly eccentric with respect to the majority of simulations.

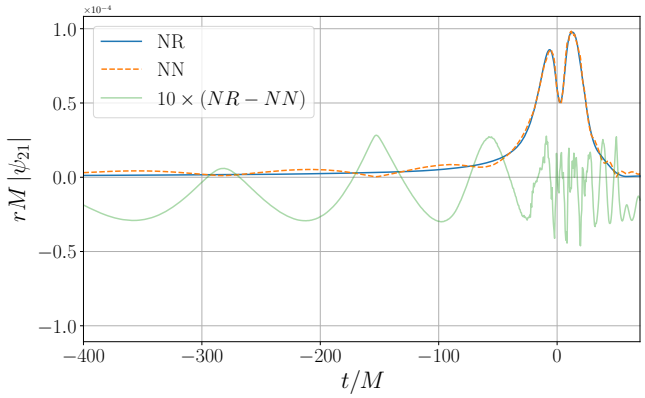

FIG. 10. NR vs NN comparison of the ψ_{21} from the simulation ID = 1508, showing a DI anomaly. $Mismatch = 2.45 \times 10^{-2}$; $q = 1.3$, $\vec{\chi}_1 = (0.0, 0.0, 0.29)$, $\vec{\chi}_2 = (0.0, 0.0, -0.07)$, and $e = 1.6 \times 10^{-4}$.

VI. DISCUSSION

To assess the quality of Numerical Relativity data, and to identify candidate problematic waveforms, we developed Waveform AnomaLy DetectOr (WALDO) wich allowed us to identify potentially anomalous waveforms both in the dominant and higher modes. We trained our model with 8046 waveforms with a U-Net neural architecture and calculated the mismatch between the NR waveforms and the NN predictions. By isolating the 1% waveforms with highest mismatch, we identified seven qualitatively different anomalies during the inspiral, merger, and ringdown stages. Table I summarizes the anomaly categories.

The present work intends to be a starting point for a more thorough investigation over neural network applied to numerical waveforms.

Except for the DI anomaly, all others ones are in the waveform merger-ringdown stages, which suggests the need to improve the *adaptive mesh refinement* method of numerical simulations [42]. The more refined the calculations during the collision of black holes, the more

Anomaly	Description
(AR) Asymptotic-Ringdown	Non-null asymptotic behavior during ringdown.
(LR) Lazy-Ringdown	Ringdown late decay.
(RR) Rippled-Ringdown	Ripples in the dominant mode ringdown amplitude.
(SR) Short-Ringdown	Ringdown with length below $70M$.
(MP) Merger-Peak	Higher amplitude around the merger peak.
(IM) Initial-Merger	Shorter amplitude in the merger beginning.
(DI) Dephased-Inspiral	Oscillatory dephasing during inspiral.

TABLE I. Seven waveform anomalies detected by WALDO.

accurate the waveform during the merger.

In our search for anomalies in the Ψ_4 radiative field modes, we did not find LR and AR anomalies, leading us to conclude that these are strain integration issues. We cannot draw the same conclusion about the RR anomaly in h_{22}, because of its small magnitude and its disappearing in Ψ_4, i.e. after double derivatives.

We focus our anomaly search on the *quantile* > 0.99 region of the mismatch, and for higher modes with $l \leq 4$.

We highlight that for our analysis the dataset need to be as homogeneous as possible in terms of astrophysical parameter space, to avoid large mismatch values when dealing with anomaly-free waveforms because of poor modeling. It is essential to remove simulations that have anomalies from the dataset and re-train the neural network to ensure that low-quality simulations do not pollute the training set. Such anomalies can impair waveform modeling [43–46] and interfere with analysis, such as the ringdown quasi-normal modes [47–49].

We propose that WALDO can be applied to any time-series, such as gravitational waves from binary neutron star and back hole-neutron star binary. Also, we suggest to evaluate the quality of new simulations for the next generations of NR codes by comparing them with waveforms from well-established catalogs in the literature.

ACKNOWLEDGMENTS

The authors thank the International Institute of Physics for hospitality and support during most of this work. TP is supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Graduate Research Fellowship. The work of RS is partly supported by CNPq under grant 310165/2021-0 and RS would like to thank ICTP-SAIFR FAPESP Grant No. 2016/01343-7. The authors thank the High Performance Computing Center (NPAD) at UFRN for providing the computational resources necessary for this work.

[1] J. Aasi *et al.* (LIGO Scientific), Class. Quant. Grav. **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].
[2] F. Acernese *et al.* (VIRGO), Class. Quant. Grav. **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].
[3] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv:2111.03606 [gr-qc].
[4] L. A. Wainstein and V. D. Zubakov, *Extraction of Signals from Noise*, Dover books on physics and mathematical physics (Prentice-Hall, Englewood Cliffs, NJ, 1962).
[5] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Class. Quant. Grav. **37**, 055002 (2020), arXiv:1908.11170 [gr-qc].
[6] G. Pratten *et al.*, Phys. Rev. D **103**, 104056 (2021), arXiv:2004.06503 [gr-qc].
[7] G. Pratten, S. Husa, C. Garcia-Quiros, M. Colleoni, A. Ramos-Buades, H. Estelles, and R. Jaume, Phys. Rev. D **102**, 064001 (2020), arXiv:2001.11412 [gr-qc].
[8] C. García-Quirós, S. Husa, M. Mateu-Lucena, and A. Borchers, Class. Quant. Grav. **38**, 015006 (2021), arXiv:2001.10897 [gr-qc].
[9] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, Phys. Rev. D **102**, 064002 (2020), arXiv:2001.10914 [gr-qc].
[10] S. Ossokine *et al.*, Phys. Rev. D **102**, 044055 (2020), arXiv:2004.09442 [gr-qc].
[11] S. Babak, A. Taracchini, and A. Buonanno, Phys. Rev. D **95**, 024010 (2017), arXiv:1607.05661 [gr-qc].
[12] Y. Pan, A. Buonanno, A. Taracchini, L. E. Kidder, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, and B. Szilágyi, Phys. Rev. D **89**, 084006 (2014), arXiv:1307.6232 [gr-qc].
[13] R. Abbott *et al.* (LIGO Scientific, Virgo), Phys. Rev. D **102**, 043015 (2020), arXiv:2004.08342 [astro-ph.HE].
[14] R. Abbott *et al.* (LIGO Scientific, Virgo), Astrophys. J. Lett. **896**, L44 (2020), arXiv:2006.12611 [astro-ph.HE].
[15] H. Abe *et al.* (KAGRA), Galaxies **10**, 63 (2022).
[16] M. Pürrer and C.-J. Haster, Phys. Rev. Res. **2**, 023151 (2020), arXiv:1912.10055 [gr-qc].
[17] M. A. Scheel, M. Boyle, T. Chu, L. E. Kidder, K. D. Matthews, and H. P. Pfeiffer, Phys. Rev. D **79**, 024003 (2009).
[18] T. Pereira, "Waveform AnomaLy DetectOr (WALDO)," (2022).
[19] 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), arXiv:1905.09300 [gr-qc].
[20] P. J. Easter, P. D. Lasky, A. R. Casey, L. Rezzolla, and K. Takami, Phys. Rev. D **100**, 043005 (2019), arXiv:1811.11183 [gr-qc].

[21] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Physical Review Letters **120** (2018), 10.1103/physrevlett.120.141103.

[22] V. Varma, D. Gerosa, L. C. Stein, F. Hébert, and H. Zhang, Phys. Rev. Lett. **122**, 011101 (2019), arXiv:1809.09125 [gr-qc].

[23] H. Shen, D. George, E. A. Huerta, and Z. Zhao, in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2019).

[24] A. Rebei, E. Huerta, S. Wang, S. Habib, R. Haas, D. Johnson, and D. George, Physical Review D **100** (2019), 10.1103/physrevd.100.044025.

[25] Y. Setyawati, M. Pürrer, and F. Ohme, Class. Quant. Grav. **37**, 075012 (2020), arXiv:1909.10986 [astro-ph.IM].

[26] L. Haegel and S. Husa, Class. Quant. Grav. **37**, 135005 (2020), arXiv:1911.01496 [gr-qc].

[27] E. Cuoco *et al.*, Mach. Learn. Sci. Tech. **2**, 011002 (2021), arXiv:2005.03745 [astro-ph.HE].

[28] S. R. Green, C. Simpson, and J. Gair, Physical Review D **102** (2020), 10.1103/physrevd.102.104057.

[29] R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis, Physical Review Research **2** (2020), 10.1103/physrevresearch.2.033066.

[30] H. Yu and R. X. Adhikari, "Nonlinear noise regression in gravitational-wave detectors with convolutional neural networks," (2021).

[31] S. Schmidt, M. Breschi, R. Gamba, G. Pagano, P. Rettegno, G. Riemenschneider, S. Bernuzzi, A. Nagar, and W. D. Pozzo, Physical Review D **103** (2021), 10.1103/physrevd.103.043020.

[32] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, Nature Physics **18**, 112 (2021).

[33] S.-C. Fragkouli, P. Nousi, N. Passalis, P. Iosif, N. Stergioulas, and A. Tefas, "Deep residual error and bag-of-tricks learning for gravitational wave surrogate modeling," (2022).

[34] J. Yan, M. Avagyan, R. E. Colgan, D. Veske, I. Bartos, J. Wright, Z. Márka, and S. Márka, Physical Review D **105** (2022), 10.1103/physrevd.105.043006.

[35] A. H. M. et al., Physical Review Letters **111** (2013), 10.1103/physrevlett.111.241104.

[36] M. B. et al., Classical and Quantum Gravity **36**, 195006 (2019).

[37] T. A. Apostolatos, C. Cutler, G. J. Sussman, and K. S. Thorne, Physical Review D **49**, 6274 (1994).

[38] R. Green, C. Hoy, S. Fairhurst, M. Hannam, F. Pannarale, and C. Thomas, Physical Review D **103** (2021), 10.1103/physrevd.103.124023.

[39] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," (2015).

[40] J. C. Duchi, E. Hazan, and Y. Singer, J. Mach. Learn. Res. **12**, 2121 (2011).

[41] C. Reisswig and D. Pollney, Classical and Quantum Gravity **28**, 195015 (2011).

[42] M. J. Berger and J. Oliger, Journal of Computational Physics **53**, 484 (1984).

[43] S. Khan, K. Chatziioannou, M. Hannam, and F. Ohme, Physical Review D **100** (2019), 10.1103/physrevd.100.024059.

[44] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, B. Szilágyi, N. W. Taylor, and A. Zenginoglu, Physical Review D **89** (2014), 10.1103/physrevd.89.061502.

[45] J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, Physical Review Letters **115** (2015), 10.1103/physrevlett.115.121102.

[46] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, Physical Review D **95** (2017), 10.1103/physrevd.95.104023.

[47] E. W. Leaver, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **402**, 285 (1985).

[48] M. Maggiore, Physical Review Letters **100** (2008), 10.1103/physrevlett.100.141301.

[49] H. Yang, D. A. Nichols, F. Zhang, A. Zimmerman, Z. Zhang, and Y. Chen, Physical Review D **86** (2012), 10.1103/physrevd.86.104006.