# Data-driven approach for modeling the temporal and spectral evolution of kilonova systematic uncertainties

Sahil Jhawar,[1,2,*] Thibeau Wouters,[3,4] Peter T. H. Pang,[4,3] Mattia Bulla,[5,6,7] Michael W. Coughlin,[8] and Tim Dietrich[1,9]

[1]*Institut für Physik und Astronomie, Universität Potsdam, Haus 28, Karl-Liebknecht-Strasse 24/25, 14476, Potsdam, Germany*
[2]*GFZ German Research Centre For Geosciences, 14473 Potsdam, Germany*
[3]*Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, Princetonplein 1, 3584 CC Utrecht, The Netherlands*
[4]*Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands*
[5]*Department of Physics and Earth Science, University of Ferrara, via Saragat 1, I-44122 Ferrara, Italy*
[6]*INFN, Sezione di Ferrara, via Saragat 1, I-44122 Ferrara, Italy*
[7]*INAF, Osservatorio Astronomico d'Abruzzo, via Mentore Maggini snc, 64100 Teramo, Italy*
[8]*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA*
[9]*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam 14476, Germany*

(Received 6 November 2024; accepted 27 January 2025; published 21 February 2025)

Kilonovae, possible electromagnetic counterparts to neutron star mergers, provide important information about high-energy transient phenomena and, in principle, also allow us to obtain information about the source properties responsible for powering the kilonova. Unfortunately, numerous uncertainties exist in kilonova modeling that, at the current stage, hinder accurate predictions. Hence, one has to account for possible systematic modeling uncertainties when interpreting the observed transients. In this work, we provide a data-driven approach to account for time-dependent and filter-dependent uncertainties in kilonova models. Through a suite of tests, we find that the most reliable recovery of the source parameters and description of the observational data can be obtained through a combination of kilonova models with time- and filter-dependent systematic uncertainties. We apply our new method to analyze AT2017gfo. While recovering a total ejecta mass consistent with previous studies, our approach gives insights into the temporal and spectral evolution of the systematic uncertainties of this kilonova. We consistently find a systematic error below 1 mag between 1 to 5 days after the merger. Our work addresses the need for early follow-up of kilonovae at earlier times, and improved modeling of the kilonova at later times, to reduce the uncertainties outside of this time window.

DOI: 10.1103/PhysRevD.111.043046

## I. INTRODUCTION

The gravitational-wave (GW) detection of the binary neutron star (BNS) merger GW170817 [1,2] by Advanced LIGO [3] and Advanced Virgo [4] GW observatories, combined with the observation of electromagnetic (EM) waves from the short gamma-ray burst GRB170817A and the kilonova AT2017gfo, has been a game changer in our understanding of, e.g., cosmology [5–14], nuclear physics [1,12,15–21], modified theories of gravity [22–24], and the chemical evolution of our Universe [25–30]. Most of these studies relied on the availability of information from multiple messengers, which in this case are, GWs and EM waves.

Although GW170817 has been the only multimessenger detection of a BNS merger, there has been observational evidence that GW190425 [31], GRB211211 [32–34], and GRB230307A [35] also originated from BNS mergers. Hence, the chances for further multimessenger detections are continuously increasing due to the increasing range of observational facilities. For a reliable interpretation of GW170817 and future events, it is necessary to compare the observational data with theoretical predictions to extract characteristic information from the sources. The employed theoretical models must be accurate to ensure an unbiased

*Contact author: jhawar@uni-potsdam.de

estimate. Otherwise, the ever-increasing number of observational data or the observation of sources with higher accuracy would potentially lead to biased constraints.

Numerous studies have addressed the accuracy of GW models, e.g., Refs. [36–40], and it has been shown that systematic uncertainties are under control for the current generation of GW detectors; yet, better models are needed for the next generation of detectors, such as the Einstein Telescope [41–43] and the Cosmic Explorer [44]. In contrast, systematic uncertainties are more severe and not yet fully understood for kilonovae, an EM transient powered by the radioactive decay of unstable heavy elements synthesized through r-process nucleosynthesis [45,46] and spanning optical, infrared, and ultraviolet frequencies. Despite numerous efforts improving the modeling of kilonovae, e.g., by moving from simplified semi-analytical models (e.g, [30,46–54]) to more complicated full 3D radiative transfer simulations (e.g., [55–60]), there are still large uncertainties in the modeling. Among the main sources of uncertainty are properties of the ejected material [61–65] and some key ingredients setting the energy available to power the kilonova, i.e., nuclear heating rates and thermalization efficiencies, and the properties of the escaping radiation, i.e., the opacities of r-process elements [66–73].

Given the significant uncertainties in kilonova modeling, it is essential to account for them during multimessenger parameter inference. In the past, most works employing the nuclear-physics and multimessenger astrophysics (NMMA) framework, which is the code infrastructure used for this article, accounted for these uncertainties by including a fixed systematic uncertainty of 1 magnitude [12,34,74,75]. This particular choice was motivated by the study of Ref. [61] where it was found that this uncertainty is sufficient to ensure that different ejecta morphologies assumed by different radiative transfer simulation codes make similar predictions, i.e., that the extracted ejecta properties are consistent within their statistical uncertainties. More recently, NMMA inference runs considered the uncertainty not as a fixed parameter, but as an additional sampling parameter [76].

Similar approaches have also been used in other works, e.g., [77–79]. In particular, Ref. [77] has updated its code infrastructure to account for such systematic uncertainty by adding a fixed systematic error standard deviation to the likelihood quadrature. A more flexible approach has been employed by Ref. [79] in which the systematic uncertainty is a sampling parameter, and Ref. [78] allowed more flexibility to account for limitations in the employed kilonova models by adding different systematic uncertainty priors across different observational bands.

However, the systematic uncertainty of a kilonova model will, in general, not only be dependent on the observational bands, i.e., filter-dependent, but will also vary over time. This time dependence appears naturally when accounting for known uncertainties in the description of opacities at early times [[68,80] e.g.] or nonlocal thermal equilibrium effects that become important after about 1 week [81,82].

In this work, we introduce a data-driven scheme for handling the time- and filter-dependent uncertainties in the light curve models. This method promises to improve the robustness of the parameter estimations. It is model-agnostic and can be applied to any EM model irrespective of the nature of the transient, e.g., also for gamma-ray-burst afterglows or supernovae.

This paper is structured as follows. Section II describes the methodological approach, including the radiative transfer models, the Bayesian inference framework, and the systematic error interpolation schemes employed. Section III provides validation tests using synthetic light curve data from different models. Section IV demonstrates the application of the methodology to AT2017gfo. Finally, Sec. V summarizes the key findings and outlines future perspectives.

## II. METHODS

### A. Kilonova light curve computation

An accurate description of the observables of kilonovae, i.e., luminosities, spectra, light curves, and polarization, requires detailed modeling of the radiation processes, incorporating the interaction between the radiation and the matter via absorption and scattering processes. In the following, we are employing results from two independent radiative-transfer codes to enable cross-validation and testing of our approach.

`POSSIS` [58,68] and `SEDONA` [83] are 3D Monte Carlo radiative transfer codes that model synthetic observable (flux and polarization spectral times series) for explosive transients such as supernovae and kilonovae. Both codes incorporate time-dependent opacities and ejecta properties, enabling them to capture the evolving nature of astrophysical events over various timescales. Inside the model grid, each cell is represented by velocity $v$, time-dependent density $\rho(t)$, time-dependent temperature $T(t)$, and the electron fraction $Y_e(t)$, starting at some reference time $t_0$.

The computation of the observables is based on the simulation of Monte Carlo photon packets diffusing through the freely expanding medium. Monte Carlo photons are created with energy and frequencies set by the specific emission process (e.g., radioactivity) and are propagated according to the opacity of the expanding medium. Those that escape the medium are then used to construct spectral-time series (from ultraviolet to infrared) at different viewing angles, from which light curves in different filters can be constructed.

#### 1. *BU2019LM*

While `POSSIS` can generally support arbitrary geometries, we focus here on the usage of a two-component model

grid; cf. Ref. [12] for more details. The first component characterizes dynamical ejecta and contains a lanthanide-rich part around the equatorial plane and a lanthanide-free part at the polar regions. The second component accounts for wind-driven ejecta and is spherically symmetric. The model, hereafter denoted with BU2019LM, is parametrized by the ejecta mass of the two components, $M_{\rm ej,dyn}$ and $M_{\rm ej,wind}$, the half-opening angle of the lanthanide-rich dynamical-ejecta component, $\phi$, and the viewing angle $\theta_{\rm obs}$.

#### 2. KA2017

Although SEDONA is a 3D code, here we focus on a grid of one-dimensional spherically symmetric models presented in Ref. [28]. The model, in the following referred to as KA2017, is parametrized by the ejecta total mass $M_{\rm ej}$, average velocity $v_{\rm ej}$, and lanthanide fraction $X_{\rm lan}$. The bulk of the freely expanding ejecta is determined by the ejecta mass $M_{\rm ej}$. The density profile of the ejecta is described using a broken power-law that transits from the gradually declining interior with $v_{\rm ej}t/r$ to the steeply dropping outer layer with $(v_{\rm ej}t/r)^{10}$. Finally, the lanthanide fraction $X_{\rm lan}$ influences the opacity and color evolution of the kilonova, where larger lanthanide fractions result in increased opacity and longer-duration emissions shifted toward the infrared.

### B. Surrogate kilonova models

The POSSIS and the SEDONA codes are computationally too expensive to be run on the fly during sampling. Therefore, we train surrogate models for the BU2019LM and KA2017 grids that are cheaper to execute during Bayesian inference.

For BU2019LM, we use the dataset of 1596 parameter combinations and their light curves generated by POSSIS to create the surrogate model. For each filter that we consider in our analyses below, we perform a singular-value decomposition (SVD) to reduce the dimensionality of the output, setting the number of SVD components to 10. Then, we train a fully connected artificial neural network that maps the values of the BU2019LM parameters to the SVD coefficients, from which the light curve can be reconstructed. The architecture of the neural network consists of three hidden layers, having 128, 256, and 128 neurons, respectively. Training is done with TensorFlow [84] and runs with the Adam optimizer [85], with a fixed learning rate of $10^{-3}$ and a batch size of 128, for 100 epochs. We rescale the input and output data with a min-max scaler before training and use 20% of the dataset as validation data to ensure that the network is not overfitting.

For KA2017, we use the publicly available[1] kilonova light curves produced with SEDONA. The full dataset contains 329 parameter combinations, which are used to create the surrogate model. A similar SVD and neural network training is performed for the KA2017 model as for BU2019LM. However, here we we used a hidden layer of 2048 neurons, followed by a dropout layer with a dropout rate of 0.6.

To quantify the performance of our surrogate models, we compute the root mean square error (RMSE). The RMSE quantifies the average deviation between model predictions and actual values, which is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}, \tag{1}$$

where $y_i$ is the actual value obtained through the radiative-transfer simulation at time $t_i$, $\hat{y}_i$ denotes the prediction from the surrogate model, and $N$ is the total number of time points. We compute this metric across our entire parameter grid and all filters and report its median value.

We find median RMSE values of 0.103 mag for BU2019LM (computed over a 14-day period) and 0.485 mag for KA2017 (computed over a 7-day period).

As discussed later in Sec. IV B regarding time- and filter-dependent uncertainties, the median $\sigma_{\rm sys}$ is 0.291 mag, greater than the surrogate model's median RMSE. This suggests that the BU2019LM surrogate model is sufficiently accurate for our study, and we are not over-confident about it.

### C. Bayesian inference

By using Bayes' theorem, the posterior $p(\vec{\theta}|d,\mathcal{H})$ on a set of parameters $\vec{\theta}$ under the hypothesis $\mathcal{H}$ and with data $d$ is given by

$$p(\vec{\theta}|d,\mathcal{H}) = \frac{p(d|\vec{\theta},\mathcal{H})p(\vec{\theta}|\mathcal{H})}{p(d|\mathcal{H})}, \tag{2}$$

or in short form

$$\mathcal{P}(\vec{\theta}) = \frac{\mathcal{L}(\vec{\theta})\pi(\vec{\theta})}{\mathcal{Z}}, \tag{3}$$

where $\mathcal{P}(\vec{\theta})$, $\mathcal{L}(\vec{\theta})$, $\pi(\vec{\theta})$, and $\mathcal{Z}$ are the posterior, likelihood, prior, and evidence, respectively. The prior describes our knowledge of the parameters before any observations. The likelihood quantifies how well the hypothesis can describe the data at a given point $\vec{\theta}$ in the parameter space. Finally, the evidence, also known as the marginalized likelihood, marginalizes the likelihood over the whole parameter space with respect to the prior, i.e.,

$$\mathcal{Z} = \int d\vec{\theta}\mathcal{L}(\vec{\theta})\pi(\vec{\theta}). \tag{4}$$

[1] https://github.com/dnkasen/Kasen_Kilonova_Models_2017.

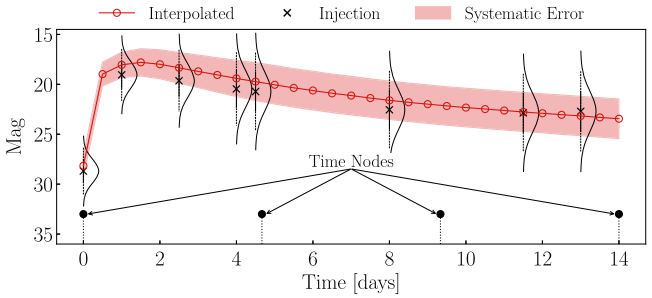


FIG. 1. The stem plot shows the placement of four time nodes at 0, 4.67, 9.34, and 14 days used for interpolation.

### D. Time-dependent systematic error

To parametrize the time-dependence of the systematic uncertainty $\sigma(t)$, we make use of a piecewise linear interpolation scheme with $N$ evenly spaced time nodes

$$\sigma_{\rm sys}(t) = \sigma_n + \frac{\sigma_{n+1}-\sigma_n}{t_{n+1}-t_n}\cdot(t-t_n), \quad \text{for } t_n \le t < t_{n+1} \quad \text{with} \quad n = 0, 1, \ldots, N-1. \tag{5}$$

Each $\sigma_n$ has a uniform prior in $[a, b)$, where $a$ and $b$ are the assumed lower and upper bound of the systematic uncertainty. Throughout our analyses, we used a prior of $\mathcal{U}(0, 2)$.

The likelihood function $\mathcal{L}(\vec{\theta})$ is given by

$$\mathcal{L}(\vec{\theta}) \propto \exp\left(-\frac{1}{2}\sum_{ij}\frac{(m_i^j - m_i^{j,\rm est}(\vec{\theta}))^2}{(\sigma_i^j)^2 + (\sigma_{{\rm sys},i})^2}\right), \tag{6}$$

where $m_i^j$ is the AB magnitude in filter $j$ at time $i$ with the corresponding measurement error $\sigma_i^j \equiv \sigma^j(t_i)$, $m_i^{j,\rm est}(\vec{\theta})$ is the estimated AB magnitude for the source parameters $\vec{\theta}$ (e.g., ejecta masses, velocities) from the model and $\sigma_{{\rm sys},i}$ is the interpolated systematic error at time $t_i$.

The described procedure is sketched in Fig. 1; where black crosses mark the mock data. The red line represents the best-fit light curve, the red-shaded region represents the systematic uncertainty in the best-fit posterior, and the Gaussian curve around them represents the denominator of Eq. (6).

### E. Time- and filter-dependent systematic error

To extend our systematic error analysis, we also implement an additional filter-dependence. This enables us to consider more general cases in which uncertainties vary in time, and the accuracy of model predictions is filter dependent, e.g., due to filter-dependent uncertainties of the opacities; cf. Ref. [80]. For this purpose, we will allow different systematic errors, $\sigma_{\rm sys}^j(t_i)$ for different observational filters. In this approach, the filters that need to be sampled independently and jointly are based on evaluating the mean absolute deviation (MAD) for each filter and comparing it to the overall MAD calculated across all filters.

The MAD is a statistical measure of the variability and dispersion of data values, and it is used here to determine the extent to which each filter contributes to the overall variability of the data. The MAD for each filter $j$ is calculated as

$$\text{MAD}_j = \frac{1}{n_j}\sum_{i=1}^{n_j}|x_{i,j} - \mu_j|, \tag{7}$$

where $x_{i,j}$ is the AB magnitude in filter $j$ at time $t_i$, $\mu_j$ is the mean of all AB magnitudes in filter $j$, and $n_j$ is the total number of data points in filter $j$. Similarly, the total MAD is calculated as

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \mu|, \tag{8}$$

where $x_i$ is the AB magnitude at time $t_i$, $\mu$ is the mean of all AB magnitudes and $n$ is the total number of data points across all available filters.

The choice of filter for independent and joint systematic error is based on the comparison of the filter-specific $\text{MAD}_j$ to MAD. Given this, the likelihood in Eq. (6) can be rewritten as

$$\mathcal{L}(\vec{\theta}) \propto \exp\left(-\frac{1}{2}\sum_{ij}\frac{(m_i^j - m_i^{j,\rm est}(\vec{\theta}))^2}{(\sigma_i^j)^2 + (\sigma_{{\rm sys},i}^j)^2}\right), \tag{9}$$

where $\sigma_{{\rm sys},i}^j$ is the interpolated systematic error at time $t_i$ and filter $j$.

Such a likelihood is equivalent to including an additional shift to the light curve by $\Delta m$, and marginalizing it with a normal distribution with a mean of 0 and variance of $\sigma_{{\rm sys},i}^j$.

## III. VALIDATION

To validate our methodology, we simulated two synthetic light curves employing the BU2019LM and KA2017 models. We use a uniform time step of 0.5 days for sampling and randomly select 45% of the data (until 20 days) to account for partially missing data due to the 'lack' of observations, e.g., through bad weather conditions or other observational limitations. To account for errors in the observations, we add a random shift to each datapoint following a Gaussian distribution with zero mean and a standard deviation of 0.1 mag.

These simulated data serve as *injections* and are taken for up to one week for KA2017 and upto two weeks for

TABLE I. Parameter values used to generate the mock light curves to test the implemented algorithm.

| Parameter | Model | |
|---|---|---|
| | BU2019LM | KA2017 |
| $D_L$ (Mpc) | 40 | 40 |
| $\log_{10} M_{\rm ej}[M_\odot]$ | ⋯ | −1.43 |
| $\log_{10} v_{\rm ej}[c]$ | | −0.74 |
| $\log_{10} X_{\rm lan}$ | | −3.38 |
| $\Phi$ (deg) | 68.69 | ⋯ |
| $\iota$ (rad) | 0.43 | |
| $\log_{10} M_{\rm ej}^{\rm dyn}[M_\odot]$ | −1.18 | |
| $\log_{10} M_{\rm ej}^{\rm wind}[M_\odot]$ | −2.25 | |

BU2019LM after the merger, i.e., later data is not injected into the parameter estimation pipeline.[2]

For each injection, we use both models (BU2019LM and KA2017) for the recovery. The employed injection parameters are inspired by the parameters of AT2017gfo and summarized in Table I. To compare the posterior distributions of two competing models, we will mainly focus on the common parameters, i.e., the luminosity distance, $D_L$, and the total ejecta mass, $\log_{10} M_{\rm ej}$, where, $M_{\rm ej} = M_{\rm ej,wind} + M_{\rm ej,dyn}$ for the BU2019LM model.

### A. Constant systematic uncertainties

We start our analysis following a similar approach as employed in our previous studies, e.g., [12,61,74], using a constant 1 mag uncertainty. The obtained results are summarized in Fig. 2, where one can see that for all injections, the assumed 1 mag uncertainty is sufficient for the recovered light curves to approximate the injected light curves reliably, even when different models are used for the injection and recovery.

Considering the recovery of the injected parameters, we find that when using the same model for the injection and the recovery, the injected parameters can be recovered reliably within the statistical uncertainties without any visible bias. However, when a different model is employed for the recovery, we find that there can be systematic biases in the recovered posteriors of the source parameters. Generally, the BU2019LM model recovers injected values for both models better than the KA2017 model, which we assume is due to the two-component ejecta, which enables more flexibility during the light curve fitting.

### B. Time-dependent uncertainties

Relaxing the assumption of a constant 1 mag uncertainty and enabling a time-dependent uncertainty, we find significantly different results regarding the accuracy of the recovered light curves. In fact, using the same model for the injection and recovery can be considered our best-case scenario, in which the model is completely accurate and accounts for all the relevant physics of the kilonova light curve with respect to the model used for injection and recovery. Therefore, one can expect the estimated systematic uncertainty $\sigma_{\rm sys}(t)$ to be minimal across the whole time range. This expectation is fully confirmed by our test and is visible in Fig. 3, where the 90% posterior light curve band is extremely tight around the injected data and the injected light curve falls within the band across all filters and times. The obtained uncertainty of $\mathcal{O}(0.1$ mag$)$ is dominated by the uncertainty added to the injection data mimicking the uncertainties in obtained observational data. Varying these uncertainties, we verified that our method is able to pick up larger uncertainties if there is a larger spread in the injected data points.

This improved recovery of the light curves also leads to posteriors recovering the injected value with a smaller spread around the injected value; being more quantitative, we find a reduction of the spread of the posterior by up to a factor of two. This is clearly visible in Fig. 3.

As illustrated in Fig. 3, when injecting KA2017 light curves, both models successfully recover the injected parameters. However, the BU2019LM model's posterior is significantly broader (with more than 5 times larger uncertainty), which clearly indicates that with sufficient flexibility in systematic errors, BU2019LM can achieve satisfactory performance. In contrast, when attempting to recover BU2019LM injections with the KA2017 model, we observe a significant discrepancy. The recovered mass deviates substantially from the combined injected masses, suggesting that the KA2017 model, being a spherically symmetric, single component ejecta model, is too limited to accurately represent kilonova light curves that have a larger variability and complexity.

### C. Filter-dependent uncertainties

Finally, we present in Fig. 4 an analysis in which we employ a time- and filter-dependent uncertainty during our recovery. Based on initial tests, in particular when studying AT2017gfo, we have found the most significant differences in the ultraviolet $u$-band and the infrared $K$-band. For this reason, we decided to allow different systematic uncertainties in these two bands and group all other bands, i.e., $g$ to $H$, together using the same uncertainty. Clearly, this particular choice is not unique, and numerous other options

[2]We decided to reduce the length of the injection of the KA2017 model to 1 week since some of the light curves show unphysical features after about 1 week, e.g., an increasing luminosity due to low signal-to-noise from the underlying simulations at late epochs. However, we have also checked our results to be robust for a length of 10 and 14 days.

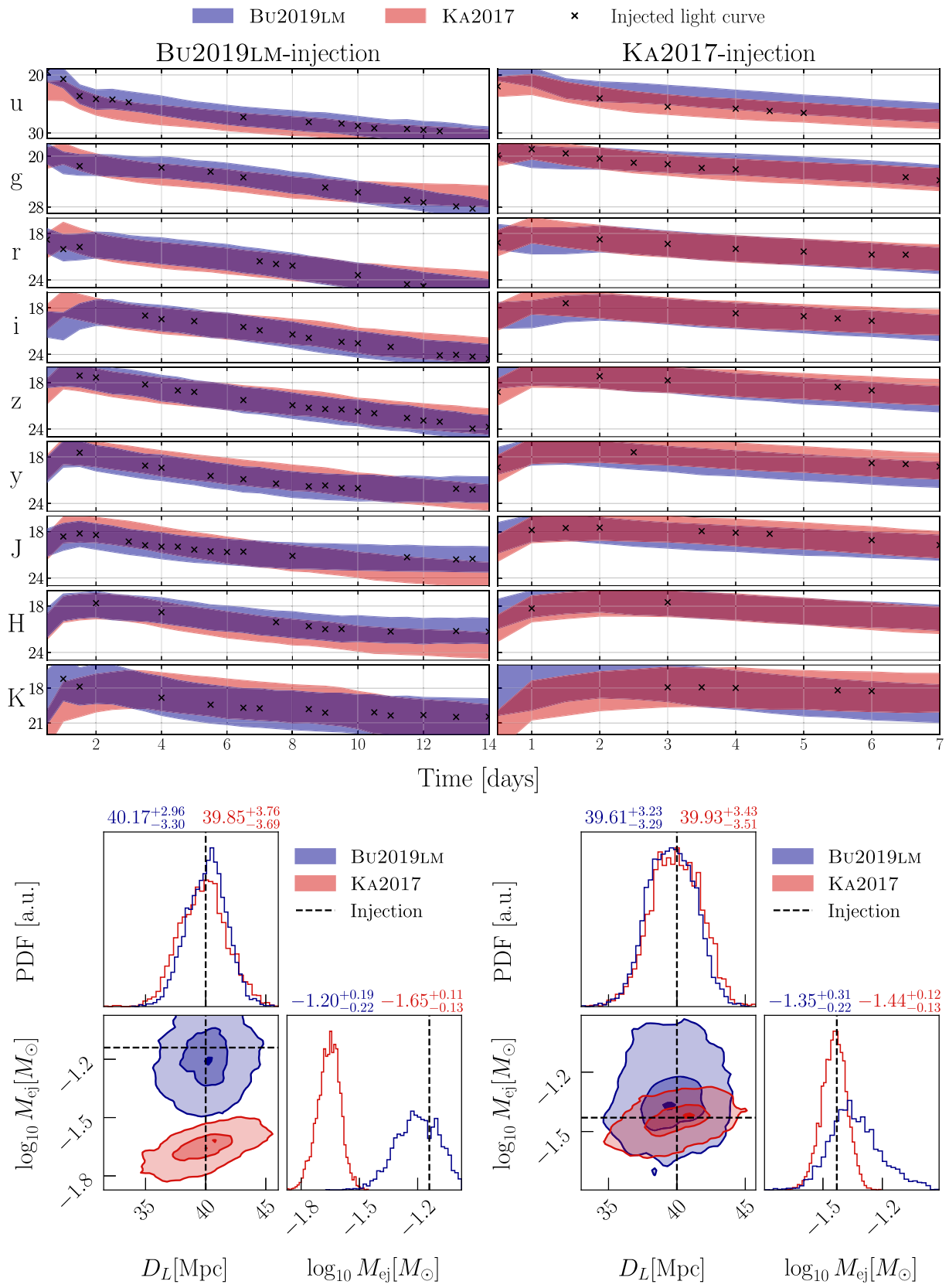


FIG. 2. Top panels: light curves for the validation test employing a constant 1-magnitude uncertainty. The crosses represent the injected light curve employed as input for the Bayesian inference. The band represents the 90% credibility region of the light curves generated from the posterior samples. Bottom panels: 2D marginalized posteriors of BU2019LM (left) and KA2017 (right) with $2\sigma$ shaded region and injected parameter.

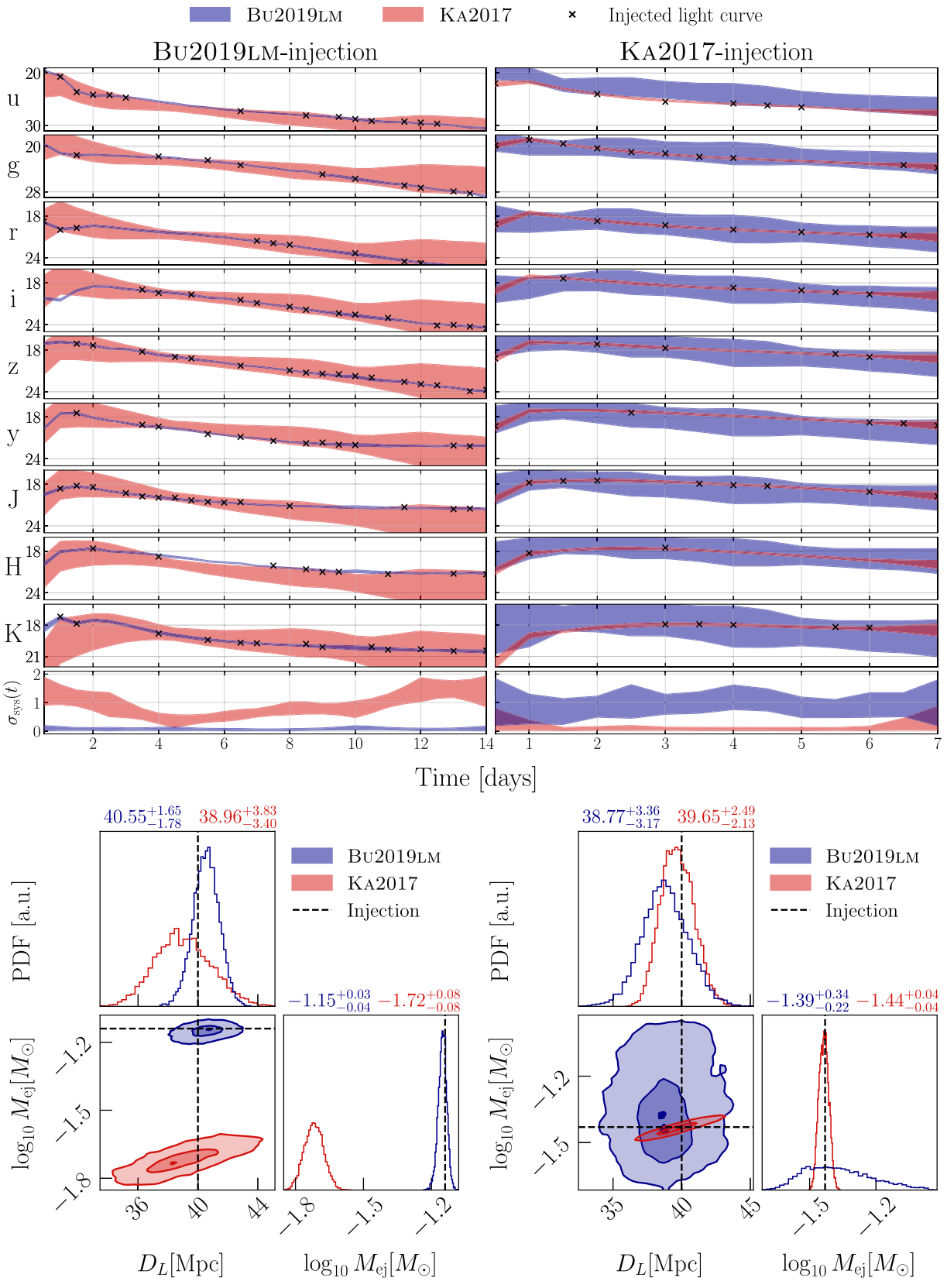


FIG. 3. Same as Fig. 2, for time-dependent systematic uncertainty. The bottom panel in the light curve plots illustrates the time-discretized systematic uncertainty, where the band represents the 90% credibility region of the reinterpolated systematic uncertainty, $\sigma_{\rm sys}(t)$.

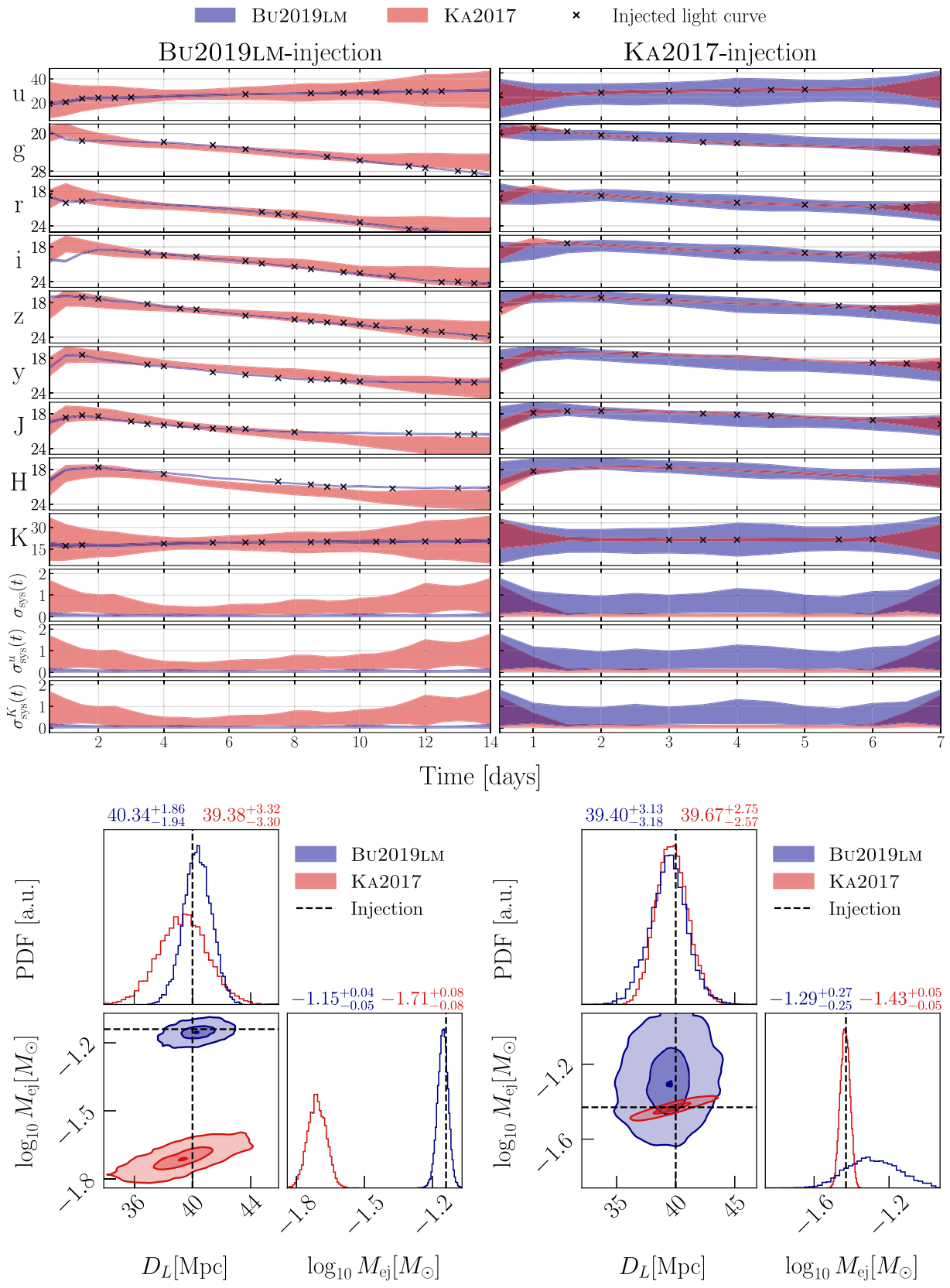


FIG. 4. Same as Fig. 3, for time- and filter-dependent systematic uncertainty. $\sigma^{u}_{\rm sys}(t)$ and $\sigma^{K}_{\rm sys}(t)$ represents systematic uncertainty for independently sampled $u$ and $K$ band, $\sigma_{\rm sys}(t)$ is for rest all the bands sampled together.

TABLE II. Posterior values with $2\sigma$ credibility, maximum log-likelihoods ($\ln \mathcal{L}^{\max}$), and maximum log-likelihood ratios ($\ln \mathcal{L}^{\max}_{\mathrm{ref}}$) values for AT2017gfo with different systematic configuration. First row: values with a constant systematic error of 1 mag. Second row: values with a free, but time- and filter-independent systematic error with a prior of $\mathcal{U}(0,2)$. Last row: values with $u$ and $K$ band systematic uncertainties sampled independently.

| Parameter | $D_L$ (Mpc) | $\Phi$ (deg) | $\iota$ (rad) | $\log_{10} M_{\mathrm{ej}}^{\mathrm{dyn}}[M_\odot]$ | $\log_{10} M_{\mathrm{ej}}^{\mathrm{wind}}[M_\odot]$ | $\ln \mathcal{L}^{\max}$ | $\ln \mathcal{L}^{\max}_{\mathrm{ref}}$ | Runtime |
|---|---|---|---|---|---|---|---|---|
| | Nodes | | | | | | | |
| Prior | $\mathcal{N}(40, 1.89)$ | $\mathcal{U}(15, 75)$ | $\mathcal{N}(0.37, 0.04)$ | $\mathcal{U}(-3, -1)$ | $\mathcal{U}(-3, -0.5)$ | … | … | … |
| … | $40.33^{+2.76}_{-2.48}$ | $60.53^{+4.77}_{-4.88}$ | $0.40^{+0.07}_{-0.09}$ | $-1.97^{+0.09}_{-0.07}$ | $-1.27^{+0.06}_{-0.05}$ | −145.162 | −51.648 | 05 m 57 s |
| … | $40.53^{+3.05}_{-2.85}$ | $60.65^{+7.19}_{-8.99}$ | $0.37^{+0.08}_{-0.08}$ | $-1.96^{+0.13}_{-0.13}$ | $-1.26^{+0.09}_{-0.08}$ | −119.275 | −25.762 | 11 m 16 s |
| 4 | $42.83^{+2.53}_{-2.78}$ | $60.65^{+3.57}_{-4.48}$ | $0.40^{+0.08}_{-0.08}$ | $-1.91^{+0.08}_{-0.10}$ | $-1.27^{+0.07}_{-0.07}$ | −98.971 | −5.457 | 20 m 15 s |
| 6 | $43.28^{+2.46}_{-2.75}$ | $60.64^{+3.50}_{-3.92}$ | $0.40^{+0.08}_{-0.08}$ | $-1.93^{+0.08}_{-0.10}$ | $-1.31^{+0.07}_{-0.08}$ | −97.729 | −4.215 | 33 m 47 s |
| 8 | $42.75^{+2.52}_{-2.60}$ | $60.30^{+3.57}_{-4.76}$ | $0.40^{+0.07}_{-0.08}$ | $-1.89^{+0.07}_{-0.09}$ | $-1.27^{+0.06}_{-0.07}$ | −93.514 | Ref. | 46 m 52 s |
| 10 | $42.83^{+2.53}_{-2.77}$ | $59.97^{+3.74}_{-4.52}$ | $0.40^{+0.08}_{-0.08}$ | $-1.92^{+0.08}_{-0.10}$ | $-1.30^{+0.08}_{-0.08}$ | −93.630 | −0.116 | 1 h 02 m 48 s |
| 12 | $42.61^{+2.59}_{-2.72}$ | $60.81^{+3.58}_{-4.47}$ | $0.40^{+0.08}_{-0.08}$ | $-1.91^{+0.08}_{-0.10}$ | $-1.28^{+0.07}_{-0.08}$ | −92.676 | 0.838 | 1 h 06 m 50 s |
| 14 | $42.54^{+2.52}_{-2.59}$ | $60.36^{+3.40}_{-3.91}$ | $0.40^{+0.08}_{-0.08}$ | $-1.92^{+0.08}_{-0.09}$ | $-1.31^{+0.07}_{-0.07}$ | −92.530 | 0.984 | 1 h 40 m 02 s |
| 16 | $42.25^{+2.56}_{-2.70}$ | $59.98^{+3.58}_{-4.31}$ | $0.40^{+0.08}_{-0.08}$ | $-1.91^{+0.07}_{-0.09}$ | $-1.31^{+0.08}_{-0.08}$ | −92.382 | 1.132 | 1 h 46 m 40 s |
| 8 | $42.56^{+2.45}_{-2.32}$ | $60.87^{+3.27}_{-3.60}$ | $0.39^{+0.08}_{-0.07}$ | $-1.91^{+0.07}_{-0.08}$ | $-1.29^{+0.06}_{-0.07}$ | −94.915 | −1.401 | 56 m 40 s |

would be possible, e.g., employing different uncertainties for all filters. However, even when grouping the bands as described above, our time- and filter-dependent uncertainty analysis has three times as many free uncertainty parameters as the analysis shown in the previous subsection, which uses purely time-dependent uncertainties, which increases the runtime of the analysis. As in the previous case, we find that using the same model for recovery as for the injections results in an accurate description of the light curve and a recovery of the source parameters with small uncertainties. The recovery is similar to the time-dependent results. Similarly, when the recovery is based on a different model than the one used for creating the injection data, we find—as before—that, in particular, the KA2017 model fails in recovering the correct injection mass of the BU2019LM injection.

## IV. ANALYZING AT2017GFO

### A. Time-dependent uncertainties

In the previous section, we tested our uncertainty quantification with injections, which confirmed the model's overall robustness but also showed the importance of using physically more complete models, e.g., nonspherical symmetric ones, to interpret observational data. To follow up on our tests, we will use the time-dependent uncertainty method to analyze real observational data, focusing on the observed kilonova AT2017gfo. Given the better performance of the BU2019LM model during our validation tests, we will focus purely on this model.

We will further use this investigation to test the influence of the number of time nodes on the obtained posteriors. For this reason, we performed parameter estimation using seven different numbers of time nodes $N \in \{4, 6, 8, 10, 12, 14, 16\}$. For comparison, we also study AT2017gfo with a constant 1-magnitude error, and a free but time- and filter-independent systematic error, $\sigma_{\mathrm{sys}} = \mathcal{U}(0, 2)$. The values are reported in the first and second rows of Table II, respectively.

We present the best-fit light curves for the analysis employing different number of time nodes in Fig. 5; cf. Fig. 6 for a corresponding analysis incorporating filter-dependent systematic uncertainties. The posteriors are shown in Fig. 7 and summarized in Table II. We find that the number of time nodes has only a small influence on the recovered source parameters and that the obtained posteriors agree with their uncertainties. In addition, we can see that the systematic uncertainty (bottom panel of Fig. 5) is largest for the early times $< 1$ day and late times $> 5$ days. We suggest that this time dependence is caused by (i) the sparseness of observational data in the early times (indicating the need for quick follow-up observations, e.g., [86]) and (ii) model limitations of BU2019LM. Regarding the latter, higher systematic uncertainties are expected from POSSIS both at early times ($\lesssim$1 day), due to the implemented opacities being computed only for low-ionization states [68,80], and at late times ($\gtrsim$5 days) when the assumed local thermodynamic equilibrium is likely to break down [81,82].

While the posteriors are consistent between different configurations, one needs to gauge which one is sufficiently

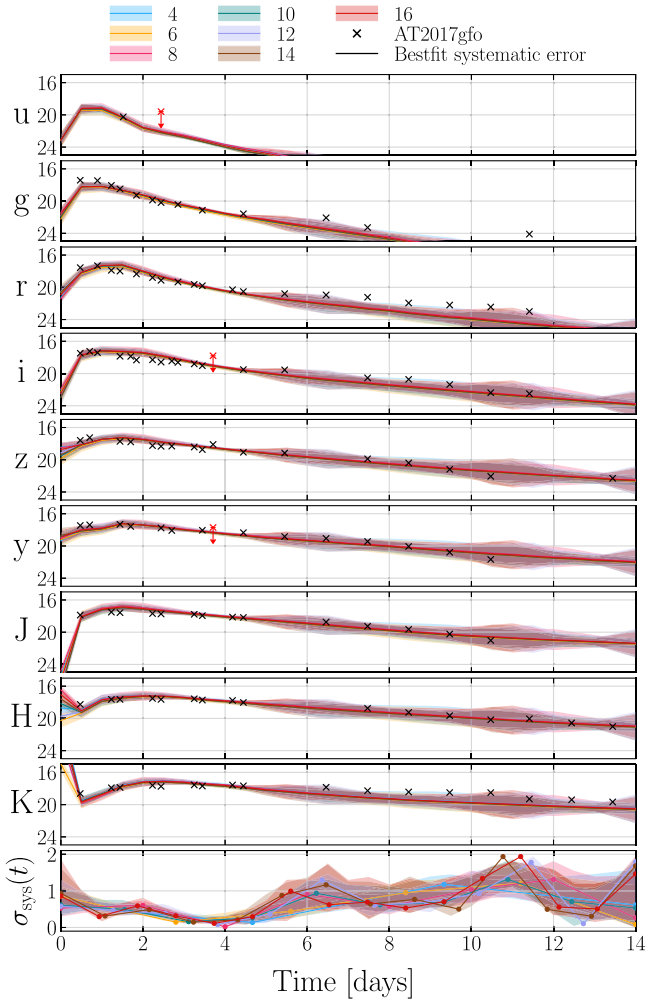


FIG. 5. Recovered light curves for AT2017gfo in different observational bands employing the time-dependent uncertainty method described in the main text. Different colors represent different numbers of employed time nodes. Observational data are marked with black crosses, while red crosses mark the nondetections. The bottom panel illustrates the time-discretized systematic uncertainty, where the circle represents the systematic uncertainty corresponding to the placement of time nodes, and the band represents the 90% highest density interval of the reinterpolated systematic uncertainty, $\sigma_{\rm sys}(t)$.

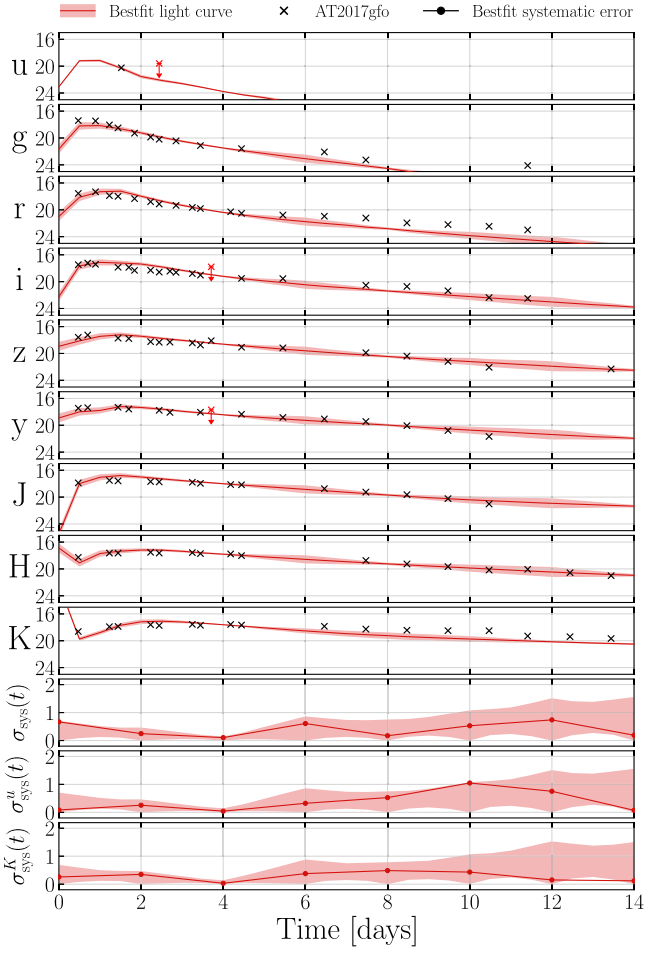


FIG. 6. Same as Fig. 5, for 8 time nodes. $\sigma^u_{\rm sys}(t)$ and $\sigma^K_{\rm sys}(t)$ represents systematic uncertainty for independently sampled $u$ and $K$ band, $\sigma_{\rm sys}(t)$ represents the systematic uncertainty for all other bands sampled together.

flexible for representing the underlying systematics. In the context of Bayesian statistics, the most straightforward choice is using the Bayes factors. However, in our use case, the systematic parameters do not represent any physical information; they only represent artificial degrees of freedom. Therefore, to exclude Occam's razor from our decision-making, we compare the maximum log-likelihood between them; cf. Table II.

As expected, the most flexible configuration with 16 time nodes achieves the highest maximum log-likelihood. However, due to the marginal increase in log-likelihood and the substantial rise in runtime beyond the eight time nodes configuration, we consider the latter as our reference. For runtime comparison, we use 10 cores on an Intel Xeon Platinum 8270 CPU for each run.

### B. Time- and filter-dependent uncertainties

For comparison, we also apply our method, employing time- and filter-dependent uncertainties to analyze AT2017gfo. As for our validation tests, we group the individual bands such that the $u$ and $K$ bands have individual uncertainties while the other bands are grouped together. We employed eight-time nodes for this analysis. The best-fit light curves are shown in Fig. 6, and the recovered source parameters are summarized in the bottom row of Table II. The uncertainty in the $u$ band is on the lower side $< 1$ mag, until 8 days despite having only one detection point, and then increases after 8 days as the model cannot perform well due to lack of data. In the $K$ band, where we have a large number of detections, the uncertainty is well constrained throughout, $< 0.5$ mag. The rest of the bands show a similar trend where the uncertainty is again on the lower side, $<1$ mag. Using filter-dependent

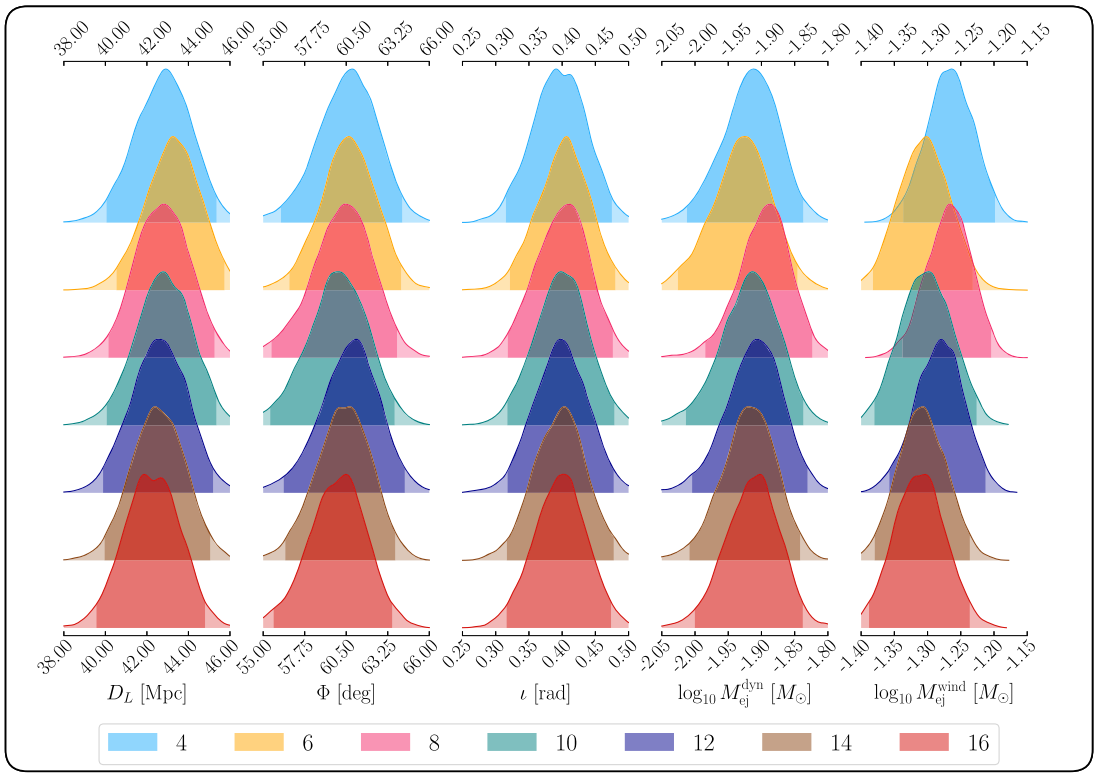


FIG. 7. Posterior distribution of the recovered source parameters when analyzing AT2017gfo with the BU2019LM model. Different colors represent different employed number of time nodes.

systematic uncertainty the maximum log-likelihood decreases marginally; however, the corresponding light curve and associated systematic uncertainty is able to capture the overall trend of the kilonova.

## V. CONCLUSIONS

In this work, we have presented a novel approach to quantify the systematic uncertainties in kilonova modeling using time-dependent and filter-dependent interpolation schemes. Our methodology allows for more robust parameter estimation by capturing the nonstationary behavior of systematic errors intrinsic to the underlying models.

Through a series of injection-recovery tests using synthetic light curves from the KA2017 and BU2019LM models, we have validated the effectiveness of our approach. These tests demonstrate that our interpolation schemes can, in most cases, successfully recover injected parameters within credible intervals.

Applying our methodology to the event AT2017gfo, we performed parameter estimation using different numbers of time nodes. While the best-fit light curves for all runs visually fit the observed data well, a maximum log-likelihood marginal increase criterion favors the model with eight-time nodes as the optimal choice.

Our work highlights the importance of properly accounting for systematic uncertainties in kilonova modeling, as they can substantially influence the inferred parameters and their uncertainties. By introducing time and filter dependence in treating systematic errors, we provide a more nuanced approach that can be adapted to various electromagnetic models and transient phenomena parameter estimation.

## ACKNOWLEDGMENTS

S. J., T. W., P. T. H. P., and T. D. acknowledge support from the Daimler and Benz Foundation for the project "NUMANJI." T. D. acknowledges support from the European Union (ERC, SMArt, 101076369). T. W. and P. T. H. P. are supported by the research programme of the Netherlands Organisation for Scientific Research (NWO). M. B. acknowledges the Department of Physics and Earth Science of the University of Ferrara for the financial support through the FIRD 2024 grant. M. W. C. acknowledges support from the National Science Foundation with Grant No. PHY-2308862 and No. PHY-2117997. The research leading to these results has received funding from the European Union's Horizon 2020 Programme under the AHEAD2020 project (Grant Agreement No. 871158).

Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

## DATA AVAILABILITY

Some of the data that support the findings of this article are openly available [58,68,83]. All data used in this article are either publicly available or generated using the NMMA code, with priors for the Bayesian inference are available within the article.


[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **119**, 161101 (2017).

[2] B. P. Abbott *et al.* (LIGO Scientific, Virgo, Fermi GBM, INTEGRAL, IceCube, AstroSat Cadmium Zinc Telluride Imager Team, IPN, Insight-Hxmt, ANTARES, Swift, AGILE Team, 1M2H Team, Dark Energy Camera GW-EM, DES, DLT40, GRAWITA, Fermi-LAT, ATCA, ASKAP, Las Cumbres Observatory Group, OzGrav, DWF (Deeper Wider Faster Program), AST3, CAASTRO, VINROUGE, MASTER, J-GEM, GROWTH, JAGWAR, CaltechNRAO, TTU-NRAO, NuSTAR, Pan-STARRS, MAXI Team, TZAC Consortium, KU, Nordic Optical Telescope, ePESSTO, GROND, Texas Tech University, SALT Group, TOROS, BOOTES, MWA, CALET, IKI-GW Follow-up, H.E.S.S., LOFAR, LWA, HAWC, Pierre Auger, ALMA, Euro VLBI Team, Pi of Sky, Chandra Team at McGill University, DFN, ATLAS Telescopes, High Time Resolution Universe Survey, RIMAS, RATIR, and SKA South Africa/MeerKAT Collaborations), Astrophys. J. Lett. **848**, L12 (2017).

[3] J. Aasi *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32**, 074001 (2015).

[4] F. Acernese *et al.* (VIRGO Collaboration), Classical Quantum Gravity **32**, 024001 (2015).

[5] B. P. Abbott *et al.* (LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, and MASTER Collaborations), Nature (London) **551**, 85 (2017).

[6] C. Guidorzi *et al.*, Astrophys. J. Lett. **851**, L36 (2017).

[7] K. Hotokezaka, E. Nakar, O. Gottlieb, S. Nissanke, K. Masuda, G. Hallinan, K. P. Mooley, and A. T. Deller, Nat. Astron. **3**, 940 (2019).

[8] E. Nakar and T. Piran, Astrophys. J. **909**, 114 (2021).

[9] M. W. Coughlin, T. Dietrich, J. Heinzel, N. Khetan, S. Antier, M. Bulla, N. Christensen, D. A. Coulter, and R. J. Foley, Phys. Rev. Res. **2**, 022006 (2020).

[10] R. Kashyap, G. Raman, and P. Ajith, Astrophys. J. Lett. **886**, L19 (2019).

[11] M. W. Coughlin, S. Antier, T. Dietrich, R. J. Foley, J. Heinzel, M. Bulla, N. Christensen, D. A. Coulter, L. Issa, and N. Khetan, Nat. Commun. **11**, 4129 (2020).

[12] T. Dietrich, M. W. Coughlin, P. T. H. Pang, M. Bulla, J. Heinzel, L. Issa, I. Tews, and S. Antier, Science **370**, 1450 (2020).

[13] M. Bulla, M. W. Coughlin, S. Dhawan, and T. Dietrich, Universe **8**, 289 (2022).

[14] G. Gianfagna, L. Piro, F. Pannarale, H. Van Eerten, F. Ricci, G. Ryan, and E. Troja, Mon. Not. R. Astron. Soc. **523**, 4771 (2023).

[15] E. Annala, T. Gorda, A. Kurkela, and A. Vuorinen, Phys. Rev. Lett. **120**, 172703 (2018).

[16] A. Bauswein, O. Just, H.-T. Janka, and N. Stergioulas, Astrophys. J. Lett. **850**, L34 (2017).

[17] F. J. Fattoyev, J. Piekarewicz, and C. J. Horowitz, Phys. Rev. Lett. **120**, 172702 (2018).

[18] C. D. Capano, I. Tews, S. M. Brown, B. Margalit, S. De, S. Kumar, D. A. Brown, B. Krishnan, and S. Reddy, Nat. Astron. **4**, 625 (2020).

[19] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **9**, 011001 (2019).

[20] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **121**, 161101 (2018).

[21] S. Huth *et al.*, Nature (London) **606**, 276 (2022).

[22] J. M. Ezquiaga and M. Zumalacárregui, Phys. Rev. Lett. **119**, 251304 (2017).

[23] T. Baker, E. Bellini, P. G. Ferreira, M. Lagos, J. Noller, and I. Sawicki, Phys. Rev. Lett. **119**, 251301 (2017).

[24] P. Creminelli and F. Vernizzi, Phys. Rev. Lett. **119**, 251302 (2017).

[25] P. S. Cowperthwaite *et al.*, Astrophys. J. Lett. **848**, L17 (2017).

[26] S. J. Smartt *et al.*, Nature (London) **551**, 75 (2017).

[27] M. M. Kasliwal *et al.*, Science **358**, 1559 (2017).

[28] D. Kasen, B. Metzger, J. Barnes, E. Quataert, and E. Ramirez-Ruiz, Nature (London) **551**, 80 (2017).

[29] D. Watson *et al.*, Nature (London) **574**, 497 (2019).

[30] S. Rosswog, J. Sollerman, U. Feindt, A. Goobar, O. Korobkin, R. Wollaeger, C. Fremling, and M. M. Kasliwal, Astron. Astrophys. **615**, A132 (2018).

[31] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Astrophys. J. Lett. **892**, L3 (2020).

[32] J. C. Rastinejad *et al.*, Nature (London) **612**, 223 (2022).

[33] G. Ryan, H. van Eerten, L. Piro, and E. Troja, Astrophys. J. **896**, 166 (2020).

[34] N. Kunert *et al.*, Mon. Not. R. Astron. Soc. **527**, 3900 (2024).

[35] A. J. Levan *et al.* (JWST Collaboration), Nature (London) **626**, 737 (2024).

[36] A. Samajdar and T. Dietrich, Phys. Rev. D **98**, 124030 (2018).

[37] A. Samajdar and T. Dietrich, Phys. Rev. D **100**, 024046 (2019).

[38] R. Gamba, M. Breschi, S. Bernuzzi, M. Agathos, and A. Nagar, Phys. Rev. D **103**, 124015 (2021).

[39] N. Kunert, P. T. H. Pang, I. Tews, M. W. Coughlin, and T. Dietrich, Phys. Rev. D **105**, L061301 (2022).

[40] A. B. Yelikar, R. O'Shaughnessy, D. Wysocki, and L. Wade, (2024).

[41] M. Punturo *et al.*, Classical Quantum Gravity **27**, 194002 (2010).
[42] B. Sathyaprakash *et al.*, Classical Quantum Gravity **29**, 124013 (2012); **30**, 079501(E) (2013).
[43] M. Branchesi *et al.*, J. Cosmol. Astropart. Phys. 07 (2023) 068.
[44] 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).
[45] L.-X. Li and B. Paczynski, Astrophys. J. Lett. **507**, L59 (1998).
[46] B. D. Metzger, Living Rev. Relativity **23**, 1 (2020).
[47] D. Grossman, O. Korobkin, S. Rosswog, and T. Piran, Mon. Not. R. Astron. Soc. **439**, 757 (2014).
[48] A. Perego, D. Radice, and S. Bernuzzi, Astrophys. J. Lett. **850**, L37 (2017).
[49] V. A. Villar *et al.*, Astrophys. J. Lett. **851**, L21 (2017).
[50] C. Barbieri, O. S. Salafia, A. Perego, M. Colpi, and G. Ghirlanda, Eur. Phys. J. A **56**, 8 (2020).
[51] K. Hotokezaka and E. Nakar, Astrophys. J. **891**, 152 (2020).
[52] J.-P. Zhu, Y.-P. Yang, L.-D. Liu, Y. Huang, B. Zhang, Z. Li, Y.-W. Yu, and H. Gao, Astrophys. J. **897**, 20 (2020).
[53] M. Nicholl, B. Margalit, P. Schmidt, G. P. Smith, E. J. Ridley, and J. Nuttall, Mon. Not. R. Astron. Soc. **505**, 3016 (2021).
[54] G. Ricigliano, A. Perego, S. Borhanian, E. Loffredo, K. Kawaguchi, S. Bernuzzi, and L. C. Lippold, Mon. Not. R. Astron. Soc. **529**, 647 (2024).
[55] D. Kasen, N. R. Badnell, and J. Barnes, Astrophys. J. **774**, 25 (2013).
[56] M. Tanaka and K. Hotokezaka, Astrophys. J. **775**, 113 (2013).
[57] 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, Mon. Not. R. Astron. Soc. **478**, 3298 (2018).
[58] M. Bulla, Mon. Not. R. Astron. Soc. **489**, 5037 (2019).
[59] J. H. Gillanders, S. J. Smartt, S. A. Sim, A. Bauswein, and S. Goriely, Mon. Not. R. Astron. Soc. **515**, 631 (2022).
[60] C. E. Collins, A. Bauswein, S. A. Sim, V. Vijayan, G. Martínez-Pinedo, O. Just, L. J. Shingles, and M. Kromer, Mon. Not. R. Astron. Soc. **521**, 1858 (2023).
[61] J. Heinzel, M. W. Coughlin, T. Dietrich, M. Bulla, S. Antier, N. Christensen, D. A. Coulter, R. J. Foley, L. Issa, and N. Khetan, Mon. Not. R. Astron. Soc. **502**, 3057 (2021).
[62] O. Korobkin *et al.*, Astrophys. J. **910**, 116 (2021).
[63] D. Tak, Z. L. Uhm, and J. H. Gillanders, Astrophys. J. **958**, 121 (2023).
[64] C. L. Fryer, A. L. Hungerford, R. T. Wollaeger, J. M. Miller, S. De, C. J. Fontes, O. Korobkin, A. Kedia, M. Ristic, and R. O'Shaughnessy, Astrophys. J. **961**, 9 (2024).
[65] D. Tak, Z. L. Uhm, and J. H. Gillanders, Astrophys. J. **967**, 54 (2024).
[66] J. Barnes, Y. L. Zhu, K. A. Lund, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower, and R. Surman, Astrophys. J. **918**, 44 (2021).
[67] Y. Zhu, J. Barnes, K. A. Lund, T. M. Sprouse, N. Vassh, G. C. McLaughlin, M. R. Mumpower, and R. Surman, EPJ Web Conf. **260**, 03004 (2022).
[68] M. Bulla, Mon. Not. R. Astron. Soc. **520**, 2558 (2023).
[69] D. Brethauer, D. Kasen, R. Margutti, and R. Chornock, Astrophys. J. **975**, 213 (2024).
[70] N. Sarin and S. Rosswog, Astrophys. J. Lett. **973**, L24 (2024).
[71] R. T. Wollaeger *et al.*, Astrophys. J. **966**, 177 (2024).
[72] B. Shenhar, O. Guttman, and E. Waxman, Mon. Not. R. Astron. Soc. **531**, 592 (2024).
[73] O. Guttman, B. Shenhar, A. Sarkar, and E. Waxman, Mon. Not. R. Astron. Soc. **533**, 994 (2024).
[74] P. T. H. Pang *et al.*, Nat. Commun. **14**, 8352 (2023).
[75] D. A. Kann *et al.*, Astrophys. J. Lett. **948**, L12 (2023).
[76] T. Hussenot-Desenonges *et al.*, Mon. Not. R. Astron. Soc. **530**, 1 (2024).
[77] N. Sarin *et al.*, Mon. Not. R. Astron. Soc. **531**, 1203 (2024).
[78] M. Breschi, R. Gamba, G. Carullo, D. Godzieba, S. Bernuzzi, A. Perego, and D. Radice, Astron. Astrophys. **689**, A51 (2024).
[79] Y. Peng, M. Ristić, A. Kedia, R. O'Shaughnessy, C. J. Fontes, C. L. Fryer, O. Korobkin, M. R. Mumpower, V. A. Villar, and R. T. Wollaeger, Phys. Rev. Res. **6**, 033078 (2024).
[80] M. Tanaka, D. Kato, G. Gaigalas, and K. Kawaguchi, Mon. Not. R. Astron. Soc. **496**, 1369 (2020).
[81] Q. Pognan, A. Jerkstrand, and J. Grumer, Mon. Not. R. Astron. Soc. **513**, 5174 (2022).
[82] Q. Pognan, A. Jerkstrand, and J. Grumer, Mon. Not. R. Astron. Soc. **510**, 3806 (2022).
[83] D. Kasen, R. C. Thomas, and P. Nugent, Astrophys. J. **651**, 366 (2006).
[84] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, (2015), software available from https://www.tensorflow.org/.
[85] D. P. Kingma and J. Ba, arXiv:1412.6980.
[86] Y. Shvartzvald *et al.*, Astrophys. J. **964**, 74 (2024).