Kilonova light-curve interpolation with neural networks

Yinglei Peng [ORCID],[1] Marko Ristić [ORCID],[2] Atul Kedia,[2] Richard O'Shaughnessy,[2] Christopher J. Fontes [ORCID],[3,4] Chris L. Fryer [ORCID],[3,5,6,7,8] Oleg Korobkin,[3,9] Matthew R. Mumpower,[3,9] V. Ashley Villar,[10] and Ryan T. Wollaeger[3,5]

[1]*Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627, USA*
[2]*Center for Computational Relativity and Gravitation, Rochester Institute of Technology, Rochester, New York 14623, USA*
[3]*Center for Theoretical Astrophysics, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[4]*Computational Physics Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[5]*Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[6]*University of Arizona, Tucson, Arizona 85721, USA*
[7]*Department of Physics and Astronomy, University of New Mexico, Albuquerque, New Mexico 87131, USA*
[8]*The George Washington University, Washington, DC 20052, USA*
[9]*Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*
[10]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, Massachusetts 02138, USA*

Kilonovae are the electromagnetic transients created by the radioactive decay of freshly synthesized elements in the environment surrounding a neutron star merger. To study the fundamental physics in these complex environments, kilonova modeling requires, in part, the use of radiative transfer simulations. The microphysics involved in these simulations results in high computational cost, prompting the use of emulators for parameter inference applications. Utilizing a training set of 22 248 high-fidelity simulations (composed of 412 unique ejecta parameter combinations evaluated at 54 viewing angles), we use a neural network to efficiently train on existing radiative transfer simulations and predict light curves for new parameters in a fast and computationally efficient manner. Our neural network can generate millions of new light curves in under a minute. We discuss our emulator's degree of off-sample reliability and parameter inference of the AT2017gfo observational data. Finally, we discuss tension introduced by multiband inference in the parameter inference results, particularly with regard to the neural network's recovery of viewing angle.

I. INTRODUCTION

On August 18, 2017, prompt observations identified gravitational wave emission (GW170817; [1,2]), shortly followed by a gamma-ray burst (short-GRB GRB170817A [3]). Extensive followup observations identified a long-duration optical/near-infrared counterpart, AT2017gfo, later identified as a "kilonova" [4–21]. A kilonova [22] is characterized by thermal emission from rapidly expanding, radioactively powered, heavy-element material ejected from the associated progenitor merger. The detection of the joint gravitational- and electromagnetic-wave emission from GW170817 and AT2017gfo has initiated an era of precision kilonova observations.

Most interpretations of kilonova observations, including kilonova candidates contributing to excess optical/infrared emission in GRB afterglows [23–29], have relied on broadband photometry. This reliance was in part owing to the relative sparsity of available spectra for AT2017gfo (and

lack of spectral observations of other kilonovae; although see, e.g., [30–32]) [24–27,29,33–45]. A comprehensive review of kilonova broadband photometry has recently been compiled and presented in Ref. [46]. Many studies interpreted their observations of AT2017gfo shortly after detection principally by comparison to simplified models for kilonovae [4–10,12,14–21,47,48] consisting of one or more groups of nonaccelerating (homologous) expanding material. Motivated by both binary merger simulations and the inability to fit observations with one component [8], at least two components are customarily employed [10,12,15–17,20], with properties loosely associated with two expected features of merger simulations: promptly ejected material (the "dynamical" ejecta), associated with tidal tails or shocked material at contact, and material driven out on longer timescales by properties of the remnant system (the "wind" ejecta) [49].

Radiative transfer models of two-component kilonovae, while more physically accurate and informative, come with a significantly higher computational cost compared to semi-analytical or one-component models. As a result of this cost, many groups have resorted to surrogate models, or emulators, for the kilonova outflow, to reduce the computational impact associated with inference using these more complex models [50–59].

In this paper we use a neural network emulator for light-curve interpolation and parameter inference. We train on a previously generated library of ∼400 two-component kilonova light-curve simulations. Our method can be easily applied to any modestly sized archive of adaptively learned astrophysical transient light-curve simulations. Like our previous study [52], our investigation is a concrete and novel realization of a widely used strategy in many areas of physics: training a surrogate model against reference simulations, then using the surrogate for model parameter inference. In this work, however, we perform a novel investigation to assess the quality of our surrogate, to find surprising differences relative to conclusions obtained with the previously described surrogate [52]. Our results raise general questions about how to appropriately characterize and propagate systematic uncertainty in surrogate models, which will be of broad interest in any area of physics where comparable surrogate approaches are used.

This paper is organized as follows. In Sec. II we discuss our simulation training library, the interpolation model and its architecture, and the associated light-curve interpolation methodology. In Sec. III we compare our emulator's performance with others employed in the literature and report inference results for observations of AT2017gfo. In Sec. IV we summarize our findings.

II. INTERPOLATION METHODOLOGY

A. Simulation description

Our two-component kilonova model consists of a lanthanide-rich equatorial dynamical ejecta component and a lanthanide-poor axial wind ejecta component as described in [60,61] and motivated by numerical simulations [49,62]. Each component is homologously expanding and parameterized by a mass and velocity such that M_d, v_d and M_w, v_w describe the dynamical and wind components' masses and averaged velocities, respectively. The morphology for the dynamical component is an equatorially centered torus, whereas the wind component is represented by an axially centered peanut component; Fig. 1 of [60] displays the torus peanut, or "TP," schematic corresponding to the morphologies employed in this work (see [61] for a detailed definition). The lanthanide-rich dynamical ejecta is a result of the *r*-process nucleosynthesis from a neutron-rich material with a low electron fraction ($Y_e \equiv n_p/(n_p + n_n)$) of $Y_e = 0.04$ with elements reaching the third *r*-process peak ($A \sim 195$), while the wind ejecta originates from higher $Y_e = 0.27$ which encapsulates elements between the first ($A \sim 80$) and second ($A \sim 130$) *r*-process peaks. The detailed breakdown of the elements in each component can be found in Table 2 of Ref. [60].

We use `SuperNu` [63], a Monte Carlo code for simulation of time-dependent radiation transport with matter in local thermodynamic equilibrium to create simulated kilonova spectra $F_{\lambda,sim}$ assuming the aforementioned two-component model. Both components are assumed to have fixed composition and morphology for the duration of each simulation. `SuperNu` uses radioactive power sources calculated from decaying the *r*-process composition from the `WinNet` nuclear reaction network [64–67]. These radioactive heating

contributions are also weighted by thermalization efficiencies introduced in Ref. [68] (see Ref. [69] for a detailed description of the adopted nuclear heating). We use detailed opacity calculations via the tabulated, binned opacities generated with the Los Alamos suite of atomic physics codes [70–72]. In the database that we use, the tabulated, binned opacities are not calculated for all elements; therefore, we produce opacities for representative proxy elements by combining pure-element opacities of nuclei with similar atomic properties [71]. Specifics of the representative elements for our composition are given in Ref. [60].

The `SuperNu` outputs are observing-angle-dependent, simulated spectra $F_{\lambda,sim}$, postprocessed to a source distance of 10 pc, in units of erg s^{-1} cm^{-2} Å^{-1}. The spectra are binned into 1024 equally log-spaced wavelength bins spanning $0.1 \leqslant \lambda \leqslant 12.8$ microns.

For the purposes of this work, we consider the light curves for the 2MASS *grizy* and Rubin Observatory *JHK* broadband filters. As we only consider anisotropic simulations in this study, unless otherwise noted, we extract simulated light curves using 54 angular bins, uniformly spaced in $\cos\theta$ over the range $-1 \leqslant \cos\theta \leqslant 1$, where the angle θ is taken between the line of sight and the symmetry axis as defined in Eq. (2). Specifically, simulations in our database cover all observing angles with a resolution ranging from $\Delta\theta = \arcsin(2/54) \simeq 2.1°$ at the equator, to $\Delta\theta = \arccos(1 - 2/54) \simeq 15.6°$ near the axes. This limiting angular resolution near the axes is comparable to the angles inferred from long-term radio observations of the gamma ray burst jet afterglow [10,16,17,73–75].

`SuperNu` Monte Carlo radiative transport results have modest but nonzero Monte Carlo error. The impact of this Monte Carlo error can be estimated both by resolution studies as well as by simple smoothness diagnostics (e.g., versus angle and time). For example, in a previous study [52] we performed Gaussian process interpolation over this same training set, inferring both an estimate of the light curve and a conservative estimate of its variance. The Monte Carlo error inherent in the underlying `SuperNu` simulations can be seen, for example, in Fig. 6 of Ref. [52] as short-angular-scale roughness on top of the overall smooth trend, with at most ∼3% deviation at the 1σ level. Later, in Sec. III C, we describe a targeted resolution study. Based on these investigations, we expect that the *statistical* error of the underlying `SuperNu` simulations is smaller than the systematic errors introduced by interpolating between these simulations as described below.

B. Training data generation

Below we describe the approach taken to generate the simulation library in Ref. [52], hereafter R22. Our training library of 22 248 kilonova light-curve simulations was constructed using iterative simulation placement guided by Gaussian process variance minimization. New simulations were placed with parameter combinations that were identified as having the largest bolometric luminosity variance by our Gaussian process regression approach. In other words, we placed new simulations in regions of parameter space where our bolometric luminosity interpolation root-mean-square uncertainty was

largest. The Gaussian process variance $s(\vec{x})^2$ is defined as

$$s(\vec{x})^2 = k(\vec{x}, \vec{x}) - k(\vec{x}, \vec{x}_a)k(\vec{x}_a, \vec{x}_{a'})^{-1}_{aa'}k(\vec{x}_{a'}, \vec{x}), \qquad (1)$$

where $\vec{x}$ is the vector of input parameters, $\vec{x}_a$ is the training data vector, the function $k(\vec{x}, \vec{x}')$ is the kernel of the Gaussian process, and the indices a, a' are used to calculate the covariance between inputs $\vec{x}$ and training data $\vec{x}_a, \vec{x}_{a'}$ such that if $a = a'$, the variance is 0. In building the simulation library, we considered only the four-dimensional space of ejecta parameters $\vec{x} = [M_d, v_d, M_w, v_w]$. Each ejecta parameter combination yields simulations calculated for 54 equally spaced viewing angles; as such, our training set of 22 248 light curves corresponds to a core set of 412 unique ejecta parameter combinations.

For this work, we use the aforementioned light curves in the original simulation library as our training set. The light curves used in this work have the same parameters as those used for our light-curve interpolation approach in R22. No additional simulations were produced for the purposes of this work; all training data came from the simulation library presented in R22.

The original training data library consists of 22 248 total light-curve simulations calculated at 264 times and 54 angular bins each. We do not perform any coordinate transformations, but rather interpolate directly in our ejecta parameter space and angle.

C. Data processing

The entirety of our 22 248 simulations is represented by a tensor, M_{abc}, containing the AB magnitudes across the bands described in Sec. II A, which correspond to a set of input parameters $\vec{x}$. M_{abc} has dimensions of $412 \times 264 \times 54$, corresponding to 412 simulations evaluated at 264 log-spaced times between 0.125 and 37.24 days for 54 viewing angles equally spaced in $\cos\theta$ for θ ranging from 0 to 180°. We do not perform any normalization of our inputs or outputs, with ejecta parameters ranging from $-3 \leqslant \log m/M_\odot \leqslant -1$ and $0.05 \leqslant v/c \leqslant 0.3$ and light curves ranging from -18 to 8 AB magnitudes.

We split our four-dimensional ejecta parameter vector $\vec{x}$ into training, validation, and test sets. We use 60% of the data for the training set, which contains information that the neural network uses to learn. Of the remaining 40%, 20% is used for the validation check, which tracks how well the network generalizes to off-sample inputs during training, and 20% goes into the test set, which is used to evaluate the network's predictions compared to known simulation data. The samples for each set are randomly drawn from $\vec{x}$ according to a uniform distribution. Neither the validation set nor the test set data is used by the neural network for learning; therefore, we use only $\sim$247 simulations for training, while the rest are used in various steps of verifying generalization (i.e., avoiding overfitting to the training data).

After splitting the data into training, validation, and test sets, we incorporate viewing angle as a fifth input parameter. As mentioned above, the viewing angles in our simulations are equally spaced in $\cos\theta$ space across 54 angular bins, as

presented in the following equation:

$$\theta = \arccos\left(1 - \frac{2(i-1)}{54}\right) \text{ for } i \text{ in } 1, 2, \ldots, 54. \qquad (2)$$

Temporarily ignoring training, validation, or test sets, our training library consists of a total of $22\,248 \times 5$ inputs, as illustrated in the following schematic matrix:

$$\vec{x}_\theta = \begin{bmatrix} m_{d,1} & v_{d,1} & m_{w,1} & v_{w,1} & \theta_1 \\ m_{d,1} & v_{d,1} & m_{w,1} & v_{w,1} & \theta_2 \\ m_{d,1} & v_{d,1} & m_{w,1} & v_{w,1} & \theta_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m_{d,1} & v_{d,1} & m_{w,1} & v_{w,1} & \theta_{54} \\ m_{d,2} & v_{d,2} & m_{w,2} & v_{w,2} & \theta_1 \\ m_{d,2} & v_{d,2} & m_{w,2} & v_{w,2} & \theta_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m_{d,412} & v_{d,412} & m_{w,412} & v_{w,412} & \theta_{54}. \end{bmatrix}$$

D. Neural network architecture and training

We use a standard feed-forward neural network called a multilayer perceptron (MLP). Figure 1 shows our MLP architecture. The input data of dimension 5 (blue block) is propagated through the hidden layers of the MLP (pink blocks). These hidden layers apply a sequence of linear and nonlinear transformations (black arrows) to progressively map the input to a higher-dimensional space. The network has six fully connected layers (pink blocks) of dimension 128, 256, 64, 256, 128, and 264, respectively, which are followed by Rectified Linear Unit (ReLU) activation functions, except for the middle two layers. The final layer reduces the dimension to a 264×1 vector, which matches the length of our light curves.

For each observing band, we train a separate neural network and compare its predictions with the simulation results. During training, the neural network's predictions for the inputs in the validation set are compared to the simulation data for those same inputs. The residual, or difference, between the two is evaluated by the mean squared error (MSE) loss function, which we use to measure the average squared difference between simulation data and the neural network prediction. We calculate the MSE according to

$$MSE = \frac{1}{264} \sum_{i=1}^{264} (y_i - \hat{y}_i)^2, \qquad (3)$$

where y_i is the corresponding simulation data at time i and $\hat{y}_i$ is the predicted value from the MLP model at the same time. We show the evolution of the training and validation losses in Fig. 2 for the g-band network.

We train a separate neural network for each of the broadband filters described in Sec. II A. Each network is trained for 1000 epochs, as this is enough time for the validation loss to convincingly stabilize at a floor value without beginning to increase, indicating overfitting. We use an Adam optimizer and train each network in batches with a batch size of 32. Our
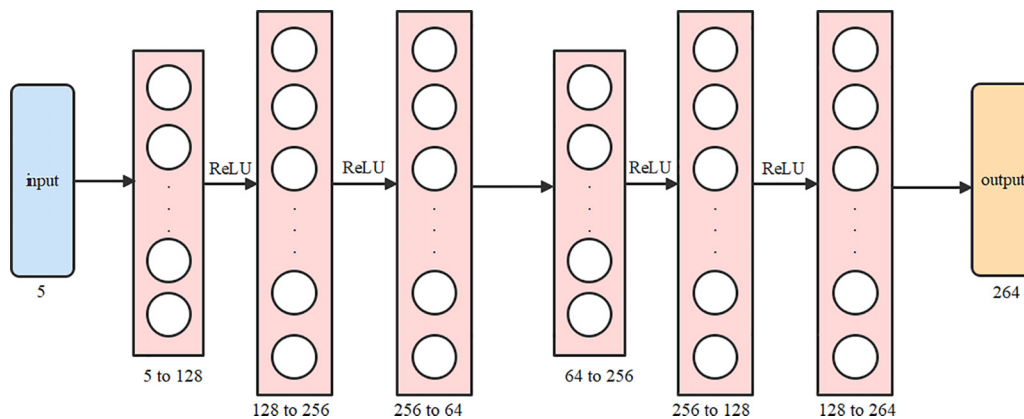
FIG. 1. A visual representation of the neural network architecture. The blue block represents our five-dimensional inputs $\bar{x}_\theta$. The pink blocks represent hidden layers, with labels under each block representing the input and output dimensions of each. Unlabeled right arrows indicate a linear mapping between layers, while those labeled "ReLU" have a Rectified Linear Unit activation function applied to their outputs. The orange block represents the neural network prediction in the form of a 264×1 vector matching the length of our broadband light curves.

initial learning rate is 2×10^{-4} with a decay rate of 5% every 10 epochs.

E. Neural network performance

Figure 2 indicates that our training loss undershoots the validation loss starting at ~ 100 epochs, but the nearly perfect recovery of the test-set light curves shown in Fig. 3 shows no indication of overfitting to the training data. For inference applications, our neural networks can generate the outcomes corresponding to five million five-dimensional inputs in about one minute. Training each neural network takes 20 minutes on a 2022 edition Macbook Pro with an M2 chip using the CPU. Since all the bands are independent and can be trained simultaneously in parallel, training all the emulators can be completed in this same 20 m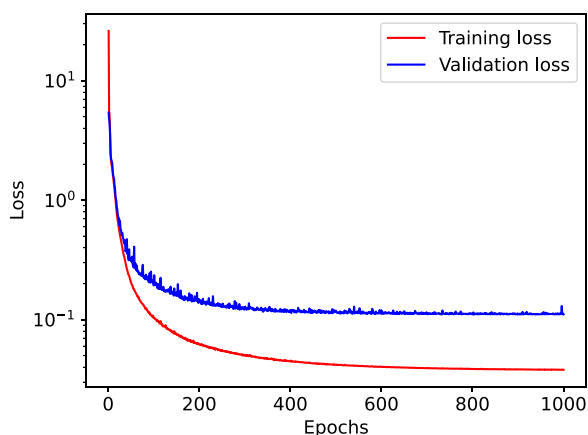inute interval. Our training time is half the value reported in Ref. [51], though it is unclear whether their reported time assumes training in parallel or in serial.

The top left plot in Fig. 3 shows a histogram of the MSE values when evaluating the simulation library parameters using the neural network. We evaluate only the 412 unique ejecta parameter combinations, fixing the viewing angle to $\theta = 0°$ in each case. In containing simulations from the training, validation, and test sets, this histogram represents the neural network's on- and off-sample fidelity. The light-curve plots in Fig. 3 show random draws from the MSE histogram for $MSE < 0.01$ (top right), $0.01 \leqslant MSE \leqslant 0.1$ (bottom left), and $MSE > 0.1$ (bottom right).

We note that, as with the emulators presented in Ref. [52], predictions for inputs with low-mass ($\log(M) \sim -3$) components or viewing angles $\theta \sim 90°$ may deviate substantially from expectations. In Monte Carlo radiative transfer simulations of a kilonova, the representation of very low-mass components and viewing angles in the plane of the binary present a significant challenge. In order to study the dynamics of the energy distribution, a finite number of particles are employed to represent photons escaping from the system, forming "packets" of energy. It is well known that at low Monte Carlo particle counts, tallies of escaping flux can even have systematic error [76]. In our simulations we mitigate this effect with implicit capture variance reduction and discrete diffusion Monte Carlo [63]. However, given a fixed number of source particles per time step, we use a metric for allocating particles per spatial cell that is sublinearly proportional to emissivity, so that the statistical quality of a spatial cell in one ejecta component is affected by the other ejecta component. Consequently, when dealing with extremely low-mass components or viewing angles that look into the high-opacity dynamical ejecta, the simulations become particularly sensitive to Poisson noise due to reduced photon count. This effect, particularly with respect to low-mass ejecta, likely arises from our `SuperNu` simulations preferably sampling photon packets from higher-energy regions of the ejecta. In future studies, we hope to enhance the simulation interpretation under these conditions by way of an increase in photon packet count.

FIG. 2. Training (red) and validation (blue) loss curves as a function of training epoch for the *g* band. The loss values reported here are mean squared error (MSE) as defined in Eq. (3). Decreasing values of loss indicate better agreement between the model and the training data. Our training and validation loss decreases over the course of the 1000 epoch training period. The validation loss appears to converge at around 500 epochs, although we continue to 1000 epochs as we do not implement an explicit convergence criterion.
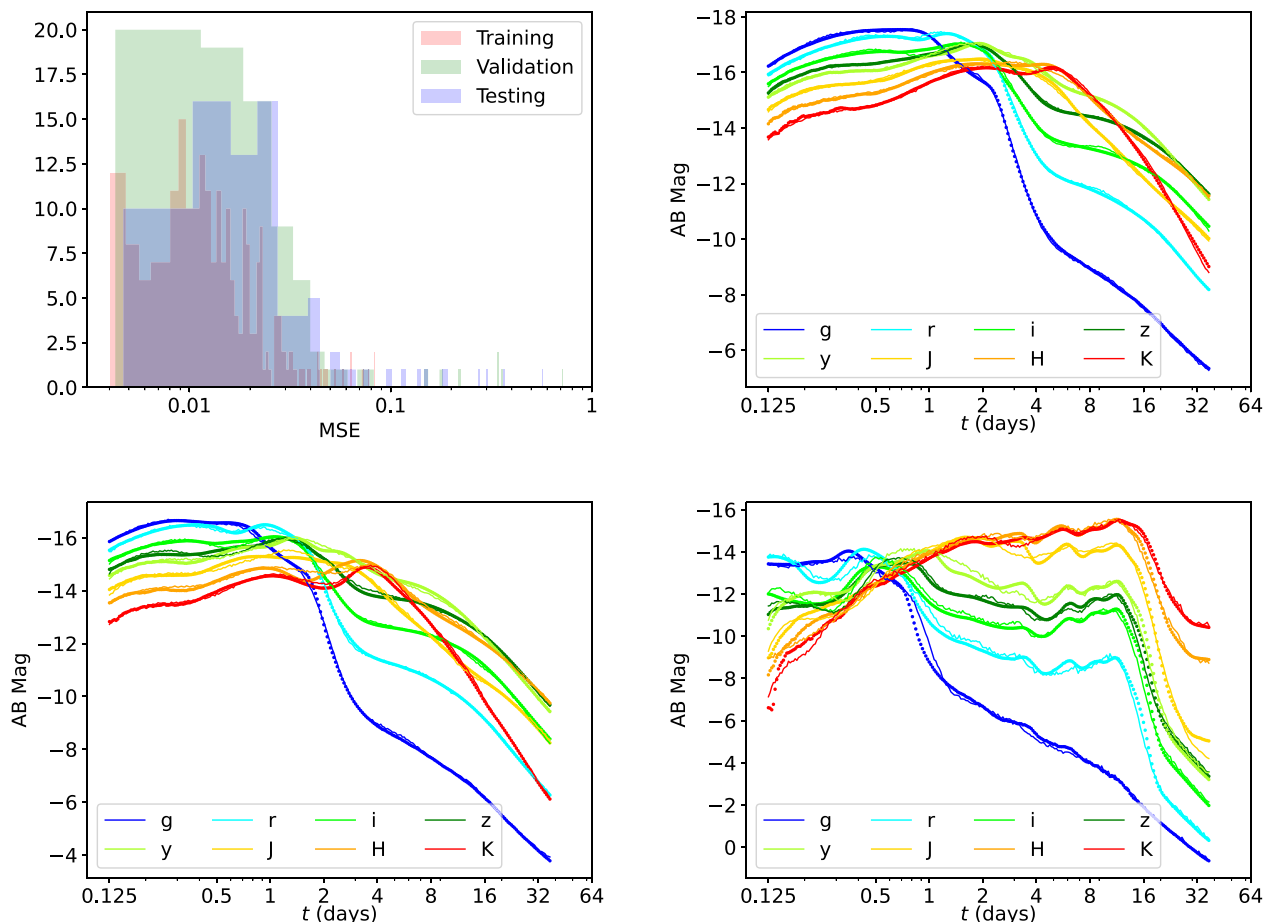
FIG. 3. Top left: A histogram of MSE values, averaged across all bands by the number of observations, which characterize the deviation of the MLP's predictions from the true simulation library light curves. For simplicity, we assume a fixed viewing angle of $\theta = 0°$ for each light curve and evaluate the MSE only for the 412 unique ejecta parameter combinations. Top right: True (points) and predicted (lines) light curves for a randomly drawn set of ejecta parameters (still assuming $\theta = 0°$) with $MSE < 0.01$. Bottom left: Same as top right, but for $0.01 \leqslant MSE \leqslant 0.1$. Bottom right: Same as top right, but for $MSE > 0.1$.

III. BAYESIAN INFERENCES WITH THE NEURAL INTERPOLATOR

A. Parameter inference methodology

As in R22, we infer the parameters of the kilonova AT2017gfo using our interpolated light curves and the AT2017gfo photometric data. The AT2017gfo data is originally presented in [4–18,20,21]. We use the RIFT framework [77] to adaptively perform the Monte Carlo integral and generate samples using a reduced χ^2 statistic. The parameter priors are the same as in R22, with uniform ejecta parameter priors of $-3 \leqslant \log m/M_\odot \leqslant -1$ and $0.05 \leqslant v/c \leqslant 0.3$ and a Gaussian angle prior with $\mu = 20$ and $\sigma = 5$ degrees. Unlike before, in this work we employ an adaptive volume Monte Carlo integrator, following closely the approach outlined in Ref. [78]. The adaptive volume integrator allows for more efficient sampling given the higher-dimensional space being explored in this work.

Each sample $\vec{x}_\theta$ is evaluated by the MLP to produce a light-curve prediction $\hat{y}$ for every one of the *grizy* JHK broadband filters. We calculate the residual between the MLP prediction $\hat{y}$ and the AT2017gfo observed data d for every band B by way

of the reduced-χ^2 statistic

$$\chi^2 = \sum_{B,i} \frac{(\hat{y}_{i,B} - d_{i,B})^2}{\sigma_i^2 + \sigma_{sys}^2}. \tag{4}$$

In our χ^2 residual calculation, we include observational uncertainties from the AT2017gfo data σ_d, as well as systematic uncertainties σ_{sys} which we use as a catch-all term to encompass all uncertainties, quantifiable or otherwise, associated with the neural network interpolation process. As outlined above and as discussed in greater quantitative detail in Sec. III C, we adopt a systematic modeling uncertainty, σ_{sys}, of 0.5 magnitudes for our inference analysis.

For inference, we adopt a purely Gaussian likelihood based on these residuals

$$\ln \mathcal{L} = -\frac{\chi^2}{2} - \frac{1}{2} \ln(2\pi)^N \sum_k \left(\sigma_d^2 + \sigma_{sys}^2 \right), \tag{5}$$

where N is the number of observations. As our inference is performed via adaptive Monte Carlo integration, the reliability of our posterior can be expressed in terms of a number of effective samples n_{eff}. (Several different conventions exist for
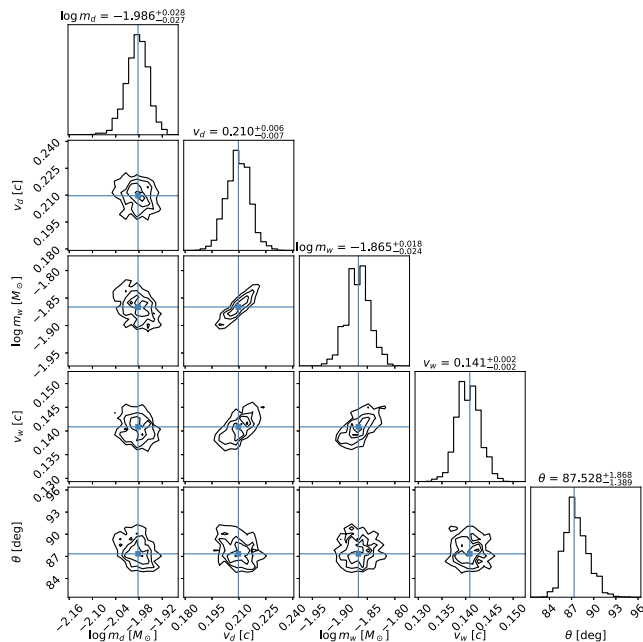
FIG. 4. Posteriors derived from a single randomly generated synthetic kilonova source are consistent with its assumed model parameters. For this analysis, we have adopted the "zero noise" realization, where the kilonova light curve is precisely equal to its expected value.
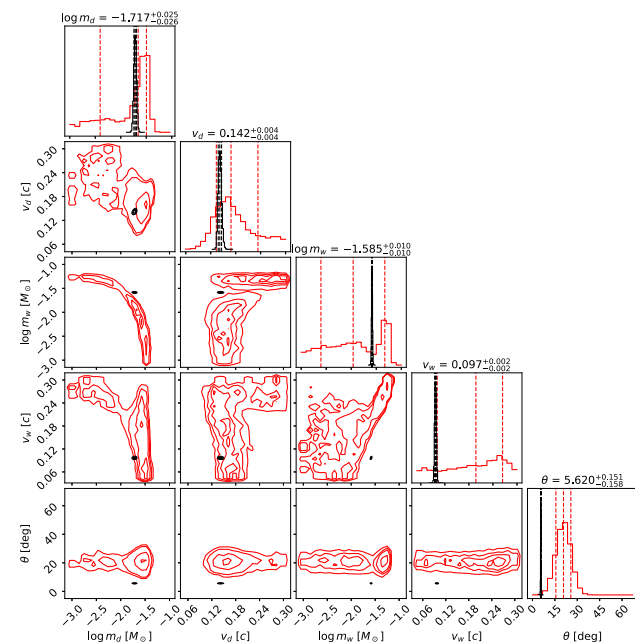

FIG. 5. In black are shown the posterior distributions for the ejecta parameters and viewing angle that most closely reconstruct the AT2017gfo observational data using the MLP. In red, we overlay the posterior distributions for the same parameters when using predictions generated by the Gaussian process presented in R22. The values at the top of each column represent the median posterior values for the inference performed in this paper using the MLP.

this number; see the Appendix of [77] for discussion.) For this study, we terminate our analyses when $n_{\text{eff}} \simeq 10^3$.

To validate our inference strategy, we constructed random synthetic sources, with kilonova model parameters drawn from our prior (albeit adopting a uniform rather than Gaussian angular prior) and using observational times and uncertainties precisely matching the AT2017gfo cadence and instruments. For our synthetic sources, the expected light curve is generated using our neural network. As demonstrated with one example in Fig. 4, our inferences are always consistent with the injected kilonova parameters. To demonstrate that our implementation retains statistical purity, we also used 100 random synthetic sources to perform a conventional probability-probability (PP) test, available in the Appendix.

While we employ the fixed $\sigma_{\text{sys}} = 0.5$ for most of our studies, in order to validate our results we also perform a few selected analyses with different choices, on the one hand adopting different discrete choices and on the other treating σ_{sys} as a continuous unknown model parameter.

The analysis in Fig. 4 demonstrates that, if the underlying model is correct, comparison with AT2017gfo-like observations should very tightly constrain each of this model's parameters. This fiducial result thus has qualitatively different behavior than our and others' prior analyses of AT2017gfo, where posterior inferences arrive at much broader posterior intervals, as discussed below. That said, the posterior shown above is consistent with the standard Fisher matrix estimate of the inverse covariance matrix $\Gamma = \Sigma^{-1}$, derived, for example, by taking (the expected value over noise realizations of) a second-order Taylor series expansion of the log-likelihood as $\ln \mathcal{L} = \ln \mathcal{L}_o - 2^{-1} \Gamma_{ab}(x - x_*)_a(x - x_*)_b$ where x_* are the

true synthetic parameters:

$$\Gamma_{ab} = \sum_{B,i} \frac{1}{\sigma_i^2 + \sigma_{\text{sys}}^2} \frac{\partial \hat{y}_B}{\partial x_a} \frac{\partial \hat{y}_B}{\partial x_b} \bigg|_{x=x_*, t=t_i}. \quad (6)$$

In this expression, only first-order derivatives appear because we assume that the model has no systematic bias such that $\langle d \rangle = \hat{y}$; however, this simple estimate also arises inevitably using the large-amplitude "linear signal approximation" [79]. This Fisher matrix can be estimated to order of magnitude by replacing the derivatives $\partial \hat{y}_B / \partial x_a$ by the ratio $\Delta y_B / \Delta x_a$, which for the mass parameters we approximate as $2/2 = 1$, so the Fisher matrix is approximately $\Gamma \simeq N/\sigma_{\text{sys}}^2$ and the posterior in each mass hyperparameter should have a one-standard-deviation scale of order $1/\sqrt{\Gamma} \simeq \sigma_{\text{sys}}/\sqrt{N} \simeq 0.5/\sqrt{333} \simeq 0.027$.

B. AT2017gfo parameter inference

The posterior distributions for our input parameters $\vec{x}_\theta$ are plotted in Fig. 5. The black posteriors represent the parameters identified in this study. As a direct comparison to the results of R22, we overplot the posterior distributions from that study in red. Despite using the same training data and comparing to the same observational data, the Gaussian process emulator and the MLP emulator recover posterior distributions with substantially different median values and distribution widths; however, we note that in R22, each time step is fit individually, whereas in this work we fit the entire light curve for a given set of input parameters. The dynamical ejecta parameters $\log m_d$
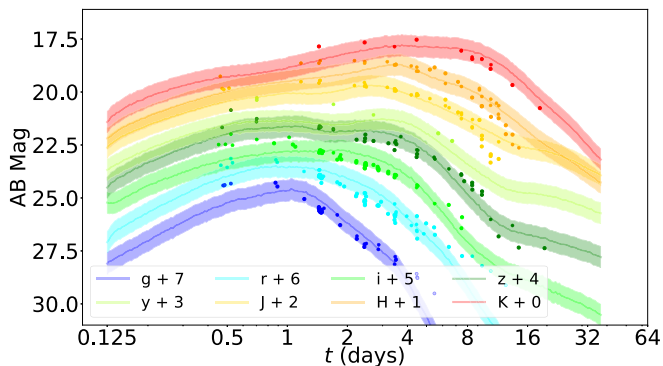
FIG. 6. Light curves generated by the MLP for the median parameters presented in Fig. 5 (lines with shaded regions) with AT2017gfo observational data overplotted (scatter points). Despite the different values recovered between this analysis and that of R22, especially for the wind ejecta parameters and θ, the MLP prediction is able to replicate the observations surprisingly well.

and v_d are similar between the two emulators, but the wind ejecta parameters $\log m_w$ and v_d, as well as the viewing angle θ, are quite different, with two of the three parameters inferred by the MLP residing outside of the GP inference 1σ limits.

We verify the fidelity of the MLP inference results by generating light curves corresponding to the parameter values identified at the top of each column in Fig. 5. These light curves are shown in Fig. 6 and indicate that, assuming a 0.5 magnitude systematic uncertainty, the inferred MLP parameters do indeed replicate the AT2017gfo data to a reasonable degree of accuracy. While the majority of data are well replicated, the early time g and r bands and the late-time j and H bands deviate slightly outside of our uncertainty bands.

The recovery of $\theta \approx 6°$ is surprising for several reasons. First, the recovered angle was modestly offset from the Gaussian adopted as our inclination prior. Although different studies find a variety of viewing angles associated with AT2017gfo [10,16,17,73–75,80], none indicate that the viewing angle is as low as our inference suggests. Second, and more importantly, for angular binning described by Eq. (2), the first angular bin encompasses all emitted photons for viewing angles ∼0–16°. Therefore, by recovering a narrowly peaked posterior around $\theta \approx 6°$, the MLP seems to indicate that it can identify angular variations within a single angular bin at a resolution much finer than what is provided by training data. In other words, these inference results are either overly constrained, or the MLP is able to identify fine angular variations in the light curves when trained on radiative transfer simulations using a coarser angular grid.

One conceivable explanation for the narrow posterior distribution seen in Fig. 5 is an underestimate of the underlying systematic error. To investigate this possibility, Fig. 7 shows the results of inferences performed when adopting dfferent choices for the white-noise systematic error parameter σ_{sys}, adopting both discrete and continuously distributed choices for this parameter. In all scenarios, we infer similar ejecta parameters for AT2017gfo, even though we allow for several magnitudes of potential systematic uncertainty. Conversely,
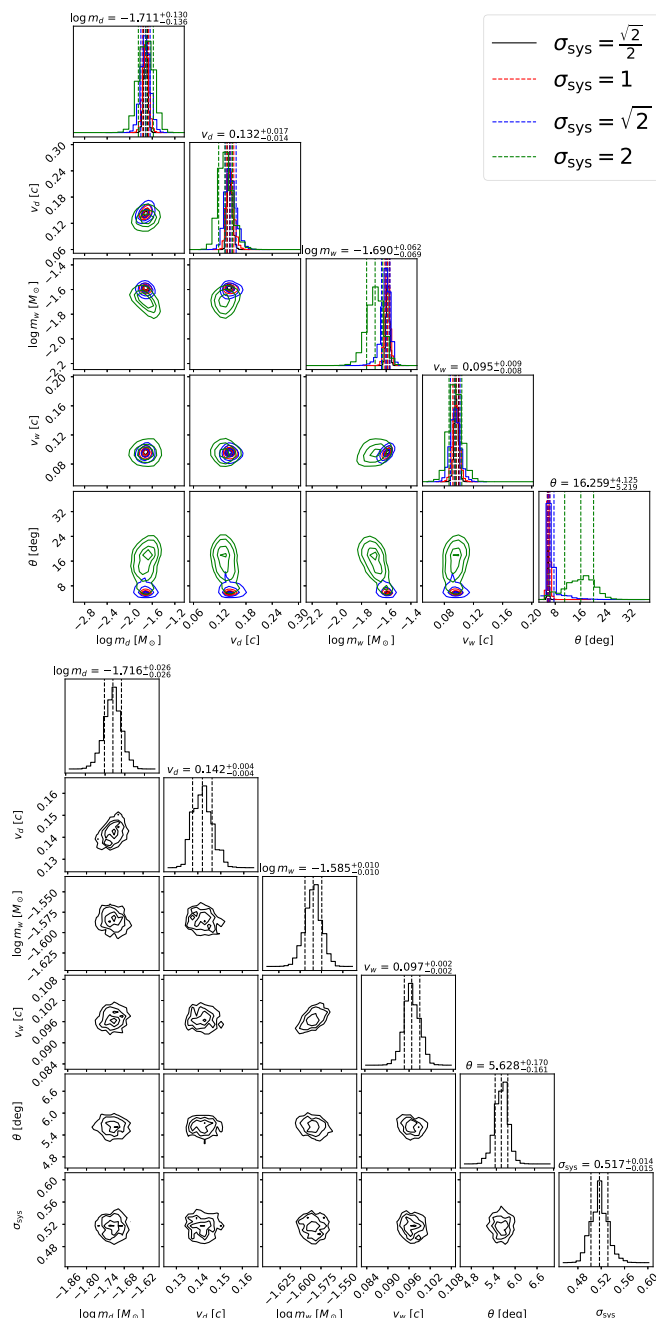

FIG. 7. Top panel: Posterior distributions using successively larger values of σ_{sys} during inference. Recovery of angles closer to the angle prior of 20° begins with $\sigma_{sys} = 2$, suggesting underestimation of systematic error. Bottom panel: Posterior distributions inferred if σ_{sys} is treated as an unknown parameter, *a priori* uniformly distributed.

this direct comparison between our models and the data directly infers a value for our systematic uncertainty parameter consistent with our fiducial choice. As such, we conclude that our inference about the ejecta parameters remains robust for all analyses, with the viewing angle inference susceptible to systematic uncertainty assumptions. However, we approximate the light curves to deviate by ∼0.35 magnitudes between 0° and 45° viewing angles [81]. This deviation for the bolometric light curve still falls below our assumed systematic
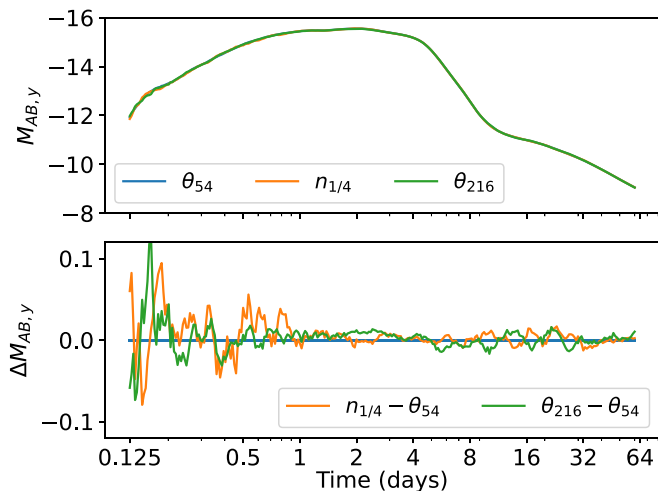
FIG. 8. Plots of `SuperNu` y-band light curves for a simulation like the ones described in Sec. II A (θ_{54}), a simulation like θ_{54}, but with one-quarter as many photon packets ($n_{1/4}$), and a simulation like θ_{54}, but with four times greater angular resolution (θ_{216}). The top panel indicates that, on a macroscopic scale, the simulations are identical. The bottom plot indicates that deviations do exist, likely attributed to statistical noise from the increase in angular bins, or matching reduction in packet count. The ejecta parameters used to create these simulations are those presented in Fig. 5.

FIG. 9. Deviations in y-band AB magnitude (ΔM_{AB}) between a `SuperNu` simulation with four times the usual angular resolution (y_{sim}) and the MLP prediction for the parameters reported in Fig. 5 (y_{MLP}) as a function of time t and angle θ. Notably, this comparison between our MLP (trained at low angular resolution) and the followup simulation (performed at high angular resolution) does *not* exhibit strong small-scale variation along the θ direction (i.e., between adjacent angular bins in the original training set and outside the original training resolution), suggesting that the MLP correctly interpolates to smaller angular scales.

C. Investigating MLP predictions near inferred AT2017gfo parameters

The extremely narrow posterior distribution in kilonova parameters and angle motivates a focused investigation of our training simulations and MLP model in the neighborhood of that posterior. As a first step, we performed followup `SuperNu` simulations at the inferred parameters using a higher angular resolution. Specifically, we increased the resolution by a factor of four to get a total of 216 angular bins. By reducing the number of photon packets in each angular bin by a factor of four, we also increase statistical noise by a factor of two. To ensure that our finer angular resolution analysis is not affected by this increase in statistical noise, we compare three separate simulations in Fig. 8. The blue curve, labeled θ_{54}, shows a `SuperNu` simulation using the parameters from Fig. 5, hereafter x_{MLP} and a standard 54-bin angular grid, as used in the training data simulations. The orange curve, labeled $n_{1/4}$, shows a `SuperNu` simulation with the same exact parameters as θ_{54}, except it uses one-quarter as many photon packets in the simulation. If the statistical noise described above were significant, the noise in $n_{1/4}$ should be much more pronounced than in θ_{54}. Finally, the green curve, labeled θ_{216}, shows a `SuperNu` simulation using the same parameters as θ_{54}, but with a 216-bin angular grid, representing a factor of four increase in resolution. All three light curves show AB magnitude in the y band as a function of time in days. As seen in Fig. 8, our followup simulations agree with one another, with small Monte Carlo error comparable to our initial estimate
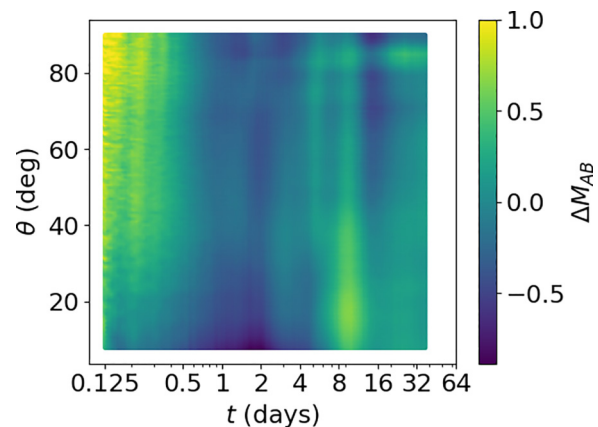
uncertainty using broadband light curves, and thus remains consistent with the baseline analysis.

and small compared to our adopted systematic uncertainty ($\sigma_{\text{sys}} = 0.5$).

We then compare our high angular resolution simulation θ_{216} to the predictions of the MLP, computing the difference between model and prediction at all times and simulation angles. Figure 9 shows the residual ΔM_{AB} in the y band when we take the absolute difference between $\hat{y}_{\text{MLP}}$ and y_{sim}, capped at a maximum difference of 1 magnitude. The residual values ΔM_{AB} are initially evaluated for the 216 discrete angular bins; for visual clarity and diagnostic power, we linearly interpolate ΔM_{AB} across time t and angle θ using the `RegularGridInterpolator` function from the `scipy.interpolate` library [82]. Figure 9 exhibits both expected and unexpected behaviors. The large mismatch at early times when photon count is low, particularly as θ approaches 90° where the higher opacity dynamical ejecta further reduces photon count, is to be expected. The MLP is not fitting the light curves in this region well as there are too few photons available in the training data. However, the low mismatch (i.e., good fitting) dark blue regions in the plot indicate some sort of structure in the MLP's underlying ability to reproduce light curves at different times and angles. While we include only the plot of ΔM_{AB} for the y band in Fig. 9, it is worth noting that the same behavior can be observed in these plots for all bands; predictable, low photon count systematics are identified in expected regions, but other, unexpected regions of increased systematic error manifest in different regions of the parameter space. Overall, however, the systematic uncertainties illustrated here are consistent with our expectations from the reported validation loss: a conservative systematic error of $\sqrt{MSE_{\text{val}}} \simeq 0.5$ reflects our overall uncertainty. For this reason, we adopted this systematic uncertainty in our parameter inferences above. This systematic uncertainty is
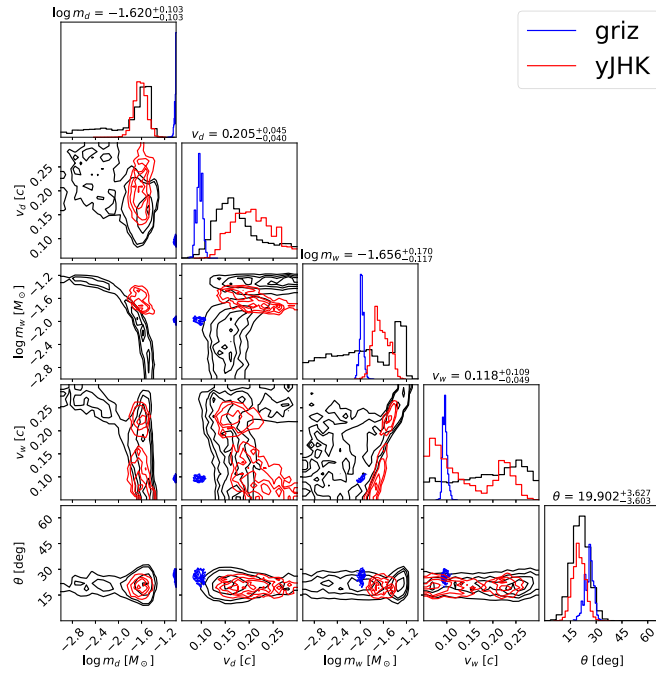
FIG. 10. Two separate parameter inference runs, with the first (in blue) using data from only the *griz* bands and the second (in red) using data from only the *yJHK* bands. The red posteriors match the R22 posteriors (in black) much more closely. Parameter values at the top of each column represent the median *yJHK* posterior values.

substantially more conservative than the uncertainties adopted in R22.

D. Inference using broadband data subsets

The investigations in Sec. III C have introduced a surprise. On the one hand, our neural network reliably reproduces its training and validation data, including followup off-sample simulations performed at higher resolution. Though not shown here, we have also confirmed that the neural network agrees well with the surrogate provided in R22, using a small sample of randomly selected ejecta parameters. On the other hand, the inferences obtained in Sec. III C by comparing all AT2017gfo kilonova observations to our MLP produce strikingly different results than R22. However, as noted in R22 and other works, most investigations have some tension between their models and the data, particularly in the bluer bands. In this section, motivated by this discrepancy, we also examine the effects on our AT2017gfo parameter inference when we use only specific subsets of the observational data. We perform two additional parameter inference runs using two categories of broadband data subsets: blue bands represented by the *griz* data and red bands represented by the *yJHK* data. The posteriors in Fig. 10 show how these band-limited results compare with each other, the results of R22, and the all-band analysis presented in Sec. III C. The most apparent result is that the angle prior is recovered in both sets of posteriors, and both cases match the R22 results well. The other interesting feature is that the red *yJHK* posterior matches the R22 results much more closely than the blue posterior. Even when using a smaller subset of the observational data,

the blue posteriors remain narrowly peaked in the parameter space, while the red posteriors become broader. The narrowness of the blue posteriors indicates that the blue broadband data determines the overall shape of the posteriors in the full broadband data inference.

The seemingly disproportionate effect of the blue broadband data on the posteriors could be attributed to the rapid evolution of the bluer bands compared to the red bands. As can be seen in Fig. 6, the evolution of the *griz* light curves is more rapid than the *yJHK* bands, with the *griz* light curves dimming by ∼4 magnitudes compared to the *yJHK* bands dimming by two magnitudes during the first 10 days of observations. As such, the *griz* light curves will be more restrictive regarding which model parameters fit the data, which is evident from the Fig. 10 posteriors.

E. Discussion

To summarize, following R22 we fit the same simulations and performed comparable inference of AT2017gfo. After assessing our training data and fit systematics, we adopted more conservative systematic uncertainties than R22. We nonetheless find dramatically narrower posteriors than R22, with inferred light curves consistent with observational predictions.

Several possible reasons for the qualitative discrepancies arise, primarily pertaining to our systematic uncertainty estimate. In R22, the Gaussian process interpolation provided a pointwise and parameter-dependent error estimate, which we qualitatively verified across the parameter space. Also, the fitting strategy adopted in R22 independently fit each time step. As a result, we expect that R22's models are unlikely to have correlated systematics in time, angle, and wavelength. For example, the R22 light curves occasionally have small but notable random discontinuities, consistent with their reported fitting uncertainty. By contrast, in this work, we fit all times and angles together within the same training set to generate a vector prediction. Our approach does not presently provide a pointwise error estimate. Therefore, for the method adopted in this work, we expect correlated errors in time and angle, but lack a method to characterize them versus those parameters or even the intrinsic kilonova parameters. Our MLP's vector light-curve prediction necessitates fitting the entire light curve for a certain band given a single set of parameters. In addition, as mentioned in Sec. III D, the blue bands are particularly constraining due to their more significant evolution over the observation period. Combining both of these effects results in only an extremely narrow region of the parameter space fitting all broadband data consistently. We did not encounter such narrow posteriors in R22 as each prediction was made for a specific time point; as such, many samples could reasonably fit an observation at any given time, resulting in broader posteriors when all times were stitched together to form the light curve.

We use the validation loss value to roughly estimate the systematic fitting error associated with our neural network outcomes: the validation curve in Fig. 2 suggests that differences of order $\sqrt{MSE} \simeq \sqrt{0.2}$ magnitudes should occur in our predictions. In practice, as illustrated in Fig. 7, we find that the average squared systematic error suggests a larger value than the validation MSE. We therefore anticipate that

our naive estimate of $\sigma_{\mathrm{sys}} = 0.5$, though well motivated by our detailed followup study, may still understate the systematic uncertainty inherent in our fitting approach, resulting in narrow posterior distributions. We also emphasize that the differences are not simply a matter of scale: the investigations performed in Fig. 7 suggest that larger white-noise systematic error cannot reconcile differences between our current analysis and previous results.

We note that we have experienced similar systematic uncertainty associated with observations in blue bands in R22 and an associated inference using simulations of spectra [56]. The systematics in those works were related to our inability to reproduce the observed blue flux at times past $\sim$2 days using our best-fit simulations. The systematics in this work, though similar in their connection to blue observations, introduce slightly different effects in our resultant inference. As we solve the bigger problem of matching our simulations to late-time blue observations, we anticipate that a more sophisticated treatment of our emulator systematics will allow us to better understand the effects of blue-band data on our inferences.

A thorough investigation of suitable fit systematics for this neural network is well beyond the scope of our study. In the meantime, the neural network is suitable for investigations such as the one presented in Fig. 10, where we can examine our models' ability to fit certain subsets of the data. In the *griz* case, we see that our models require over 0.1 $M_\odot$ of slow-moving dynamical ejecta to fit the blue data. But we expect dynamical ejecta to be less massive and faster, thus potentially suggesting a missing modeling component.

IV. CONCLUSIONS

We present a neural network architecture that is useful for the interpolation of kilonova light curves. We report on the network's training and validation loss as a metric of successful training, as well as present examples of off-sample light-curve recovery. We use the neural network to infer the parameters of the AT2017gfo kilonova and compare to previous inference performed in Ref. [52]. We find that the inference results are quite different from those previously obtained, but the light curves generated by the recovered parameters align well with the observational data. In particular, we investigate the neural network's ability to seemingly infer narrow regions of the angle space despite being trained on light-curve data that should not allow for such specific inference. Given a detailed analysis of the mismatch between the neural network's predictions and a simulation with higher-resolution angular data, we find that the network's pointwise systematic errors are consistent with our error estimate. However, our investigations also suggest that the systematic errors are correlated, not independent, in time and angle, in a way that is not captured by our model for systematic uncertainties. In other words, we have discovered that the neural network's goodness-of-fit varies appreciably across the time-angle space. While some of these variations are expected, others form interesting features that we cannot readily explain. We leave the analysis of the interpretability of these features for a future investigation.

We also show that the systematic uncertainty may be more complex than assumed in our simple uncorrelated (white
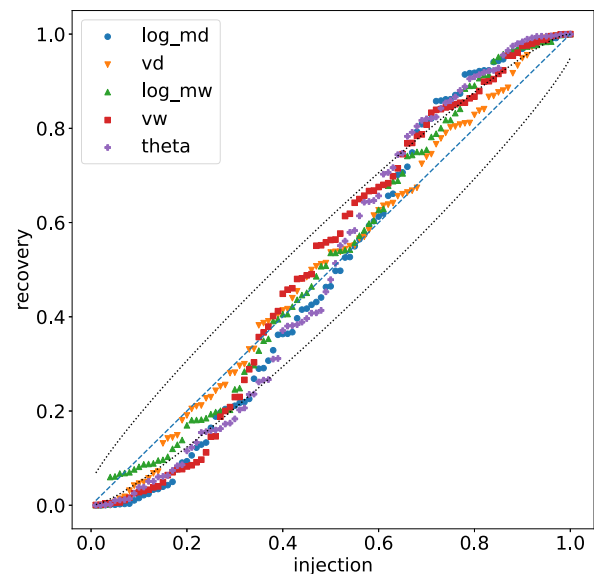

FIG. 11. Probability-probability plot to validate recovery of synthetic kilonova light curves, following [77]. Points show the PP plot data, with each simulation parameter indicated as in the legend. The dotted curves show the expected value and estimated 90% credible binomial interval.

noise) error model. This was not the case in R22 due to the interpolation uncertainty, which naturally stemmed from the Gaussian process methodology. In recovering different parameters for AT2017gfo using two emulators trained on the same library of simulations, we highlight the importance of quantifiable uncertainty analysis in using emulators for robust inference. As we do not present a way to handle correlated uncertainties in this work, a detailed uncertainty analysis, along with the resultant effects on parameter inference, will be necessary in future work.

APPENDIX: VALIDATING INFERENCE METHOD

To validate the statistical purity of our brute-force inference technique and our understanding of the noise model, we constructed a standard probability-probability (PP) plot test [83,84]. Our description follows the notation and narrative used in Ref. [77]. For each source k, with true parameters

λ_k, we calculate the fraction of its posterior distribution with parameter λ_α below the true source value $\lambda_{k,\alpha}$ [$\hat{P}_{k,\alpha}(<\lambda_{k,\alpha})$]. After reindexing the sources so that $\hat{P}_{k,\alpha}(\lambda_{k,\alpha})$ increases with k for some fixed α, a plot of k/N versus $\hat{P}_k(\lambda_{k,\alpha})$ can be compared with the expected diagonal result ($P(< p) = p$) and binomial uncertainty interval. Figure 11 shows the PP plot derived using kilonova light curves generated with our neural network interpolator. In these analyses, we adopt precisely the same observation cadence and uncertainties as AT2017gfo. As in our fiducial analysis of AT2017gfo, we adopt $\sigma_{\text{sys}} = 0.5$. Each synthetic observation incorporates both observational and (white noise) systematic uncertainty, added in quadrature consistent with our assumed likelihood. The PP plot in Fig. 11, being sufficiently consistent with the binomial credible interval, suggests that the brute-force Monte Carlo inference strategy adopted in this work suffices for our

purposes: in short, that the qualitative extent and character of the posteriors shown in our figures are reasonably correct, such that the considerable tension between our analysis and previous inferences accurately reflects the posterior. We have specifically chosen to present results from a brute-force inference technique to circumvent debates about our choice of implementation or our method of assessing convergence. With AT2017gfo, we have confirmed that the choice of integrator also doesn't qualitatively change our answer: alternative brute-force Monte Carlo integrator implementations produce similar results. That said, the PP plot above is clearly not as diagonal as would expected for a well-developed and calibrated Bayesian inference algorithm applied to this problem: its S-shape features suggests either modest overdispersion in our synthetic error model or modest underdispersion in our posterior distributions.

[1] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, GW170817: Observation of gravitational waves from a binary neutron star inspiral, Phys. Rev. Lett. **119**, 161101 (2017).

[2] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, Multi-messenger observations of a binary neutron star merger, Astrophys. J. Lett. **848**, L12 (2017).

[3] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya *et al.*, Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A, Astrophys. J. Lett. **848**, L13 (2017).

[4] I. Andreoni, K. Ackley, J. Cooke, A. Acharyya, J. R. Allison, G. E. Anderson, M. C. B. Ashley, D. Baade, M. Bailes, K. Bannister *et al.*, Follow up of GW170817 and its electromagnetic counterpart by Australian-led observing programmes, Publ. Astron. Soc. Aust. **34**, e069 (2017).

[5] I. Arcavi, G. Hosseinzadeh, D. A. Howell, C. McCully, D. Poznanski, D. Kasen, J. Barnes, M. Zaltzman, S. Vasylyev, D. Maoz, and S. Valenti, Optical emission from a kilonova following a gravitational-wave-detected neutron-star merger, Nature (London) **551**, 64 (2017).

[6] D. A. Coulter, R. J. Foley, C. D. Kilpatrick, M. R. Drout, A. L. Piro, B. J. Shappee, M. R. Siebert, J. D. Simon, N. Ulloa, D. Kasen *et al.*, Swope Supernova Survey 2017a (SSS17a), the optical counterpart to a gravitational wave source, Science **358**, 1556 (2017).

[7] M. R. Drout, A. L. Piro, B. J. Shappee, C. D. Kilpatrick, J. D. Simon, C. Contreras, D. A. Coulter, R. J. Foley, M. R. Siebert, N. Morrell *et al.*, Light curves of the neutron star merger GW170817/SSS17a: Implications for r-process nucleosynthesis, Science **358**, 1570 (2017).

[8] P. S. Cowperthwaite, Jr., E. Berger, V. A. Villar, B. D. Metzger, M. Nicholl, R. Chornock, P. K. Blanchard, W. Fong, R. Margutti, M. Soares-Santos *et al.*, The electromagnetic counterpart of the binary neutron star merger LIGO/Virgo GW170817. II. UV, optical, and near-infrared light curves and comparison to kilonova models, Astrophys. J. **848**, L17 (2017).

[9] M. C. Díaz, L. M. Macri, D. Garcia Lambas, C. Mendes de Oliveira, J. L. Nilo Castellón, T. Ribeiro, B. Sánchez, W. Schoenell, L. R. Abramo, S. Akras *et al.*, Observations of the first electromagnetic counterpart to a gravitational-wave source by the TOROS Collaboration, Astrophys. J. Lett. **848**, L29 (2017).

[10] P. A. Evans, S. B. Cenko, J. A. Kennea, S. W. K. Emery, N. P. M. Kuin, O. Korobkin, R. T. Wollaeger, C. L. Fryer, K. K. Madsen, F. A. Harrison *et al.*, Swift and NuSTAR observations of GW170817: Detection of a blue kilonova, Science **358**, 1565 (2017).

[11] L. Hu, X. Wu, I. Andreoni, M. C. B. Ashley, J. Cooke, X. Cui, F. Du, Z. Dai, B. Gu, Y. Hu *et al.*, Optical observations of LIGO source GW 170817 by the Antarctic Survey Telescopes at Dome A, Antarctica, Sci. Bull. **62**, 1433 (2017).

[12] M. M. Kasliwal, E. Nakar, L. P. Singer, D. L. Kaplan, D. O. Cook, A. Van Sistine, R. M. Lau, C. Fremling, O. Gottlieb, J. E. Jencson *et al.*, Illuminating gravitational waves: A concordant picture of photons from a neutron star merger, Science **358**, 1559 (2017).

[13] V. M. Lipunov, E. Gorbovskoy, V. G. Kornilov, N. Tyurina, P. Balanutsa, A. Kuznetsov, D. Vlasenko, D. Kuvshinov, I. Gorbunov, D. A. H. Buckley *et al.*, MASTER optical detection of the first LIGO/Virgo neutron star binary merger GW170817, Astrophys. J. Lett. **850**, L1 (2017).

[14] E. Pian, P. D'Avanzo, S. Benetti, M. Branchesi, E. Brocato, S. Campana, E. Cappellaro, S. Covino, V. D'Elia, J. P. U. Fynbo *et al.*, Spectroscopic identification of r-process nucleosynthesis in a double neutron-star merger, Nature (London) **551**, 67 (2017).

[15] S. J. Smartt, T. W. Chen, A. Jerkstrand, M. Coughlin, E. Kankare, S. A. Sim, M. Fraser, C. Inserra, K. Maguire, K. C. Chambers *et al.*, A kilonova as the electromagnetic counterpart to a gravitational-wave source, Nature (London) **551**, 75 (2017).

[16] N. R. Tanvir, A. J. Levan, C. González-Fernández, O. Korobkin, I. Mandel, S. Rosswog, J. Hjorth, P. D'Avanzo, A. S. Fruchter, C. L. Fryer *et al.*, The emergence of a lanthanide-rich kilonova following the merger of two neutron stars, Astrophys. J. Lett. **848**, L27 (2017).

[17] E. Troja, L. Piro, H. van Eerten, R. T. Wollaeger, M. Im, O. D. Fox, N. R. Butler, S. B. Cenko, T. Sakamoto, C. L. Fryer *et al.*, The X-ray counterpart to the gravitational-wave event GW170817, Nature (London) **551**, 71 (2017).

[18] Y. Utsumi, M. Tanaka, N. Tominaga, M. Yoshida, S. Barway, T. Nagayama, T. Zenko, K. Aoki, T. Fujiyoshi, H. Furusawa *et al.*, J-GEM observations of an electromagnetic counterpart to the neutron star merger GW170817, PASJ **69**, 101 (2017).

[19] S. Valenti, D. J. Sand, S. Yang, E. Cappellaro, L. Tartaglia, A. Corsi, S. W. Jha, D. E. Reichart, J. Haislip, and V. Kouprianov, The discovery of the electromagnetic counterpart of GW170817: Kilonova AT 2017gfo/DLT17ck, Astrophys. J. Lett. **848**, L24 (2017).

[20] A. S. Pozanenko, M. V. Barkov, P. Yu. Minaev, A. A. Volnova, E. D. Mazaeva, A. S. Moskvitin, M. A. Krugov, V. A. Samodurov, V. M. Loznikov, and M. Lyutikov, GRB 170817A associated with GW170817: Multi-frequency observations and modeling of prompt gamma-ray emission, Astrophys. J. Lett. **852**, L30 (2018).

[21] B. J. Shappee, J. D. Simon, M. R. Drout, A. L. Piro, N. Morrell, J. L. Prieto, D. Kasen, T. W. S. Holoien, J. A. Kollmeier, D. D. Kelson *et al.*, Early spectra of the gravitational wave source GW170817: Evolution of a neutron star merger, Science **358**, 1574 (2017).

[22] B. D. Metzger, G. Martínez-Pinedo, S. Darbha, E. Quataert, A. Arcones, D. Kasen, R. Thomas, P. Nugent, I. V. Panov, and N. T. Zinner, Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei, Mon. Not. R. Astron. Soc. **406**, 2650 (2010).

[23] D. A. Perley, B. D. Metzger, J. Granot, N. R. Butler, T. Sakamoto, E. Ramirez-Ruiz, A. J. Levan, J. S. Bloom, A. A. Miller, A. Bunker *et al.*, GRB 080503: Implications of a naked short gamma-ray burst dominated by extended emission, Astrophys. J. **696**, 1871 (2009).

[24] N. R. Tanvir, A. J. Levan, A. S. Fruchter, J. Hjorth, R. A. Hounsell, K. Wiersema, and R. L. Tunnicliffe, A 'kilonova' associated with the short-duration γ-ray burst GRB130603B, Nature (London) **500**, 547 (2013).

[25] W. Fong, E. Berger, B. D. Metzger, R. Margutti, R. Chornock, G. Migliori, R. J. Foley, B. A. Zauderer, R. Lunnan, T. Laskar *et al.*, Short GRB 130603B: Discovery of a jet break in the optical and radio afterglows, and a mysterious late-time x-ray excess, Astrophys. J. **780**, 118 (2014).

[26] Z.-P. Jin, X. Li, Z. Cano, S. Covino, Y.-Z. Fan, and D.-M. Wei, The light curve of the macronova associated with the long-short burst GRB 060614, Astrophys. J. **811**, L22 (2015).

[27] Z.-P. Jin, K. Hotokezaka, X. Li, M. Tanaka, P. D'Avanzo, Y.-Z. Fan, S. Covino, D.-M. Wei, and T. Piran, The macronova in GRB 050709 and the GRB-macronova connection, Nat. Commun. **7**, 12898 (2016).

[28] W. Fong, R. Margutti, R. Chornock, E. Berger, B. J. Shappee, A. J. Levan, N. R. Tanvir, N. Smith, P. A. Milne, T. Laskar *et al.*, The afterglow and early-type host galaxy of the short GRB 150101B at $z = 0.1343$, Astrophys. J. **833**, 151 (2016).

[29] M. M. Kasliwal, O. Korobkin, R. M. Lau, R. Wollaeger, and C. L. Fryer, Infrared emission from kilonovae: The case of the nearby short hard burst GRB 160821B, Astrophys. J. Lett. **843**, L34 (2017).

[30] J. H. Gillanders, E. Troja, C. L. Fryer, M. Ristic, B. O'Connor, C. J. Fontes, Y.-H. Yang, N. Domoto, S. Rahmouni, M. Tanaka *et al.*, Heavy element nucleosynthesis associated with a gamma-ray burst, arXiv:2308.00633.

[31] Y.-H. Yang, E. Troja, B. O'Connor, C. L. Fryer, M. Im, J. Durbak, G. S. H. Paek, R. Ricci, C. R. Bom, J. H. Gillanders *et al.*, A lanthanide-rich kilonova in the aftermath of a long gamma-ray burst, Nature (London) **626**, 742 (2024).

[32] A. J. Levan, D. B. Malesani, B. P. Gompertz, A. E. Nugent, M. Nicholl, S. R. Oates, D. A. Perley, J. Rastinejad, B. D. Metzger, S. Schulze *et al.*, A long-duration gamma-ray burst of dynamical origin from the nucleus of an ancient galaxy, Nature Astron. **7**, 976 (2023).

[33] S. Ascenzi, M. W. Coughlin, T. Dietrich, R. J. Foley, E. Ramirez-Ruiz, S. Piranomonte, B. Mockler, A. Murguia-Berthier, C. L. Fryer, N. M. Lloyd-Ronning, and S. Rosswog, A luminosity distribution for kilonovae based on short gamma-ray burst afterglows, Mon. Not. R. Astron. Soc. **486**, 672 (2019).

[34] S. B. Pandey, Y. Hu, A. J. Castro-Tirado, A. S. Pozanenko, R. Sánchez-Ramírez, J. Gorosabel, S. Guziy, M. Jelinek, J. C. Tello, S. Jeong *et al.*, A multiwavelength analysis of a collection of short-duration GRBs observed between 2012 and 2015, Mon. Not. R. Astron. Soc. **485**, 5294 (2019).

[35] E. Troja, G. Ryan, L. Piro, H. van Eerten, S. B. Cenko, Y. Yoon, S. K. Lee, M. Im, T. Sakamoto, P. Gatkine *et al.*, A luminous blue kilonova and an off-axis jet from a compact binary merger at $z = 0.1341$, Nat. Commun. **9**, 4089 (2018).

[36] G. P. Lamb, N. R. Tanvir, A. J. Levan, A. de Ugarte Postigo, K. Kawaguchi, A. Corsi, P. A. Evans, B. Gompertz, D. B. Malesani, K. L. Page *et al.*, Short GRB 160821B: A reverse shock, a refreshed shock, and a well-sampled kilonova, Astrophys. J. **883**, 48 (2019).

[37] E. Troja, A. J. Castro-Tirado, J. Becerra González, Y. Hu, G. S. Ryan, S. B. Cenko, R. Ricci, G. Novara, R. Sánchez-Rámirez, J. A. Acosta-Pulido *et al.*, The afterglow and kilonova of the short GRB 160821B, Mon. Not. R. Astron. Soc. **489**, 2104 (2019).

[38] Z.-P. Jin, X. Li, H. Wang, Y.-Z. Wang, H.-N. He, Q. Yuan, F.-W. Zhang, Y.-C. Zou, Y.-Z. Fan, and D.-M. Wei, Short GRBs: Opening angles, local neutron star merger rate, and off-axis events for GRB/GW association, Astrophys. J. **857**, 128 (2018).

[39] Z.-P. Jin, S. Covino, N.-H. Liao, X. Li, P. D'Avanzo, Y.-Z. Fan, and D.-M. Wei, A kilonova associated with GRB 070809, Nature Astron. **4**, 77 (2020).

[40] E. O. Ofek, S. B. Cenko, A. Gal-Yam, D. B. Fox, E. Nakar, A. Rau, D. A. Frail, S. R. Kulkarni, P. A. Price, B. P. Schmidt *et al.*, GRB 060505: A possible short-duration gamma-ray burst in a star-forming region at a redshift of 0.09, Astrophys. J. **662**, 1129 (2007).

[41] B. Yang, Z.-P. Jin, X. Li, S. Covino, X.-Z. Zheng, K. Hotokezaka, Y.-Z. Fan, T. Piran, and D.-M. Wei, A possible macronova in the late afterglow of the long-short burst GRB 060614, Nat. Commun. **6**, 7323 (2015).

[42] B. Zhang, B.-B. Zhang, E.-W. Liang, N. Gehrels, D. N. Burrows, and P. Mészáros, Making a short gamma-ray burst from a long ine: Implications for the nature of GRB 060614, Astrophys. J. **655**, L25 (2007).

[43] E. Troja, C. L. Fryer, B. O'Connor, G. Ryan, S. Dichiara, A. Kumar, N. Ito, R. Gupta, R. T. Wollaeger, J. P. Norris *et al.*, A

nearby long gamma-ray burst from a merger of compact objects, Nature (London) **612**, 228 (2022).

[44] J. C. Rastinejad, B. P. Gompertz, A. J. Levan, W.-F. Fong, M. Nicholl, G. P. Lamb, D. B. Malesani, A. E. Nugent, S. R. Oates, N. R. Tanvir *et al.*, A kilonova following a long-duration gamma-ray burst at 350 Mpc, Nature (London) **612**, 223 (2022).

[45] J. Yang, S. Ai, B.-B. Zhang, B. Zhang, Z.-K. Liu, X. I. Wang, Y.-H. Yang, Y.-H. Yin, Y. Li, and H.-J. Lü, A long-duration gamma-ray burst with a peculiar origin, Nature (London) **612**, 232 (2022).

[46] E. Troja, Eighteen years of kilonova discoveries with Swift, Universe **9**, 245 (2023).

[47] V. A. Villar, J. Guillochon, E. Berger, B. D. Metzger, P. S. Cowperthwaite, M. Nicholl, K. D. Alexander, P. K. Blanchard, R. Chornock, T. Eftekhari *et al.*, The combined ultraviolet, optical, and near-infrared light curves of the kilonova associated with the binary neutron star merger GW170817: Unified data set, analytic models, and physical implications, Astrophys. J. Lett. **851**, L21 (2017).

[48] A. Perego, D. Radice, and S. Bernuzzi, AT 2017gfo: An anisotropic and three-component kilonova counterpart of GW170817, Astrophys. J. Lett. **850**, L37 (2017).

[49] M. Shibata and K. Hotokezaka, Merger and mass ejection of neutron star binaries, Annu. Rev. Nucl. Part. Sci. **69**, 41 (2019).

[50] J. Heinzel, M. W. Coughlin, T. Dietrich, M. Bulla, S. Antier, N. Christensen, D. A. Coulter, R. J. Foley, L. Issa, and N. Khetan, Comparing inclination-dependent analyses of kilonova transients, Mon. Not. R. Astron. Soc. **502**, 3057 (2021).

[51] M. Almualla, Y. Ning, M. Bulla, T. Dietrich, M. W. Coughlin, and N. Guessoum, Using neural networks to perform rapid high-dimensional kilonova parameter inference, arXiv:2112.15470.

[52] M. Ristic, E. Champion, R. O'Shaughnessy, R. Wollaeger, O. Korobkin, E. A. Chase, C. L. Fryer, A. L. Hungerford, and C. J. Fontes, Interpolating detailed simulations of kilonovae: Adaptive learning and parameter inference applications, Phys. Rev. Res. **4**, 013046 (2022).

[53] K. Lukošiute, G. Raaijmakers, Z. Doctor, M. Soares-Santos, and B. Nord, KilonovaNet: Surrogate models of kilonova spectra with conditional variational autoencoders, Mon. Not. R. Astron. Soc. **516**, 1137 (2022).

[54] S. Anand, P. T. H. Pang, M. Bulla, M. W. Coughlin, T. Dietrich, B. Healy, T. Hussenot-Desenonges, T. Jegou du Laz, M. M. Kasliwal, N. Kunert *et al.*, Chemical distribution of the dynamical ejecta in the neutron star merger GW170817, arXiv:2307.11080.

[55] N. M. Ford, N. Vieira, J. J. Ruan, and D. Haggard, Kilonovae: Exploring kilonova spectral features with autoencoders, Astrophys. J. **961**, 119 (2024).

[56] M. Ristić, R. O'Shaughnessy, V. A. Villar, R. T. Wollaeger, O. Korobkin, C. L. Fryer, C. J. Fontes, and A. Kedia, Interpolated kilonova spectra models: Examining the effects of a phenomenological, blue component in the fitting of AT2017gfo spectra, Phys. Rev. Res. **5**, 043106 (2023).

[57] A. Kedia, M. Ristic, R. O'Shaughnessy, A. B. Yelikar, R. T. Wollaeger, O. Korobkin, E. A. Chase, C. L. Fryer, and C. J. Fontes, Surrogate light curve models for kilonovae with comprehensive wind ejecta outflows and parameter estimation for AT2017gfo, Phys. Rev. Res. **5**, 013168 (2023).

[58] P. Darc, C. R. Bom, B. M. O. Fraga, and C. D. Kilpatrick, Kilonova spectral inverse modelling with simulation-based inference: An amortized neural posterior estimation analysis, arXiv:2311.09471.

[59] S. Saha, M. J. Williams, L. Datrier, F. Hayes, M. Nicholl, A. K. H. Kong, M. Hendry, I. K. S. Heng, G. P. Lamb, E.-T. Lin, and D. Williams, Rapid generation of kilonova light curves using conditional variational autoencoder, Astrophys. J. **961**, 165 (2024).

[60] R. T. Wollaeger, C. L. Fryer, E. A. Chase, C. J. Fontes, M. Ristic, A. L. Hungerford, O. Korobkin, R. O'Shaughnessy, and A. M. Herring, A broad grid of 2D kilonova emission models, Astrophys. J. **918**, 10 (2021).

[61] O. Korobkin, R. T. Wollaeger, C. L. Fryer, A. L. Hungerford, S. Rosswog, C. J. Fontes, M. R. Mumpower, E. A. Chase, W. P. Even, J. Miller *et al.*, Axisymmetric radiative transfer models of kilonovae, Astrophys. J. **910**, 116 (2021).

[62] O. Just, V. Vijayan, Z. Xiong, S. Goriely, T. Soultanis, A. Bauswein, J. Guilet, H. T. Janka, and G. Martínez-Pinedo, End-to-end kilonova models of neutron star mergers with delayed black hole formation, Astrophys. J. Lett. **951**, L12 (2023).

[63] R. T. Wollaeger and D. R. van Rossum, Explosive outflows: Opacity regrouping, Astrophys. J. Suppl. Series **214**, 28 (2014).

[64] C. Winteler, R. Käppeli, A. Perego, A. Arcones, N. Vasset, N. Nishimura, M. Liebendörfer, and F. K. Thielemann, Magnetorotationally driven supernovae as the origin of early galaxy r-process elements? Astrophys. J. **750**, L22 (2012).

[65] O. Korobkin, S. Rosswog, A. Arcones, and C. Winteler, On the astrophysical robustness of the neutron star merger r-process, Mon. Not. R. Astron. Soc. **426**, 1940 (2012).

[66] M. Reichert, C. Winteler, O. Korobkin, A. Arcones, J. Bliss, M. Eichler, U. Frischknecht, C. Fröhlich, R. Hirschi, M. Jacobi *et al.*, The nuclear reaction network WinNet, Astrophys. J. Suppl. Series **268**, 66 (2023).

[67] M. Reichert, C. Winteler, O. Korobkin, A. Arcones, J. Bliss, M. Eichler, U. Frischknecht, C. Fröhlich, R. Hirschi, M. Jacobi *et al.*, WinNet: v1.0.1., Zenodo (2023), https://zenodo.org/records/8220550.

[68] J. Barnes, D. Kasen, M.-R. Wu, and G. Martínez-Pinedo, Radioactivity and thermalization in the ejecta of compact object mergers and their impact on kilonova light curves, Astrophys. J. **829**, 110 (2016).

[69] R. T. Wollaeger, O. Korobkin, C. J. Fontes, S. K. Rosswog, W. P. Even, C. L. Fryer, J. Sollerman, A. L. Hungerford, D. R. van Rossum, and A. B. Wollaber, Impact of ejecta morphology and composition on the electromagnetic signatures of neutron star mergers, Mon. Not. R. Astron. Soc. **478**, 3298 (2018).

[70] C. J. Fontes, H. L. Zhang, J. Abdallah, Jr., R. E. H. Clark, D. P. Kilcrease, J. Colgan, R. T. Cunningham, P. Hakel, N. H. Magee, and M. E. Sherrill, The Los Alamos suite of relativistic atomic physics codes, J. Phys. B: At. Mol. Opt. Phys. **48**, 144014 (2015).

[71] C. J. Fontes, C. L. Fryer, A. L. Hungerford, R. T. Wollaeger, and O. Korobkin, A line-binned treatment of opacities for the spectra and light curves from neutron star mergers, Mon. Not. R. Astron. Soc. **493**, 4143 (2020).

[72] K. Olsen, C. J. Fontes, C. L. Fryer, A. L. Hungerford, R. T. Wollaeger, O. Korobkin, and Y. Ralchenko, NIST-LANL lanthanide opacity database (ver. 1.2) (2023), https://nlte.nist.gov/OPAC/.

[73] K. P. Mooley, A. T. Deller, O. Gottlieb, E. Nakar, G. Hallinan, S. Bourke, D. A. Frail, A. Horesh, A. Corsi, and K. Hotokezaka, Superluminal motion of a relativistic jet in the neutron-star merger GW170817, Nature (London) **561**, 355 (2018).

[74] A. Bauswein, S. Blacker, V. Vijayan, N. Stergioulas, K. Chatziioannou, J. A. Clark, N.-U. F. Bastian, D. B. Blaschke, M. Cierniak, and T. Fischer, Equation of state constraints from the threshold binary mass for prompt collapse of neutron star mergers, Phys. Rev. Lett. **125**, 141103 (2020).

[75] E. Troja, H. van Eerten, B. Zhang, G. Ryan, L. Piro, R. Ricci, B. O'Connor, M. H. Wieringa, S. B. Cenko, and T. Sakamoto, A thousand days after the merger: Continued x-ray emission from GW170817, Mon. Not. R. Astron. Soc. **498**, 5643 (2020).

[76] P. Camps and M. Baes, The failure of Monte Carlo radiative transfer at medium to high optical depths, Astrophys. J. **861**, 80 (2018).

[77] J. Wofford, A. B. Yelikar, H. Gallagher, E. Champion, D. Wysocki, V. Delfavero, J. Lange, C. Rose, V. Valsan, S. Morisaki *et al.*, Improving performance for gravitational-wave parameter inference with an efficient and highly-parallelized algorithm, Phys. Rev. D **107**, 024040 (2023).

[78] V. Tiwari, C. Hoy, S. Fairhurst, and D. MacLeod, Fast non-Markovian sampler for estimating gravitational-wave posteriors, Phys. Rev. D **108**, 023001 (2023).

[79] M. Vallisneri, Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects, Phys. Rev. D **77**, 042001 (2008).

[80] D. Finstad, S. De, D. A. Brown, E. Berger, and C. M. Biwer, Measuring the viewing angle of GW170817 with electromagnetic and gravitational waves, Astrophys. J. Lett. **860**, L2 (2018).

[81] M. Ristić, Smarter simulation placement of kilonova light curve models for computationally inexpensive surrogate model creation, Thesis, Rochester Institute of Technology, 2020.

[82] 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, Nature Methods **17**, 261 (2020).

[83] S. Cook, A. Gelman, and D. Rubin, Validation of software for Bayesian models using posterior quantiles, J. Comput. Graphical Stat. **15**, 675 (2006).

[84] T. Sidery, B. Aylott, N. Christensen, B. Farr, W. Farr, F. Feroz, J. Gair, K. Grover, P. Graff, C. Hanna *et al.*, Reconstructing the sky location of gravitational-wave detected compact binary systems: Methodology for testing and comparison, Phys. Rev. D **89**, 084060 (2014).