An online framework for fitting fast transient light curves

Tyler Barna [1]⋆ Brandon Reed,[2] Igor Andreoni,[3,4,5] Michael W. Coughlin [1] Tim Dietrich,[6,7]
Steven L. Groom,[8] Theophile Jegou du Laz,[9] Peter T. H. Pang,[10,11] Josiah N. Purdum[12]
and Ben Rusholme [8]

[1]*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA*
[2]*Department of Physics and Astronomy, University of Minnesota – Duluth, Duluth, Minnesota 55812, USA*
[3]*Joint Space-Science Institute, University of Maryland, College Park, MD 20742, USA*
[4]*Astrophysics Science Division, NASA Goddard Space Flight Center, MC 661, Greenbelt, MD 20771, USA*
[5]*Department of Astronomy, University of Maryland, College Park, MD 20742, USA*
[6]*Institute of Physics and Astronomy, Theoretical Astrophysics, University Potsdam, Haus 28, Karl-Liebknecht-Str. 24/25, D-14476, Potsdam, Germany*
[7]*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, Potsdam D-14476, Germany*
[8]*IPAC, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125, USA*
[9]*Division of Physics, Mathematics, and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA*
[10]*Nikhef – National Institute for Subatomic Physics, Science Park, NL-1098 XG Amsterdam, the Netherlands*
[11]*Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University, Princetonplein 1, NL-3584 CC Utrecht, the Netherlands*
[12]*Caltech Optical Observatories, California Institute of Technology, Pasadena, CA 91125, USA*

ABSTRACT

The identification of extragalactic fast optical transients (eFOTs) as potential multimessenger sources is one of the main challenges in time-domain astronomy. However, recent developments have allowed for probes of rapidly evolving transients. With the increasing number of alert streams from optical time-domain surveys, the next paradigm is building technologies to rapidly identify the most interesting transients for follow-up. One effort to make this possible is the fitting of objects to a variety of eFOT light curve models such as kilonovae and γ-ray burst (GRB) afterglows. In this work, we describe a new framework designed to efficiently fit transients to light curve models and flag them for further follow-up. We describe the pipeline's workflow and a handful of performance metrics, including the nominal sampling time for each model. We highlight as examples ZTF20abwysqy, the shortest long gamma-ray burst discovered to date, and ZTF21abotose, a core-collapse supernova initially identified as a potential kilonova candidate.

Key words: methods: data analysis – software: development – (transients:) gamma-ray bursts – (transients:) neutron star mergers – (transients:) black hole - neutron star mergers.

1 INTRODUCTION

The detection of GW170817 (Abbott et al. 2017a) and its associated electromagnetic transients AT2017gfo (Smartt et al. 2017; Coulter et al. 2017; Abbott et al. 2017d) and GRB170817A (Goldstein et al. 2017; Savchenko et al. 2017; Abbott et al. 2017c) has accelerated the field of multimessenger astronomy. In particular, this single event has increased our knowledge of the neutron star equation of state (Bauswein et al. 2017; Margalit & Metzger 2017; Annala et al. 2018; Coughlin et al. 2018; Most et al. 2018; Radice et al. 2018; Lai, Zhou & Xu 2019; Coughlin et al. 2019a, b; Dietrich et al. 2020), the Hubble constant (Abbott et al. 2017b; Hotokezaka et al. 2018; Dietrich et al. 2020; Coughlin et al. 2020a, b), and *r*-process nucleosynthesis (Chornock et al. 2017; Cowperthwaite et al. 2017; Pian et al. 2017; Rosswog et al. 2017; Smartt et al. 2017; Coulter et al. 2017; Kasliwal et al. 2019; Watson et al. 2019). However, with

the third LIGO–Virgo observing run (O3) ending without a viable optical counterpart to a binary neutron star or neutron star–black hole merger candidate (e.g. Andreoni et al. 2019, 2020; Goldstein et al. 2019; Gomez et al. 2019; Lundquist et al. 2019; Coughlin et al. 2019c; Ackley et al. 2020; Anand et al. 2020; Antier et al. 2020; Gompertz et al. 2020; Kasliwal et al. 2020), improving upon the technology we use to efficiently identify transient candidates for follow-up is necessary as we look to the fourth Laser Interferometer Gravitational-Wave Observatory–Virgo (Ligo-Virgo) observing run (O4) and beyond.

During the down-time between the end of O3 and the start of O4, a new framework for fast transient identification known as Zwicky Transient Facility Realtime Search and Triggering (ZTFReST; Andreoni et al. 2021) was developed. This framework has already identified at least seven confirmed afterglows and early identification of a number of these sources has made it possible to carry out multiwavelength follow-up observations. The development of this framework was motivated by searches for serendipitous kilonovae in

⋆ E-mail: tylerpbarna@gmail.com

optical survey data, which rapidly fade in the optical and therefore require online frameworks to have hope for successful identifications.

Following ZTFReST, the Nuclear-physics and Multi-Messenger Astrophysics (NMMA) framework (Pang et al. 2022) was introduced. There are a number of transient classes that possess similar features to kilonova at early times, making the optimization of follow-up observations difficult. To address this, NMMA was developed, which allows for Bayesian inference on electromagnetic (EM) observations to classify objects. In the event that another multimessenger event occurs, NMMA also enables researchers to perform combined analyses of EM and gravitational wave (GW) signals to constrain the Hubble constant and neutron star equation of state (EoS).

In this paper, we describe an improvement of the ZTFReST automated infrastructure through the inclusion of multimodal EM light curve fitting using NMMA rather than the optional kilonova model fitting as described in Section 2 of Andreoni et al. (2021). In addition, we have developed an extension to the pipeline that allows for the light curve analyses to be easily accessed by the broader collaboration in a Slack channel as well as access to these analyses via Fritz. We describe the pipeline in Section 2. Example results and an assessment of NMMA is shown in Section 3. We summarize our conclusions and future outlook in Section 4.

2 PIPELINE FRAMEWORK

This online framework builds upon ZTFReST[1] and NMMA[2] in order to deliver automated fitting of candidates on a regular basis. Additional information regarding the methods underlying ZTFReST and NMMA can be found in Andreoni et al. (2021) and Pang et al. (2022), respectively.

While NMMA and ZTFReST enable powerful analysis, they still require manual input in order to initiate fitting new objects. Moreover, the existing framework does not have native support for conducting multiple analyses of the same object with different models. This online framework seeks to make use of the tools provided by NMMA and ZTFReST to automatically perform fits of new candidates to multiple models in a systematic fashion. This reduces the amount of time that needs to be spent initializing analysis of new candidates and allows for researchers to focus on choosing the most promising candidates for follow-up.

2.1 Candidates

The goal is to identify rare transients like kilonovae and orphan afterglows (which occur when a GRB is not observed prior to the afterglow; this is discussed further in Section 2.3) using multiband, and potentially, multiwavelength or multimessenger data. The former is enabled by the use of alert streams of all-sky optical surveys such as ZTF (Bellm et al. 2019; Graham et al. 2019; Masci et al. 2019; Dekany et al. 2020) and in the future, the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST; Ivezić et al. 2019). In the case of ZTFReST (Andreoni et al. 2021), which relies on the ZTF alert stream (Patterson et al. 2018), the alert stream from both public and private ZTF surveys (Bellm et al. 2019) is used. Fig. 1 shows the distribution of daily candidates between 2021 July and 2022 October from this alert stream.

To identify rare and fast transients, we require a framework by which the nature of the transients could be determined based on
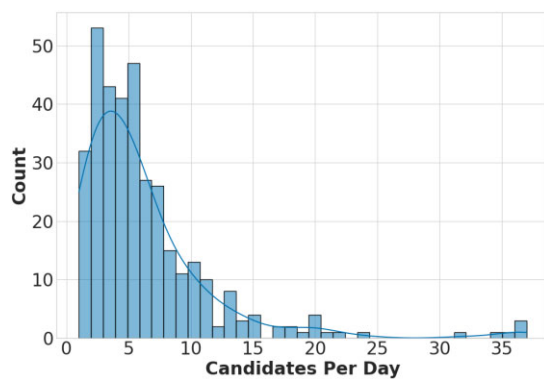

Figure 1. A histogram of the number of candidates each day from 2021 July to 2022 October. There were usually between 2 and 10 candidates analysed each day. The plotted line represents a kernel density estimation of the histogram data. The median number of daily candidates was 5.

time-series of photometric, multiband light curves. In this way, their magnitude and colour evolution, in particular those that are rapidly fading and/or reddening with no history of variability, can identify them as being of significant interest. For the most interesting objects, potentially spectroscopic classification and a well-sampled, multiwavelength light curve to characterize the system is then desired.

To do so, we perform parameter estimation of the light curves using NMMA, which is designed for Bayesian inference on multi-messenger signals (Dietrich et al. 2020; Pang et al. 2021, 2022; Tews et al. 2021; Huth et al. 2022) and has been used to analyse GW170817 (Dietrich et al. 2020); AT2020scz, the shortest long GRB (LGRB) ever confirmed (Ahumada et al. 2021a); GRB 211211A (Kunert et al. 2023); and GRB 221009A (Kann et al. 2023), a hyper-luminous GRB, amongst others.

We use the framework to obtain posteriors along with the log Bayesian evidence for the models presented below. A higher evidence value suggests that the model is a more plausible descriptor of the data. Conversely, a smaller value indicates that a model is less likely to accurately describe the data. Additionally, the log-likelihood ratio of one model against another can be found by calculating the difference between the maximum log-likelihood of the two models. A positive value of this odds ratio would suggest the first model is a better fit, whereas a negative value would imply the latter model fits the data better.

2.2 Data processing

The bulk of our computing is done on the high-performance computing (HPC) clusters maintained by the Minnesota Supercomputing Institute (MSI). Data from the ZTF (Bellm et al. 2019; Graham et al. 2019; Masci et al. 2019; Dekany et al. 2020) alert stream is initially downloaded on to a system at the California Institute of Technology (CalTech) that is referred to as schoty. The pipeline triggers a job every half hour on MSI to download data from the schoty directory that contains ZTF observations. Upon detecting new light curve files, the pipeline will trigger a series of fits to four different models, described in Section 2.3, for each candidate.

After completing fits of all detected light curves, the results will be sent back to schoty. To make the fits easier to review, the pipeline will remotely execute a command to initiate a Slack bot script on schoty. This bot posts plots of each of the fits for the daily candidates as well as their posteriors and Bayes evidence values on to a channel

---

[1]The ZTFReST repository can be found here.
[2]The NMMA repository can be found here.

on the GROWTH MMA Slack server. Since 2021 July, this bot has successfully posted fits on a regular basis with only brief interruptions due to MSI or ZTF maintenance.

Generally, this entire process takes on the order of 6 h at the current level of detail. The rationale behind tuning the number of live points to reduce the total execution time is to allow for automating follow-up decisions in the future; this would require the fits to be completed some time before observing targets for the next night are set. There are several projects currently in-progress by members of the NMMA collaboration investigating methods by which the time required for fitting light curves can be reduced. One collaborator is investigating the use of machine-learning (ML) on a collection of simulated light curves with known parameters to train an algorithm that would associate features of a light curve with certain parameter values. More generally, ML techniques represent an area of development that could enhance the speed and utility of the NMMA framework.

2.3 Light curve models

There are a number of astrophysical processes that produce fast transients. In this work, we will focus on a handful of extragalactic astrophysical processes. For each model, both the explosion time and the distance is allowed to vary. We include Milky Way like host extinction (Fitzpatrick 1999) with $R_V = 3.1$ and $E(B - V) = 0.1$ mag.

The first are kilonovae, the optical counterparts to binary neutron star or neutron star–black hole system mergers generated by the *r*-process material that is produced in these events (Metzger 2017). In our analysis, we use a POSSIS-based (Bulla 2019) grid of kilonova models spanning the plausible binary neutron star parameter space (Dietrich et al. 2020). There are four parameters, the dynamical ejecta mass M_{ej}^{dyn}, the disc wind ejecta mass M_{ej}^{wind}, the half-opening angle Φ, and the observation angle Θ_{obs}. Internal code for the pipeline refers to this model as BU2019M.

The second process are GRB afterglows. We use AFTERGLOWPY (Ryan et al. 2020), an open-source computational tool modelling forward shock synchrotron emission from relativistic blast waves as a function of jet structure and viewing angle. The model parameters are the isotropic kinetic energy, $E_{K,iso}$; the jet collimation angle, θ_c; the viewing angle, θ_v; the circumburst constant density, n; the spectral slope of the electron distribution, p; the fraction of energy imparted to both the electrons, ϵ_e, and to the magnetic field, ϵ_B, by the shock. This model is also called TRPI2018 in the code.

The third process are supernovae. We use a model for SN Ib/csupernovae (Levan et al. 2005) from the SNCOSMO package (Barbary et al. 2016) with the stretch and scale set to match the intrinsic (dereddened, rest frame) *R*-band luminosity of SN 1998bw at maximum light. The absolute magnitude is the free parameter with the largest effect on the quality of the fit. This is referred to as the `nugent-hyper` model within the pipeline.

The fourth process are *shock cooling* supernovae using a model that follows Piro, Haynie & Yao (2021). After shock breakout, the radiation of shock heated material expands and cools, known as shock cooling emission. The model considers extended material with mass M_e and radius R_e, which is imparted with an energy E_e as the shock passes through it. Internally, this is called the `Piro2021` model. Generally, this paper refers to the `nugent-hyper` model as SNe Ib/c or simply supernovae, while the `Piro2021` model is referred to as shock cooling or shock cooling SNe.

In addition to the three non-kilonova models presented here, there are several other source categories that can imitate kilonova light curves, such as M dwarf flares and cataclysmic variables; modelling

these sources with NMMA and integrating them into this pipeline is an area for future work.

3 RESULTS

We evaluate the reliability of NMMA in recovering the different types of transients discussed in Section 2.3 through the use of simulated light curves. We also highlight the performance of the pipeline for two notable examples from the literature. Following this, the general daily performance of the automated pipeline infrastructure is summarized.

3.1 Model recovery

An important consideration for the utility of the framework is the predictive power of the fitting – that is to say, how often it can accurately identify a candidate of a given type as such. Ideally, a kilonova would always be best fit by a kilonova model as compared to other analysed models, with the same being true of other transient event types.

To evaluate this, we used an existing NMMA tool that generates an injection file of parameter values that fall within the defined prior space of the provided model. We then use this injection file to NMMA to create a simulated light curve. By using a kilonova model to create an injection, we can simulate a kilonova light curve that could be plausibly observed. From there, we can apply the same multimodel analysis described above to evaluate how well NMMA is able to recover the correct model.

We created a sample of 100 simulated light curves for each model using the same priors used for this pipeline. To minimize the dependence on observation cadences, we imposed a half day cadence in observations for all light curves and considered data from a single ZTF filter, the *g* band. A realistic detection limit of 21.5 mag was also imposed; all detections above this limit up to 21 d from initial detection was considered in the analysis. We then fit each light curve in the sample to each of the analysis models and evaluated the relative quality of the fits. This is accomplished by calculating the odds ratios of the true model for each light curve and the three other models. If all three of these odds ratios are positive, that indicates that the true model is the best fit for the light curve.

After noticing a high rate of failure for fitting shock cooling light curves, further investigation revealed that a majority of the shock cooling light curves only had a handful of detected data points, resulting in the analysis failing due to insufficient data. This prompted the addition of a method to enforce minimum observing conditions for injected light curves; a flag was added such that, if an injection file produced a light curve that did not have at least three detections within the first 3 d of observation, the injection was regenerated until one was produced that met the defined observing condition. In the case of the shock cooling model, this resulted in a noticeable difference in the distribution of the physical parameters associated with the model, implying that there are certain areas of the parameter space that correspond to shock cooling that would not be observable with a limiting magnitude of 21.5.

We found that kilonova light curves were accurately identified 100 per cent of the time; this high rate of model recovery is likely due to the even and frequent cadence of the observations in the data. The characteristic rise of a kilonova is not necessarily likely to be observed in realistic observing conditions due to its short duration, but its presence in this sample is another potential factor causing the identification of kilonova to be so reliable. SNe Ib/c were also identified correctly 100 per cent of the time, which is most probably

Table 1. Results of model recovery and analysis with NMMA for the sample of simulated light curves described in Section 3.1. Each row is the distribution of what per cent of each true model type was identified as the model listed in the column, such that each row will sum to 100 per cent. The diagonal cells show what per cent of each model was accurately identified as the correct model based on the Bayes evidence of the NMMA fits. Off-diagonal cells for each row will show what percentage of each model was incorrectly identified as a different transient type.

True model	Kilonova (per cent)	GRB afterglow (per cent)	Supernova (per cent)	Shock cooling (per cent)
Kilonova	100	0	0	0
GRB afterglow	5	90	3	2
Supernova	0	0	100	0
Shock cooling	0	28	36	36

a confluence of the simplicity of the `nugent-hyper` model and characteristic light curve shape.

GRB afterglows are the first model with a sub-100 per cent recovery rate, being correctly identified 90 per cent of the time. The 10 per cent of incorrect identifications are spread across the three other models, with 5 per cent being identified incorrectly as kilonovae and 3 per cent and 2 per cent being identified as SNe Ib/c and shock cooling SNe, respectively.

The shock cooling model had the lowest rate of recovery, with only 36 per cent of light curves being correctly identified as such. An equal number of shock cooling SNe were misidentified as SNe Ib/c, and the remaining 28 per cent were incorrectly identified as GRB afterglows. The summarized results of this model recovery can be found in Table 1.

3.2 Time dependence of model recovery

A complimentary topic of interest is how the quality of fitting might change over time, especially at early times where there is a paucity of data. To investigate this further, a similar process was used as described in Section 3.1, but with the added complexity of performing a series of successive fits on each light curve with increasing amounts of data in consideration. In other words, the same series of fits would be conducted with the first 2 d of data from each light curve, then for the first 3 d of data, and so on going out to 10 d from the initial observation of each light curve. A minimum of 2 d was selected to allow for a sufficient number of detections such that NMMA could complete its analyses. This generates a large volume of analyses, over 10 000 for a sample of 100 light curves of each of the models. This, in turn, produces a large volume of data, both in total size and number of files, so the results are consolidated into a *.csv* file with the best-fit light curve and its accompanying parameters for each analysis.

Upon initial evaluation of the rate of correct identifications over time, it became apparent that the kilonova model was being identified correctly essentially 100 per cent of the time, even when only provided the minimum number of data points. This could be due to a number of factors: as mentioned above, by observing kilonova early on, the light curve captures their characteristic rise, which may result in consistent identification early on. Most data from ZTF will be far less consistent in their observation cadence and it's unlikely that they will be possible to capture the early rise of kilonova with ZTF, so this set of light curves is not necessarily informative of how well NMMA will be able to perform under realistic scenarios. Another possibility is simply Occam's Razor – the kilonova model is less complex in comparison to the GRB model, so it is possible the kilonova is preferentially chosen by the Bayesian inference performed by NMMA; we are actively investigating the potential source of this.

Regardless, in order to evaluate the performance of NMMA in a more realistic manner, we make use of a feature already present in NMMA, which samples light curves on ZTF-like cadences. It also simulates variance in the ZTF limiting magnitude each night. To account for these changes, the observing condition for generation was loosened to only require a minimum of three observations within 5 d of the initial observation; this represents a source that is well-observed by ZTF but still plausible. Even with this relaxed requirement, some models saw a significant increase in the number of attempts required in order to generate a valid light curve. This was most apparent with the shock cooling model, which could require upwards of 200 attempts before generating a light curve that met the conditions. This reinforces the notion that areas of the parameter space, while physically possible, may result in objects that are unlikely to be observable by ZTF.

From this new sample of light curves, the Bayes evidence was aggregated at each time step for each combination of injected light curve and fit model. The median value of the Bayes evidence for each model at each time step was taken, and then the odds ratio was calculated between the true model and the other models in consideration. This can be seen in Fig. 2. The primary takeaway from this metric is the increase in the median odds ratio over time for all models; with additional data, objects are better fit by their true model than by other models. This metric also suggested that all four models were fairly consistent in being accurately identified from early times, with the exception of the shock cooling model. However, evaluating the results by finding the model with the highest odds ratio for each light curve at each time step demonstrated a different result.

The kilonovae were still identified fairly reliably with the minimum amount of data (2 d), suggesting that the characteristic rise may play a role in the rate of identification. There is a small, temporary dip in the rate of correct identifications when considering a maximum time of 3 d. This could be a result of small-number statistics, as this analysis only contains 100 light curves for each model, but it could also indicate there's an intermediate amount of data where kilonovae can appear to resemble supernovae analytically.

Supernovae demonstrated a high rate of misidentification with the minimal amount of data but rapidly increase in identification with an additional day of data, reaching 100 per cent identification by the time there was 5 d of data in consideration. It took somewhat longer for GRB afterglows to reach a majority being correctly identified, though it still demonstrates a positive trend. Regardless of the amount of data, all shock cooling SNe were misidentified, with a majority being classified as the standard SN model.

The low rate of recovery for the shock cooling model suggests that, while they were incorrectly identified, the individual odds ratios between the best fit and the other fits are fairly small. The median odds ratio values in the shock cooling sub-plot in Fig. 2 supports this, with none of the odds ratios exceeding 2.5 prior to the 10th day of observations, though they do trend positive. This suggests that there is some lower cut-off for the odds ratio below which one cannot confidently assert the likelihood of one model over another.
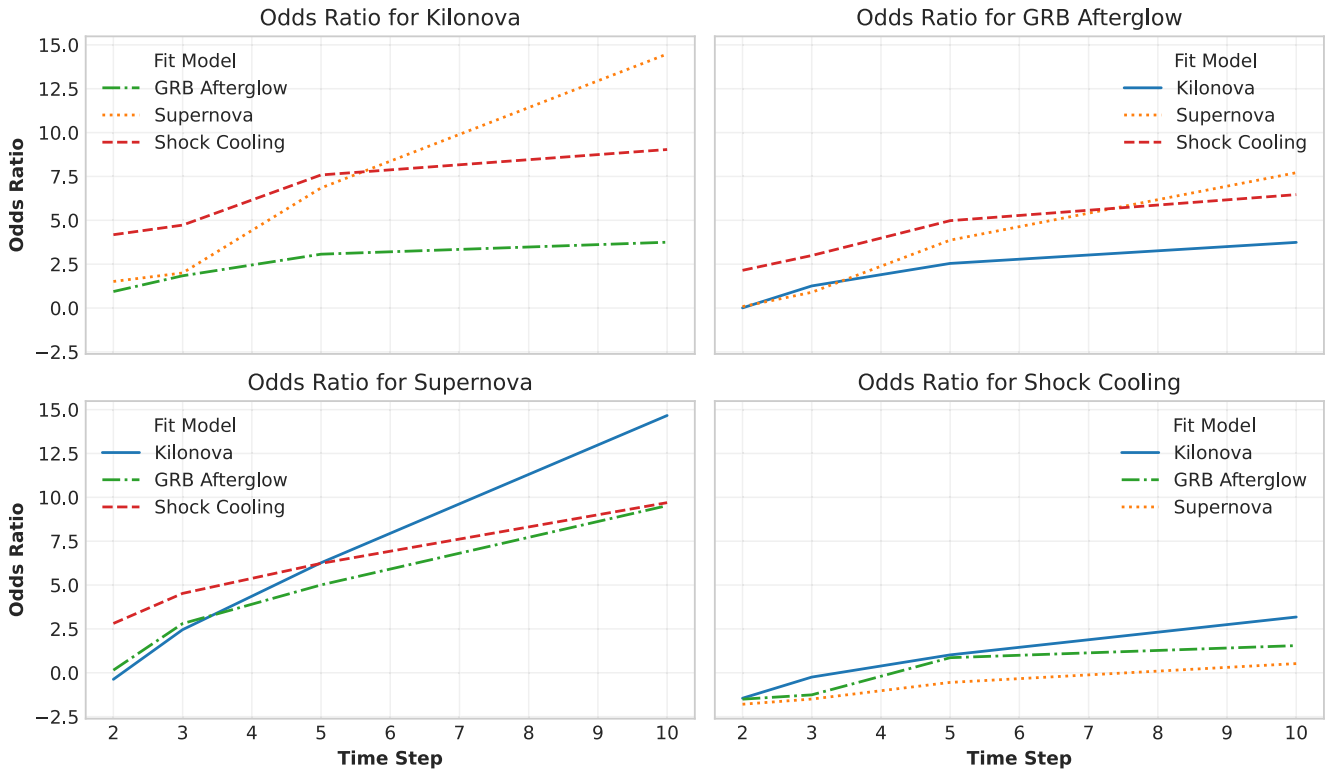
Figure 2. The median odds ratios of simulated light curves over time. Each sub-plot corresponds to the sample of light curves generated with the model listed in the sub-plot title (i.e. the true model). The different lines in each sub-plot represent the median odds ratio over time between the Bayes evidence of fits to the true model and fits to the other models. Generally, all models show a trend of odds ratios increasing over time, suggesting that NMMA analysis more strongly prefers the correct model as more data are considered.

While the rate of identifying kilonovae was fairly high, even with the minimal amount of data, it's worth noting the number of other candidates that were incorrectly identified as kilonovae. With the minimum amount of data, there were a larger proportion of light curves that were incorrectly identified as kilonovae compared to the number of correctly identified kilonovae. It is worth noting that this is a sample of an equal number of each model; the actual number of kilonovae in comparison to other transient types would be significantly lower with a more realistic sample of transients.

Inspecting the odds ratios between the best and next best fits for both the true kilonovae and the imposter kilonovae, many of the imposter kilonovae had odds ratios that trended worse than the odds ratios of the true kilonovae, which represents a possible avenue for evaluating the merits of sources in the event multiple objects in a given night are best fit by a kilonova model. Confusion matrices representing the transient identifications over time are shown in Fig. 3.

3.3 Multi-armed bandit

Given the time-sensitive nature of follow-up observation, we explored using concepts from statistics to optimize decision making. We have implemented a multi-armed bandit (MAB) algorithm, which is a strategy that uses a defined reward to make the optimal selection from a number of different possible choices. Slivkins (2024) provides an extensive introduction to various MAB algorithms.

In our case, the MAB selects the most likely kilonova from an ensemble of candidates. For a given observing scenario, we include nine candidates: one kilonova, one GRB afterglow, and one supernova, generated using the same models as detailed in previous sections.

Light curves for each observing scenario are sampled on a ZTF-realistic cadence using the same functionality described above, though we impose a broader requirement of eight g-band detections within 18 d of the first detection to allow for the MAB to be run over a longer period of time. Again, we use NMMA to fit each light curve to each model; the results of these analyses are then used to evaluate a reward function.

The reward is defined as the odds ratio of the kilonova model and the best non-kilonova model as well as an upper confidence bound (UCB) term that allows for exploration of the parameter space. Essentially, this term prevents the algorithm from exclusively selecting a specific object unless the odds ratio continues to improve with additional observations or is markedly higher than odds ratios for other objects at early times. The reward equation is given by

$$R_n(t) = \left(L(t)_{n,KN} - L(t)_{n,non-KN} \right) + \sqrt{\frac{2\ln(t)}{N_n(t)}}, \qquad (1)$$

where t is the time-step being evaluated and the subscript n differentiates between candidates. The left most term is the odds ratio of the kilonova model fit and the next best model fit. The right most term is the UCB, where the numerator is two times the natural log of the current time-step and the denominator is the number of times that object n has been selected by the MAB.

Generally, the reward will be positive if the kilonova model is the best fit and it will be negative if it is not. If multiple candidates
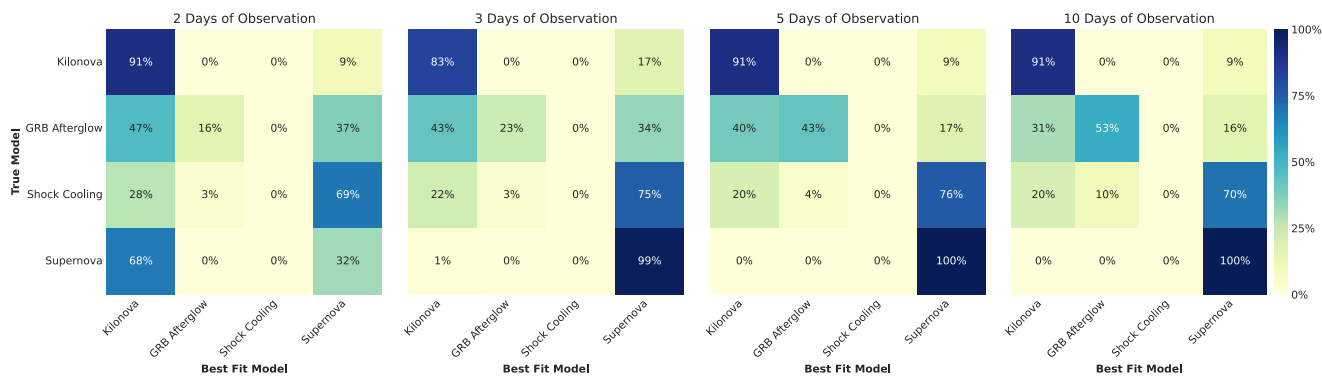
Figure 3. A collection of confusion matrices denoting the distribution of the best fit models versus the true models of simulated light curves described in Section 3.2. In each confusion matrix, the diagonal elements represent correct identifications and the off-diagonal elements are incorrect identifications; each row should sum to 100 per cent, representing the collection of simulated light curves for that model, and each column denotes the relative per cent of light curves identified as being best fit by a given model. Each matrix corresponds to fits done with data for each light curve from their initial detection up to the number of days listed in their respective titles.

have a positive reward, the candidate with the highest reward will be selected. In this case, 'selection' refers to being chosen for follow-up observation. The candidate with the highest reward is chosen for observation during the next observing period. Ideally, the MAB will quickly identify the correct candidate and continue to observe the kilonova, while the UCB term will allow for the consideration of other candidates if the odds ratios between the best candidates are similar.

Similar to Section 3.2, each light curve is simulated in its entirety before being concatenated into an initial set of observations, namely when there are a minimum of three observations for each candidate, representing time-step 0. After analysing these initial light curves, the reward is evaluated for each candidate. For the selected candidate, any observations made in the interval spanning the next day from the full light curve are added to the concatenated light curve and the analysis is conducted with the additional data.

To evaluate the performance of the MAB algorithm, we ran the algorithm on 100 observing scenarios, using a time-step of 2 d for calculation of the reward function and object selection and a total of seven time-steps. We found that the MAB was fairly effective in selecting kilonova candidates early and often. The kilonova was correctly selected in all 100 observing scenarios at the first time step; at the final time step, the kilonova was correctly selected in 93 of the 100 observing scenarios. Across all time-steps, kilonovae were selected for observation 92.6 per cent of the time. In 80 of the 100 observing scenarios, the kilonova was selected at every time-step.

At all time-steps where the kilonova was not selected, the GRB afterglow was selected instead; there were no cases of supernovae being selected at any time-step. The rewards for the objects observed at each time-step are shown in Fig. 4. This figure contextualizes the lack of additional supernova observations given the significantly lower average reward when compared to the other candidate types.

While the average GRB afterglow reward exceeds the average kilonova reward at time-step 2 and beyond, the rate of GRB afterglow selection peaked at 12 per cent (also at time-step 2), with an average selection rate of 7.4 per cent across all time-steps. Because of the relatively low selection rate for GRB afterglows, those that are selected would tend to have higher reward values than the average kilonova reward.

Even in cases where a GRB afterglow was chosen, the kilonova was still selected for a majority of the time-steps. Of the 20 observing
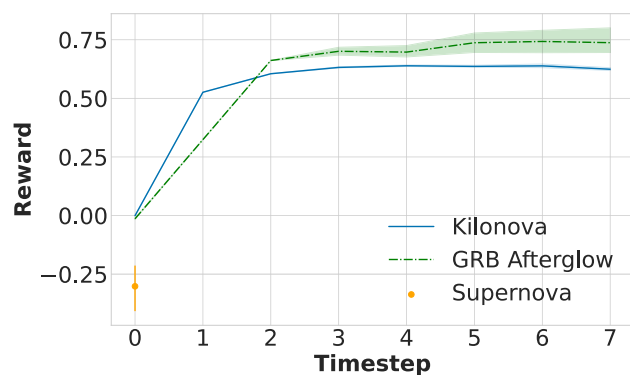

Figure 4. The MAB reward, as calculated by equation (1), for the ensemble of observing scenarios, where the lines show the average value of the rewards for objects observed at that time-step and the shaded areas represent the 95 per cent confidence interval. No supernovae were selected for further observation by the algorithm, so the rewards for supernovae are plotted as a single point at time-step 0.

scenarios where the kilonova was not chosen every time-step, the GRB afterglow was selected more than the kilonova only 7 times, and there were no scenarios where the GRB afterglow was selected at every time-step. For GRB afterglows that were selected at least once, the average number of selections (out of a possible 7) was 2.6, as compared to an average of 6.5 selections for all kilonova selected at least once. This suggests that, even for scenarios where differentiating between a kilonova and an imposter via photometric analysis might be difficult, the MAB algorithm is unlikely to select the incorrect candidate for follow-up observation more than the true kilonova.

This initial study of the utility of MAB algorithms for optimizing follow-up observation is promising, and we plan to integrate MAB functionality into the NMMA framework to allow for future analysis. Particularly, we plan to evaluate its performance across larger samples of candidates, with additional classes of eFOTs, and using data from multiple filters.
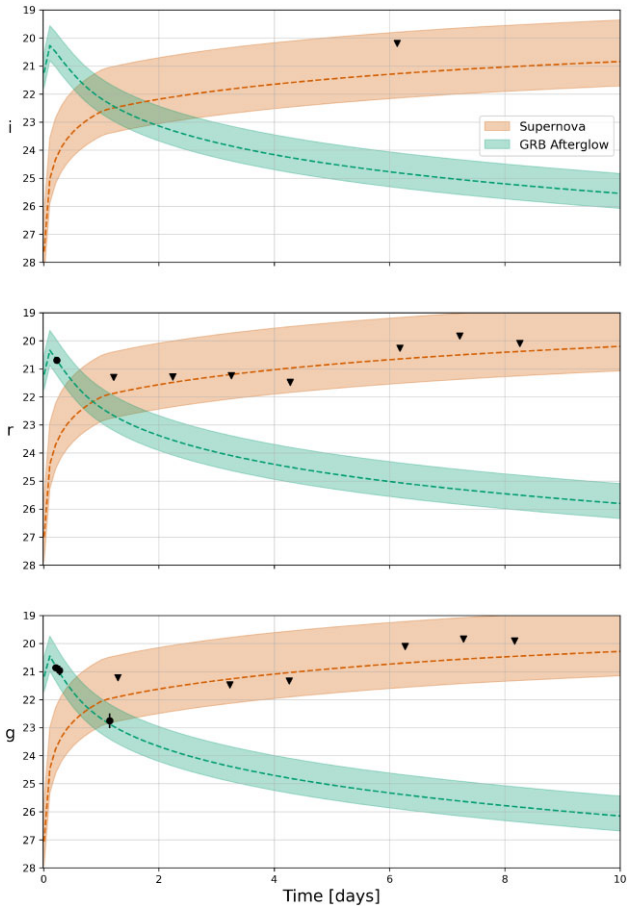
Figure 5. The light curve of the ZTF20abwysqy and analyses of the object fit to the supernova and GRB afterglow models; non-detections are denoted as triangles, while detections are represented by circles. The dashed lines trace the best fits for each model, and shaded regions correspond to one standard deviation in estimated luminosity distance among the samples calculated for a given model fit. This object was initially suspected to be a supernova, but additional observations suggested it was the optical companion of a GRB. Its classification as a such was later confirmed in a follow-up study.

3.4 ZTF20abwysqy

ZTF20abwysqy, also known as AT2020scz, is an afterglow that was identified as the optical companion of the GRB 200826A, which is noted in Ahumada et al. (2021b) as the shortest LGRB discovered, suggesting it may straddle the minimum conditions for which a successful collapsar can occur. While this was originally detected by *Fermi*, follow-up observations with ZTF allowed for the identification of the optical counterpart (Andreoni et al. 2021).

As detailed in Andreoni et al. (2021), ZTFReST can create light curves both from ZTF alerts and forced photometry. These light curves can then be fit by the pipeline. While X-ray observations with *Swift* would have ultimately identified ZTF20abwysqy as an afterglow, the use of forced photometry with ZTFReST allowed it to be recognized as a rapidly fading source more promptly and motivated follow-up observation.

Analysis of ZTF20abwysqy was conducted using this pipeline for both the supernova and GRB afterglow models; the results of these analyses is shown in Fig. 5. Visual inspection of the fits to ZTF20abwysqy suggests that the non-detections cannot entirely rule out the possibility of it being a supernova, there is a notable difference in the Bayes evidence values between the two. The odds ratio of the

Bayes evidence values between the GRB afterglow model and the supernova model is 19.77, which seems to favour the GRB afterglow model in comparison to the supernova model. In comparison, the simulated light curves for GRB afterglows, shown in the upper-right sub-plot in Fig. 2, had a median odds ratio of around 7.5 with roughly the same amount of observations available for analysis (though the fits of ZTF20abwysqy benefit from limiting magnitude measurements in the *i* and *r* bands as well as an additional data point in the *r* band). This suggests that this pipeline and NMMA would have supported the idea of ZTF20abwysqy being a GRB afterglow rather than a supernova at the time these observations were occurring, which could have motivated additional follow-up observation.

3.5 ZTF21abotose

The primary goal of ZTFReST is to find kilonovae and GRBs, but there are cases where other transients might appear to match the form of a kilonova or GRB. Objects like ZTF21abotose exemplify both an 'impostor' among candidates and a potential avenue for expansion of the pipeline.

From initial ZTF observations, ZTF21abotose, also referred to as SN2021ugl, originally appeared to be a kilonova, but spectroscopic analysis of the object revealed it was actually a Type IIb supernova (Ridley et al. 2021), which occurs as a result of core collapse of a massive star. More specifically, it was the signature of cooling following a shock breakout. When the core of a star collapses in on itself, a shock propagates outward from the inner core. Once this shock makes it to the surface of the star, it creates a spike in luminosity separate from the ongoing supernova (Burrows, Hayes & Fryxell 1995; Campana et al. 2006; Janka et al. 2007).

In the past, it has been challenging to observe shock breakout events, as they occur at the earliest stages of a core collapse supernova. Observations of shock breakouts have largely been a result of serendipitous observation at the time of explosion (Soderberg et al. 2008). Now, with automated all-sky surveys like those conducted with ZTF, the observation of shock breakouts have become much more prevalent. In fact, they represent one of the primary classes of transients that are detected by ZTFReST outside of the intended targets.

Analysis of ZTF21abotose with this pipeline was conducted for both the shock cooling and kilonova models, the results of which are shown in Fig. 6. Generally, there is significant overlap between the best fits for both the kilonova and shock breakout models; only the final observed data point in the *g* band falls outside the projection of the kilonova model fit, though this does not necessarily rule it out as a potential kilonova. Inspection of the associated odds ratio between the Bayes evidence values of the shock breakout and kilonova models is 1.55, which is notably lower than that of the one found for ZTF20abwysqy above. While the odds ratio for ZTF21abotose slightly favours the shock cooling model over the kilonova model, this analysis does not strongly support one model over the other.

Though detections of Type IIb SNe and other transients that resemble kilonovae and GRBs are troublesome for our current objectives, their prevalence in the ZTF data suggests a potential for ZTFReST to be employed in aiding in the search of other transient objects.

3.6 General performance

Between 2021 July and 2022 October, the pipeline executed on 354 d with a total of 2149 daily candidates, of which 517 (roughly one quarter) were unique candidates. Candidates were observed for
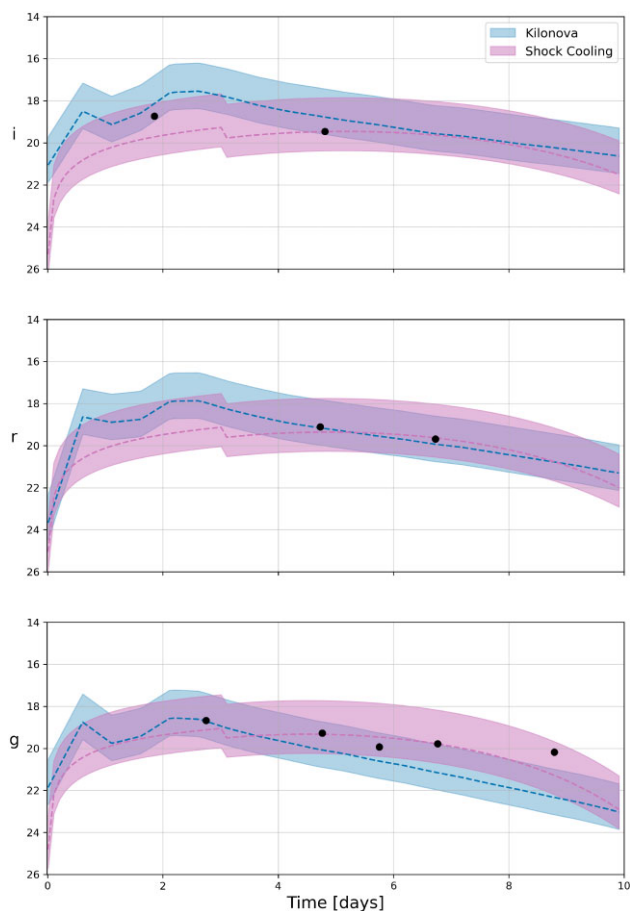
Figure 6. The light curve of ZTF21abotose and fits to the kilonova and shock cooling supernova models. As in Fig. 5, the lines for each model show the best fit, and the shaded regions represent one standard deviation of the estimated luminosity distance. This object was difficult to classify based on photometric analysis alone, but spectroscopic analysis ultimately confirmed its classification as a Type IIb supernova.

an average of just over four nights, with the maximum number of observations being 16. The average (median) number of candidates analysed each day was 6.1 (5.0), and the average (median) number of unique candidates was 2.8 (2.0). A histogram of the number of daily candidates during the analysis period is shown in Fig. 1.

Of the four models implemented on the MSI system, the shock cooling model remained non-functional for the duration of the analysis period. The precise reason for this is unclear since it is possible to run fits with the shock cooling model locally and investigation is ongoing. Overall, 71 per cent of candidates had a successful fit to at least one model. That remaining 29 per cent could have been not fit for several reasons. Some candidates do not meet the filtering requirements we impose, and even if they pass the filtering requirements, it is possible that NMMA may not be able to converge on a solution.

During the course of pipeline development, a manual time limit of 8 h was imposed on all model fits so as to allow fits to be reviewed before the next night of observing. Fig. 7 shows the distribution of sampling times on a per-model basis, where sampling time refers to the amount of time in seconds required to complete an analysis using NMMA; there is a noticeable difference in the time required to fit the supernova model as compared to the other models. The time required to complete analysis of a given candidate is a meaningful metric by

which to evaluate the feasibility of using this framework to inform follow-up observation. Excessively long sampling times would result in analyses not being completed prior to the next night of observing, reducing the utility of the framework. However, the results of the analysis period suggests the framework is able to complete most individual analyses within an hour.

In total, the pipeline used roughly 4760 computing hours during the analysis period, which corresponds to an average (parallel) computing time of just over 13 h d^{-1}. The supernova, GRB afterglow, and kilonova models made up 4 per cent, 37 per cent, and 59 per cent of the computing time, respectively. Fig. 7 shows that the median computing time for the supernova model fits is on the order of minutes, whereas the other two models generally take almost two orders of magnitude longer to complete. The supernova model is a less complex model with fewer parameters, so NMMA requires much less time for likelihood evaluations and converges faster.

As shown in Fig. 8, the cumulative computing time tracked fairly closely with the cumulative number of candidates analysed with the exception of the supernova model. Beyond the first month, the supernova model computing time grew very little, making up an increasingly small proportion of the cumulative computing time. This figure also demonstrates the framework's relatively low usage of computing, requiring an average of just under 13.5 h of computing time per day, which is generally run in parallel so as to complete daily analyses within a few hours.

Each of the models had similar maximum and minimum Bayes evidence values. However, the distribution of these Bayes evidence values differ. Fig. 9 demonstrates the relationship between sampling times and the resulting Bayes evidence. There was not a particularly strong correlation between the Bayes evidence of supernova model fits and their sampling times, but significantly more of fits of this model had very negative Bayes evidence values. This makes sense; while the lower dimensionality of the supernova model resulted in shorter sampling times, the other two models are more likely to produce better Bayes evidence values because there are more parameters by which to fit the data; additionally, the ZTFReST algorithm is designed for fast transients, which supernova are not.

The kilonova and GRB afterglow models had a similar distribution of sampling times and Bayes evidence values; both indicate a general trend of longer sampling times resulting in more negative Bayes evidence values, with the distribution of sampling times for the kilonova model extending slightly further than that of the GRB afterglow model.

4 DISCUSSION

In this paper, we have described a useful extension of the ZTFReST pipeline that can be generalized for use with other data sets and potential future functionality. We have evaluated the general performance of this pipeline as well as the performance of NMMA. As O4 continues and we look to future observing runs, this will provide an opportunity to identify additional needs for enabling the automatic fitting of fast transient light curves to inform follow-up observations of intriguing sources.

The pipeline has demonstrated the ability to perform fits on a daily basis without additional oversight. That being said, there are a number of open questions and potential functionalities that can be developed for the pipeline.

One of the principal questions that must be answered as development continues is how comparison between different fits can be quantified in a meaningful way. In order to properly inform automated follow-up observations, we must investigate what constitutes
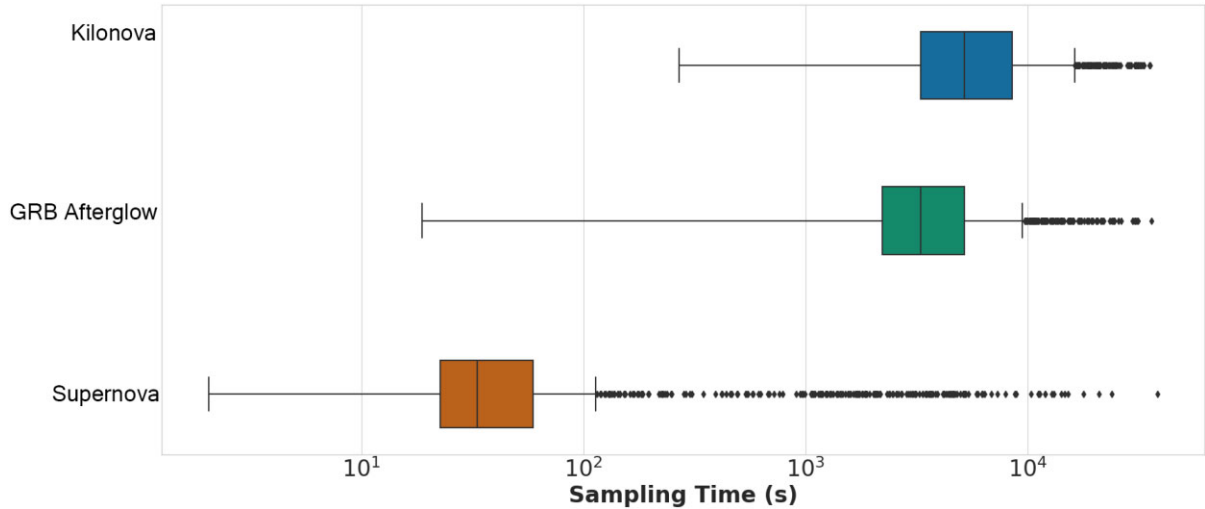
Figure 7. Distribution of fit sampling times by model. The *x*-axis corresponds to the time in seconds required to complete an analysis of a given model. Each box represents the inner quartile range of the data, with the line inside this box being the calculated median, and outliers are plotted as individual points. Note the log scaling of the *x*-axis; the median sampling time for the supernova model is roughly two orders of magnitude less than the median values of the GRB and kilonova models, due in part to the relative simplicity of the supernova model.
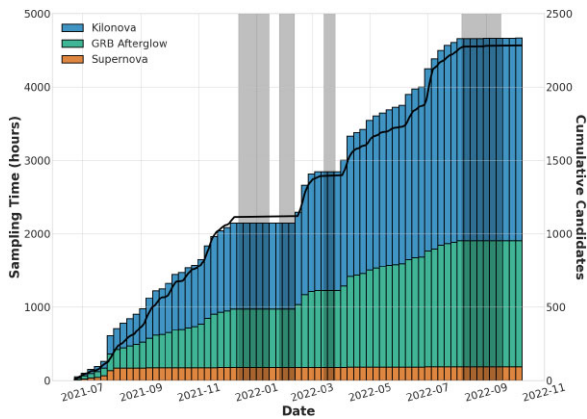
Figure 8. The bar plot (left axis) represents the cumulative sampling time, in hours, on MSI between 2021 July and 2022 October. The total sampling time was roughly 4670 h. The line (right axis) represents the cumulative number of candidates, including candidates that were analysed on more than one day. The shaded regions on the plot represent periods where the pipeline was not consistently receiving new data. For the left most three instances, this was due to ZTF being down for maintenance or service. The right most instance was a result of issues with the connection between MSI and schoty.

Figure 9. A contour density plot of the sampling times compared to the Bayes evidence by model. Generally, the supernova model completed fitting within minutes, whereas the median sampling times for the kilonova and GRB afterglow models were on the order of hours. However, the supernova model demonstrated a much larger range in values for the Bayes evidence compared to the other models. Note that, while the contours may visually extend past the x-intercept, no fits took negative time to complete. This is also true for the y-intercept – the upper limit of the Bayes evidence is 0.

a significant enough difference in quality of fits between models to motivate additional observation of a target. By expanding on the model recovery analyses discussed in Sections 3.1 and 3.2 to generate a more expansive set of light curves, it should be possible to quantify this difference.

The use of NMMA in this pipeline demonstrates its ability to perform comparative analysis of ZTF light curves to tentatively identify their transient type to motivate follow-up observation. Features of this pipeline would be beneficial to integrate into NMMA to increase its utility as a multipurpose analytical tool for MMA science. NMMA currently supports combined analysis of multiple transient models on a single light curve, but work is being done to enable users to provide multiple models as part of a single command that will then perform independent analysis of the light curve to the
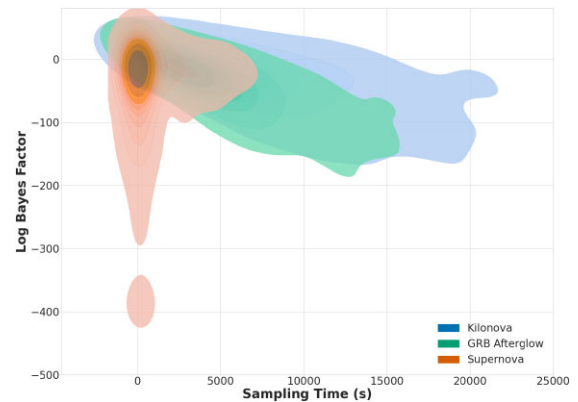
models. By integrating features from this pipeline, NMMA could then rank the relative quality of the models in fitting the light curve. Extending this further, it would then be possible to develop a built-in method that performs this comparative analysis on multiple light curves and evaluate which is most likely to be the model of interest.

The work outlined in Section 3.2 also lays the groundwork for another addition to the NMMA framework. By generating a large sample of light curves for various common transient sources and using NMMA to analyse their relative fit quality as additional observations are considered, this could form the basis for a ML training set using one of the several ML libraries available for PYTHON to optimize the selection of a candidate from a set of multiple light curves that would maximize the odds ratio for the transient type of interest to the user. This is a valuable prospect when one is limited

in the number of candidates that can be triggered in a given night for follow-up observation.

During the course of this project, it became apparent that this body of analyses would be beneficial to make available to the broader collaboration outside of members of the Growth MMA Slack server. As a result, an extension to the pipeline was made such that the analyses are now made available on Fritz, which is an open-source platform for time domain astronomers. Along these same lines, work is being done to develop an NMMA API to enable NMMA analyses to be requested an executed through Fritz. After adding the aforementioned features to NMMA motivated by this pipeline, they could also be made available through the API to enable users to leverage the power of NMMA in analysing light curves and informing follow-up observations while using Fritz. Extending the API even further, it should be possible to automatically perform these comparative analyses on new objects as they're posted to Fritz. Adding these features to NMMA will be a boon to researchers and enable more productive time-domain astronomy.

Most of the outstanding questions are statistical in nature. With additional time and data, more definitive evaluations of fit quality will be able to be made automatically as we determine statistically significant measures. One consideration is the current level of manual review conducted prior to processing by the pipeline. As of this writing, only a handful of the most promising targets are considered. It may prove beneficial to relax this filtering process and allow the pipeline to evaluate less promising targets. Statistical analysis of the difference in fit quality between these two groups could reveal a measure that could be added to the pipeline to automate this filtering in the future.

DATA AVAILABILITY

Data regarding job completion statistics are available upon request; observational data reported in this article are publicly available on the Transient Name Server.

REFERENCES

Abbott B. P. et al., 2017a, Phys. Rev. Lett., 119, 161101
Abbott B. P. et al., 2017b, Nature, 551, 85
Abbott B. P. et al., 2017c, ApJ, 848, L13
Abbott B. P. et al., 2017d, ApJ, 850, L39
Ackley K. et al., 2020, A&A, 643, A113
Ahumada T. et al., 2021a, Nat. Astron., 5, 1179
Ahumada T. et al., 2021b, Nat. Astron., 5, 917
Anand S. et al., 2021, Nat. Astron., 5, 46
Andreoni I. et al., 2019, ApJ, 881, L16
Andreoni I. et al., 2020, ApJ, 890, 131
Andreoni I. et al., 2021, ApJ, 918, 63
Annala E., Gorda T., Kurkela A., Vuorinen A., 2018, Phys. Rev. Lett., 120, 172703
Antier S. et al., 2020, MNRAS, 497, 5518
Barbary K. et al., 2016, Astrophysics Source Code Library, record ascl:1611.017
Bauswein A., Just O., Janka H.-T., Stergioulas N., 2017, ApJ, 850, L34
Bellm E. C. et al., 2019, PASP, 131, 068003
Bulla M., 2019, MNRAS, 489, 5037
Burrows A., Hayes J., Fryxell B. A., 1995, ApJ, 450, 830
Campana S. et al., 2006, Nature, 442, 1008
Chornock R. et al., 2017, ApJ, 848, L19
Coughlin M. W. et al., 2018, MNRAS, 480, 3871
Coughlin M. W., Dietrich T., Margalit B., Metzger B. D., 2019a, MNRAS, 489, L91
Coughlin M. W. et al., 2019b, MNRAS, 492, 863
Coughlin M. W. et al., 2019c, ApJ, 885, L19
Coughlin M. W. et al., 2020a, Phys. Rev. Research, 2, 022006
Coughlin M. W. et al., 2020b, Nat. Commun., 11, 4129
Coulter D. A. et al., 2017, Science, 358, 1556
Cowperthwaite P. S., Berger E., Villar V. A. et al., 2017, ApJ, 848, L17
Dekany R. et al., 2020, PASP, 132, 038001
Dietrich T., Coughlin M. W., Pang P. T. H., Bulla M., Heinzel J., Issa L., Tews I., Antier S., 2020, Science, 370, 1450
Fitzpatrick E. L., 1999, PASP, 111, 63
Goldstein A. et al., 2017, ApJ, 848, L14
Goldstein D. A. et al., 2019, ApJ, 881, L7
Gomez S. et al., 2019, ApJ, 884, L55
Gompertz B. P. et al., 2020, MNRAS, 497, 726
Graham M. J. et al., 2019, PASP, 131, 078001
Hotokezaka K., Nakar E., Gottlieb O., Nissanke S., Masuda K., Hallinan G., Mooley K. P., Deller A., Nat. Astron., 3, 940
Huth S. et al., 2022, Nature, 606, 276
Ivezić Ž. et al., 2019, ApJ, 873, 111
Janka H. T., Langanke K., Marek A., Martínez-Pinedo G., Müller B., 2007, Phys. Rep., 442, 38
Kann D. A. et al., 2023, ApJ, 948, L12
Kasliwal M. M. et al., 2019, MNRAS, 510, L7
Kasliwal M. M. et al., 2020, ApJ, 905, 145
Kunert N. et al., 2023, MNRAS, 527, 3900
Lai X., Zhou E., Xu R., 2019, Eur. Phys. J. A, 55, 60
Levan A. et al., 2005, ApJ, 624, 880
Lundquist M. J. et al., 2019, ApJ, 881, L26
Margalit B., Metzger B., 2017, L19 ApJ, 850
Masci F. J. et al., 2019, PASP, 131, 018003
Metzger B. D., 2017, Living Rev. Rel., 20, 3
Most E. R., Weih L. R., Rezzolla L., Schaffner-Bielich J., 2018, Phys. Rev. Lett., 120, 261103

Pang P. T. H., Tews I., Coughlin M. W., Bulla M., Broeck C. V. D., Dietrich T., 2021, ApJ, 922, 14

Pang P. T. H. et al., 2022, Nat. Commun., 14, 8352

Patterson M. T. et al., 2018, PASP, 131, 018001

Pian E. et al., 2017, Nature, 551, 67 EP

Piro A. L., Haynie A., Yao Y., 2021, ApJ, 909, 209

Radice D., Perego A., Zappa F., Bernuzzi S., 2018, ApJ, 852, L29

Ridley E. et al., 2021, TNSAN, 220, 1

Rosswog S., Feindt U., Korobkin O. et al., 2017, Class. Quant. Grav., 34, 104001

Ryan G., Eerten H. v., Piro L., Troja E., 2020, ApJ, 896, 166

Savchenko V. et al., 2017, ApJ, 848, L15

Slivkins A., 2019, Foundations and Trends in Machine Learning, 12, 1

Smartt S. J. et al., 2017, Nature, 551, 75 EP

Soderberg A. M. et al., 2008, Nature, 453, 469

Tews I., Pang P. T. H., Dietrich T., Coughlin M. W., Antier S., Bulla M., Heinzel J., Issa L., 2021, ApJ, 908, L1

Watson D. et al., 2019, Nature, 574, 497

This paper has been typeset from a TEX/LATEX file prepared by the author.