

MALLORN: many artificial LSST light curves based on observations of real nuclear transients

Dylan Magill¹*, M. Nicholl¹, V. Anilkumar², S. van Velzen², X. Sheng¹, T. S. Mai³, H. V. Tran³, N. P. Doan³, T. Moore^{1,4}, S. Srivastav⁵, D. R. Young¹, C. R. Angus¹ and J. Weston¹

¹*Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK*

²*Leiden Observatory, Leiden University, Postbus 9513, NL-2300 RA Leiden, the Netherlands*

³*School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast BT9 5BN, UK*

⁴*Space Telescope Science Institute, 3700 San Martin Dr, Baltimore, MD 21218, USA*

⁵*Astrophysics sub-Department, Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK*

Accepted 2026 March 3. Received 2026 January 21; in original form 2025 December 4

ABSTRACT

The Vera C. Rubin Observatory's 10-yr Legacy Survey of Space and Time (LSST) is expected to produce a hundredfold increase in the number of transients we observe. However, there are insufficient spectroscopic resources to follow up on all of the wealth of targets that LSST will provide. As such it is necessary to be able to prioritize objects for follow-up observations or inclusion in sample studies based purely on their LSST photometry. We are particularly keen to identify tidal disruption events (TDEs) with LSST. TDEs are immensely useful for determining black hole parameters and probing our understanding of accretion physics. To assist in these efforts, we present the Many Artificial LSST Light curves based on the Observations of Real Nuclear transients (MALLORN) data set and the corresponding classifier challenge for identifying TDEs. MALLORN comprises 10 178 simulated LSST light curves, constructed from real Zwicky Transient Facility (ZTF) observations of 64 TDEs, 727 nuclear supernovae and 1407 AGN with spectroscopic labels using Gaussian process fitting, empirically motivated spectral energy distributions from `SNCosmo` and the baseline from the Rubin Survey Simulator. Our novel approach can be easily adapted to simulate transients for any photometric survey using observations from another, requiring only the limiting magnitudes and an estimate of the cadence of observations. The MALLORN Astronomical Classification Challenge, launched on Kaggle on 2025 October 15, will allow competitors to test their photometric classifiers on simulated LSST data to find TDEs and improve upon their capabilities prior to the start of LSST.

Key words: Data methods – Machine learning – Time domain astronomy – Transient sources.

1 INTRODUCTION

The Vera C. Rubin Observatory's 10-yr Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) is expected to revolutionize time-domain astronomy. It is predicted to produce a hundredfold increase in the number of transients (such as supernovae and tidal disruption events) that we are able to discover, thanks to its large mirror size (8.4 m) allowing it to detect fainter and more distant transients compared to previous wide-field survey telescopes.

However, we have insufficient spectroscopic resources to follow up on every photometrically discovered transient that LSST is able to identify, even with dedicated spectroscopic follow-up from facilities including ESO 4MOST, operating the Time Domain Extragalactic Survey (TiDES; C. Frohmaier et al. 2025), the Son of X-shooter (SOXS; P. Schipani et al. 2018) instrument on the

New Technology Telescope (NTT), and US NOIRLab telescopes such as Gemini (C. M. Mountain, F. C. Gillett & J. M. Oschmann 1998) and SOAR (V. L. Krabbendam et al. 2004). As spectroscopic observations require much longer exposure times compared to photometry, and most spectrographs can only focus on one object at a time, we will struggle to get spectra for even a few percent of the objects that LSST will discover. Therefore, in order to benefit from the wealth of targets that LSST will provide, it will be necessary to identify which objects to prioritize for follow-up observations and to conduct analysis of larger samples based on photometry alone.

Selection of targets based on photometry is particularly important for identifying examples of rare transient classes. One rare transient of great physical interest are tidal disruption events (TDEs). TDEs occur whenever an unfortunate star is ripped apart by the immense gravitational forces it experiences if it is scattered onto an orbit that passes too close to a supermassive black hole (S. Gezari 2021). TDEs are a relatively recent astronomical discovery, first predicted in the 1970s (J. G. Hills 1975; V. V. Lidskii & L. M.

* E-mail: dmagill14@qub.ac.uk

Ozernoi 1979) and first observed in optical wavelengths in the early 2010s (S. Velzen et al. 2011; S. Gezari et al. 2012).

TDEs are of key astrophysical significance because they can be used to identify the mass of the disrupting supermassive black hole (SMBH). This can be applied to smaller SMBHs which may not otherwise be detectable for a dynamical approach. This allows us to investigate scaling relations and black hole growth at the lower mass end of the SMBH mass function (P. Ramsden et al. 2025). They are also sensitive probes of galaxy dynamics/evolution (K. D. French et al. 2020), and accretion physics (e.g. M. Guolo et al. 2025), and may be detectable to very high redshifts due to their high luminosity in the ultraviolet (M. Karmen et al. 2025).

Our understanding of TDEs is currently limited by the small sample size of observed objects (~ 100). Fortunately, it has been predicted that LSST will detect roughly a few thousand TDEs per year (S. Velzen et al. 2011; K. Bricman & A. Gomboc 2020; K. B. Bricman et al. 2023). The significant uncertainty in this number is due to the uncertainty in SMBH mass distribution, which the volumetric TDE rate is heavily dependent on. Additionally, the fraction of these TDEs that is actually identified is limited by the availability of spectroscopic follow-up resources to classify discovered transients.

Machine learning classifiers are now being utilized to photometrically select TDEs. The most prominent current TDE classifiers are the NEural Engine for Discovering Luminous Events (NEEDLE; X. Sheng et al. 2024), TDEscore (R. Stein et al. 2024), Finding Luminous Exotic Extragalactic Transients (FLEET; S. Gomez et al. 2020, 2023) and the Automatic Learning for the Rapid Classification of Events (ALeRCE) light-curve classifier (M. Pavez-Herrera et al. 2025). These classifiers typically select for blue transients in the nuclei of centrally concentrated galaxies. The most common contaminants in the search for TDEs are active galactic nuclei (AGN) and Type Ia supernovae, which make up the bulk of the observed nuclear transients (R. Hložek et al. 2023; Y. Dgany et al. 2023).

To evaluate the performance of multiple classifiers and to allow for them to improve upon their capabilities, the scientific community has developed the concept of classification challenges for classifier algorithms. The data set provided to the competitor is split into two parts: a training set and a testing set. The competitor is able to train their classifier on the training portion of the data set in which the types of objects are known, and will then use their classifier to attempt to identify the objects in the testing set. The competitor will then submit their results to the organizers of the classification challenge and have their performance evaluated.

Recent astronomical classification challenges include the Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC; A. I. Malz et al. 2019; R. Kessler et al. 2019; R. Hložek et al. 2023) and the Extended LSST Astronomical Time-Series Classification Challenge (ELAsTiCC; M. Lokken, A. Gagliano & ELAsTiCC Team 2023). PLAsTiCC created 3.5 million simulated objects for LSST using SNANA (R. Kessler et al. 2009) and asked competitors to provide likelihoods for each object of belonging to each transient class. PLAsTiCC focused on producing successful classifications for all objects in order to equally weight the various science goals of LSST. ELAsTiCC was the successor competition that converted the data into a stream of alerts fed into the LSST brokers (platforms to host and add value to real-time detections of transient and variable objects) to allow for classifier testing on those platforms. Additional examples

of previous classifier challenges for other science goals include the Supernova Photometric Classification Challenge (SNPCC; R. Kessler et al. 2010), Mapping Dark Matter (T. Kitching et al. 2011), Observing Dark Worlds (D. Harvey et al. 2014) and Galaxy Challenge (S. Dieleman, K. W. Willett & J. Dambre 2015).

PLAsTiCC and ELAsTiCC both used models to simulate the data, some of which were based on coarse analytic approximations, and so may not be fully accurate representations of real transients. Training ML algorithms on models rather than data therefore runs the risk of resulting in poor performance when deployed in the real world.

To aid in the preparation for LSST, we have produced the Many Artificial LSST Light curves based on the Observations of Real Nuclear transients (MALLORN) data set. This is a simulated photometric data set of nuclear transients for LSST, with realistic cadence and luminosities, that has been developed using real observations from the Zwicky Transient Facility (ZTF; E. C. Bellm et al. 2019). This data set is being shared as a classifier challenge to allow others in the field to test and improve upon their photometric classifiers. This will allow the scientific community to better prepare for the start of LSST and improve upon our collective ability to effectively use the wealth of data that LSST will provide.

In this paper we describe the method used to create simulated LSST light curves, the construction of the MALLORN data set and what we have released in the classification challenge.¹

2 SIMULATED LIGHT CURVE PRODUCTION

This section will detail the process of producing a simulated LSST light curve from a real ZTF light curve. More details on the production at scale and specific choices made for each class of object are provided in Section 3. For MALLORN, this method used specific data from ZTF and specific survey characteristics appropriate to LSST. However, provided you have the limiting magnitudes of the instrument and an estimate of the cadence for the target survey, this method should be applicable to produce simulated data for any new photometric survey.

The MALLORN Data Production Google Colab Notebook² details the process of going from a real ZTF light curve to a simulated LSST light curve. The methods are outlined as clearly as possible to ease the process for anyone who wishes to replicate this method for any future use.

2.1 Initial light curve selection

The data set used to produce MALLORN consisted of all of the ZTF forced photometry of nuclear transients from 2017 to 2023. This sample was created by applying a filter to the AMPEL archive of ZTF alerts (J. Nordin et al. 2019). This so-called Nuclear Filter was originally written and employed by S. Velzen et al. (2019), E. Hammerstein et al. (2021), and S. Velzen et al. (2021) to select nuclear transients in real-time from the ZTF alert stream.

To pass the Nuclear Filter an object must have a high likelihood of a galactic host, at least one detection brighter than 20 mag and an angular distance to the core of less than 0.5 arcsec. For more information on this filter, please see S. Reusch (2024). The sample

¹<https://www.kaggle.com/competitions/mallorn-astronomical-classification-challenge/overview>

²<https://colab.research.google.com/drive/1oy96r29Zs4U5Hl-THsZPCnOQbuz21hl5>

produced by this filter contained a total of 11 190 objects. There are no hostless transients present in the sample.

We first run the ZTF forced photometry pipeline for all transients in our sample. This returns the time, count rate and zero point for each observation, in the ZTF g and r bands. We use the zero point to convert each observation to a flux density measurement in microjanskys (μJy).

The data set was cross-matched with the `Transient Name Server`³ using the ZTF object names and coordinates (RA and Dec) to determine the spectroscopic type and redshift where available. This produced 791 objects for which the type and redshift were known. Further manual cross-matching was able to identify an additional 38 TDEs within the ZTF data set, bringing the total number of confirmed TDEs included to 64. This sample contains both TDEs in galaxies without AGN and TDEs in galaxies with AGN. Further cross-matching with version 8 of the `Milliquas` catalogue (E. W. Flesch 2023) identified 1407 additional AGN within the ZTF data set. There were some objects misclassified as Type II superluminous supernovae on TNS that were corrected to AGN based on the contaminants identified in P. J. Pessi et al. (2025).

Following the approach outlined in E. Hammerstein et al. (2021), a baseline correction is carried out on the light curve. We define the baseline as the inverse-variance weighted average of the flux a hundred days before the rise of the transient event. This correction step subtracts this average value from the complete light curve and propagates the errors, bringing the mean of the pre-transient event flux to zero. In addition, the reduced χ^2 values of the baseline are brought to one by scaling the uncertainties of the flux measurement.

We next correct for Galactic extinction using the `dustmaps` (G. M. Green 2018) and `extinction` (K. Barbary 2016) packages. The extinction coefficient, $E(B - V)$, for the specified RA and Dec is determined using the Schlafly & Finkbeiner dustmap (E. F. Schlafly & D. P. Finkbeiner 2011). Using the $E(B - V)$ value and the effective wavelength for the respective filter (g or r band) sourced from the Spanish Virtual Observatory filter service (C. Rodrigo & E. Solano 2020), the extinction at a given wavelength (A_λ) can then be calculated using the extinction laws from E. L. Fitzpatrick (1999) and assuming a $R_V = 3.1$ for Milky Way extinction (J. A. Cardelli, G. C. Clayton & J. S. Mathis 1989). The de-extincted flux values (F_λ) can then be calculated from the initial extincted flux ($F_{\lambda, \text{Ext}}$) following:

$$F_\lambda = F_{\lambda, \text{Ext}} 10^{\frac{A_\lambda}{2.5}} \quad (1)$$

2.2 Gaussian process fit

We then take the baseline- and extinction-corrected light curves, and select the band (g or r) with the most observations. Although ZTF also uses i -band for some observations, these data are significantly shallower and at a lower cadence than the data in g and r , and are insufficient for the fitting process. The peak time for the selected band is found by determining the time which corresponds to the greatest flux value. A Gaussian Process (GP) fit is then applied to the data for the selected band. This produces a continuous light curve, interpolating between any gaps in the observations. There is strong precedent in the literature for the application of Gaussian process fitting to the photometry for

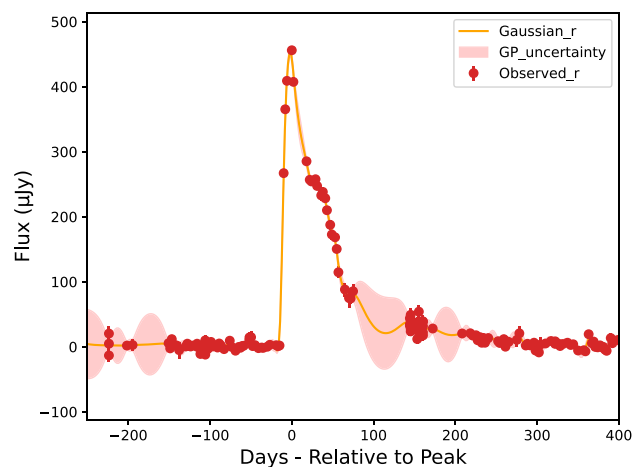


Figure 1. Plot of flux (μJy) against time (relative to peak) showing the r -band observations for SN II ZTF22abcfics/SN2022ryv and the GP fit to those observations. The GP fit reliably follows the trend within the data points, matching the rise to peak and subsequent decline of the supernova. During large gaps in observations the uncertainty in the GP fit is inflated. We show the GP uncertainty only for visualization and use a different method to determine the uncertainty in our simulated LSST observations.

transient objects (D. Foreman-Mackey et al. 2017; K. Boone 2019; M. V. Kornilov, T. A. Semenikhin & M. V. Pruzhinskaya 2023), though see H. F. Stevance & A. Lee (2023) for limitations of light-curve GP fitting.

The kernel used for the GP fit consists of three components combined as:

$$K = K_C \times K_M + K_{WN}. \quad (2)$$

The first is a constant kernel (K_C), which defines amplitude and range of possible parameters. Another component is the Matérn kernel (K_M), which is a generalized radial basis function (RBF) kernel – parametrized by a given length scale and used to determine the similarity of two data points. This sets the characteristic time-scale over which the flux changes. The advantage of a Matérn kernel over a standard RBF kernel is that it contains an additional smoothness parameter which allows for more efficient computation. The order of the Matérn kernel used is $3/2$. The final component is a white noise kernel (K_{WN}), which accounts for possible variation due to noise in the data. An example of a GP fit to a ZTF light curve is shown in Fig. 1.

To verify the validity of the GP fit, we calculate the mean squared error (MSE) of the fit. To do this, the flux is first normalized so that the amplitude ranges from 0 to 1 and then a set fraction of data points are randomly removed. The GP fit is then re-ran and used to create simulated points at the positions of the removed points. The MSE is computed from the differences between the original points and the simulated points (K. Benidis et al. 2022). An MSE below 0.2 is generally considered a good fit (T. Hastie, R. Tibshirani & J. Friedman 2009). Running on a test of 200 randomly selected objects from the ZTF nuclear transient data set, the GP fits returned a median value of 0.14.

When generating simulated light curves, we do not allow fits which produce an MSE value greater than 0.3 to pass to the next stage of light-curve production, as we deem this to be an unacceptable fit. When this occurs, the fitting process is repeated with altered parameters for a limited number of attempts until the fit either passes the criteria or fails to do so and the attempt

³<https://www.wis-tns.org/>

at creating that simulated light curve is abandoned. This prevents unphysical light curves from poor GP fits appearing in our final data set.

Using a multidimensional GP fit which considered data from both bands to fit to one band was also tested. This produced little to no improvement to the fit. Therefore we chose to apply the GP fit only to the band with the most observations. This has the advantage that the other unused band can later be used to verify the fits and our ability to extrapolate to other wavelengths (see Section 2.4).

In order to more appropriately handle the stochastic nature of AGN light curves, a damped random walk (DRW) time-series model was used to produce a continuous fit to these data. A DRW can be defined by an exponential covariance matrix (kernel) and two main parameters: the long-term variability amplitude, which defines the largest possible fluctuation range of a light curve, and the signal decorrelation time-scale, which defines the minimum time-scale beyond which the luminosity differences over increasing time intervals are no longer correlated with the length of those intervals. In application, this is broadly similar to an order 1/2 Matérn kernel.

The DRW/Matérn-1/2 kernel is more suited for the stochastic nature of AGN light curves as it does not contain the smoothness parameter present in the Matérn-3/2 kernel and consequently produces a better match to the multiple, stochastic rises and falls present in AGN light curves, as opposed to the isolated and relatively smooth rise and decline present in the light curves of SNe and TDEs.

There are multiple examples in the literature of DRWs being used to model the variability in the light curves of quasars and AGN (B. C. Kelly, J. Bechtold & A. Siemiginowska 2009; C. L. MacLeod et al. 2010; Ž. Ivezić & C. MacLeod 2014). The approach used here is based on that of X. Sheng, N. Ross & M. Nicholl (2022), which employs the `EzTao` toolkit (W. Yu & G. T. Richards 2022). `EzTao` is an efficient package for simulating, modelling, and recovering the parameters of DRW and higher order time-series models for AGN variability, implemented using Gaussian processes. Solely the DRW kernel is then used in a GP fit to produce a continuous light curve for the AGN. An example of this fit is shown in Fig. 2.

Our Gaussian process fit yields inflated uncertainties, particularly during the gaps between observations. This behaviour arises from the use of a zero-mean function with a stationary kernel, which models the transient as a deviation from the mean (Anilkumar et al., in preparation). As shown in Fig. 1, these effects are most pronounced further away from the transient peak. However, this does not impact our subsequent analysis, as data point errors are determined using the `Rubin Survey Simulator` (see Section 2.5) rather than the GP uncertainty.

2.3 Rescaling luminosity for simulated light curves

Since LSST observes to a deeper limiting magnitude than ZTF (e.g. in r -band, $m_{\text{lim,ZTF}} = 20.6$, $m_{\text{lim,LSST}} = 24.7$), a transient at a given redshift would be detected with high confidence in LSST both earlier and later in time than the same transient in ZTF. Simulating this would require an extrapolation in time, which would require a more complicated model not anchored in the ZTF data.

To avoid this issue, we instead place each simulated light curve at a distance where the signal-to-noise in LSST would approximately match the observed signal-to-noise ratio of the observed

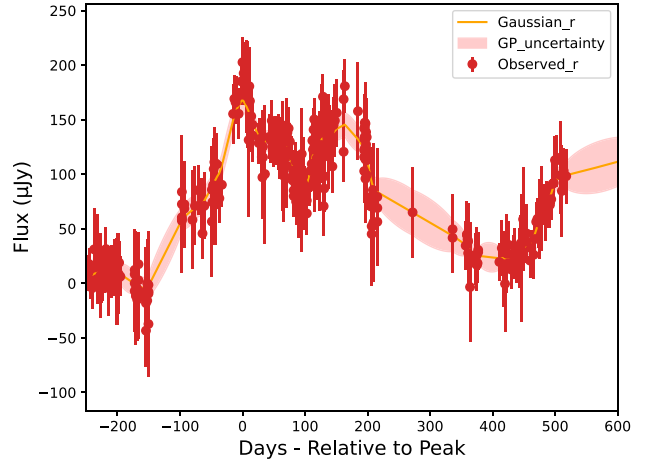


Figure 2. Plot showing the Damped Random Walk (DRW) fit to the observed AGN ZTF light curve (ZTF20acxfcmr). The DRW more effectively handles the stochasticity of an AGN light curve. During large gaps in observations the uncertainty in the GP fit is inflated. We show the GP uncertainty only for visualization and use a different method to determine the uncertainty in our simulated LSST observations.

light curve in ZTF. Effectively, this shifts the simulated population to higher redshifts where the bulk of LSST transients will be detected, while also preserving the shape of the observed luminosity distribution from ZTF. We note that this will result in few low-redshift transients in our simulated data set; however such objects are likely to be identifiable already with existing methods and surveys, complemented by the additional colour information from LSST. It is for the more distant objects that we need to improve light-curve classifiers.

The real distance to each ZTF object is calculated from its known redshift, using `astropy.cosmology` with `WMAP9` from G. Hinshaw et al. (2013). The following equation is then used to calculate a simulated distance at which the observed flux from the object is at the same factor above the LSST detection limit as the original flux was above the ZTF detection limit:

$$D_{\text{LSST}} = D_{\text{ZTF}} \times \sqrt{\frac{F_{\text{lim,ZTF}}}{F_{\text{lim,LSST}}}} \times \sigma. \quad (3)$$

A random scatter component ($\sigma = 0.9\text{--}1.1$) is included within this equation to enhance the variety of simulated objects produced for the data set. This simulated LSST distance for each object is then used to determine the simulated LSST redshift using the same cosmology calculator. The simulated flux is rescaled following:

$$F_{\text{LSST}} = F_{\text{ZTF}} \times \frac{F_{\text{lim,ZTF}}}{F_{\text{lim,LSST}}} \times \delta. \quad (4)$$

To further enhance the variety of simulated objects produced for the data set, a scatter component (δ) is selected from a normal distribution, which can be tuned differently if desired to represent the diversity inherent to different transient classes. For discussion on the values used in our data challenge, see Section 3.3.

Finally, the effect of time dilation on the observed light curve was taken into account using the known ZTF redshift (z_{ZTF}) and the simulated LSST redshift (z_{LSST}):

$$(t - t_{\text{peak}})_{\text{LSST}} = (t - t_{\text{peak}})_{\text{ZTF}} \times \frac{1 + z_{\text{LSST}}}{1 + z_{\text{ZTF}}}. \quad (5)$$

2.4 Using SNCosmo models to create data for missing bands

One of the main challenges in simulating LSST data from ZTF observations is that ZTF observes predominantly in two bands (g and r – though with occasional i -band observations), whereas the LSSTCam (SLAC National Accelerator Laboratory & NSF-DOE Vera C. Rubin Observatory 2025) will observe in six bands (u , g , r , i , z , and y). Therefore, it is necessary to simulate data for the other four bands not observed by ZTF, and important that this is done in such a way as to still be grounded in the real transient observations. Since the surveys share some bands, we will henceforth differentiate observations/simulations in ZTF and LSST using the subscripts ‘Z’ and ‘L’.

To accomplish this, we used the `SNCosmo` package (K. Barbary et al. 2025), which allows us to generate empirically motivated spectral energy distribution (SED) models for different types of transients at a given redshift and time since explosion. For each ZTF transient with an available spectroscopic type, we select an appropriate `SNCosmo` model, and use it to generate two time-dependent SEDs – one at the observed redshift and one at the simulated LSST redshift.

From these time-dependent SED models, we can then compute the colour evolution and k -correction in each LSST band with respect to our reference band (g_Z or r_Z); i.e. $r_Z - u_L$, $r_Z - g_L$, $r_Z - r_L$, etc. These colour curves can then be added to the (time-dilated) GP fit of the reference band to produce k -corrected simulated light curves for all six LSST bands using:

$$F_L(t'') = F_Z(t') 10^{(m_Z(t') - m_L(t''))/2.5}. \quad (6)$$

F_Z and F_L are the fluxes in the ZTF reference band and the simulated LSST band respectively. $m_Z - m_L$ is the colour computed with `SNCosmo`. All of the variables are functions of time, with $t' = (t - t_{\text{peak}})_{\text{ZTF}}$ and $t'' = (t - t_{\text{peak}})_{\text{LSST}}$; these quantities are related through equation (5).

These colour curves are applied only in a restricted temporal region around the transient. This is necessary as it prevents the colour curves from exaggerating random noise in ZTF around zero flux, which may produce a ‘detection’ in another band if the colour term is large enough. The colour curves are set to zero outside of this range so that each curve follows that of the reference band. This is a reasonable approximation as we would expect photometry in all bands to fall within a standard deviation of the baseline before the transient has occurred and after it has faded. For most transient types, we define this range as $50 \text{ d} \times (1 + z)$ prior to the peak to $150 \text{ d} \times (1 + z)$ after the peak. For particularly long-lived transients this range is extended. For SN Ic-BL, SN IIn, and SLSN, the defined range is $100 \text{ d} \times (1 + z)$ prior to the peak to $300 \text{ d} \times (1 + z)$ after the peak. For TDEs, the defined range is $100 \text{ d} \times (1 + z)$ prior to the peak to $400 \text{ d} \times (1 + z)$ after the peak. The range limit is not applied to AGN light curves, as they are expected to vary continuously and as such may have colour evolution throughout their entire light curve.

This process therefore allows us to produce a full six-band continuous k -corrected light curve based on a real ZTF light curve and realistic SEDs for a given transient type. An example is shown in Fig. 3.

To verify the validity of this approach, we also simulate data in the ZTF band with fewer data points (i.e. the band not used for the GP fit) at the ZTF (true) redshift, and compare to the real observations for that band. As shown in Fig. 4, the simulated

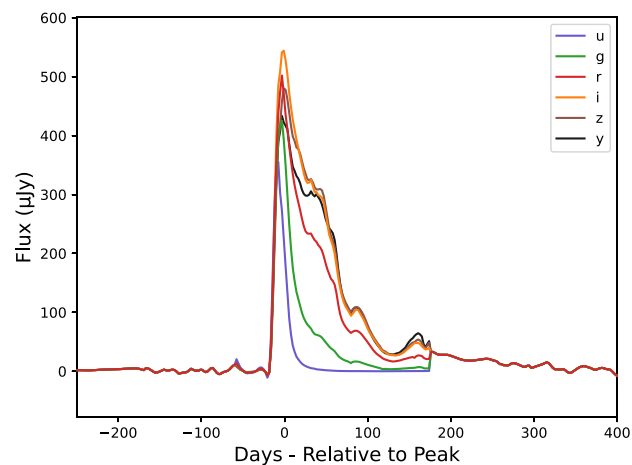


Figure 3. Six band plot of continuous k -corrected light curves for the SN II ZTF22abcfcis/SN2022ryv. The r -band is produced via a GP fit of the observed data, whilst the other bands are produced using the colour differences generated from the `SNCosmo` models. The instances of more erratic variation in the z and y bands are the result of less data for the model for those bands, but are well within the bounds of the error values for their respective bands.

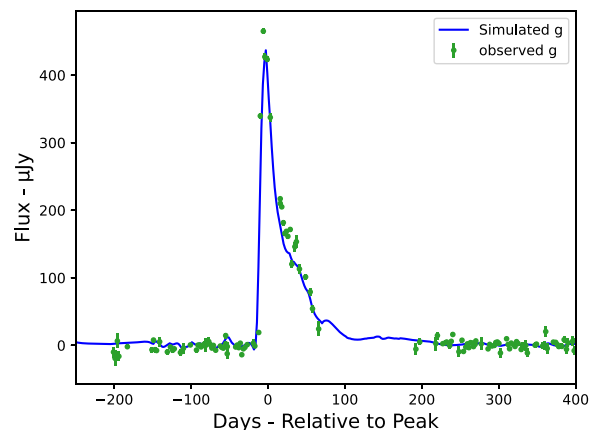


Figure 4. Plot comparing the simulated continuous g band light curve generated via the observed redshift `SNCosmo` colour differences to the observed g -band data for the SN II ZTF22abcfcis/SN2022ryv. The simulated data is shown to provide a reasonable match to the observed data, verifying the validity of this approach.

continuous g -band light curve provides an acceptable match to the real ZTF g -band observations.

2.5 Implementing LSST cadence

In order to create realistic and representative simulated transients for LSST, we sample our model light curves using the LSST cadence Baseline v5.0.0 from the Rubin Survey Simulator (P. Yoachim 2025; P. Yoachim et al. 2025a, b).

For each model light curve, a random RA and Dec are chosen within the LSST footprint, and a random date for the light curve peak is chosen within the 10 yr survey period. The cadence simulation is then used to determine the visit dates for each filter. These visit dates can then be matched with the corresponding values from the continuous light curves for each band.

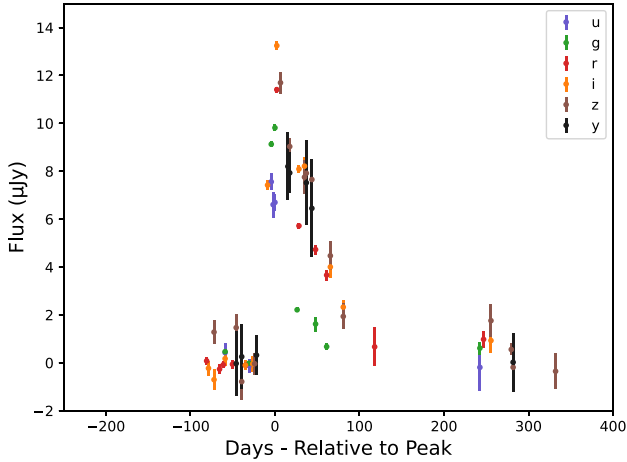


Figure 5. Example of a final simulated LSST light curve based on the real ZTF light curve for SN II ZTF22abcfcis/SN2022ryv. The luminosities have been subjected to some random scatter within measurement uncertainties to mimic the expected scatter from noise in LSST.

Often, these random dates fall in a period of solar conjunction, or when the field is in the ‘off-season’ of the LSST rolling cadence. To avoid generating useless light curves which would be egregiously difficult to classify, we apply safety checks to ensure that the transient is actually detected in the simulation. We define a detection as having 10 data points across all bands with a signal-to-noise ratio $> 5\sigma$. If the transient is not detected, the matching is repeated with newly generated co-ordinates and date.

We then apply Galactic extinction to the simulated LSST light curves. This is carried out in an inverse of the method described in Section 2.1. The effective wavelengths for the LSSTCam used to compute the extinction for each LSST band were: $u = 3641\text{\AA}$, $g = 4704\text{\AA}$, $r = 6155\text{\AA}$, $i = 7504\text{\AA}$, $z = 8695\text{\AA}$ & $y = 10056\text{\AA}$. These values were sourced from the SVO Filter Service (C. Rodrigo & E. Solano 2020).

The Rubin Survey Simulator baseline contains 5σ detection limits for each observation. Given the faintness of our targets, we assume that all measurements will be sky-noise limited. We therefore estimate the errors on each photometric point by converting these limits from AB magnitude to flux and multiplying by 0.2.

Using a normal distribution with width equal to these estimated measurement uncertainties, scatter is added to each light-curve point. This mimics the expected scatter due to observational noise. This is the final stage in producing a simulated LSST light curve from an existing ZTF light curve, an example of which is shown in Fig. 5. This process should be replicable for any future photometric survey. All that is required are the light curves from one survey, the cadence of the new survey, and the limiting magnitudes for each survey.

Though most of the data from LSST will be obtained in the wide-fast-deep (WFD) fields, through the use of the Rubin Survey Simulator, we can also produce light curves with deep drilling field (DDF) cadence. By selecting co-ordinates within one of the deep drilling fields, the simulation will provide visit dates and depths for that position, producing a light curve with a much greater cadence. An example is shown in Fig. 6. This will occasionally happen naturally in our simulation via the random selection of RA and Dec, or can be artificially forced by selecting a DDF position.

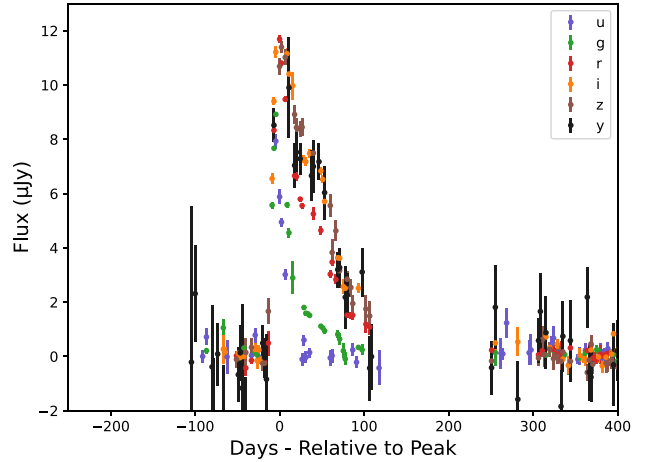


Figure 6. Plot showing a simulated LSST light curve based on the real ZTF observations of the SN II ZTF22abcfcis/SN2022ryv. This light curve has been simulated within the Extended Chandra Deep Field South (ECDFS) and thus given Deep Drilling Field (DDF) cadence, resulting in a much more densely observed light curve than in Fig. 5.

3 MALLORN DATA SET PRODUCTION

3.1 Zwicky Transient Facility data

The total number of ZTF objects used to generate simulated LSST light curves was 2198. Of this, roughly two-thirds are AGNs and just less than a quarter are SNe Ia. The remainder are made up of a variety of other transients. The following types of transients are within the ZTF data set and are used within the production of MALLORN: SN Ia, SN Ia-91T-like, SN Ia-91bg-like, SN Ia-02cx-like, SN Ia-pec, SN Ib, SN Ib/c, SN Ic, SN Ic-BL, SN II, SN IIb, SN IIn, superluminous supernovae (SLSN) Types I and II, TDE and AGN.

Although present in small numbers within the ZTF nuclear transient data set, no galactic transients (such as cataclysmic variables) or variable stars were included within the MALLORN data set. The contextual information LSST and the alert brokers will provide (e.g. host galaxy associations; D. Young 2023) in combination with the nature of the light curves for such objects should be sufficient to exclude most of these objects – and as such it was unnecessary for us to simulate them within the MALLORN data set.

In constructing our simulation, some initial checks on input light-curve quality were applied. If the light curve had less than 30 data points it was screened out. These measures predominantly affected the AGN matched via the `Milliquas` crosscheck, some of which were close to the ZTF detection limit. This is particularly prevalent for more distant AGN. For other object types, the selection on spectroscopic labels from TNS ensured that most were bright and well detected.

Further safety checks are present in the GP fit to ensure the validity of the fit. The fit will rerun with slightly altered parameters if it fails to reach 90 per cent of the observed peak flux, repeating until within the range of this threshold value or continuing on after the fifth attempt. This primarily affects narrow light curves where the GP might smooth the sharp peak. An MSE value is calculated for each of the GP fits calculated using the method described in Section 2.2. If the MSE value produced is greater than 0.3, the fit is considered failed and the object is discarded. This was quite a rare occurrence, but was a necessary measure to

ensure the validity of the GP fits. From the test of 200 randomly selected objects 85.4 per cent passed this requirement.

There is an additional safety measure for the DRW model, which excludes AGN with light curves having a duration of less than thirty days. The MSE approach was not viable to assess the DRW fit due to the different nature of variability in the AGN light curves. Removing only a small fraction of the points can result in a significant change to the shape of light curve and a large MSE value, resulting in a large quantity of AGN fits not passing the MSE check, even though the light curve produced is still physically reasonable for a stochastically varying AGN. A large number were visually assessed when testing the DRW fit which verified the validity of the approach. With the stochastic nature of AGN variability and the large number of examples present within the data set, an exact fit to any particular light curve is not as important as it was for other transient types.

3.2 SNCosmo models

A variety of models were used to produce the SED models required to generate the colour differences, with examples shown in Fig 7. For thermonuclear supernovae, the ‘hsiao’ model (E. Y. Hsiao et al. 2007) was used for SNe Ia, as it provided good coverage across all six bands. For SNe Ia-91T-like the ‘nugent-sn91t’ model was used (D. Stern et al. 2004) and for SNe Ia-91bg the ‘nugent-sn91bg’ model was used (P. Nugent, A. Kim & S. Perlmutter 2002). There is no matching model for SN Ia-02cx-like (or Iax), so the ‘nugent-sn91t’ model was used as an approximate fit for this class. The early-time colours are similar since SNe Iax spectroscopically resemble SNe Ia-91T-like at early times within about a month of explosion (S. W. Jha 2017). Similarly for SN Iap-pec (an ill-defined term covering any anomalous SN Ia) there is no matching SNCosmo model and so the ‘hsiao’ model was used to provide an approximate fit for the colours.

For core collapse supernovae, there are a greater variety of models available on SNCosmo. For consistency, it was decided that we would prioritize the ‘v19’ models from M. Vincenzi et al. (2019). For each class, there were multiple options to choose from, each based on an individual well-observed event. Where possible, we included multiple models for each type and selected one at random for each MALLORN simulated object, in order to enhance the variety present within the data set. It was necessary to exclude some models from this to ensure the production of quality six-band light curves. These models were excluded predominantly due to poor coverage in the z and y bands. For the full SNCosmo model selection details see Table 1.

There is no model or viable appropriate substitute within SNCosmo for TDEs or AGN. For these it was necessary to construct blackbody models – an excellent approximation for TDEs in particular. For TDEs a time-evolving blackbody model with a linear temperature change was implemented. A TDE peak temperature is selected at random from a list of ZTF-based temperatures. The average of the TDE temperature distribution is 1.3×10^4 K with a standard deviation of 5×10^3 K and a maximum temperature of 3.2×10^4 K. Following S. Velzen et al. (2021), a linear temperature evolution list was produced via multiband modelling of the g and r bands of the ZTF TDEs, supplemented by UV data from other surveys if available.⁴ The value for the linear change in temperature over time (K d^{-1}) which most closely matches the

Table 1. SNCosmo models chosen for each transient type.

Transient type	Selected models
SN Ia	hsiao
SN Ia-91T	nugent-sn91t
SN Ia-91bg	nugent-sn91bg
SN Ia-02cx	nugent-sn91t
SN Ia-pec	hsiao
SN Ib	v19-2004gq, v19-2007uy, v19-2012au
SN Ib/c	v19-2005bf, v19-2006ep
SN Ic	v19-2004gt, v19-2011bm
SN Ic-BL	v19-1998bw, v19-2002ap, v19-2007ru
SN II/IIP	v19-1987a, v19-2009ib, v19-2009kr, v19-2013ab, v19-2013am, v19-2016x, v19-1999dn, v19-2006t, v19-2008aq, v19-2008bo, v19-2011dh, v19-2011ei, v19-2011fu, v19-2011hs
SN Iib	v19-2006aa, v19-2011ht
SN Iin	v19-2011ht
SLSN-II	Linear Temperature Blackbody
SLSN-I	(M. Nicholl, J. Guillochon & E. Berger 2017; S. Gomez et al. 2024)
TDE	Linear Temperature Blackbody (S. Velzen et al. 2021)
AGN	Constant Temperature Blackbody

selected peak temperature is then chosen from the list of ZTF-based temperatures. This linear temperature change, which can be positive or negative, is then applied to the temperature used in the model for 100 d after the peak date of the light curve. After this 100 d a constant temperature is used. A maximum temperature of 40 000 K and minimum temperature of 8000 K were implemented to prevent the linear temperature change from producing unphysical TDEs. While we could have used the best-fitting temperature and evolution for the ZTF TDEs, randomly selecting a temperature for each simulated TDE was considered preferable as it treats all of the objects the same and helps to enhance the variety within the data set.

For AGN, a constant-temperature blackbody model is used. AGN spectra are traditionally approximated with a power law (C. M. Urry & P. Padovani 1995), however a blackbody also provides a reasonable approximation in the optical range. The AGN temperatures are produced via the same multiband modelling method as the TDE temperatures using ZTF AGN, and are again selected at random to enhance variety. The average of the AGN temperature distribution is 1.5×10^4 K with a standard deviation of 2.3×10^4 K, with the standard deviation inflated by some blue outlier AGN. We note that AGN often show lags between the light curves in different bands. Bluer bands change first, with the corresponding change in the redder bands following later. This is likely due to accretion occurring at the hotter, innermost edge of the disc and energy generated there takes at least a light travel-time to reach the cooler, outer parts of the disc (R. D. Blandford & C. F. McKee 1982; B. M. Peterson 1993). Typical observed time lags from u to y band are of the order of a few days (e.g. B. J. Shappee et al. 2014; M. M. Fausnaugh et al. 2016; H. Netzer 2022). Therefore while these are neglected in our SED model, they have little effect on the observed light curves when passed through the LSST cadence. Treating these delays properly would require constraints on the SMBH mass, which we lack for most of our sample.

The other classes with no built-in examples present in SNCosmo are the SLSNe. For SLSNe-I, we used a blackbody model with UV absorption below 3000 Å, based on the magnetar model described in M. Nicholl et al. (2017), coupled with a

⁴<https://github.com/sjoertvv/manyTDE>

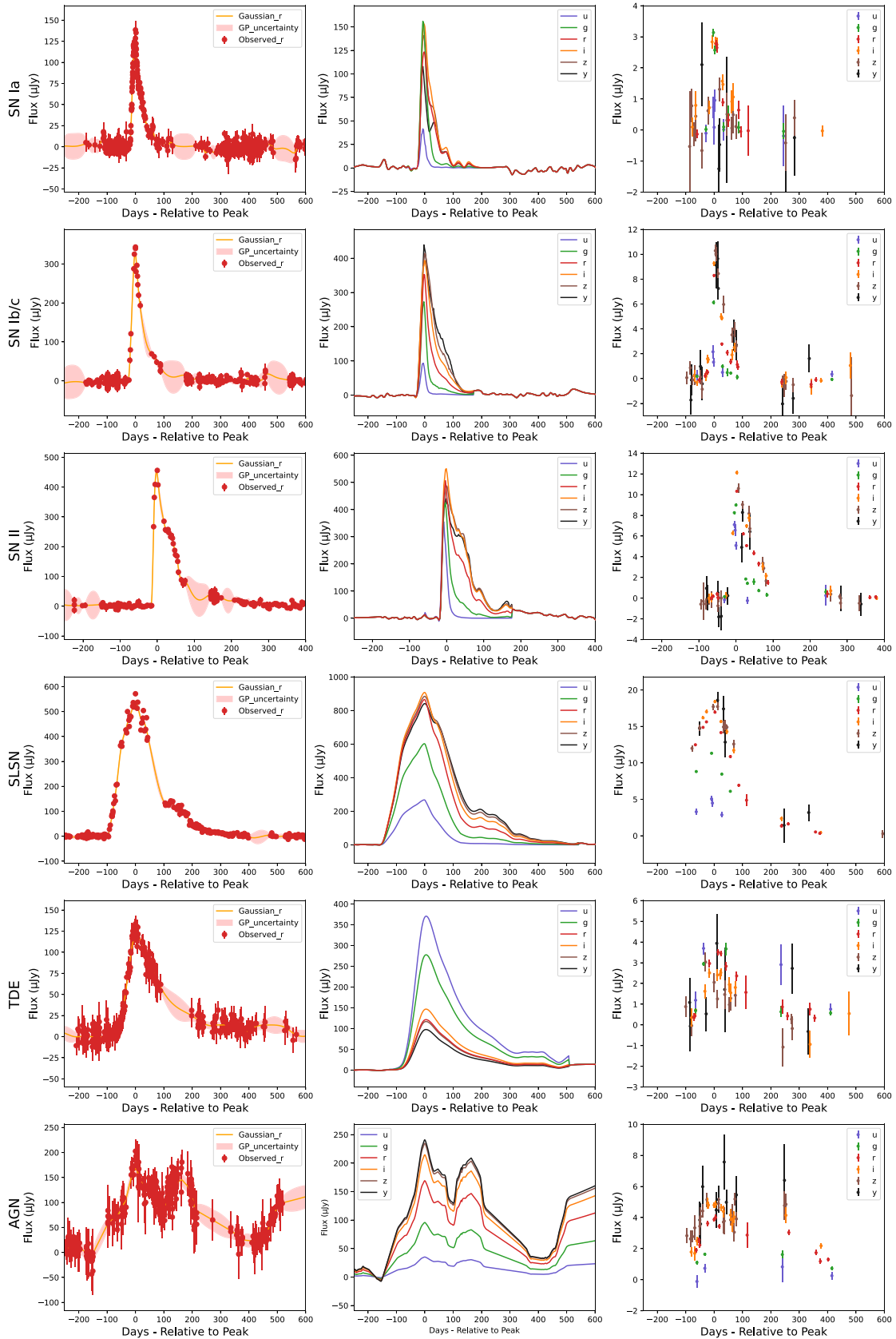


Figure 7. Plots showing the GP fit to the observed data points (LHS), the 6-band continuous SNCosmo plots (centre) and an example of that object simulated at an example LSST luminosity and observing cadence. From top to bottom, the plots are for SN Ia, SN Ibc, SN II, SLSN, TDE, and AGN. For SNe & TDEs the SNCosmo models are used to generate colours in a defined temporal region when the transient occurs. As AGN are expected to vary continuously the SNCosmo model is used to generate colours for the entire light curve.

Table 2. Summary of all quality selection criteria for light-curve generation.

Metric	Criteria
GP fit quality*	MSE < 0.3
Cadence and detection quality	> 10 5σ detections across all six bands
Bright enough to classify	Peak flux > $2.5 \times m_{\text{lim},r}$ and $z < 5$

*AGN exempted from this check.

linear temperature change similar to our TDEs. For this model, a random start and end temperature are selected (from between 10 000–15 000 K and 5000–8000 K, respectively) based on the values in S. Gomez et al. (2024). A cooling rate (K d^{-1}) is then calculated by dividing the difference between these temperatures by a randomly selected cooling time (30–100 d). This linear temperature change is then applied for 70 d after peak. For SLSNe-II the brightest SN IIn model, ‘v19-2011ht’, was selected, as SLSNe-II are thought to be the high end luminosity tail of SNe IIn (A. Gal-Yam 2019).

SNe & TDEs which occur in an AGN are not modelled differently with SNCosmo than other non-AGN SNe & TDEs. As the light curves are produced from difference imaging, it is assumed that the AGN will have been included in the template image and effectively subtracted from subsequent images. Consequently, the light curves and the colour evolution of the transient should not be affected by the presence of an AGN provided that the AGN is at a similar brightness at the time of the transient as it was at the time of the template image. In reality, some AGN will likely have coincidentally faded or brightened between the template image and observations of the transient, however modelling this is beyond the current scope of MALLORN.

3.3 Large scale production

For each real ZTF object within the source data set, it was possible to produce several distinct LSST light curves, enabling us to upsample the data set.

The σ and δ parameters described in Section 2.3 provide variation in the peak absolute and apparent magnitudes. Using a standard deviation of 0.3 for δ provided a good compromise between adding variation and avoiding excessively bright or faint transients. We get variety in colour via using different templates for core-collapse SNe and different temperatures for TDEs, AGN, and SLSNe. The LSST cadence supplies sufficient variety that objects produced from the same source with a different position on the sky and at a different date are distinct from each other as shown in Fig. 8.

Through the combination of the random scatter component in the distance calculation and the peak magnitude scatter component, it was possible for objects to be generated at luminosities that were not viable for detection. To prevent these objects from leaking into the MALLORN data set, a safety measure was implemented which removed objects that peaked below the LSST limiting magnitude. This mostly excluded AGN whose source objects were already close to the ZTF limit.

This allowed for the MALLORN data set to reach a potential total of roughly 20 000 objects. Due to the quality cuts a final total of 10 178 light curves were produced and included in the MALLORN data set. A summary of all of the light-curve quality criteria is outlined in Table 2.

Determining the redshifts of the observed objects is vital for many of the science goals for LSST. Ideally the redshift is determined spectroscopically, however this is not possible to obtain for all galaxies in LSST or even just all transient hosts. LSST instead aims to use its broad-band photometry and predictions from galaxy SEDs to determine a photometric redshift value. This approach allows for a redshift value to be determined for a much larger quantity of objects, however it is less accurate than a spectroscopic redshift (M. L. Graham et al. 2018; E. Jones et al. 2024). For the objects in the MALLORN data set a simulated photometric redshift is created. The ‘true’ simulated redshift is known for all objects. The uncertainty in LSST photometric redshifts is expected to follow M. L. Graham et al. (2018):

$$z_{\text{err}} = 0.02 \times (1 + z). \quad (7)$$

A simulated photometric redshift for each MALLORN object is drawn from a normal distribution centred on the real redshift value with a standard deviation equal to z_{err} . This approximates the photometric redshift and its uncertainty that will be available in LSST data releases, and may be an important feature in transient photometric classification. While the above relation will hold for most LSST galaxies, misidentified photometric redshift estimates can produce large outliers. For the first two years of LSST it is expected that 15 per cent of photometric redshifts will be incorrect, with this fraction reducing to 4 per cent by the end of the 10-yr survey (M. L. Graham et al. 2018). Conversely, ~ 90 per cent of photometric redshifts are expected to be accurate. Therefore, we do not currently include failed photometric redshifts within the MALLORN data set. Dealing with inaccurate photometric redshifts is an important issue, but in some ways a separate step to the photometric classification that the MALLORN Astronomical Classification Challenge is focused on improving. Furthermore, given the subtlety of this issue, inclusion of inaccurate photometric redshifts would likely disproportionately affect the performance of non-astronomers in the MALLORN Astronomical Classification Challenge.

Storing our simulated objects, and associated classifications with light curves, requires a unique identifier for each one. To produce a unique object ID for each of the simulated light curves produced, three Sindarin words are chosen at random and combined. Sindarin is a fictional language J.R.R. Tolkien developed for his Elves in ‘*The Lord of the Rings*’ (J. R. R. Tolkien 1954), which felt appropriate given the data set is named after the golden-leaved mallorn trees of Lothlórien. The Sindarin dictionary⁵ chosen contained 375 words, therefore allowing for roughly 5×10^7 possible combinations which was more than sufficient for the roughly 2×10^4 light curves generated for the MALLORN data set. The simulations were run on the Queen’s University Belfast high-performance computing facility Kelvin2.

4 DATA CHALLENGE STRUCTURE

The MALLORN Astronomical Classification Challenge is hosted on Kaggle.⁶ This site has been used previously for other astronomical classifier challenges (T. Kitching et al. 2011; D. Harvey et al. 2014; R. Kessler et al. 2019), providing the opportunity for

⁵Sindarin Dictionary – <https://www.ambar-eldaron.com/english/downloads/sindarin-english.pdf>

⁶<https://www.kaggle.com/>

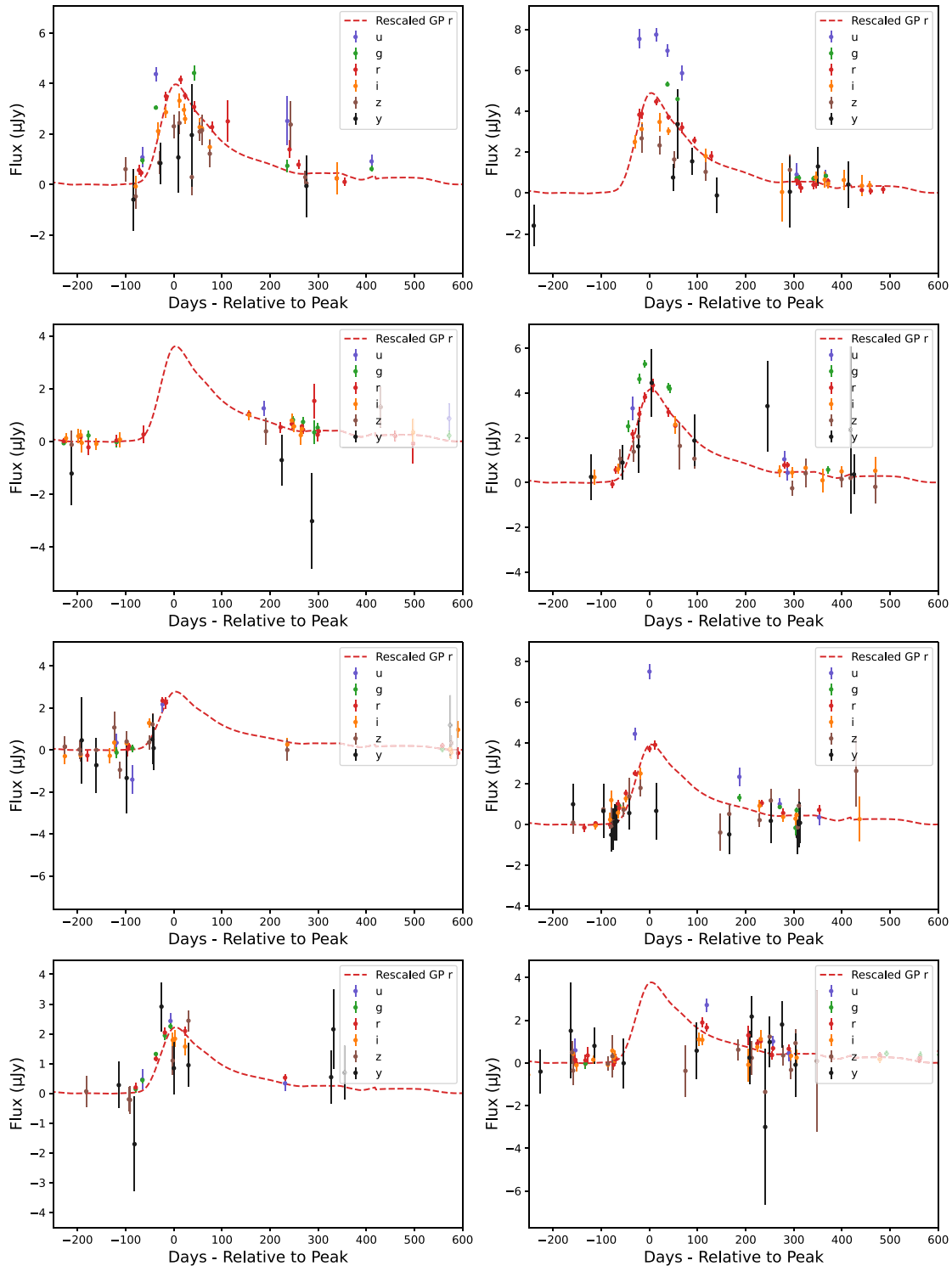


Figure 8. Eight simulated LSST light curves produced from the same real ZTF TDE (ZTF21abcgnqn). The LSST cadence and the scatter parameters used to generate the light curves provide sufficient variety to allow for multiple distinct light curves to be created from the same ZTF object. The GP fit to the r band, rescaled and extinguished at the same values as the points, is included as a dashed red line to show the continuous light curve of the transient.

both astronomers and members of the general public to participate in producing and testing the effectiveness of classifiers.

For the purposes of the classifier challenge the data set is split into training set, public testing set, and private testing set files. For the training set, the spectral type and true redshift are known,

thereby providing all the information that would be required to train a machine learning classifier (or other alternative) on that portion of the MALLORN data. This comprises 30 per cent of the overall data. The remaining 70 per cent is the testing set, on which the responses to the classifier challenge will be judged. For

this portion of the data, the user will not know the spectral type and only an estimated photometric redshift is provided. The testing set is split into two portions – public and private. For the public testing set the user will be able to see the overall performance of their classifier when they submit, therefore allowing them to make changes to improve their classifier and resubmit another response to the classifier challenge. For the private testing set, the user’s performance is not shown to them. The purpose of this is to prevent a user from overfitting by iteratively training to improve their score only on the public test set. The inclusion of a private testing set keeps the classifier challenge fair and helps to ensure that the classifiers tested on it will be effective at classifying actual LSST TDEs and not just fine-tuned to spot the TDEs included in the MALLORN public testing data set.

The light-curve files contain information on the object ID, date (MJD), flux (μJy), flux error (μJy), and filter. Multiple objects are assigned to one light-curve file. There are training set and testing set log files which contain information on the object ID, redshift, redshift error, extinction coefficient $E(B - V)$, spectral type, and an English translation providing the meaning of the object ID. The training set log is set to have spectroscopically determined redshifts, and as such has the actual simulated redshift value and no error value. Additionally, the spectral type is available for the training set. For the testing sets, the photometrically determined redshifts and their corresponding error values are used and the spectral type is not included. The $E(B - V)$ value is included to allow for the light curves to be de-extincted easily by the user. It was chosen to include these values directly rather than the RA and Dec to simplify the process of handling the data for non-astronomers who may wish to participate in the MALLORN classifier challenge. There are additional master log files detailing other parameters generated in production which are kept private.

To further lower the minimum barrier to entry for the classifier challenge, the Using MALLORN Data Google Colab Notebook⁷ details the process of loading in and de-extincting a light curve from the files and loading the relevant information for that light curve from the log files. The purpose of this notebook is primarily to lower the barrier of entry for non-astronomers and better enable their participation in the classifier challenge. This is of particular importance as there may be many who are talented in machine learning and computer science who could otherwise be put off from the challenge by its inaccessibility. Of the top three best performing entries to PLASTiCC, only the top entry, detailed in K. Boone (2019), was from an astronomer, with the other top performing entries from members of the general public (R. Hložek et al. 2023). We aim to make this classifier challenge as accessible as possible to maximize the potential gains in classifier performance prior to the start of LSST and therefore bolster our ability to identify TDEs with LSST.

Users are welcome to identify other types of transient within the data (and some participants may be specifically interested in learning how to find other classes in LSST), but the final score in the classifier challenge will only be based on the ability of the classifier to identify TDEs within the MALLORN data set. Responses will be evaluated using the F1 score (D. M. W. Powers 2011). This is a common machine learning evaluation metric which assesses the precision and recall scores of a model. Precision is the accuracy of the model, evaluated based on the ratio

of true positive values retrieved and the total number of positive values retrieved. Recall is the sensitivity of the model, determined from the ratio between the number of correctly predicted positive values and the total number of positive values in the data set. The F1 score itself can be interpreted as the harmonic mean of these two values, with an F1 score of 1 corresponding to the best possible performance and a score of 0 to the worst. The precision and recall values have equal weighting in the calculation of the F1 score. The F1 score is calculated using the following formula:

$$\text{F1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}} \quad (8)$$

In the above equation TP is the number of true positives (correct positive predictions), FN is the number of false negatives (incorrect negative predictions) and FP is the number of false positives (incorrect positive predictions). The competitor submitted classifications will be compared against the known types from the data production log files to determine the F1 score. Using the F1 score to select the most successful model disfavours models that overpredict the amount of TDEs in order to inflate their completeness. Whichever competitor achieves the highest F1 score will be determined to be the winner of the MALLORN Classification Challenge. The MALLORN Astronomical Classification Challenge was live as of the time of writing, having launched on 2025 October 15, and closed on 2026 January 30.

The competitor who produces the highest scoring classifier will win a €1000 prize. Whilst second and third place will receive €250 and €100, respectively. In addition to the monetary reward the best performing competitors will be invited to add their classifier to a platform which is being developed to host the best TDE classifiers. The benefit of this being that their classifier, along with others, will contribute to and receive credit for many future TDE discoveries.

5 DISCUSSION AND CONCLUSIONS

Having produced our simulated data set of $\sim 10\,000$ LSST transients, we now assess what their light curves imply for transient science, and particularly TDE science. During the creation of the MALLORN data set it became clear that some LSST bands will be much more useful for transient discovery than others, largely due to the differences in detection limits and the cadence strategy chosen. The g , r , and i bands will be the most useful, with frequent observations allowing for effective analysis of time evolution and detection limits which are capable of providing good signal-to-noise detections of fainter transients.

The u band has a sufficient detection limit to provide solid detections for fainter transients. This band is expected to be particularly important for the photometric identification of TDEs, which are inherently blue (S. Gezari 2021). Supernovae have been shown to peak at much redder wavelengths with increasing redshift, whereas TDEs continue to peak at blue wavelengths with increasing redshift (S. Velzen et al. 2021). Therefore, these u band observations will likely play a key role in the photometric identification of TDEs. The utility of this band is however limited by its less-frequent use in the LSST cadence.

For the z and y bands, the combination of infrequent sampling and shallower detection limits significantly restrict their utility for transient discovery and likely for photometric classification. This is particularly prominent for y band observations of fainter transients, which are much noisier than in other bands (see e.g. Figs 5 and 7). For much brighter objects, these filters will likely

⁷<https://colab.research.google.com/drive/1N7Q1bxc2gxBuOv2eD3fTYsrxoLC2dQAP>

provide good IR coverage of objects, but they will be of very limited use in the discovery and photometric classification of fainter transients.

We analyse the likelihood of detecting TDEs in LSST using our MALLORN data set and the metrics outlined in K. B. Bricman et al. (2023). The `some_color` metric requires at least one detection pre-peak ($t < t_{\text{peak}} - 10$ d), at least three detections in at least three different filters within 10 days of peak, and at least two detections in at least two different filters 10–30 days after peak. The `some_color_pu` (‘plus u ’) metric requires at least one detection pre-peak ($t < t_{\text{peak}} - 10$ d), at least one detection in the u band and at least one detection in another filter within 10 d of the peak and at least one detection in another filter 10–30 d after peak. These criteria give a rough estimate of whether a TDE can be selected photometrically, and whether there is enough information to extract physical insight. A detection within 10 d of the peak is required for a reliable estimate of the black hole mass (K. B. Bricman et al. 2023). Detections within 10–30 d after peak are necessary to estimate the fallback rate and hence the mass and impact parameter of the star which produced the TDE (J. Guillochon & E. Ramirez-Ruiz 2013).

Using a simulated population of TDEs [1000 at varying redshifts, based on a ‘normal’ TDE, PS1-10jh (S. Gezari et al. 2012), and a ‘fast and faint’ TDE, iPTF16fnl (N. Blagorodnova et al. 2017)] and `Baseline v1.5`, K. B. Bricman et al. (2023) evaluated how many TDEs would pass these detection metrics in LSST. 7.5 per cent of TDEs were found to pass the `some_color` metric and 1.5 per cent were found to pass the `some_color_pu` metric. These results were found to be heavily dependent on redshift, with mainly low-redshift TDEs passing the metrics, particularly in the case of the ‘fast and faint’ TDEs.

We recalculate these metrics for MALLORN to update the predictions for TDE performance, as our simulations included a much more representative set of TDEs and the latest LSST cadence. Defining a successful detection as a 5σ observation, 6.5 per cent of the TDEs in MALLORN pass the `some_color` metric whilst 2.7 per cent pass the `some_color_pu` metric. The results are broadly similar, with the updated version of the baseline accounting for some of the discrepancy. Additionally, the measures taken to ensure a minimum level of quality in MALLORN may slightly inflate the amount of TDEs which pass the metric in comparison to K. B. Bricman et al. (2023). However, it is worth noting that our simulation largely ignores low-redshift TDEs due to how we rescale the ZTF objects. Including more TDEs at $z \ll 0.1$ (though a small fraction of LSST TDEs) would somewhat increase these metrics. A correlation between redshift and metric performance was also observed in MALLORN, as shown in Fig. 9. The fraction of TDEs which pass both metrics decreases with increasing redshift, particularly in the case of the `some_color_pu` metric, which has no successes at $z \gtrsim 1$.

There are various types of transients (such as kilonovae and SNe Ibn) not included within the MALLORN data set, as they were not present in the ZTF nuclear transient data set that was used to generate it. Given that these types of transients did not appear in the extensive ZTF sample, they are clearly rare in galaxy nuclei and so we do not expect them to be a significant contaminant in the search for TDEs. Ambiguous Nuclear Transients (ANTs; P. Wiseman et al. 2025) are also not included as a class in the MALLORN data set, due to their rarity. However, our data set does include some TDEs which occurred in AGN.

There are some limitations of the MALLORN data set that could be improved upon if this method was to be repeated for

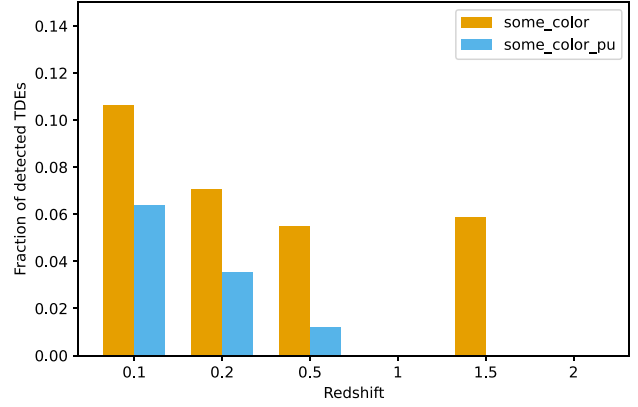


Figure 9. Bar chart showing the fraction of TDEs in MALLORN to pass the `some_color` and `some_color_pu` metrics from K. B. Bricman et al. (2023). The fraction of TDEs able to pass both metrics is shown to decrease with increasing redshift, particularly in the case of the `some_color_pu` metric.

a future survey. Many of the `SNCosmo` models lack good SED model coverage of the i , z , and y bands. By carefully selecting models we were able to create reasonable colour curves and k -corrections, but improved coverage in redder bands (which LSST will provide) would allow for this to be more accurate in future. Bright and nearby objects that LSST observes will likely have additional behaviour prior to the rise visible in the ZTF data and potentially more activity after the light curve fades to the ZTF detection limit. The MALLORN data set focused on being grounded in real observations, and consequently we avoided extrapolating to produce the simulated data required to produce these light curves. We believe that these brighter transients should be easier to classify, whereas the fainter objects we are focusing on will be more difficult. Therefore our data set will be more challenging for current classifiers and provide a greater avenue for improvement in advance of the start of LSST.

The ZTF data used in the creation of MALLORN is almost unbiased above $m = 18$, as the Bright Transient Survey (BTS; C. Fremling et al. 2020) will have obtained spectra for all transients brighter than this. However, for objects fainter than this, there is a bias in that bluer objects are more likely to have a spectroscopic classification (and therefore make it into our training set), resulting from the selection criteria for additional observations of TDE candidates defined in Y. Yao et al. (2023). The practical implication of this is to slightly increase the fraction of events in the MALLORN data set that are TDEs, or that have colours at peak similar to TDEs. Since the goal of the challenge is to improve TDE photometric classifiers, the inclusion of more TDEs and likely TDE contaminants is unlikely to have too negative an impact on this science.

The selection of temperature at random rather than matching the observed temperatures for all chosen objects may result in an inaccurate distribution of temperature in the TDEs generated in MALLORN. This would be the case if there is a strong correlation between temperature and luminosity, in which case we may inaccurately simulate some cooler distant TDEs that would be harder to classify. There is some evidence for a correlation between luminosity and temperature, but with a very large intrinsic scatter that suggests this effect will be modest within our data set (Y. Yao et al. 2023). Despite this potential issue, this approach was still necessary in order to produce multiple simulated objects from

one source object. The linear temperature change over time is not randomly selected, the corresponding value for the nearest peak temperature is selected to ensure physically reasonable temperature distributions.

The class imbalances in time-domain transient astronomy present some inherent challenges, which may make classification difficult – especially for some traditional computer science approaches. A simulated data set grounded in observational reality helps to overcome this issue, and is therefore beneficial to both fields, providing a strong example of the benefit such interdisciplinary approaches can create.

The MALLORN data set and Astronomical Classification Challenge will allow us to better prepare in advance of LSST, and contribute to a more effective and immediate start to transient research once the survey begins. The method developed for the production of MALLORN is highly replicable. It is possible to use it to generate simulated data for any future photometric survey from a previous one, given the limiting magnitudes of both and an expected cadence. By publicly providing our method via Google Colab (and releasing on GitHub after the challenge) we expect that this could be useful to prepare for other future surveys, such as the *Roman* High-Latitude Time-Domain Survey.

DATA AVAILABILITY

All the data produced as part of this work has been made publicly available via Kaggle (<https://www.kaggle.com/competitions/mallorn-astronomical-classification-challenge/overview>).

The spectroscopic types of the private data set will remain hidden during the duration of the data challenge and will be made available upon its completion. Google Colab notebooks have been made available that describe how the data was produced (<https://colab.research.google.com/drive/1oy96r29Zs4U5HI-THsZPCnOQBuz21h15>) and how to use the data (<https://colab.research.google.com/drive/1N7Q1bxc2gxBuOv2eD3FTYsrxoLC2dQAP>).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

Thanks to M. Kowalski and S. Reusch for running the AMPEL Nuclear Filter and creating the ZTF nuclear transient data set used to create the MALLORN data set. DM acknowledges a studentship funded by the Leverhulme Interdisciplinary Network on Algorithmic Solutions. MN is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 948381). We thank the anonymous referees for providing comments on the manuscript.

REFERENCES

- Barbary K., 2016, *extinction v0.3.0*, Software, Zenodo
- Barbary K. et al., 2025, *SNCosmo*, Software, Zenodo
- Bellm E. C. et al., 2019, *PASP*, 131, 018002
- Benidis K. et al., 2022, *ACM Comput. Surv.*, 55, 6
- Blagorodnova N. et al., 2017, *ApJ*, 844, 46
- Blandford R. D., McKee C. F., 1982, *ApJ*, 255, 419
- Boone K., 2019, *AJ*, 158, 257
- Bricman K., Gomboc A., 2020, *ApJ*, 890, 73
- Bricman K. B., van Velzen S., Nicholl M., Gomboc A., 2023, *ApJS*, 268, 13
- Cardelli J. A., Clayton G. C., Mathis J. S., 1989, *ApJ*, 345, 245
- Dgany Y., Arcavi I., Makrygianni L., Pellegrino C., Howell D. A., 2023, *ApJ*, 957, 57
- Dieleman S., Willett K. W., Dambre J., 2015, *MNRAS*, 450, 1441
- Fausnaugh M. M. et al., 2016, *ApJ*, 821, 56
- Fitzpatrick E. L., 1999, *PASP*, 111, 63
- Flesch E. W., 2023, *Open J. Astrophys.*, 6, 49
- Foreman-Mackey D., Agol E., Ambikasaran S., Angus R., 2017, *AJ*, 154, 220
- Fremling C. et al., 2020, *ApJ*, 895, 32
- French K. D., Wevers T., Law-Smith J., Graur O., Zabludoff A. I., 2020, *Space Sci. Rev.*, 216, 32
- Frohmaier C. et al., 2025, *ApJ*, 992, 158
- Gal-Yam A., 2019, *ARA&A*, 32, 58
- Gezari S., 2021, *Annu. Rev. Astron. Astrophys.*, 37, 40
- Gezari S. et al., 2012, *Nature*, 485, 217
- Gomez S., Berger E., Blanchard P. K., Hosseinzadeh G., Nicholl M., Villar V. A., Yin Y., 2020, *ApJ*, 904, 74
- Gomez S., Villar V. A., Berger E., Gezari S., van Velzen S., Nicholl M., Blanchard P. K., Alexander K. D., 2023, *ApJ*, 949, 113
- Gomez S. et al., 2024, *MNRAS*, 535, 471
- Graham M. L., Connolly A. J., Ivezić Ž., Schmidt S. J., Jones R. L., Jurić M., Daniel S. F., Yoachim P., 2018, *AJ*, 155, 1
- Green G. M., 2018, *J. Open Source Softw.*, 3, 695
- Guillochon J., Ramirez-Ruiz E., 2013, *ApJ*, 767, 25
- Guolo M. et al., 2025, preprint ([arXiv:2510.26774](https://arxiv.org/abs/2510.26774))
- Hammerstein E. et al., 2021, *ApJ*, 908, L20
- Harvey D., Kitching T. D., Noah-Vanhoucke J., Hamner B., Salimans T., Pires A. M., 2014, *Astron. Comput.*, 5, 35
- Hastie T., Tibshirani R., Friedman J., 2009, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edn.. Springer, Berlin
- Hills J. G., 1975, *Nature*, 254, 295
- Hinshaw G. et al., 2013, *ApJS*, 208, 19
- Hložek R. et al., 2023, *ApJS*, 267, 25
- Hsiao E. Y., Conley A., Howell D. A., Sullivan M., Pritchett C. J., Carlberg R. G., Nugent P. E., Phillips M. M., 2007, *ApJ*, 663, 1187
- Ivezić Ž., MacLeod C., 2014, in Mickaelian A. M., Sanders D. B., eds, *Proc. IAU Symp. 304, Multiwavelength AGN Surveys and Studies*. Cambridge Univ. Press, Cambridge, p. 395
- Ivezić Ž. et al., 2019, *ApJ*, 873, 111
- Jha S. W., 2017, in Alsabti A. W., Murdin P., eds, *Handbook of Supernovae*. Springer, Cham, p. 375
- Jones E., Do T., Boscoe B., Singal J., Wan Y., Nguyen Z., 2024, *ApJ*, 964, 130
- Karmen M. et al., 2025, *ApJ*, 990, 149
- Kelly B. C., Bechtold J., Siemiginowska A., 2009, *ApJ*, 698, 895
- Kessler R. et al., 2009, SNANA: A Public Software Package for Supernova Analysis, Technical Report. Publ. Astron. Soc. Pac., available at: <http://www.darkenergysurvey.org>
- Kessler R. et al., 2010, *PASP*, 122, 1415
- Kessler R. et al., 2019, *PASP*, 131, 094501
- Kitching T. et al., 2011, *Ann. Appl. Stat.*, 5, 2231
- Kornilov M. V., Semenikhin T. A., Pruzhinskaya M. V., 2023, *MNRAS*, 526, 1822
- Krabbendam V. L., Heathcote S., Schumacher G., Schwarz H. E., Sebring T. A., Warner M., 2004, in Oschmann J. M., Jr, ed., *Proc. SPIE Conf. Ser. Vol. 5489, Ground-based Telescopes*. SPIE, Bellingham, p. 615
- Lidskii V. V., Ozernoi L. M., 1979, *Sov. Astron. Lett.*, 5, 16
- Lokken M., Gagliano A., ELASTICC Team, 2023, American Astronomical Society Meeting Abstracts #241, p. 358.04
- MacLeod C. L. et al., 2010, *ApJ*, 721, 1014
- Malz A. I. et al., 2019, *AJ*, 158, 171
- Mountain C. M., Gillett F. C., Oschmann J. M., 1998, in Stepp L. M., ed., *Proc. SPIE Conf. Ser. Vol. 3352, Advanced Technology Optical/IR Telescopes VI*. SPIE, Bellingham, p. 2

- Netzer H., 2022, *MNRAS*, 509, 2637
- Nicholl M., Guillochon J., Berger E., 2017, *ApJ*, 850, 55
- Nordin J. et al., 2019, *A&A*, 631, A147
- Nugent P., Kim A., Perlmutter S., 2002, *PASP*, 114, 803
- Pavez-Herrera M. et al., 2025, *A&A*, 696, A153
- Pessi P. J. et al., 2025, *A&A*, 695, A142
- Peterson B. M., 1993, *PASP*, 105, 247
- Powers D. M. W., 2011, *Internat. J. Mach. Learn. Tech.*, 2, 37
- Ramsden P., Nicholl M., Mcgee S. L., Mummery A., 2025, *MNRAS*, 541, 1218
- Reusch S., 2024, PhD thesis, Humboldt-Universität zu Berlin
- Rodrigo C., Solano E., 2020, in XIV.0 Scientific Meeting (virtual) of the Spanish Astronomical Society. p. 182
- SLAC National Accelerator Laboratory, NSF-DOE Vera C. Rubin Observatory, 2025, *The LSST Camera (LSSTCam)*
- Schipani P. et al., 2018, Technical Report, *SOXS: a wide band spectrograph to follow up transients*
- Schlafly E. F., Finkbeiner D. P., 2011, *ApJ*, 737, 103
- Shappee B. J. et al., 2014, *ApJ*, 788, 48
- Sheng X., Ross N., Nicholl M., 2022, *MNRAS*, 512, 5580
- Sheng X. et al., 2024, *MNRAS*, 531, 2474
- Stein R. et al., 2024, *ApJ*, 965, L14
- Stern D. et al., 2004, *ApJ*, 612, 690
- Stevance H. F., Lee A., 2023, *MNRAS*, 518, 5741
- Tolkien J. R. R., 1954, *The Lord of the Rings*. George Allen and Unwin, London
- Urry C. M., Padovani P., 1995, *PASP*, 107, 803
- van Velzen S. et al., 2011, *ApJ*, 741, 73
- van Velzen S. et al., 2019, *ApJ*, 872, 198
- van Velzen S. et al., 2021, *ApJ*, 908, 4
- Vincenzi M., Sullivan M., Firth R. E., Gutiérrez C. P., Frohmaier C., Smith M., Angus C., Nichol R. C., 2019, *MNRAS*, 489, 5802
- Wiseman P. et al., 2025, *MNRAS*, 537, 2024
- Yao Y. et al., 2023, *ApJ*, 955, L6
- Yoachim P., 2025, *lsst-sims/sims_featureScheduler_runs5.0: Initial Release*, Zenodo
- Yoachim P. et al., 2025a, *lsst/rubin_sim: v2.2.4*, Zenodo
- Yoachim P., Jones L., Eric H. Neilsen J., MacBride S., Bechtol K., Becker M. R., Ross, Humna, 2025b, *lsst/rubin_scheduler: v3.12.0*, Zenodo
- Young D., 2023, *Sherlock. Contextual classification of astronomical transient sources*, Zenodo
- Yu W., Richards G. T., 2022, *Astrophysics Source Code Library*, record ascl:2201.001

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.