Predicting binary neutron star postmerger spectra using artificial neural networks

Dimitrios Pesios[1,*] Ioannis Koutalios[1,2,†] Dimitris Kugiumtzis[3,‡] and Nikolaos Stergioulas[1,§]

[1]*Department of Physics, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece*
[2]*Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands*
[3]*Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece*

Gravitational waves in the postmerger phase of binary neutron star mergers may become detectable with planned upgrades of existing gravitational-wave detectors or with more sensitive next-generation detectors. The construction of template banks for the postmerger phase can facilitate signal detection and parameter estimation. Here, we investigate the performance of an artificial neural network in predicting simulation-based waveforms in the frequency domain (restricted to the magnitude of the frequency spectrum and to equal-mass models) that depend on three parameters that can be inferred through observations, neutron star mass, tidal deformability, and the gradient of radius versus mass. Compared to a baseline study using multiple linear regression, we find that the artificial neural network can predict waveforms with higher accuracy and more consistent performance in a cross-validation study. We also demonstrate, through a recalibration procedure, that future reduction of uncertainties in empirical relations that are used in our hierarchical scheme will result in more accurate predicted postmerger spectra.

I. INTRODUCTION

During the previous decade, we witnessed the birth of a new branch of observational astronomy that does not rely on the electromagnetic spectrum to derive observations of the Universe and its components [1]. Instead, it uses the properties of space-time itself to reveal hidden aspects of astronomical objects by means of gravitational-wave (GW) interferometric detectors. These space-time properties are exerted by violent events occurring in the Universe, such as the coalescence of binary black hole (BBH) or binary neutron star (BNS) systems, leading to the generation of detectable gravitational waves. In the first three observing runs, 90 events have been reported by the LIGO-Virgo-KAGRA Collaboration, the majority of them concerning BBH systems and only two of them concerning BNS systems [2]. More GW candidate events are accumulating in the ongoing O4 phase that started in May 2023. As current detectors are being improved and third-generation detectors are being designed, a greater number of signals, and possibly even new types, are expected to be detected in the near future [3–5].

One of the anticipated discoveries, which so far has remained undetected [6,7], is the gravitational-wave signal from the postmerger phase of BNS mergers. However, these signals may be identified using planned upgrades that extend beyond the fifth observing run (O5) and become more probable with third-generation or dedicated high-frequency detectors [8–17].

The postmerger phase may last up to a few tens of milliseconds and is characterized by the emission of significant amounts of gravitational radiation at distinct frequencies of a few kilohertz. Identifying specific postmerger frequencies could set strong constraints on the radius of neutron stars and, consequently, on their equation of state (EOS), which is a major goal in high-energy astrophysics; see [18–26], and references therein.

Figure 1 shows an example of the GW strain during the postmerger phase and the corresponding Fourier spectral amplitude for a particular EOS (DD2) and component mass $1.3M_\odot$ (the same for both stars), extracted from the CoRe v2 dataset [27]. During merger, linear nonaxisymmetric oscillation modes, nonlinear combination tones, and other transient effects are excited in the remnant; see, e.g., [28–35]. The main oscillation mode is the $l = m = 2$ f mode, commonly[1] referred to as f_{peak}. For comprehensive

*Contact author: dipesios@auth.gr
†Contact author: koutalios@mail.strw.leidenuniv.nl, iokoutal@physics.auth.gr
‡Contact author: dkugiu@auth.gr
§Contact author: niksterg@auth.gr

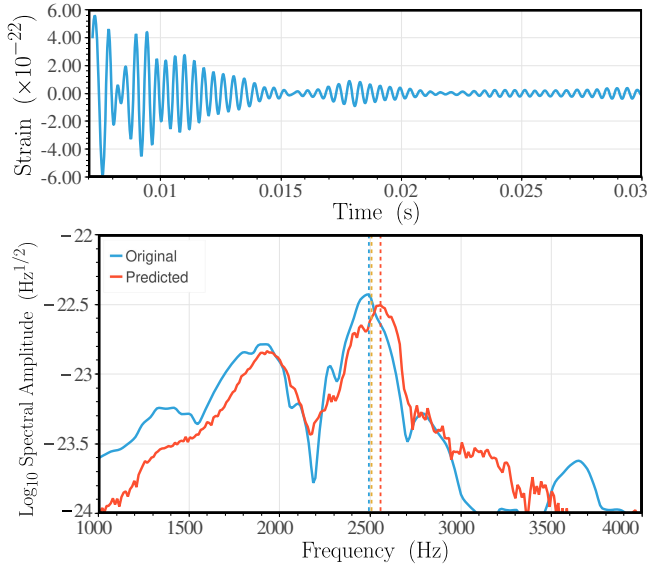[1]In some cases, it is also denoted as f_2.

FIG. 1.   Plus polarization of the postmerger strain, h_+, in the time domain (top panel) and the frequency domain (bottom panel) for a $1.3 M_\odot$ equal-mass BNS merger with the DD2 EOS. The source is assumed to be at 50 Mpc. The spectrum with our ANN-based method (red curve) agrees well with the original spectrum (blue curve). The blue dashed vertical line and the red one indicate the frequencies where each spectrum attains its peak value correspondingly, whereas the orange one indicates the predicted frequency of the empirical relation.

reviews of the various characteristics of the postmerger spectrum, see [18,20,22,25,36,37].

The postmerger GW spectrum (typically in the range between 1.5 and 4 kHz) depends on the masses of the two binary components and the EOS—see, e.g., [22,29,31,35,38–43]—and includes information that may allow the inference of various neutron star properties—see, e.g., [8,38,44–48]. This could be achieved, for instance, using empirical relations that relate f_{peak} to radii of nonrotating neutron stars in the inspiral phase [20,29,31,42,49–53].

A prerequisite for extracting source parameters from future detections of the postmerger signal is the availability of a specialized template bank that depends on the two component masses and the EOS. However, BNS simulations in numerical relativity are time-consuming, and, at the moment, only a few hundred different simulations have been performed. Even in the case of BBH mergers, where a larger number of numerical simulations are available, regression methods have been employed to construct template banks to reach the number of templates required for accurate detection and parameter estimation; see [54] and references therein. In the case of BNS mergers, the problem is more challenging. Ideally, we would like to construct detailed template banks that describe and parametrize the full inspiral-merger-postmerger waveform and encode the impact of the EOS.

In [55], a matched-filtering approach for detecting the postmerger signal is described. The template bank consists of the analytic waveform model proposed in [50]. Additionally, in [56], the impact of transient noise artifacts with high-frequency components on the sensitivity of the search for GWs in the postmerger phase was considered. In [57], a frequency-domain analytical model was explored using wavelets in combination with empirical relations, where also a recalibration was applied and its feasibility for detecting the postmerger phase with third-generation detectors was studied in [16]. In addition, [58] presented another analytic model in the frequency domain, based on a parametrized Lorentzian function, for the complete coalescence process. Apart from individual detections, the possibility of combining information from a number of weak postmerger signals (too weak to be detectable individually) has been explored in [53], concluding that first indications for the presence of postmerger GW emission could come with the final upgrades of the second-generation detectors. Lastly, in [59], a morphology-independent method was presented to characterize the fate of the postmerger remnant.

A large number of detections of the inspiral phase of BNS mergers using third-generation detectors is expected to lead to tight EOS constraints; see [60] and references therein. However, using only the inspiral phase, one cannot probe EOS properties at densities higher than those encountered in the core of isolated neutron stars before they merge. In contrast, the detection of GWs in the postmerger phase will allow us to probe higher densities and test for the presence of phase transitions or exotic components [61–67] or deviations from general relativity [68–77].

Easter *et al.* [78] introduced a methodology comprising a hierarchical model of two main steps to predict postmerger spectra in the frequency domain. Here, we construct a similar model, but, in contrast to [78], we do not use an empirical relation for the compactness $C = M/R$, but, instead, we use the inverse gradient dR/dM of the mass-radius relation (where M is the gravitational mass and R the circumferential radius of a relativistic star). Furthermore, we extended the training set from 35 spectra in [78] to a total of 87 spectra. Apart from improved results with this hierarchical multivariate regression scheme, we are also interested in exploring the application of artificial neural networks (ANNs) as a regression scheme for predicting postmerger spectra. To this end, we train a feed-forward ANN on the same data and perform extensive comparisons between the two methods. Although, at the moment, the training set comprises a limited number of models, we obtain comparable or slightly better results with the ANN regression, showing that this is a very promising method to be applied in future searches, when many more simulation results will be available. Furthermore, using a recalibration of the results, we show that if the uncertainties

of the employed empirical relations will be reduced in the future (due to better EOS constraints obtain through GW observations in the inspiral phase), this will greatly improve the accuracy of our proposed method.

This work is organized as follows: In Sec. II, we present the two regression-based techniques used to construct predictive models, along with some statistical and machine learning (ML)-based concepts. In Sec. III, we outline the experiments conducted to assess the performance of the models, such as comparing the resulting fitting factor histograms before and after recalibration and a necessary cross-validation sensitivity analysis. Finally, in Sec. IV, we make some final remarks and draw conclusions. Some complementary results are depicted in collective spectra plots in the Appendix for the two models.

We note that in certain expressions, such as M/R, we use the standard convention of setting $c = G = 1$.

II. PREREQUISITES

Here, we introduce the applied machine learning and statistical techniques used to construct and analyze the predictive models of our work. Waveform model-based prediction is useful to construct templates that cover a wide range of possible system parameters. The statistical aspect of predictive modeling is primarily concerned with minimizing the error of a model in order to make predictions of future outcomes as accurately as possible.

A. Regression techniques for prediction

We mainly leverage two approaches: a classical multi-input and multitarget linear regression approach to formulate a multivariate linear model and, second, a more general function-approximating approach, based on an artificial neural network model, with various enhancements. Both approaches can fall into the regression-based description, since they obey the same mathematical equation [as in Eq. (1) below] connecting input and output; however, it is implemented differently in the two approaches.

1. Multivariate linear regression (MLR)

Multivariate linear regression is an extension of multiple linear regression allowing for more than one output variables, which may be correlated [79,80]. The multivariate linear model is

$$\underset{(n\times m)}{Y} = \underset{(n\times l+1)}{X} \underset{(l+1\times m)}{B} + \underset{(n\times m)}{E}, \qquad (1)$$

where Y is a matrix of n observations on m response variables (regressands), X is a model (design) matrix with columns for $l + 1$ regressors, typically including a column of ones for the regression constant, B is a matrix of regression coefficients, with one column for each response variable, and E is a matrix of residuals, where the rows E_i,

$i = 1, \ldots, n$, have all mean 0 and the same error covariance matrix, and they are uncorrelated to each other.

Assuming that $X'X$ is nonsingular and the residuals are Gaussian, it can be shown that the maximum-likelihood estimate of B in the multivariate linear model is equivalent to the least squares estimate for the individual responses:

$$\hat{B} = (X'X)^{-1}X'Y, \qquad (2)$$

where the hat indicates the estimated regression values of the coefficients. If the responses are assumed to be uncorrelated, the multivariate linear regression can be treated as many independent and stacked multiple linear regressions.

2. Artificial neural networks as universal approximators

Artificial neural networks, in contrast, are more general problem-solving models. For the multivariate regression problem, ANNs estimate any function $f: R^k \to R^m$, having as input a vector of k regressors and as output a vector of m responses. The formulation of a regression solution using an ANN bears a mathematical expression, which is complicated, with the complication increasing with the number of layers. In this expression, crucial nonlinearity is introduced by means of specialized activation functions applied to the outputs of individual neurons or layers within the network.

Although an ANN may consist of more than three layers of neurons, it has been mathematically shown that any functional relationship can, in practice, be implemented with a three-layer ANN [81–83]. Theoretically, according to the universal approximation theorem, a network with just one hidden layer and nonlinear activation functions should be sufficient to model any function mapping, but, in practice, additional layers are included. In our implementation of ANN, we use four layers and make further enhancements to improve network performance.

The first enhancement we used was a *learning rate* (LR) scheduling scheme [84–86], involving a nonmanual tuning of the LR value. Typically, we want the LR value not to be too low, since the neural network would learn slowly, but also not too high, since this would result in divergence issues. The strategy we followed suggests starting with a low LR (warm-up), which gradually increases (LR annealing), followed by a slow LR decrease for a certain number of steps when the loss function reaches a plateau, to achieve the best possible loss value (local minima). Moreover, introducing a warm-up period in the LR can reduce the initial variance of the training procedure and further stabilize it.

Another enhancement, which is common in statistics, is *standardization* (that is, subtracting the mean and dividing with the standard deviation) of the input values fed into the neural network (and the same for the output values during training) and the reverse procedure of bringing the

predicted output values back into the original domain. For each variable, the mean and standard deviation are computed from the input or output training samples and divided by the respective standard deviation. The same mean and standard deviation are then used in the reverse procedure on the output values.

Lastly, one can use *early stopping* while training the ANN, a technique that stops the training procedure when a low plateau is reached in the corresponding validation loss, calculated using a separate set other than the training one. In this way, the prediction accuracy and the training time can improve. Furthermore, to ensure that the ANN is able to generalize and mitigate overfitting effects in deployment, other techniques can be employed, such as noise addition in the output of constituting layers [87] and random dropout of nodes [88,89].

B. Statistical design of experiments

In this subsection, we briefly describe some statistical concepts used in our work to efficiently design statistical experiments, such as hierarchical or multilevel regression models and unbalanced designs. Reference [90] provides a thorough and general discussion on the subject.

1. Hierarchical or multilevel models

Some data settings require a hierarchy of regressions, the so-called hierarchical linear models (also known as multilevel models), for instance, by regressing X on Y and Y on Z (a two-level approach).

These kinds of designs are called hierarchical for two reasons [91]: first, due to the structure of the data and, second, because of the model itself, which may demonstrate its own hierarchy with within-group regression parameters controlled by hyperparameters of some upper-level model. Depending on whether these parameters are considered as random variables or not, we distinguish between random-effect and fixed-effect models.

2. Balanced versus unbalanced design

During the design of a regression solution, data usually form groups based on a common characteristic, as frequency series belonging to the same EOS as far as this work is concerned.

When there is an equal number of observations in each group, we can infer that the design matrix of the regression will be homogeneous and its values (in our case columns) evenly distributed, namely, *balanced*. Conversely, when there is not an equal number of observations in each group, the design will be inhomogeneous, something that will lead to an *unbalanced* design matrix.

Balanced designs are often preferred over unbalanced ones, due to their higher statistical power and reliable test statistic. However, in many practical problems, this is not a strong constraint on the applicability of a solution, such as linear regression.

C. The fitting factor

To compare the predicted waveform, denoted $h_1(t)$, with the corresponding spectra of the original waveform $h_2(t)$ using our methodology, we use a noise-weighted index called the fitting factor (FF), which denotes their overlap, defined as

$$\mathrm{FF}(h_1, h_2) \equiv \frac{\langle h_1 | h_2 \rangle}{\sqrt{\langle h_1 | h_1 \rangle \langle h_2 | h_2 \rangle}}, \qquad (3)$$

where the (approximate) inner product is (as introduced in [78])

$$\langle h_1 | h_2 \rangle \equiv 4 \int_0^\infty df \frac{|\tilde{h}_1(f)||\tilde{h}_2(f)|}{S_h(f)} \qquad (4)$$

and $\tilde{h}(f)$ is the Fourier transform of a waveform $h(t)$. In Eq. (4), $S_h(f)$ is the noise power spectral density (PSD), averaged over a sufficiently long time segment. We will assume that the PSD is a stationary, colored Gaussian noise model. A good match (or overlap) between the two spectra is indicated with values of FF close to 1.

We note that this definition of FF was introduced in [78], since we are dealing with only the amplitudes (and not the phases) of the Fourier transform, and differs from the full definition introduced in [92].

D. Cross-validation

Cross-validation (CV), or out-of-sample testing, is one of the most popular approaches for model evaluation [93]. Usually, the k-fold CV is partitioning the dataset into k roughly equal parts, training the model with $k - 1$ of them and testing it using the one excluded. This procedure is repeated k times, covering, in this way, the entire dataset.

A subcase of k-fold CV is the so-called *leave-one-out* CV (LOO-CV), when k equals the cardinality of the dataset. This approach is employed in [78]. However, various trade-offs emerge using a k-fold CV, and many of them are still under debate [94–96] with one of the most famous being the bias-variance trade-off with respect to k value. Thus, in this work, we will conduct a detailed k-fold cross-validation study, varying k from 2 to 87.

III. EXPERIMENTS

In this section, we describe the experiments conducted to assess the robustness and efficiency of our proposed extended methodology. First, the dataset is described. Then, implementation-specific details about the MLR and ANN regression models are given, and, finally, the two models are compared.

TABLE I. Description of the training dataset of 87 waveforms. The first column lists the EOS name, the second column lists the number of different models for each EOS, the third column gives the main reference for each case, and the last column lists the component masses of the individual equal-mass models for each EOS. For the APR4, H4, and SLy, the Core v2 database was used above $1.325 M_\odot$.

EOS	Waveforms	References	Component masses (in $M_\odot$)
ALF2	10	Rezzolla and Takami [45] and CoRe v2 [27]	1.2, 1.225, 1.25, 1.275, 1.3, 1.325, 1.35, 1.3505, 1.375, 1.3755
APR4	7	Rezzolla and Takami [45]	1.2, 1.225, 1.25, 1.275, 1.3, 1.325, 1.35
BHBlp	4	CoRe v2 [27]	1.25, 1.3, 1.35, 1.4
BLh	4	CoRe v2 [27]	1.3, 1.3325, 1.364, 1.4
DD2	7	CoRe v2 [27]	1.2, 1.25, 1.3, 1.35, 1.364, 1.4, 1.5
ENG	1	CoRe v2 [27]	1.3495
GNH3	7	Rezzolla and Takami [45]	1.2, 1.225, 1.25, 1.275, 1.3, 1.325, 1.35
H4	13	Rezzolla and Takami [45] and CoRe v2 [27]	1.2, 1.225, 1.25, 1.275, 1.3, 1.325, 1.3495 1.35, 1.3505, 1.3715, 1.3725, 1.3735, 1.3795
LS220	4	CoRe v2 [27]	1.2, 1.35, 1.364, 1.4
MPA1	8	Soultanis *et al.* [97]	1.2, 1.25, 1.3, 1.35, 1.4, 1.45, 1.5, 1.55
MS1	2	CoRe v2 [27]	1.3495, 1.351
MS1b	8	CoRe v2 [27]	1.35, 1.3505, 1.375, 1.3805, 1.381, 1.5, 1.6, 1.7
SFHo	2	CoRe v2 [27]	1.35, 1.364
SLy	10	Rezzolla and Takami [45] and CoRe v2 [27]	1.2, 1.225, 1.25, 1.275, 1.3, 1.325, 1.35, 1.351, 1.3575, 1.364

A. Dataset description

For both regression-based models, we have to provide data for both the regressors (input) and the regressands (output). The columns of the design matrix are populated with data pertaining to three regressors [$k = 3$ in Eq. (1)], which are the component mass M in an equal-mass ($q = 1$) binary system, the tidal coupling constant κ_2^τ (expressing tidal deformability), and the derivative dR/dM, which corresponds to the inverse gradients of the $M(R)$ relation of nonrotating equilibrium models for a given EOS. In the case of the MLR model, the design matrix includes a first row consisting of unit values. The response matrix Y is formed by the amplitudes of the postmerger waveforms in the Fourier domain.

In Table I, we list a detailed description of the 87 different waveforms used in our study. Most of the waveforms are from the CoRe v2 database [27] (47 waveforms) and the Rezzolla and Takami catalog [45] (32 waveforms), while eight waveforms are from Soultanis *et al.* [97]. All waveforms were scaled to a common distance of 50 Mpc truncated at their maximum amplitude to separate the inspiral and postmerger phases. The postmerger part was then transformed to the Fourier domain, and the resulting spectra were partially aligned using the empirical relation of Eq. (5).

Lastly, we confined the frequency range of the resulting spectra to the astrophysically relevant domain of 1–4 kHz, leaving some tolerance due to the shift-unshift procedures

(depending on the maximum displacement of the spectrum). The total number of frequency bins is $m = 370$.

B. Alignment of postmerger spectra

A common feature of all postmerger spectra (irrespective of the source parameters) is that they are dominated by the f_{peak} frequency. In [32], this feature was used to align the spectra to a common reference frequency, which was useful for applying a principal component analysis. Similarly, the hierarchical method of [78] includes an alignment step, which needs to be reversed after the regression is performed. Here, we also align the spectra to a common reference frequency, which we take to be the average value of f_{peak} in our dataset. For given values of the tidal coupling constant κ_2^τ and component mass M for an equal-mass system, we predict f_{peak} by the empirical relation

$$f_{\text{peak}}(\kappa_2^\tau, M) = 4 \frac{\beta_1}{M} \ln\left(\frac{\beta_0}{8\kappa_2^\tau}\right). \quad (5)$$

Equation (5) is based on the empirical relation Eq. (41) in [52] but is written here differently and using the known relation between mass-weighted tidal deformability $\tilde{\Lambda}$ and κ_2^τ, as well as the relation between chirp mass M_{chirp} and component mass M for equal-mass systems. The fitting values for β_0 and β_1 were recalculated, since a single composite variable was used in [52].
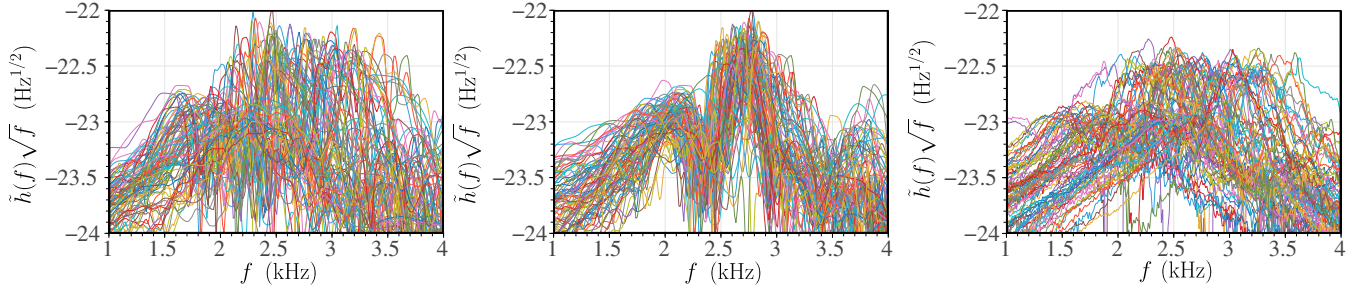
FIG. 2. Left panel: collective plot of the 87 different postmerger spectra used in both the MLR regression and the ANN method. Middle panel: the same spectra, but partially aligned to a common f_{peak} frequency, using the empirical relation of Eq. (5) (departures from a perfect alignment are due to uncertainties in the empirical relation). Right panel: corresponding predicted spectra, using the ANN-based method, after recalibration (see the text for details).

The left panel in Fig. 2 shows a collective plot of the 87 different postmerger spectra we use in our study. The value of f_{peak} can be as low as ~1.5 kHz or as high as ~3.5 kHz. The middle panel in Fig. 2 shows the same spectra but aligned using the empirical relation in Eq. (5). Because of the uncertainties of the empirical relation, the alignment is only partial; i.e., the spectra are not perfectly aligned according to their actual peak values. However, the empirical relation allows the alignment to be reversed after the regression is performed.

The alignment shifts the minimum and maximum frequency of each available spectrum. To be able to use the data as input training data, we truncate the spectra to a common minimum and maximum frequency. However, our sampling rate is sufficient so that this truncation does not affect the main part of the spectrum, where the main postmerger frequency is found.

The partially aligned spectra, in the middle panel in Fig. 2, serve as training data for both the MLR and ANN regression models presented in Sec. III C. The right panel in Fig. 2 shows the corresponding predicted spectra (with the alignment reversed), using the ANN-based method and after applying the recalibration procedure discussed in Sec. III D.

C. Regression models

In this subsection, we describe in detail the implementation of the two different regression-based models on the dataset of 87 waveforms.

1. MLR-based model

The implementation of this model is quite straightforward, since the model produced is determined solely by the regression coefficient matrix $\hat{B}$. Taking into account the fact that multivariate linear regression is equivalent with separate and independent multiple linear regressions (multi-input, single output), provided that responses are not correlated, one can multiply $\hat{B}$ with any chosen column

vector of four elements (that is, $[1, dR/dM, M, \kappa_2^\tau]$) to obtain an output vector that will represent the strain.

For the training, we consider the spectral amplitude in linear scale (in the various figures, we show the base-10 logarithm of the spectra). For the regressors, that is, the columns of the design matrix X, we performed a standardization of the values along each row of X.

2. ANN-based model

We used a four-layer (three hidden plus output), feed-forward ANN, with sigmoid transfer functions (activations) in the hidden layers (comprising between 200 and 400 nodes each) and linear in its input and output ones. Additionally, an Adam optimizer with a batch size of six samples was used and training was completed in at most 100 epochs, due to early stopping. The optimizer itself introduces a small stochasticity in the final FF values between different trainings. We implemented the algorithm using the TensorFlow framework [98]. The ANN architecture is summarized in Table II. To ensure the generalization of the proposed architecture, we also

TABLE II. Artificial neural network architecture, invoking the summary function of the TensorFlow library [98].

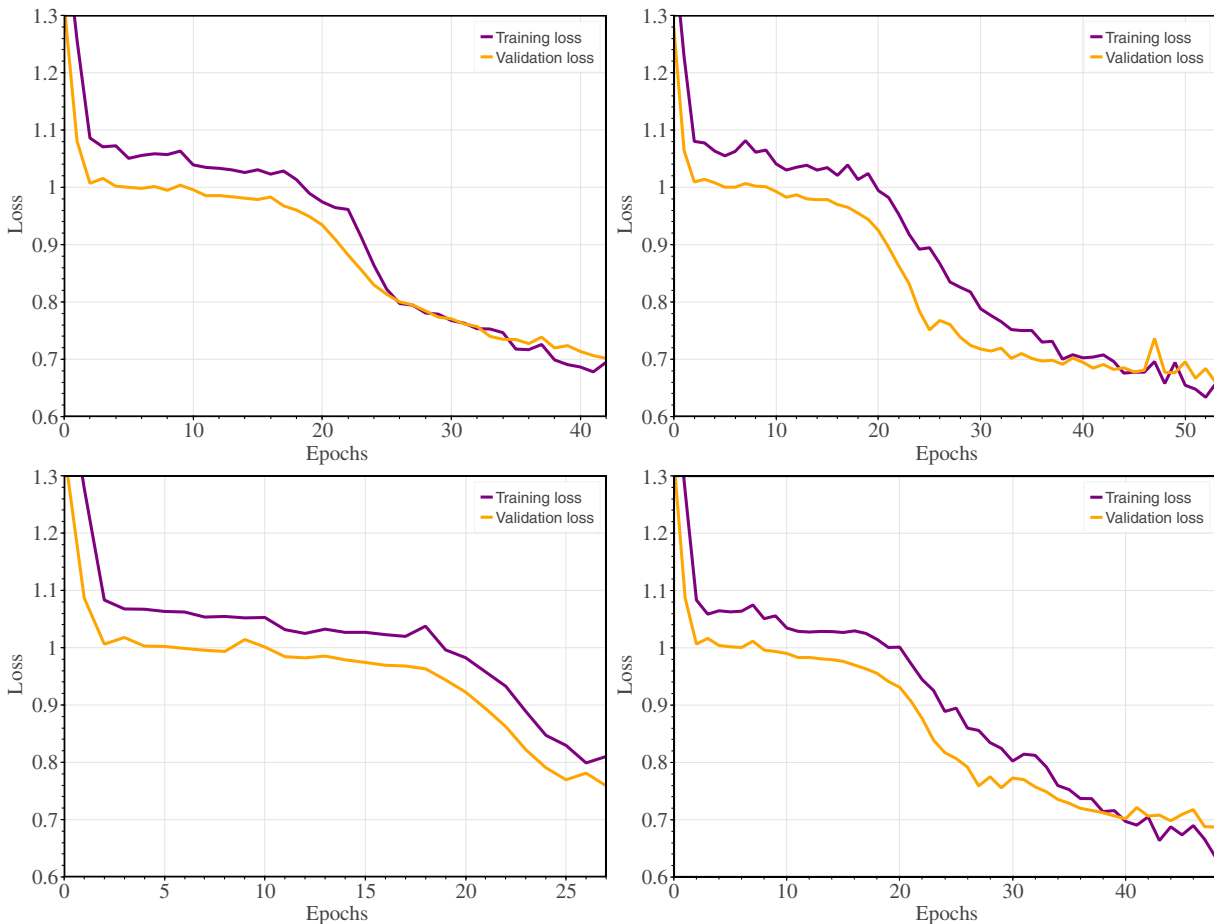| Layer	Type	Shape	Activation	Params
1	Gaussian noise (0.1)	(None, 3)	$\cdots$	0
2	Dense	(None, 200)	Linear	800
3	Gaussian noise (0.05)	(None, 200)	$\cdots$	0
4	Dropout (0.15)	(None, 200)	$\cdots$	0
5	Dense	(None, 400)	Sigmoid	80400
6	Gaussian noise (0.1)	(None, 400)	$\cdots$	0
7	Dropout (0.15)	(None, 400)	$\cdots$	0
8	Dense	(None, 400)	Sigmoid	160400
9	Gaussian noise (0.1)	(None, 400)	$\cdots$	0
10	Dropout (0.05)	(None, 400)	$\cdots$	0
11	Dense	(None, 370)	Linear	148370

FIG. 3.   Training and validation loss as a function of the number of epochs, shown for four different shufflings of the training set. The training was stopped when the validation loss rose above the training loss for seven consecutive epochs.

integrated layers adding Gaussian noise, as well as dropout layers (which also introduce small stochasticity). The respective settings for these additional layers are also depicted in Table II. Similarly, Gaussian noise was added to the validation set.

During training, we used a learning rate scheduler, which starts with a low LR (for warm-up) and increases the LR linearly in each step of the training process. The warm-up period lasted approximately four epochs. Finally, we scaled the input of the network using an established standardization technique and inverted the result of the output (unscale). This standard scaling was also used for both the regressors and the regressands of the training set and of the validation set as well.

To train the network, we gave pairs of individual columns of the design matrix X (regressors) in combination with the respective columns of the regressands matrix Y, which represented a specific spectrum. To be more precise, we feed the network with X's column vectors omitting the first entry (with constant value of 1), since the bias is already included in the ANN architecture. The hyper-parameters were optimized heuristically.

Furthermore, to avoid any bottleneck in the architecture of the ANN, the consecutive layers have an increasing number of nodes, except the last one having the number of nodes equal to the frequency bins of the spectra.

In Fig. 3, we depict exemplary learning curves for this ANN model to diagnose its behavior. The validation loss curve follows the training loss curve and lies mostly below it, implying that the dataset is representative [81,99], allowing for a good fit, from an ML perspective. We can also observe in these curves that the loss value does not exhibit a significant percentage decrease, because it happens to be in the same range of initialized network weights (we used the default initialization in TensorFlow) since we did not use any initialization function. Lastly, the training procedure is stopped when the validation loss curve diverges for seven epochs consecutively.

Before feeding the spectra into the ANN for the training procedure, we performed the same partial alignment (shift) procedure as for MLR. This renders structure in the training dataset (with respect to the regressands), not captured by the three variables we chose in the design matrix. This may not have been necessary, if we had a
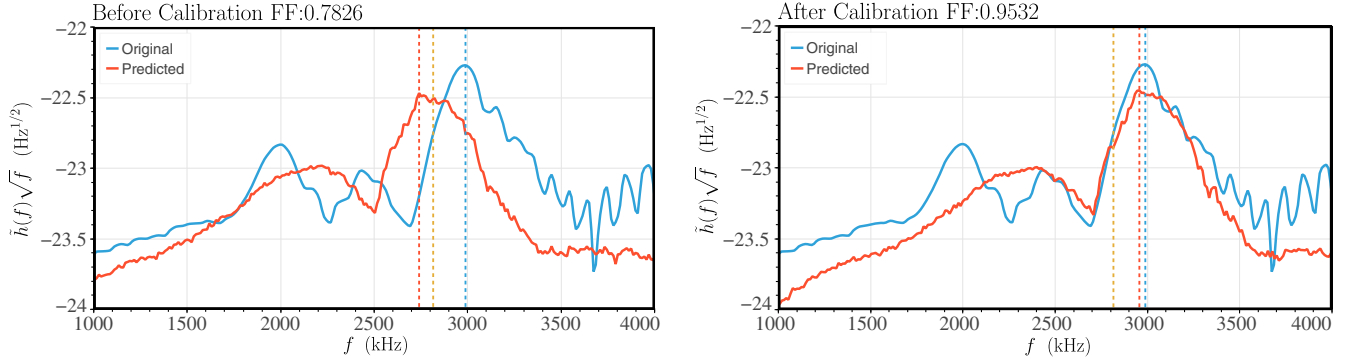
FIG. 4.　Recalibration of the predicted postmerger spectrum for a $1.364M_\odot$ equal-mass BNS merger with EOS LS220. The left panel shows the ANN-based prediction before calibration, where the main postmerger peak is misaligned with respect to the original spectrum. The right panel shows the calibrated spectrum, which has a much better alignment, leading to a significantly higher fitting factor of 0.953.

larger training set and additional variables in the design matrix (e.g., including the maximum mass and the radius at the maximum mass). For example, in the time-domain model of [97], a larger number of parameters was used to achieve high fitting factors.

As a final remark, the ANN produces a predicted spectrum, by nonlinearly entangling information from all frequency bins, which is not the case for the MLR model, where at each frequency bin, a linear regression, independent of other frequency bins, is performed. The fact that both methods yield comparable results in many examples shows that secondary peaks can be captured well and that the loss function of the ANN method is not entirely dominated by the peak frequency.

D. Recalibration of predicted spectra

Since we are using an empirical relation for the partial alignment of the spectra, the inverse procedure induces an error in the predicted spectrum that depends on the accuracy of the empirical relation. To test the impact of this error, we performed a "recalibration," as in [57]. In this way, one can reverse, to some extent, the horizontal or vertical misalignment of a predicted spectrum with respect to the original one. This is useful, as it shows the real potential of our method for predicting postmerger spectra in the future, when more training samples will be available and, in addition, the empirical relation will become more accurate, due to tightening EOS constraints derived from astrophysical observations (see, e.g., [100]).

In [57], several parameters were used for recalibration. Here, we use only the known peak frequency of the postmerger spectra in the training set. We, therefore, define as Δf_{peak} the difference between the peak frequency of the original spectrum and the peak frequency of the predicted spectrum after calibration. Then, we define the likelihood function

$$\mathcal{L}(\Delta f_{\text{peak}}) = \sum_{f_{\text{min}}}^{f_{\text{max}}} [h_{\text{o}}(f) - h_{\text{p}}(f - \Delta f_{\text{peak}})]^2, \quad (6)$$

where h_{o} is the original strain spectrum and h_{p} the predicted strain spectrum, while f_{min} and f_{max} define the frequency range for recalibration. This likelihood is basically a mean-squared-error function that is minimized close to the value of Δf_{peak}, through the maximum-likelihood estimation technique. This is equivalent to Bayesian inference using uninformed priors. In practice, we use 60 frequency values in the range 1–4 kHz. An indicative example of recalibration is shown in Fig. 4.

In the Appendix, we display collective plots of predicted versus original spectra, produced using a *k*-fold CV with $k = 4$ (each plot shows a hold-out case, not included in the corresponding training sample)[2]; see Figs. 7–9. In all figures, the calibrated cases are shown. The predicted spectra are, in most cases, close to the predicted ones, even in the low- and high-frequency tails in this part of the postmerger spectrum. There is good alignment of the main frequency peaks, and in most cases the shape of one or more secondary peaks is also predicted with satisfactory accuracy. In some cases, the main peak in the original spectra is quite narrow, which is due to a long-lived f_2 oscillation. This feature cannot yet be captured by our three-parameter model, and the predicted spectra show a somewhat wider peak, with smaller maximum amplitude. This could be remedied in future improved models, by including more parameters.

E. Distribution of fitting factors

In the left column in Fig. 5, the calculated histograms of the fitting factors FF are depicted in the uncalibrated case for the MLR and ANN models, respectively. We set the

---

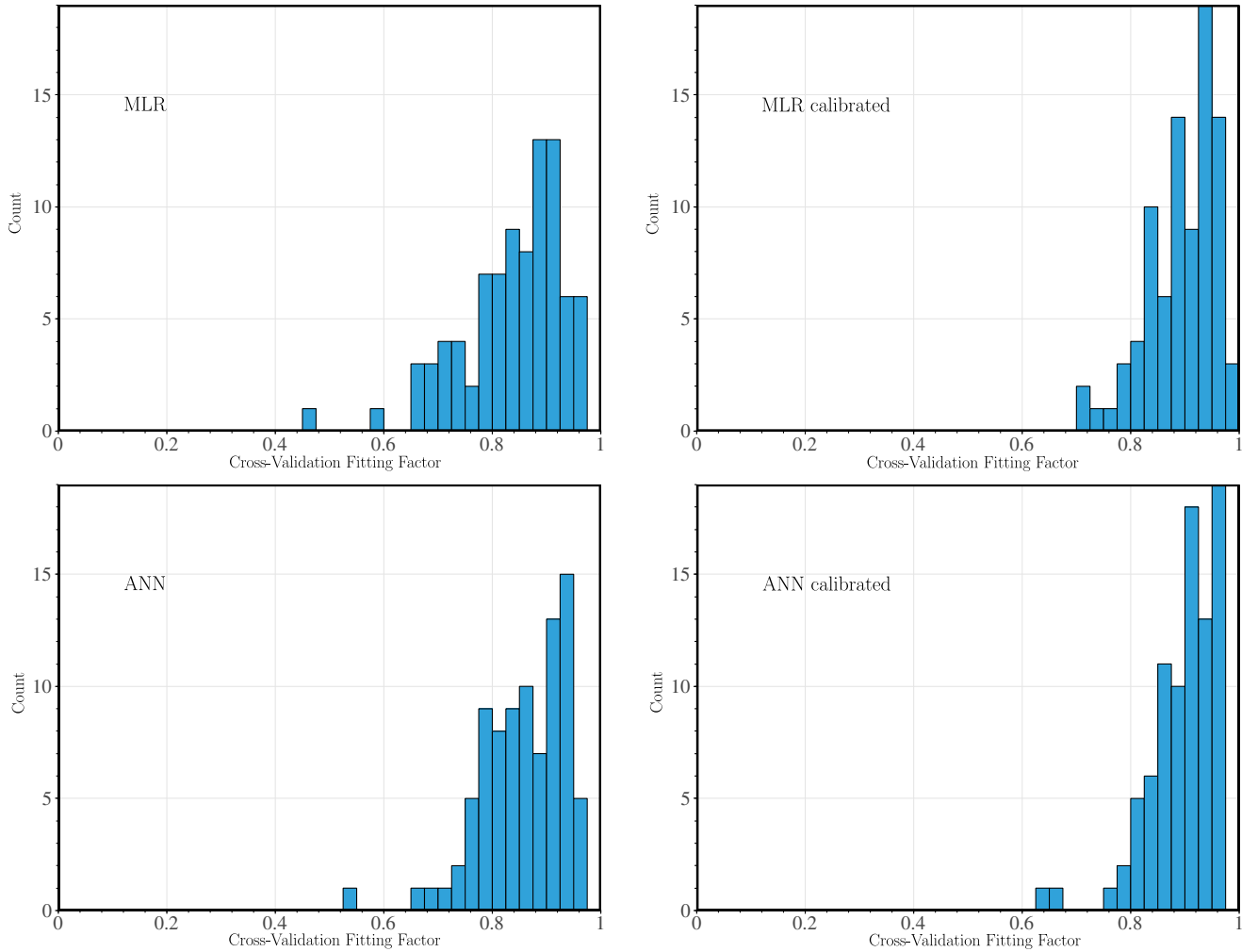[2]The choice of $k = 4$ is common in such investigations.

FIG. 5. Histograms of fitting factors for predicted postmerger spectra when using the MLR (top row) and the ANN model (bottom row), for $k = 4$ cross-validation. The left column corresponds to the cases without recalibration, while in the right column recalibration was applied (see the text for details). The distribution of the fitting factors obtained with the ANN models is more concentrated toward one than for the MLR model.

number of bins to 40 in each case and the number k of CV folds equal to 4.

In the uncalibrated case, we observe that the fitting factor distributions of the two models have a similar shape, but the histogram is more concentrated to higher values for the ANN model. Specifically, for the MLR model, there are 18 fitting factor values below 0.8, whereas for the ANN model, this reduces to only 11. Moreover, the mode of the ANN histogram is closer to 1, and there are fewer outliers.

On the other hand, after recalibration both histograms greatly improve, as shown in the right column in Fig. 5. The mode increases, the histogram becomes narrower with values closer to one, and there are even fewer outliers, compared to the uncalibrated case. Comparing the MLR-calibrated histogram of fitting factors with the ANN-calibrated one, we observe that

the latter is more concentrated closer to one, except for two outliers.

In addition to the comparison of the fitting factor histograms, we also investigate the behavior of the models with respect to changing the value of k in the CV k-fold. In Fig. 6, the histograms of the fitting factors are displayed as separate violin plots. We observe that, irrespective of individual cases of k, the overall behavior of the ANN models is slightly better than that of the MLR, since the mean FF value across all k's is higher and less dispersed. Regarding the dispersion of the FF values in the MLR case, we observe that the interquartile range decreases slightly as the number of folds k increases and becomes minimum when using leave-one-out cross-validation, leading to concentration of values in higher histogram bins (bias-variance trade-off). The ANN model is more stable in terms of dispersion of FF values and exhibits similar behavior
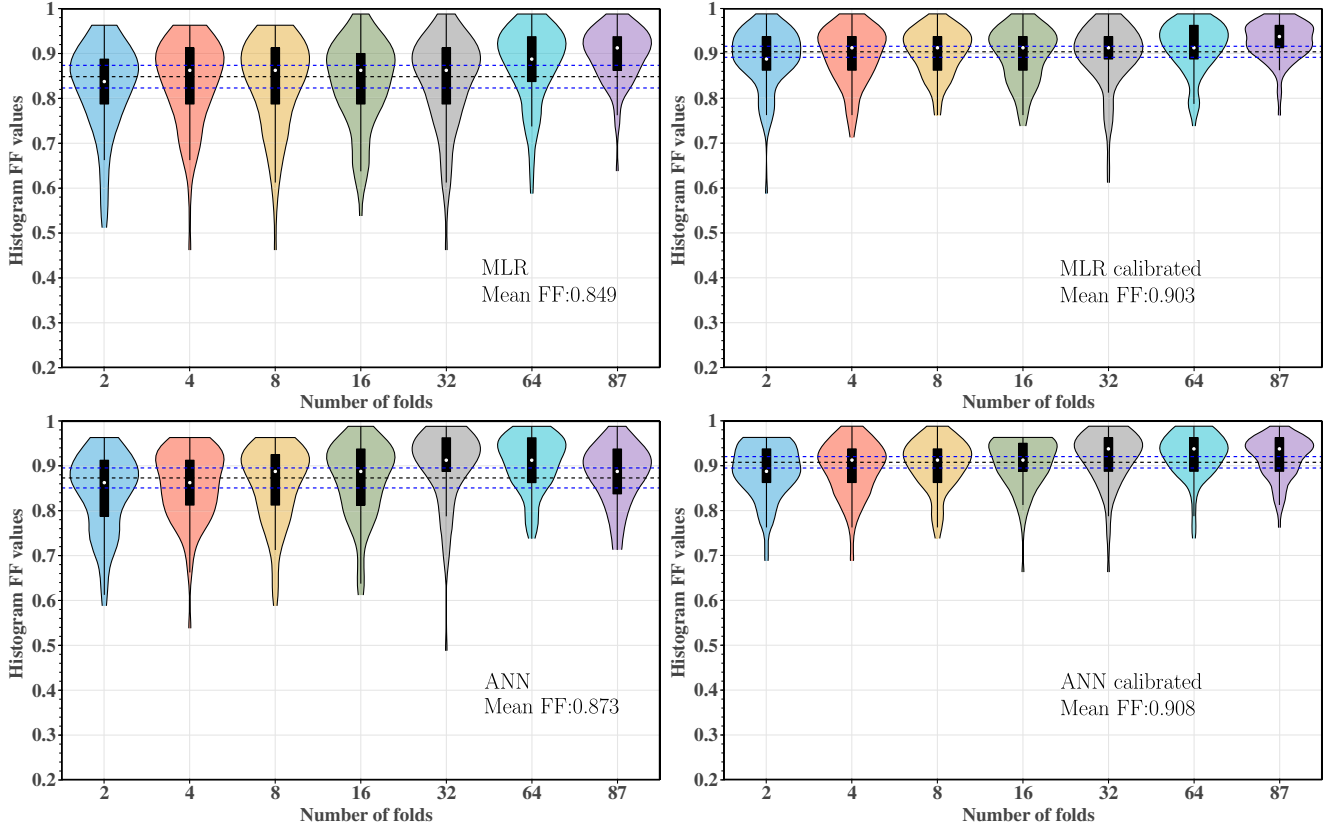
FIG. 6.   Violin plot representations of the distribution of fitting factors for the MLR model (top row) and the ANN one (bottom row) for different number of folds k in k-fold cross-validation. The left column corresponds to the cases without recalibration, while in the right column recalibration was applied (see the text for details). The mean fitting factor among all the number of folds is shown as a dashed black horizontal line, and the two dashed blue horizontal lines represent one standard deviation. The highest mean fitting factor (0.908) is obtained with the ANN-calibrated model.

when varying k. The MLR model is significantly more susceptible to outliers than the ANN model. We note that, to make the comparisons, we first performed a stable shuffling in the training dataset for both the regressors and the regressands.

IV. CONCLUSIONS

In this work, we presented two robust methods for predicting the GW postmerger spectrum for binary neutron star mergers. One method uses multivariate least-squares regression, whereas the second method uses artificial neural networks.

Concerning the multivariate least-squares regression, our work improves on the previous work by Easter *et al.* [78] in the following points: (a) In the design matrix, we use the inverse gradient dR/dM of specific points in the $M(R)$ plots instead of the compactness $C = M/R$. (b) We use a refined empirical relation for the postmerger peak frequency in terms of other characteristics of neutron stars, presented in Vretinaris *et al.* [52], instead of a different relation used in [78]. (c) We use an extended dataset of 87 different spectra, compared to 35 in [78]. These changes

lead to higher fitting factors of the predicted spectra. Our approach is robust, taking into consideration that we used an unbalanced design matrix.

In addition, we demonstrate that a trained ANN can predict the postmerger spectra with higher mean fitting factors compared to the MLR model. Specifically, we used a four-layer feed-forward ANN (with the Adam optimizer), comprising between 200 and 400 nodes in each layer and with additional Gaussian noise layers and dropout ones (which introduce a small stochasticity).

We evaluated the accuracy of the two methods in predicting postmerger spectra using the standard fitting factor and performed several k-fold cross-validations. For both the $k = 4$ case (which corresponds approximately to a 75%–25% train/test split) as well as the $k = 87$ case (which corresponds to the leave-one-out method), we find that the histogram of fitting factors is comparable between the two methods, when calibration is performed. However, we note a slightly higher mean fitting factor (across all k-folds) when using the ANN-based approach.

The distribution of fitting factors when using the recalibrated spectra is significantly improved with respect to the uncalibrated predictions. Essentially, the recalibrated

results remove, to some degree, the uncertainties introduced by the use of the empirical relation in aligning the spectra during training. We note that postmerger spectra are realistically expected to be observed with third-generation detectors, which will also significantly constrain the EOS using information from the inspiral phase; see, e.g., the anticipated constraints in [60], when assuming 500 BNS observations with third-generation detectors. In parallel, the available collections of postmerger BNS spectra produced by numerical-relativity simulations is expected to be significantly enlarged in the time it will take for third-generation detectors to become operational. These two improvements will allow us to significantly reduce the uncertainty of the empirical relation, and, thus, our calibrated results point toward the anticipated accuracy of our method at the time when actual observations will be obtained.

Possible future improvements include the extension of the design matrix by adding, e.g., quadratic terms or additional physical characteristics. Furthermore, the design matrix could take into account specific EOS information, leading to categorical regression.

The methods presented here can be used to create template banks of postmerger spectra, which will be useful for detecting this phase after a BNS event. On the other hand, the methods could be inverted, allowing the estimation of the parameters included in the design matrix, given a potential observation (see also [78]). We are planning to investigate this application in forthcoming work.

In the future, a significantly larger number of available simulations will allow us to also investigate in detail the level of systematic errors introduced by, e.g., the choice of space-time formulation (full general relativity or conformal flatness approximation), the EOS formulation (tabulated finite-temperature or hybrid), and the resolution of the simulation.

In this initial investigation, we have examined a basic configuration for the BNS simulations, focusing solely on hydrodynamics while disregarding the influence of magnetic fields, neutrino transport, and other forms of dissipative effects or effective viscosities [101–112]. It will be important to conduct comprehensive analyses that incorporate all relevant physical factors to have a faithful representation of the postmerger spectra. In addition, more work is needed to produce a large number of training spectra with unequal masses. Finally, future studies will need to include the potential impact of departures from general relativity on the postmerger spectra produced during BNS mergers.

APPENDIX: COLLECTIVE PLOTS OF MODELS PREDICTIONS

In Figs. 7–9, individual predictions of the ANN-based model are depicted for each EOS and mass using four-fold CV.

In particular, in Fig. 7, we present the predictions for the Easter *et al.* dataset [78] of numerical simulations, except for the last row, that is, H4 and SLy EOS for the component mass $1.35 M_\odot$, where we used simulations from the CoRe dataset [27]. Similarly, Figs. 8 and 9 collectively represent predictions for spectra that were taken from the CoRe dataset [27] and Soultanis *et al.* [97], respectively. In these plots, the blue and red dashed vertical lines correspond to the peak value of the original and the predicted spectra, respectively, whereas the orange dashed vertical line represents the prediction of the empirical relation Eq. (5).

Finally, Figs. 10–12 show the corresponding spectra when using the MLR-based approach.
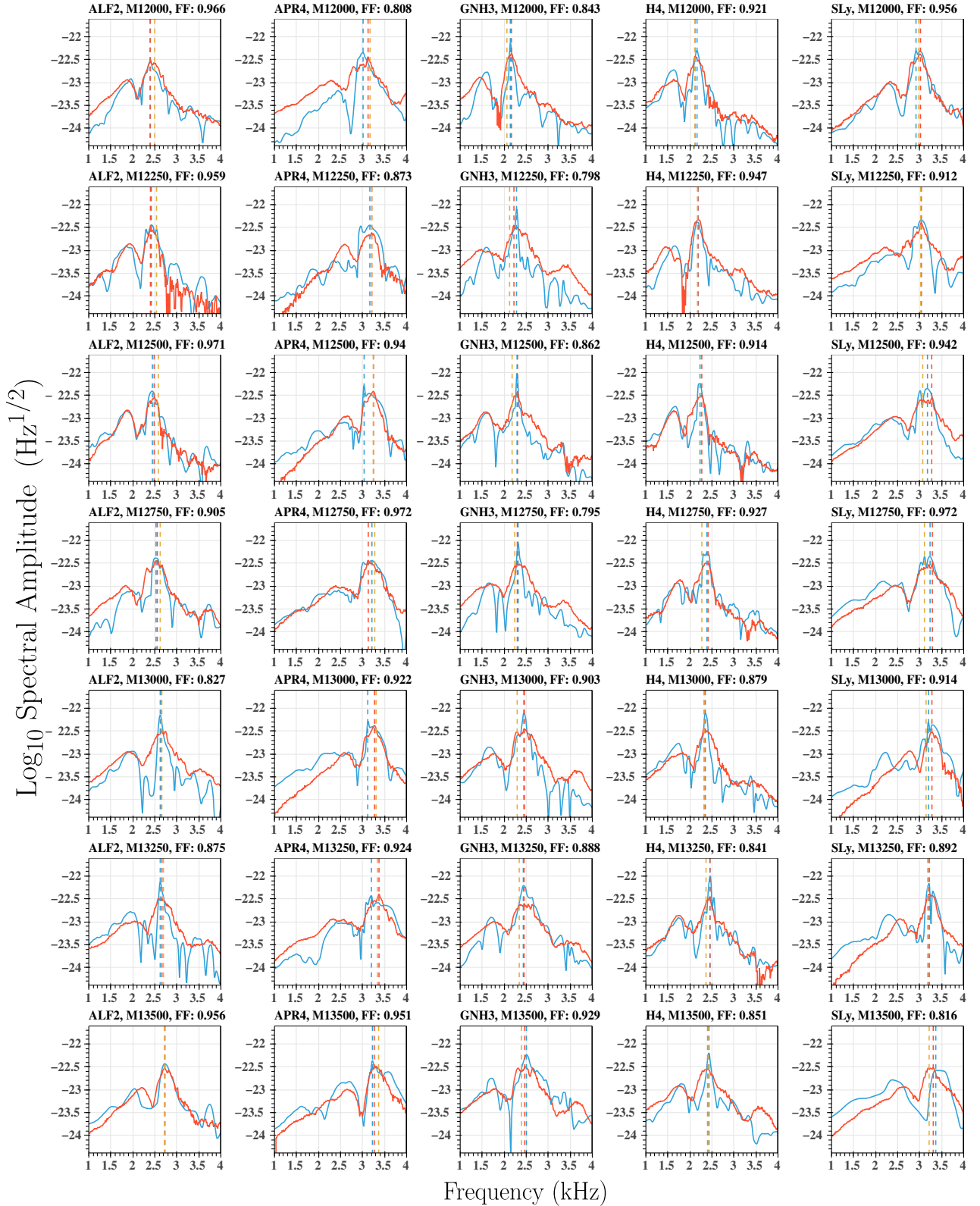
FIG. 7. Original gravitational-wave spectra of BNS mergers (blue curves) depicted along with their ANN-based model predicted ones (red curves) for the Rezzolla *et al.* dataset of numerical simulations [78], except for the H4 and SLy cases with component mass of $1.35 M_\odot$, which were chosen from the CoRe database [27]. The blue and red dashed vertical lines correspond to the peak value of the original and the predicted spectra, correspondingly. The orange dashed vertical line represents the prediction of the empirical relation Eq. (5).
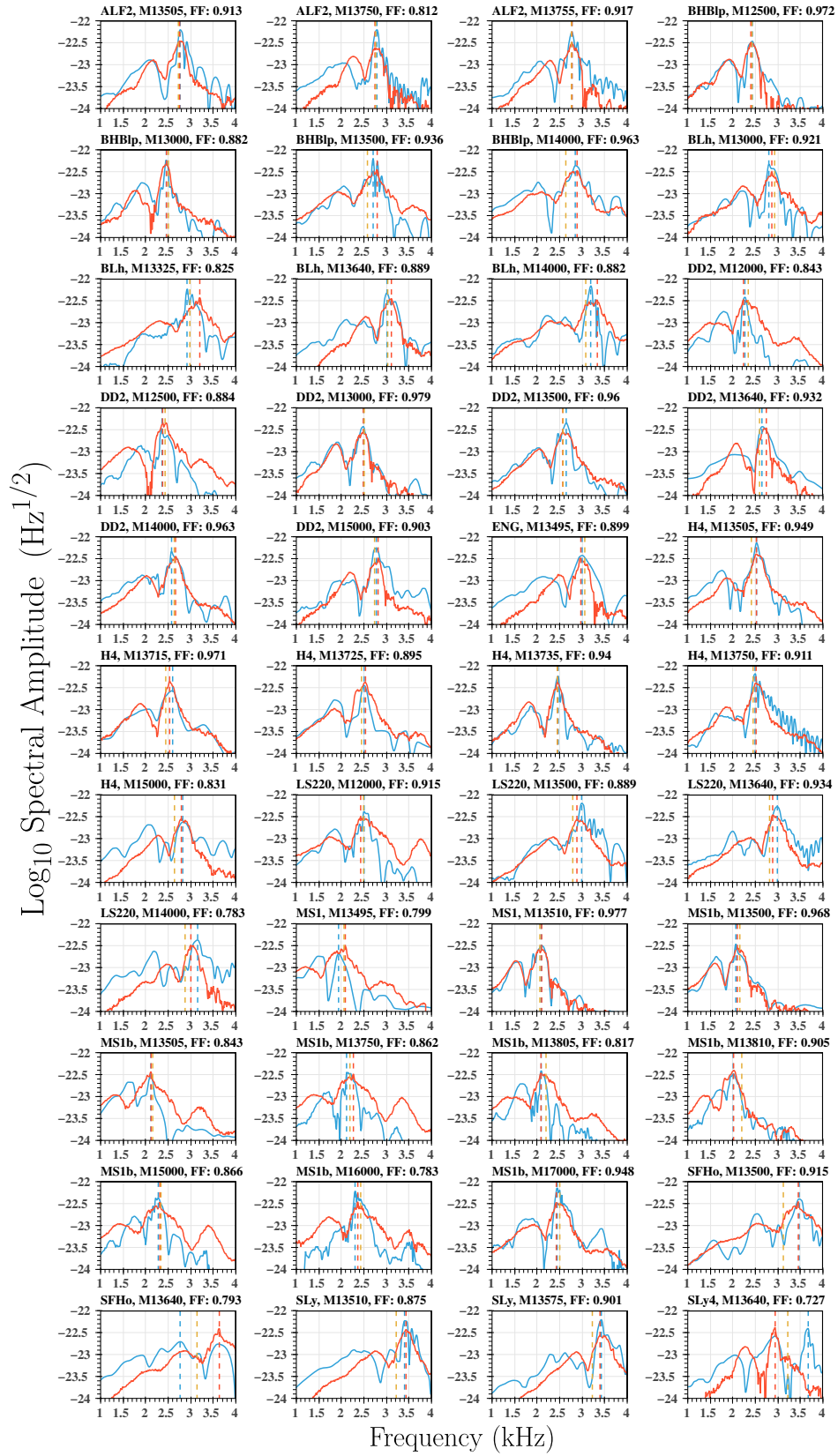
FIG. 8.   The same as Fig. 7, but for models included in the CoRe [27] dataset.
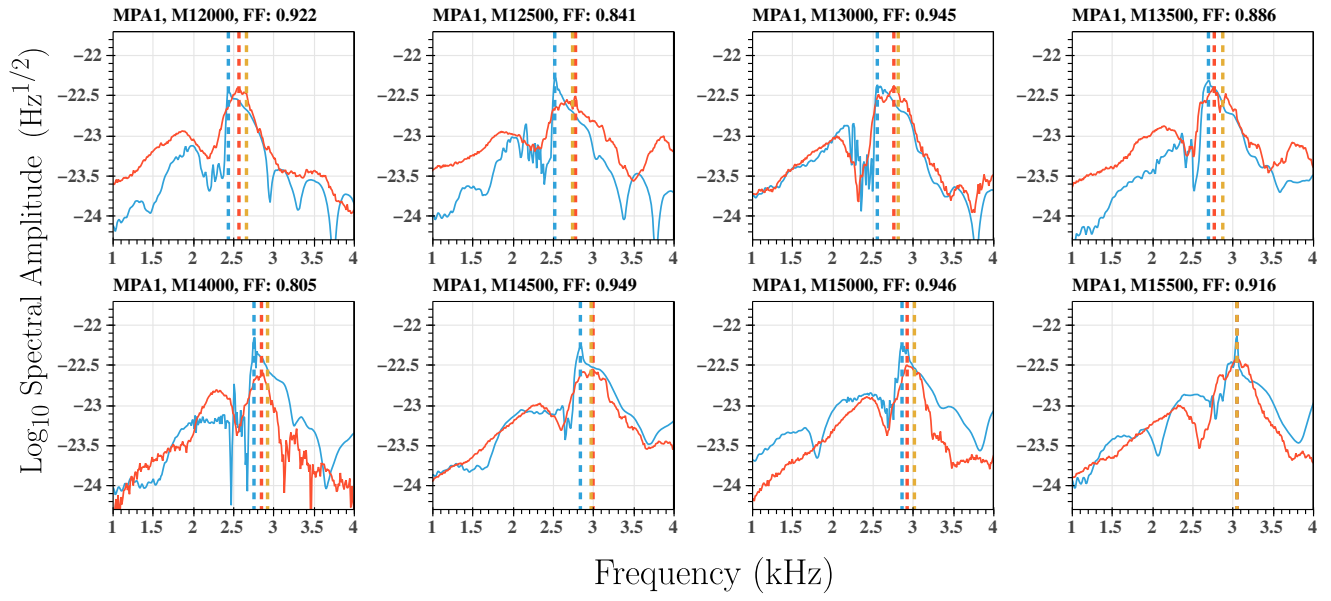
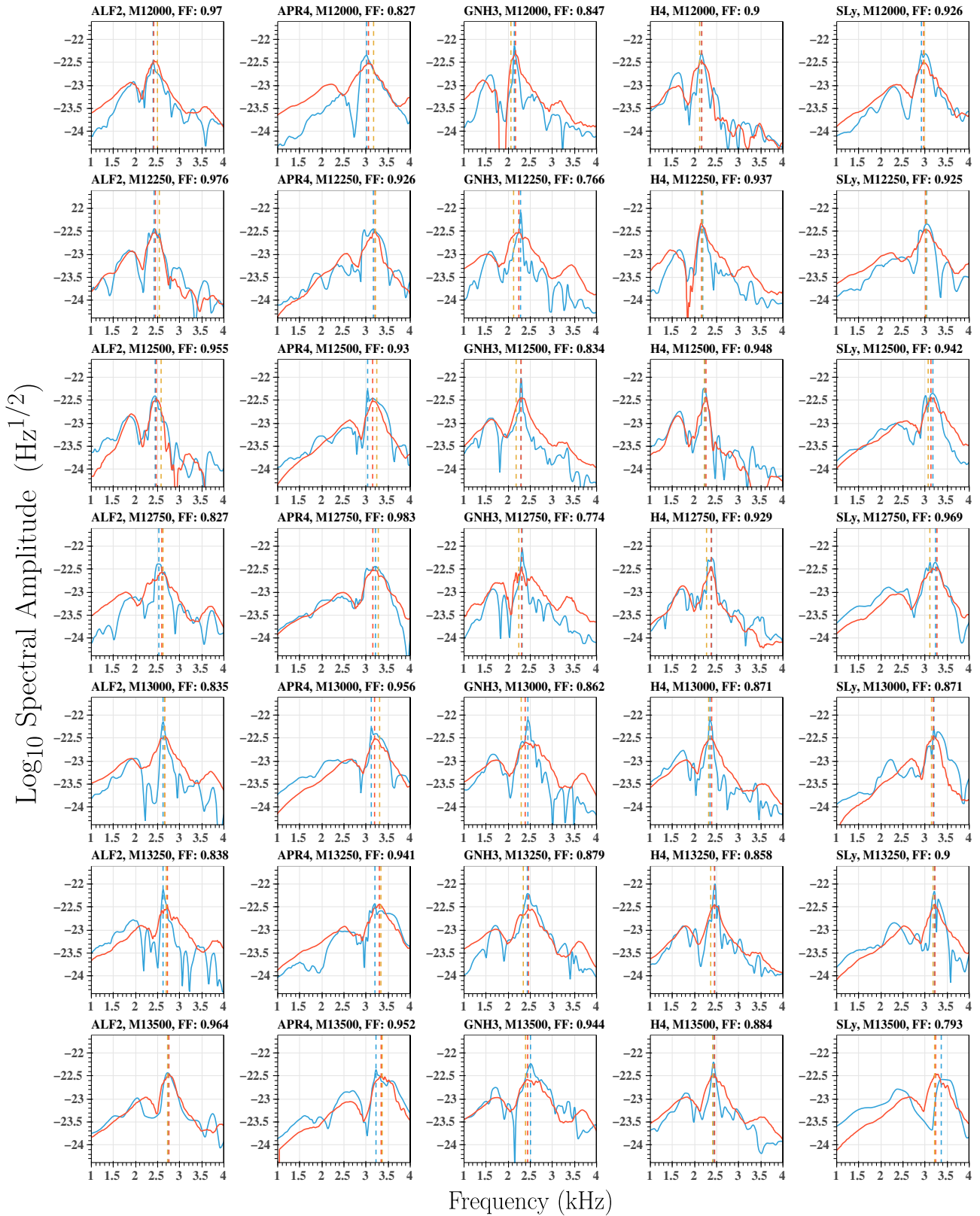FIG. 9. The same as Fig. 7, but for models from Soultanis *et al.* [97].

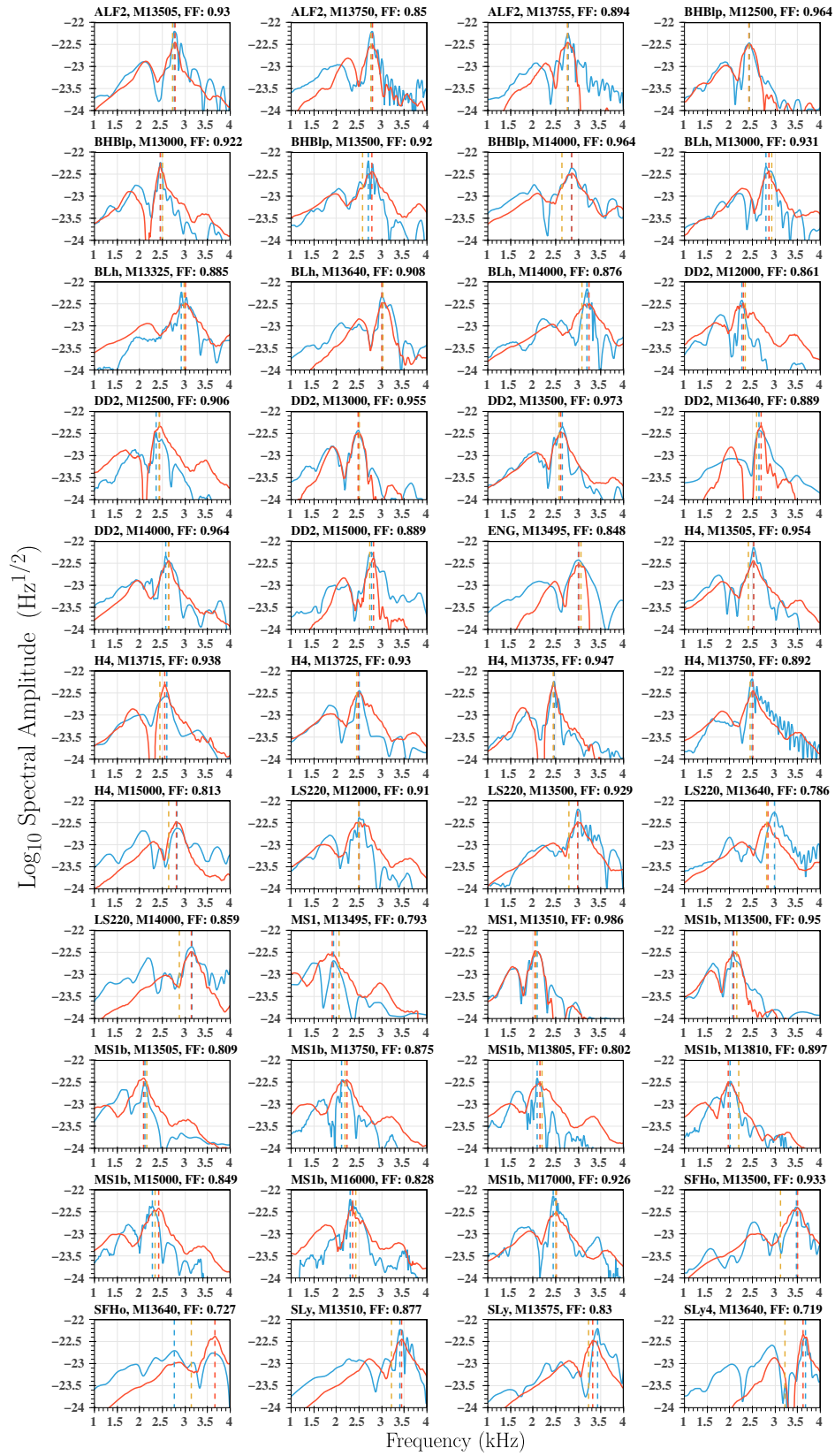FIG. 10.   The same as Fig. 7, but using the MLR-based model.

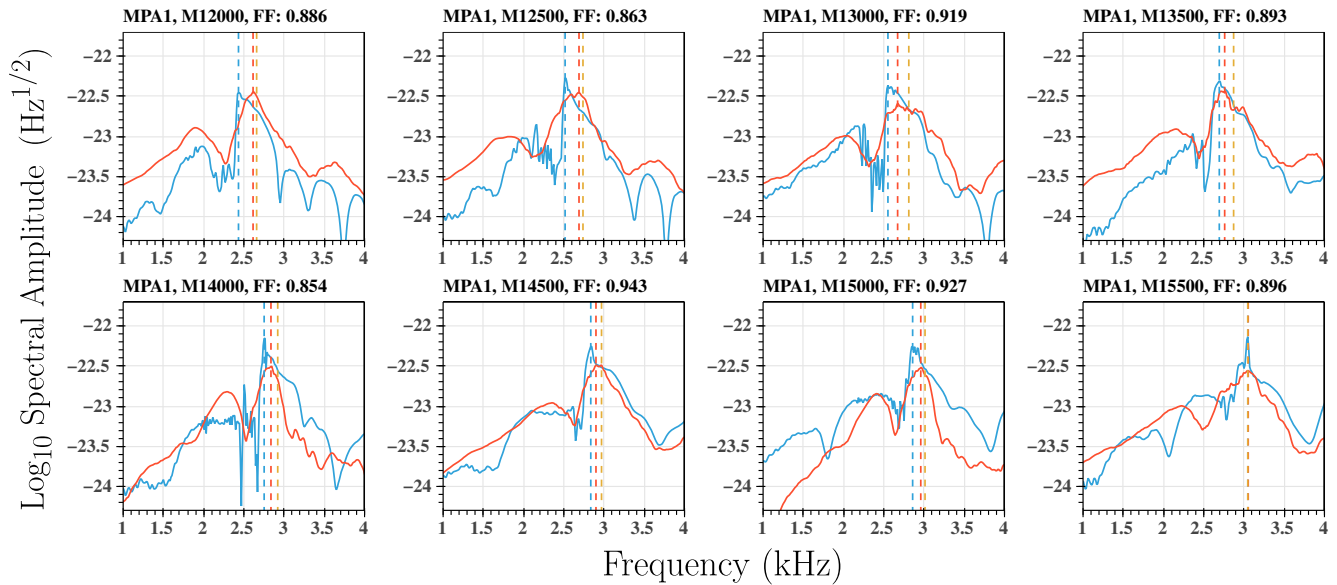FIG. 11. The same as Fig. 8, but using the MLR-based model.

FIG. 12. The same as Fig. 9, but using the MLR-based model.

[1] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116,** 061102 (2016).

[2] B. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Phys. Rev. X **13,** 041039 (2023).

[3] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, and Virgo Collaborations), Living Rev. Relativity **23,** 3 (2020).

[4] M. Bailes *et al.*, Nat. Rev. Phys. **3,** 344 (2021).

[5] M. Arimoto *et al.*, Prog. Theor. Exp. Phys. **2023,** 10A103 (2021).

[6] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Astrophys. J. Lett. **851,** L16 (2017).

[7] B. Grace, K. Wette, and S. Scott, arXiv:2403.11392.

[8] J. Clark, A. Bauswein, L. Cadonati, H.-T. Janka, C. Pankow, and N. Stergioulas, Phys. Rev. D **90,** 062004 (2014).

[9] K. Chatziioannou, J. A. Clark, A. Bauswein, M. Millhouse, T. B. Littenberg, and N. Cornish, Phys. Rev. D **96,** 124035 (2017).

[10] A. Torres-Rivas, K. Chatziioannou, A. Bauswein, and J. A. Clark, Phys. Rev. D **99,** 044014 (2019).

[11] B. Abbott *et al.*, Classical Quantum Gravity **34,** 044001 (2017).

[12] N. Sarin and P. D. Lasky, Publ. Astron. Soc. Aust. **39,** e007 (2022).

[13] K. Ackley, V. B. Adya, P. Agrawal, P. Altin, G. Ashton, M. Bailes, E. Baltinas, A. Barbuio, D. Beniwal, C. Blair *et al.*, Pub. Astron. Soc. Aust. **37,** e047 (2020).

[14] D. Martynov *et al.*, Phys. Rev. D **99,** 102004 (2019).

[15] V. Srivastava, D. Davis, K. Kuns, P. Landry, S. Ballmer, M. Evans, E. D. Hall, J. Read, and B. S. Sathyaprakash, Astrophys. J. **931,** 22 (2022).

[16] M. Breschi, R. Gamba, S. Borhanian, G. Carullo, and S. Bernuzzi, arXiv:2205.09979.

[17] T. Zhang, H. Yang, D. Martynov, P. Schmidt, and H. Miao, Phys. Rev. X **13,** 021019 (2023).

[18] A. Bauswein, N. Stergioulas, and H.-T. Janka, Eur. Phys. J. A **52,** 56 (2016).

[19] L. Baiotti and L. Rezzolla, Rep. Prog. Phys. **80,** 096901 (2017).

[20] A. Bauswein and N. Stergioulas, J. Phys. G **46,** 113002 (2019).

[21] L. Baiotti, Prog. Part. Nucl. Phys. **109,** 103714 (2019).

[22] J. L. Friedman and N. Stergioulas, Int. J. Mod. Phys. D **29,** 2041015 (2020).

[23] S. Bernuzzi, Gen. Relativ. Gravit. **52,** 108 (2020).

[24] T. Dietrich, T. Hinderer, and A. Samajdar, Gen. Relativ. Gravit. **53,** 27 (2021).

[25] N. Sarin and P. D. Lasky, Gen. Relativ. Gravit. **53,** 59 (2021).

[26] M. Breschi, S. Bernuzzi, D. Godzieba, A. Perego, and D. Radice, Phys. Rev. Lett. **128,** 161102 (2022).

[27] A. Gonzalez *et al.*, Classical Quantum Gravity **40,** 085011 (2023).

[28] N. Stergioulas, A. Bauswein, K. Zagkouris, and H.-T. Janka, Mon. Not. R. Astron. Soc. **418,** 427 (2011).

[29] K. Hotokezaka, K. Kiuchi, K. Kyutoku, T. Muranushi, Y.-i. Sekiguchi, M. Shibata, and K. Taniguchi, Phys. Rev. D **88,** 044026 (2013).

[30] A. Bauswein and N. Stergioulas, Phys. Rev. D **91,** 124056 (2015).

[31] K. Takami, L. Rezzolla, and L. Baiotti, Phys. Rev. D **91,** 064001 (2015).

[32] J. A. Clark, A. Bauswein, N. Stergioulas, and D. Shoemaker, Classical Quantum Gravity **33**, 085003 (2016).

[33] D. Radice, S. Bernuzzi, and C. D. Ott, Phys. Rev. D **94**, 064011 (2016).

[34] R. D. Pietri, A. Feo, J. A. Font, F. Löffler, F. Maione, M. Pasquali, and N. Stergioulas, Phys. Rev. Lett. **120**, 221101 (2018).

[35] J. Fields, A. Prakash, M. Breschi, D. Radice, S. Bernuzzi, and A. da Silva Schneider, Astrophys. J. Lett. **952**, L36 (2023).

[36] A. Bauswein, N.-U. F. Bastian, D. Blaschke, K. Chatziioannou, J. A. Clark, T. Fischer, H.-T. Janka, O. Just, M. Oertel, and N. Stergioulas, in *Proceedings of the Xiamen-CUSTIPEN Workshop on the Equation of State of Dense Neutron-Rich Matter in the Era of Gravitational Wave Astronomy*, American Institute of Physics Conference Series Vol. 2127 (American Institute of Physics, College Park, MD, 2019), p. 020013.

[37] D. Radice, S. Bernuzzi, and A. Perego, Annu. Rev. Nucl. Part. Sci. **70**, 95 (2020).

[38] M. Shibata, Phys. Rev. Lett. **94**, 201101 (2005).

[39] M. Shibata, K. Taniguchi, and K. Uryū, Phys. Rev. D **71**, 084021 (2005).

[40] R. Oechslin and H.-T. Janka, Phys. Rev. Lett. **99**, 121102 (2007).

[41] A. Bauswein and H.-T. Janka, Phys. Rev. Lett. **108**, 011101 (2012).

[42] A. Bauswein, H.-T. Janka, K. Hebeler, and A. Schwenk, Phys. Rev. D **86**, 063001 (2012).

[43] K. Takami, L. Rezzolla, and L. Baiotti, Phys. Rev. Lett. **113**, 091104 (2014).

[44] F. A. Rasio and S. L. Shapiro, Astrophys. J. **401**, 226 (1992).

[45] L. Rezzolla and K. Takami, Phys. Rev. D **93**, 124051 (2016).

[46] F. Foucart, R. Haas, M. D. Duez, E. O'Connor, C. D. Ott, L. Roberts, L. E. Kidder, J. Lippuner, H. P. Pfeiffer, and M. A. Scheel, Phys. Rev. D **93**, 044019 (2016).

[47] C.-J. Haster, K. Chatziioannou, A. Bauswein, and J. A. Clark, Phys. Rev. Lett. **125**, 261101 (2020).

[48] K. Topolski, S. Tootle, and L. Rezzolla, Astrophys. J. **960**, 86 (2024).

[49] S. Bernuzzi, T. Dietrich, and A. Nagar, Phys. Rev. Lett. **115**, 091101 (2015).

[50] S. Bose, K. Chakravarti, L. Rezzolla, B. Sathyaprakash, and K. Takami, Phys. Rev. Lett. **120**, 031102 (2018).

[51] M. Breschi, S. Bernuzzi, F. Zappa, M. Agathos, A. Perego, D. Radice, and A. Nagar, Phys. Rev. D **100**, 104029 (2019).

[52] S. Vretinaris, N. Stergioulas, and A. Bauswein, Phys. Rev. D **101**, 084039 (2020).

[53] A. W. Criswell, J. Miller, N. Woldemariam, T. Soultanis, A. Bauswein, K. Chatziioannou, M. W. Coughlin, G. Jones, and V. Mandic, Phys. Rev. D **107**, 043021 (2023).

[54] Y. Setyawati, M. Pürrer, and F. Ohme, Classical Quantum Gravity **37**, 075012 (2020).

[55] A. Królak, P. Jaranowski, M. Bejger, P. Ciecieląg, O. Dorosh, and A. Pisarski, Classical Quantum Gravity **40**, 215008 (2023).

[56] F. H. Panther and P. D. Lasky, Mon. Not. R. Astron. Soc. **523**, 2928 (2023).

[57] M. Breschi, S. Bernuzzi, K. Chakravarti, A. Camilletti, A. Prakash, and A. Perego, Phys. Rev. D **109**, 064009 (2024).

[58] A. Puecher, T. Dietrich, K. W. Tsang, C. Kalaghatgi, S. Roy, Y. Setyawati, and C. V. D. Broeck, Phys. Rev. D **107**, 124009 (2023).

[59] M. C. Tringali, A. Puecher, C. Lazzaro, R. Ciolfi, M. Drago, B. Giacomazzo, G. Vedovato, and G. A. Prodi, Classical Quantum Gravity **40**, 225008 (2023).

[60] F. Iacovelli, M. Mancarella, C. Mondal, A. Puecher, T. Dietrich, F. Gulminelli, M. Maggiore, and M. Oertel, Phys. Rev. D **108**, 122006 (2023).

[61] S. Blacker, N.-U. F. Bastian, A. Bauswein, D. B. Blaschke, T. Fischer, M. Oertel, T. Soultanis, and S. Typel, Phys. Rev. D **102**, 123023 (2020).

[62] E. R. Most, L. J. Papenfort, V. Dexheimer, M. Hanauske, S. Schramm, H. Stöcker, and L. Rezzolla, Phys. Rev. Lett. **122**, 061101 (2019).

[63] A. Bauswein, N.-U. F. Bastian, D. B. Blaschke, K. Chatziioannou, J. A. Clark, T. Fischer, and M. Oertel, Phys. Rev. Lett. **122**, 061102 (2019).

[64] E. R. Most, L. Jens Papenfort, V. Dexheimer, M. Hanauske, H. Stoecker, and L. Rezzolla, Eur. Phys. J. A **56**, 59 (2020).

[65] A. Prakash, D. Radice, D. Logoteta, A. Perego, V. Nedora, I. Bombaci, R. Kashyap, S. Bernuzzi, and A. Endrizzi, Phys. Rev. D **104**, 083029 (2021).

[66] M. Breschi, G. Carullo, and S. Bernuzzi, arXiv:2301.09672.

[67] S. Blacker, H. Kochankovski, A. Bauswein, A. Ramos, and L. Tolos, Phys. Rev. D **109**, 043015 (2024).

[68] E. Barausse, C. Palenzuela, M. Ponce, and L. Lehner, Phys. Rev. D **87**, 081506 (2013).

[69] C. Palenzuela, E. Barausse, M. Ponce, and L. Lehner, Phys. Rev. D **89**, 044024 (2014).

[70] M. Shibata, K. Taniguchi, H. Okawa, and A. Buonanno, Phys. Rev. D **89**, 084005 (2014).

[71] L. Sagunski, J. Zhang, M. C. Johnson, L. Lehner, M. Sakellariadou, S. L. Liebling, C. Palenzuela, and D. Neilsen, Phys. Rev. D **97**, 064016 (2018).

[72] R. F. P. Mendes, N. Ortiz, and N. Stergioulas, Phys. Rev. D **104**, 104036 (2021).

[73] W. E. East and F. Pretorius, Phys. Rev. D **106**, 104055 (2022).

[74] H.-J. Kuan, K. Van Aelst, A. T.-L. Lam, and M. Shibata, Phys. Rev. D **108**, 064057 (2023).

[75] M. Bezares, R. Aguilera-Miret, L. ter Haar, M. Crisostomi, C. Palenzuela, and E. Barausse, Phys. Rev. Lett. **128**, 091103 (2022).

[76] K. V. Staykov, D. D. Doneva, L. Heisenberg, N. Stergioulas, and S. S. Yazadjiev, Phys. Rev. D **108**, 024058 (2023).

[77] H.-J. Kuan, A. T.-L. Lam, D. D. Doneva, S. S. Yazadjiev, M. Shibata, and K. Kiuchi, Phys. Rev. D **108**, 063033 (2023).

[78] P. J. Easter, P. D. Lasky, A. R. Casey, L. Rezzolla, and K. Takami, Phys. Rev. D **100**, 043005 (2019).

[79] A. J. Izenman, Multivariate regression, in *Modern Multivariate Statistical Techniques: Regression, Classification,*

and *Manifold Learning* (Springer, New York, 2008), pp. 159–194.

[80] J. Fox and S. Weisberg, *An R Companion to Applied Regression*, 3rd ed. (Sage, Thousand Oaks, CA, 2019).

[81] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, Adaptive Computation and Machine Learning (MIT Press, Cambridge, MA, 2016).

[82] K. Hornik, M. Stinchcombe, and H. White, Neural Netw. **2**, 359 (1989).

[83] K. Hornik, Neural Netw. **4**, 251 (1991).

[84] M. D. Zeiler, arXiv:1212.5701.

[85] Ç. Gülçehre and Y. Bengio, arXiv:1412.7419.

[86] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, arXiv:1412.7419.

[87] C. M. Bishop, Neural Comput. **7**, 108 (1995).

[88] J. Sietsma and R. J. Dow, Neural Netw. **4**, 67 (1991).

[89] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, J. Mach. Learn. Res. **15**, 1929 (2014), https://jmlr.org/papers/v15/srivastava14a.html.

[90] D. C. Montgomery, *Design and Analysis of Experiments*, 7th ed. (Wiley, New York, 2008).

[91] A. Gelman and J. Hill, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Analytical Methods for Social Research (Cambridge University Press, Cambridge, England, 2007), p. 625.

[92] T. A. Apostolatos, Phys. Rev. D **52**, 605 (1995).

[93] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics (Springer, New York, 2001).

[94] Y. Bengio and Y. Grandvalet, J. Mach. Learn. Res. **5**, 1089 (2004).

[95] Y. Zhang and Y. Yang, J. Econom. **187**, 95 (2015).

[96] R. Kohavi, in *Proceedings of the 14th International Joint Conference on Artificial Intelligence—Volume 2*, IJCAI'95 (Morgan Kaufmann Publishers Inc., San Francisco, 1995), pp. 1137–1143.

[97] T. Soultanis, A. Bauswein, and N. Stergioulas, Phys. Rev. D **105**, 043020 (2022).

[98] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems (2015), software available from https://www.tensorflow.org/.

[99] M. J. Anzanello and F. S. Fogliatto, Int. J. Indus. Ergon. **41**, 573 (2011).

[100] P. T. H. Pang *et al.*, Nat. Commun. **14**, 8352 (2023).

[101] K. Kiuchi, K. Kyotoku, Y. Sekiguchi, M. Shibata, and T. Wada, Phys. Rev. D **90**, 041502 (2014).

[102] B. Giacomazzo, J. Zrake, P. Duffell, A. I. MacFadyen, and R. Perna, Astrophys. J. **809**, 39 (2015).

[103] K. Kiuchi, P. Cerdá-Durán, K. Kyotoku, Y. Sekiguchi, and M. Shibata, Phys. Rev. D **92**, 124034 (2015).

[104] D. Radice, Astrophys. J. Lett. **838**, L2 (2017).

[105] K. Kiuchi, K. Kyotoku, Y. Sekiguchi, and M. Shibata, Phys. Rev. D **97**, 124039 (2018).

[106] R. Ciolfi, W. Kastaun, J. V. Kalinani, and B. Giacomazzo, Phys. Rev. D **100**, 023005 (2019).

[107] L. Sun, M. Ruiz, S. L. Shapiro, and A. Tsokaros, Phys. Rev. D **105**, 104028 (2022).

[108] C. Palenzuela, S. Liebling, and B. Miñano, Phys. Rev. D **105**, 103020 (2022).

[109] L. R. Werneck, Z. B. Etienne, A. Murguia-Berthier, R. Haas, F. Cipolletta, S. C. Noble, L. Ennoggi, F. G. Lopez Armengol, B. Giacomazzo, T. Assumpção, J. Faber, T. Gupte, B. J. Kelly, and J. H. Krolik, Phys. Rev. D **107**, 044037 (2023).

[110] P. L. Espino, G. Bozzola, and V. Paschalidis, Phys. Rev. D **107**, 104059 (2023).

[111] F. Foucart, Living Rev. Comput. Astrophys. **9**, 1 (2023).

[112] M. Miravet-Tenés, D. Guerra, M. Ruiz, P. Cerdá-Durán, and J. A. Font, arXiv:2401.02493.