







Impact of Rubin Observatory Cadence Choices on Supernovae Photometric Classification

Catarina S. Alves¹ , Hiranya V. Peiris^{1,2} , Michelle Lochner^{3,4} , Jason D. McEwen⁵, Richard Kessler^{6,7} , and
The LSST Dark Energy Science Collaboration

¹ Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK; catarina.alves.18@ucl.ac.uk

² Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, AlbaNova University Center, Stockholm SE-10691, Sweden

³ Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town, 7535, South Africa

⁴ South African Radio Astronomy Observatory, 2 Fir Street, Black River Park, Observatory, 7925, South Africa

⁵ Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK

⁶ Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

⁷ Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA

Received 2022 October 30; revised 2023 February 4; accepted 2023 February 6; published 2023 March 23

Abstract

The Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST) will discover an unprecedented number of supernovae (SNe), making spectroscopic classification for all the events infeasible. LSST will thus rely on photometric classification, whose accuracy depends on the not-yet-finalized LSST observing strategy. In this work, we analyze the impact of cadence choices on classification performance using simulated multiband light curves. First, we simulate SNe with an LSST baseline cadence, a nonrolling cadence, and a presto-color cadence, which observes each sky location three times per night instead of twice. Each simulated data set includes a spectroscopically confirmed training set, which we augment to be representative of the test set as part of the classification pipeline. Then we use the photometric transient classification library `snmachine` to build classifiers. We find that the active region of the rolling cadence used in the baseline observing strategy yields a 25% improvement in classification performance relative to the background region. This improvement in performance in the actively rolling region is also associated with an increase of up to a factor of 2.7 in the number of cosmologically useful Type Ia SNe relative to the background region. However, adding a third visit per night as implemented in presto-color degrades classification performance due to more irregularly sampled light curves. Overall, our results establish desiderata on the observing cadence related to classification of full SNe light curves, which in turn impacts photometric SNe cosmology with LSST.

Unified Astronomy Thesaurus concepts: [Cosmology \(343\)](#); [Supernovae \(1668\)](#); [Astronomy software \(1855\)](#); [Open source software \(1866\)](#); [Astronomy data analysis \(1858\)](#); [Light curve classification \(1954\)](#); [Classification \(1907\)](#)

1. Introduction

Supernovae (SNe) are used for diverse astrophysical and cosmological studies, such as measurements of the universe’s accelerated expansion (e.g., Perlmutter et al. 1995; Riess et al. 1998; Astier et al. 2006; Kessler et al. 2009a; Betoule et al. 2014; Scolnic et al. 2018a; Abbott et al. 2019; Brout et al. 2022). For most cosmological analyses, SNe were spectroscopically classified to ensure a pure Type Ia sample, but this will be impossible for the large SNe sample expected from the Vera C. Rubin Observatory’s Legacy Survey of Space and Time (LSST; LSST Science Collaboration et al. 2009, 2017; Ivezić et al. 2019). Thus, LSST will rely on photometric classification, utilizing spectroscopically confirmed SNe samples to train classifiers.

The Supernova Photometric Classification Challenge (Kessler et al. 2010a) and the Photometric LSST Astronomical Time-Series Classification Challenge⁸ (PLAsTiCC; PLAsTiCC team et al. 2018; Kessler et al. 2019) catalyzed the development of

photometric classifiers in preparation for the Dark Energy Survey (Dark Energy Survey Collaboration & Flaugher 2005) and LSST (LSST Science Collaboration et al. 2009, 2017; Ivezić et al. 2019), respectively. Many of the resulting recent classifiers rely on machine learning methods, such as neural networks (Charnock & Moss 2017; Muthukrishna et al. 2019; Möller & de Boissière 2020; Villar et al. 2020; Boone 2021; Qu & Sako 2022), boosted decision trees (Boone 2019; Alves et al. 2022), and self-attention mechanisms (Allam & McEwen 2021; Pimentel et al. 2022).

Accurate classification requires representative training sets; the feature-space distributions of the training set should be similar to those of the test set (e.g., Lochner et al. 2016). However, classifiers are usually trained with either simulated data sets, which may suffer from model misspecification, or spectroscopically confirmed events, which are nonrepresentative of the test set due to selection effects. Several methods have been proposed to address the second problem, predominantly based on data augmentation techniques (e.g., Revsbech et al. 2017; Boone 2019; Pasquet et al. 2019; Carrick et al. 2021). This previous work has demonstrated that the bias introduced by nonrepresentative training sets can be corrected.

Another crucial factor that impacts the accuracy of photometric classification is the survey observing strategy (Alves et al. 2022; Lochner et al. 2022). Over the course of 10 yr, LSST will

⁸ <https://www.kaggle.com/c/PLAsTiCC-2018/>

repeatedly observe the southern sky every few days in multiple passbands. Its observing strategy encompasses diverse aspects such as the survey footprint, season length, inter- and intranight gaps, cadence of repeat visits in different passbands, and exposure time per visit. Changes in how LSST observes the sky can improve the scientific output of the survey; however, observing strategy optimization is challenging due to the diverse goals of LSST (LSST Science Collaboration et al. 2009; Ivezić et al. 2019).

Recently, the Survey Cadence Optimization Committee (2022) Phase 1 report (hereafter: **Ph1R**) narrowed down the choice of possible observing strategies and recommended new simulations⁹ to respond to the findings of the previous optimization work (e.g., LSST Science Collaboration et al. 2017; Scolnic et al. 2018b; Gonzalez et al. 2018; Laine et al. 2018; Lochner et al. 2018; Olsen et al. 2018; Bianco et al. 2019; Jones et al. 2020; Alves et al. 2022; Lochner et al. 2022) and enable further optimization. In particular, it is not yet decided whether LSST will use a rolling cadence¹⁰ or whether it will visit each sky pointing two or three times per night.

In this work, we study the impact of these key observing strategy choices on photometric classification accuracy. We focus on the rolling cadence and the intranight observing strategy, since we expect these factors to have the greatest impact on the efficacy of light-curve classification.

Our work builds on Alves et al. (2022) by studying the performance of photometric SN classification for light curves simulated with different LSST observing strategies for the first 3 yr of the survey; we chose this time frame because early science drivers are one of the highest priorities for the next set of cadence decisions. First, we simulated multiband light curves using the SuperNova ANALYSIS package¹¹ (SNANA; Kessler et al. 2009b). These simulated data sets included a nonrepresentative spectroscopically confirmed training set that was biased toward brighter events. Next, we followed the classification approach of Alves et al. (2022), using the photometric transient classification library `snmachine`¹² (Lochner et al. 2016; Alves et al. 2022) to build a classifier based on wavelet features obtained from Gaussian process (GP) fits. We also included the host galaxy photometric redshifts and their uncertainties as features. The simulated training set was augmented to be representative of the photometric redshift distribution per SNe class, the cadence of observations, and the flux uncertainty distribution of the test set.

In Section 2 we describe the LSST observing strategies and the framework that we used to generate our SNe data sets. Our classification and augmentation methodologies that relied on `snmachine` are presented in Section 3. Section 4 focuses on our results and their implications for observing strategy. We conclude in Section 5.

2. Simulation of LSST SNe

2.1. Overview

In this work we simulated LSST-like SN light curves for the first 3 yr of the survey using three observing strategies:

⁹ The Jupyter Notebook in https://github.com/lst-pst/survey_strategy/blob/main/fbs_2.0/SummaryInfo_v2.1.ipynb provides a short summary and details of the observing strategies simulated.

¹⁰ In a rolling cadence strategy, LSST observes a part of the sky at a higher cadence than the rest. After a fixed period, usually 1 yr, the rolling moves to a different part of the sky.

¹¹ <https://snana.uchicago.edu/>

¹² <https://github.com/LSSTDESC/snmachine>

`baseline_v2.0`, `noroll_v2.0`, and `presto_gap2.5_mix_v2.0`. These observing strategies were created with the Feature-Based Scheduler (FBS; Naghib et al. 2019), which is the default scheduler for LSST. We then used the infrastructure developed for PLAsTiCC to simulate the light curves of the SNe with realistic sampling and noise properties (Kessler et al. 2019). We describe the observing strategies in Section 2.2, the SNe models in Section 2.3, and the simulations infrastructure in Section 2.4.

Following the PLAsTiCC team et al. (2018) and Kessler et al. (2019), we simulated the Wide-Fast-Deep (WFD) survey, which is the main survey of LSST (containing $\sim 98\%$ of our simulated events), and the Deep Drilling Fields (DDF) survey, which covers small patches of the sky with more frequent and deeper observations. The properties of each survey mode depend on the observing strategy, but since the release of PLAsTiCC the footprints of the DDFs have changed considerably and the three observing strategies simulated in this work used the DDF locations presented in Table 2 of Jones et al. (2020). Since their DDF sequence is generally the same, we focused our analysis on the implications of the observing strategy for the WFD survey. The DDF events are still included in the training set, as they improve our augmentation procedure (see Section 3.3), but are not included in the test set.

We used 0.2% of the simulations to construct a nonrepresentative spectroscopically confirmed training set. The training set was biased toward brighter events, with a median redshift ~ 0.3 . The relatively small training set mimics the available data from current and near-term spectroscopic surveys at the start of LSST science operations. Following Kessler et al. (2019), we loosely based the training set on the planned magnitude-limited 4 m Multi-Object Spectroscopic Telescope Time Domain Extragalactic Survey (Swann et al. 2019).

2.2. Observing Strategies

Rubin’s Survey Cadence Optimization Committee¹³ has been formed to make recommendations for the observing strategy with inputs from the community. Following their recommendations, a new set of LSST observing strategy simulations created with FBS was released to respond to the findings of previous optimizations (**Ph1R**), including an update of the LSST baseline observing strategy: `baseline_v2.0`. Under the updated baseline, the telescope observes each field twice with a gap of approximately 15 minutes during twilight and 33 minutes during the rest of the night; these visit pairs are in different passbands. In the extragalactic (i.e., dust-extinction limited) WFD, the sky is divided into two regions: an “active” area that is observed more often (rolling at 90% strength) and a “background” area. This two-band rolling cadence is defined by decl. and shown in Figure 1. In this observing strategy simulation, the telescope observes in a rolling cadence between the years 1.5 and 8.5 of the survey to ensure that the first and last years have uninterrupted coverage of the entire sky (**Ph1R**). In this work, we simulated the first 3 yr of the survey, and therefore only half of the light-curve observations were performed with the rolling cadence.

A key aim of the new LSST observing strategy simulations is to evaluate whether a rolling cadence is suitable, as demonstrated by science metrics (**Ph1R**). In this work, we

¹³ For further details see <https://www.lsst.org/content/charge-survey-cadence-optimization-committee-scoc>.

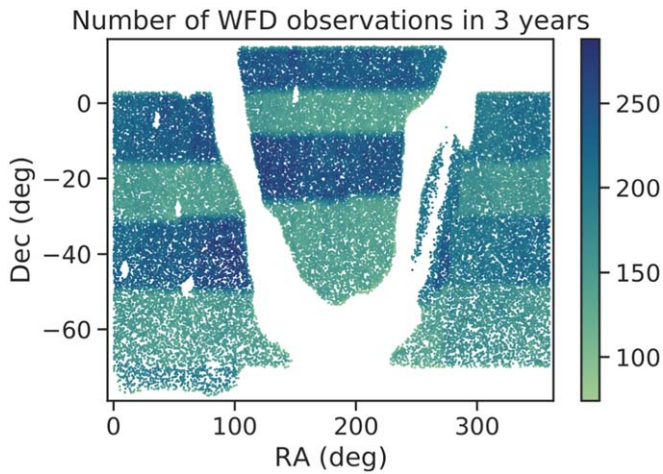


Figure 1. Footprint of the baseline cadence with the number of WFD observations in the first 3 yr (1.5 yr of nonrolling followed by 1.5 yr of rolling cadence). The dark bands correspond to the “active” area of the rolling cadence, which is observed at a higher cadence, and the light bands correspond to the “background.”

studied the impact of the rolling cadence on the photometric classification of SNe by comparing the baseline observing strategy with a similar strategy without the rolling (`no-roll_v2.0`, hereafter referred to as `no-roll`). Since the rolling starts in year 1.5 of the baseline simulation, we restricted this comparative analysis to the events observed between years 1.5 and 3 on both simulations. We refer to this subset of the baseline data set as `Y1.5-3 baseline`; we used `Y0-3 baseline` when considering the entire 3 yr. The simulations are otherwise identical.

Another aim of the new observing strategies is to investigate modifications of the intranight cadence. In this work, we studied the impact of adding a third visit per night in a passband that had been previously observed; this addition is motivated by expected improvements to the performance of early classification and fast transient detection. The `presto-color` family (Bianco et al. 2018, 2019) encompasses a number of variations of the third visit inclusion, such as different intranight gaps between the observations (e.g., 1.5 and 4 hr between the first pair of observations and the third), whether the initial pair of visits is in consecutive passbands ($g+r$, $r+i$ or $i+z$) or mixed passbands ($g+i$, $r+z$ or $i+y$), and whether to obtain the visit triplet every night or every other night (`Ph1R`). SNe do not vary significantly during a single night, so the difference in intranight gaps between 1.5 and 4 hr has minimal impact. We thus chose a `presto-color` cadence whose third visit has an intermediate value of 2.5 hr for the intranight gap (`presto_gap2.5_mix_v2.0`, hereafter referred to as `presto-color`). Since the total number of visits per pointing is fixed, adopting a `presto-color` cadence results in longer internight gaps; further, each field is observed for fewer nights in total. Similarly to baseline, the rolling starts in year 1.5 of the `presto-color` cadence; thus we compare baseline and `presto-color` for the entire first 3 yr of LSST. For more details on the simulations, see `Ph1R` and the descriptions in the associated Jupyter Notebook.¹⁴

¹⁴ https://github.com/lst-pst/survey_strategy/blob/191bfac9915fc8e7e363d457e119297320f3c591/fbs_2.0/SummaryInfo_v2.1.ipynb

Table 1
Breakdown of the Number of SNe per Class and Observing Strategy Used in This Work

| Y1.5-3 baseline | | |
|-----------------|---------------------------|-----------------------|
| SN class | N_{training} (%) | N_{test} (%) |
| SN Ia | 1738 (64%) | 755330 (62%) |
| SN Ibc | 231 (8%) | 57854 (5%) |
| SN II | 759 (28%) | 405139 (33%) |
| Total | 2728 (100%) | 1218323 (100%) |
| Y0-3 baseline | | |
| SN class | N_{training} (%) | N_{test} (%) |
| SN Ia | 3421 (65%) | 1563427 (62%) |
| SN Ibc | 460 (9%) | 119703 (5%) |
| SN II | 1392 (26%) | 818241 (33%) |
| Total | 5273 (100%) | 2501371 (100%) |
| No-roll | | |
| SN class | N_{training} (%) | N_{test} (%) |
| SN Ia | 1787 (63%) | 875487 (63%) |
| SN Ibc | 240 (9%) | 66389 (5%) |
| SN II | 801 (28%) | 445651 (32%) |
| Total | 2828 (100%) | 1387527 (100%) |
| Presto-color | | |
| SN class | N_{training} (%) | N_{test} (%) |
| SN Ia | 3243 (63%) | 1287360 (65%) |
| SN Ibc | 483 (9%) | 93776 (5%) |
| SN II | 1422 (28%) | 597310 (30%) |
| Total | 5148 (100%) | 1978446 (100%) |

Note. Left: events simulated between years 1.5 and 3 of the survey. Right: events simulated between years 0 and 3 of the survey.

2.3. SNe Models

Following Alves et al. (2022), we focused on classifying SN Ia, SN Ibc, and SN II, which have been found to be difficult transient classes to distinguish (Hložek et al. 2020). We simulated each class in a similar manner to Kessler et al. (2019), using models from Kessler et al. (2010b), Guy et al. (2010), Kessler et al. (2013), Villar et al. (2017), Guillochon et al. (2018), and Pielert et al. (2018). However, similarly to Lokken et al. (2023), we did not include the `SNIbc-MOSFiT` model because it produces unphysical light curves. We also adjusted the relative fraction of simulated core-collapse SNe (CC SNe) to follow Table 3 of Shivvers et al. (2017). Additionally, due to the lack of SN Iib models in Kessler et al. (2019), we redistributed their fraction among the other stripped envelope SNe (SN Ib and SN Ic); see Table 2 of Appendix A for the relative rates used to simulate CC SNe in this work. Table 1 shows the resulting number of SNe per class for each observing strategy.

2.4. Framework for Generating Simulations

Our SNe simulations were built on top of the observing strategy cadences produced by `FBS`¹⁵ previously discussed in

¹⁵ The observing strategies are hosted in https://epyc.astro.washington.edu/~lynnej/opsim_downloads/fbs_2.0/.

Section 2.2 (Naghieb et al. 2019); this scheduler decides the passband to use and the direction to point the telescope to using a Markovian decision process, while accounting for interruptions, such as telescope maintenance downtime. Despite the FBS outputs containing a record of each simulated pointing of the survey, for generating light curves it is more convenient to compute all the observations of each event and iterate over the events. Therefore, we used the python package `OpSimSummary`¹⁶ (Biswas et al. 2020, 2022) to reorder the observations. This package also translates the FBS output into the appropriate format for use with the SNe simulation code from SNANA (Kessler et al. 2009b), which we used to generate realistic light curves in the LSST passbands. We broadly followed the methodology described in Kessler et al. (2019), which relies on SNANA to generate simulated data sets of SNe and associated metadata (e.g., host galaxy photometric redshift and its uncertainty). SNANA uses models of the SN sources, observing conditions, observing strategy, and instrumental noise to generate light curves. Then it applies triggers to select the observations that would be seen by LSST. Following Kessler et al. (2019), we applied the SNANA transient trigger to keep only events with at least two detections in our data sets; SNANA uses the DES-SN detection model from Kessler et al. (2015) to decide which observations are flagged as detected. See Figure 13 of Kessler et al. (2019) for a summary of the SNANA simulation stages.

Following Kessler et al. (2019)’s usage of SNANA, we truncated the 10 yr survey to the first 3 yr, removed season fragments with less than 30 days, and used the cosmological parameters $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $w_0 = -1$, and $H_0 = 70$. However, we used an updated version of the code,¹⁷ which included improvements for the K-corrections for events at the highest simulated redshift. We made two further changes from Kessler et al. (2019) to improve the realism of our simulations, as follows. While Kessler et al. (2019) used a pixel-flux saturation of 3,900,000 photoelectrons pixel⁻¹, we used the more realistic value of 100,000 photoelectrons pixel⁻¹. We also corrected the code to ensure that any observations in the same band in a given night are coadded and count as a single observation. We provide our SNANA input files for each observing strategy simulation at doi:10.5281/zenodo.7552490.

3. Photometric Classification

We followed the approach of Alves et al. (2022) to photometrically classify the SNe simulated from each observing strategy. In that work, we benchmarked our classification approach against the winning PLAsTiCC entry (Boone 2019) and showed that our classification results were generalizable; they hold if we replace our classification predictions with the predictions of Boone (2019). Here we used the photometric transient classification library `snmachine` (Lochner et al. 2016; Alves et al. 2022) and updated it to handle the output files of SNANA (FITS files). In the sections below, we describe the main steps of the approach and any modifications relative to Alves et al. (2022).

3.1. Light-curve Preprocessing

Following Alves et al. (2022), we preprocessed the simulated light curves to only include the observing season in which the

SNe is detected. To isolate this season for each event, we removed all observations 50 days before the first detection and 50 days after the last. Next, we divided the remainder light curve into sequences of observations without internight gaps >50 days; we selected our preprocessed light curve as the sequence of observations that contained the largest number of detections. Finally, we translated the light curve so the first observation was at time zero. The longest resulting light curves, as measured between the first and last observations, lasted for 274, 253, and 295 days, respectively, for `baseline`, `no-roll`, and `presto-color`.

3.2. GP Modeling of Light Curves

We used GP regression (e.g., MacKay 2003; Rasmussen & Williams 2005) to model each light curve. Following Boone (2019) and Alves et al. (2022), we fitted two-dimensional GPs in time and wavelength; we applied a null mean function and a Matérn 3/2 kernel for the GP covariance. We fixed the length scale of the wavelength dimension to 6000 Å and used maximum likelihood estimation to optimize the time dimension length scale and amplitude per event. We implemented the GPs with the python package `George`¹⁸ (Ambikasaran et al. 2014). We note that Stevance & Lee (2022) investigated possible improvements to using GPs for SNe light-curve fitting. We leave these extensions to future work on SNe classification.

3.3. Augmentation

We applied the methodology developed in Alves et al. (2022) to augment the training set of each simulated observing strategy to be representative of its respective test set in terms of the photometric redshift distribution per SNe class, the cadence of observations, and the flux uncertainty distribution. We delineate below the departures from the augmentation procedure described in Section 4 of Alves et al. (2022) and defer the details to Appendix B.

We augmented the training-set SNe to generate synthetic events at a different redshift from the original; this approach relied on using two-dimensional GP models of the training-set events to generate the synthetic light curves. Since we removed an SN model (as mentioned in Section 2.3), the redshift distribution of the events changed with respect to Alves et al. (2022). Consequently, we used a different distribution to produce the augmented training sets, as detailed in Appendix B.

Following our previous work, we generated 15,440 WFD synthetic events for each SNe class. Figure 2 shows that for the Y0-3 `baseline`, the photometric redshift distribution of the augmented training set is closer to the test set than the original training set. Although the distribution does not match exactly, it is sufficiently close to expect minimal impact on performance. Ensuring identical distributions could require introducing an undesirable amount of fine-tuning to the methodology, and hence we have not attempted to obtain a closer match. Comparable figures for the other observing strategies are shown in Figure 14 of Appendix B.

We also tuned the distribution of the number of observations and their flux uncertainty for the synthetic events of each observing strategy. We drew the target number of observations for each light curve from a Gaussian mixture model based on

¹⁶ <https://github.com/LSSTDESC/OpSimSummary>

¹⁷ In this work we used SNANA version v11_04i.

¹⁸ george.readthedocs.io/

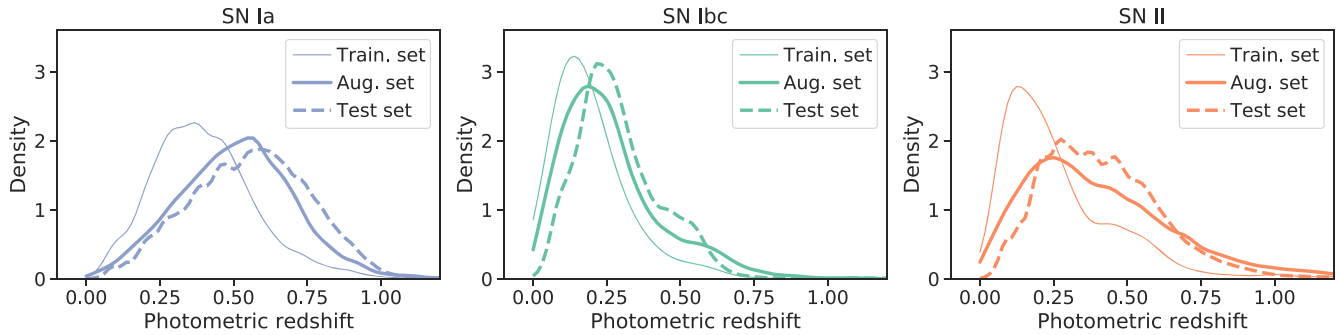


Figure 2. Host galaxy photometric redshift distribution per supernova class for $\Upsilon 0-3$ baseline, where SN Ia, SN Ibc, and SN II are shown, respectively, on the left, middle, and right panels. The training-set distribution (solid line) is not representative of the test set (dashed line), but the augmented training set (bold solid line) is close to the desired test distribution.

the test set; Table 3 of Appendix B shows the parameters used for each observing strategy. For the flux uncertainty, we followed Boone (2019) and Alves et al. (2022) and combined the uncertainty predicted by the GP in quadrature with a value drawn from the flux uncertainty distribution of the test set. Table 3 of Appendix B shows the parameters of the Gaussian mixture model used to fit the flux uncertainty distribution of each passband and observing strategy.

3.4. Feature Extraction

For photometric classification we used the host galaxy photometric redshift, its uncertainty, and model-independent wavelet coefficients obtained from the GP fits as features. The redshift features mentioned above were directly obtained from the metadata associated with each event. For the wavelet features, we followed the feature extraction procedure of Lochner et al. (2016) and Alves et al. (2022), which we briefly summarize in the following paragraph.

To perform a wavelet decomposition on the light curves we sampled them onto a time grid. We used the two-dimensional GP that models each light curve to interpolate between the observations. For uniformity, we used the same time grid for all the observing strategies; the time range of the grid corresponds to the maximum light-curve duration of the events, 295 days. Following Alves et al. (2022) we chose 292 grid points to sample the events approximately once per day. Next, we performed a two-level wavelet decomposition using a stationary wavelet transform and the `symlet` family of wavelets.¹⁹ These decomposition choices resulted in 7008 redundant wavelet coefficients per event. Following Lochner et al. (2016) and Alves et al. (2022), we reduced the dimensionality of the wavelet space to 40 components using principal component analysis (Pearson 1901; Hotelling 1933). We used the augmented training set of each observing strategy to construct the dimensionality-reduced wavelet space; the test-set events were projected onto the corresponding wavelet space.

3.5. Classification

We used `snmachine` to build a photometric classifier trained on the augmented training set. We used gradient boosting decision trees (GBDTs; Friedman 2002), classifiers whose predictions are based on ensembles of decision trees.

¹⁹ Using the `PyWavelets` (Lee et al. 2019a) package as part of `snmachine`.

We trained the classifier for each observing strategy separately, using dedicated augmented training sets (Section 3.3) and features (Section 3.4). The GBDT classifier hyperparameters were optimized following the procedure described in Section 3.4 of Alves et al. (2022); Table 5 of Appendix B shows the values of the hyperparameters per observing strategy.

3.5.1. Performance Evaluation

We used the PLAsTiCC weighted log-loss metric (PLAsTiCC team et al. 2018; Malz et al. 2019) to optimize the photometric classifiers and to evaluate their performance. Following the PLAsTiCC challenge, we gave the same weight to each SN class.

Confusion matrices are commonly used to assess the performance of classifiers (see, e.g., Hložek et al. 2020). To produce a confusion matrix, we first assigned each test-set event to its most probable class. For ease of comparison between different classes and observing strategies, we normalized the resulting confusion matrices by dividing each entry by the true number of SNe in each class. In this setting, a perfect classification results in the identity matrix.

We measured the classification performance using the recall (also called completeness/sensitivity) and precision of each SNe class. These are defined as

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

and

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (2)$$

where in a binary classification setting TP, FN, and FP are, respectively, the number of true positives, false negatives, and false positives.

The computational performance of this procedure and an estimate of the resources needed for reproducing this analysis are discussed in Appendix C.

4. Results and Implications for Observing Strategy

Here we present our results on the impact of rolling cadence and the intranight gap on SNe classification performance. We perform a comparative analysis for these two cases relative to the baseline strategy in Section 4.1. In our previous work (Alves et al. 2022) we found that light-curve length (time difference between the first and last observation after the light-curve preprocessing described in Section 3.1) and internight

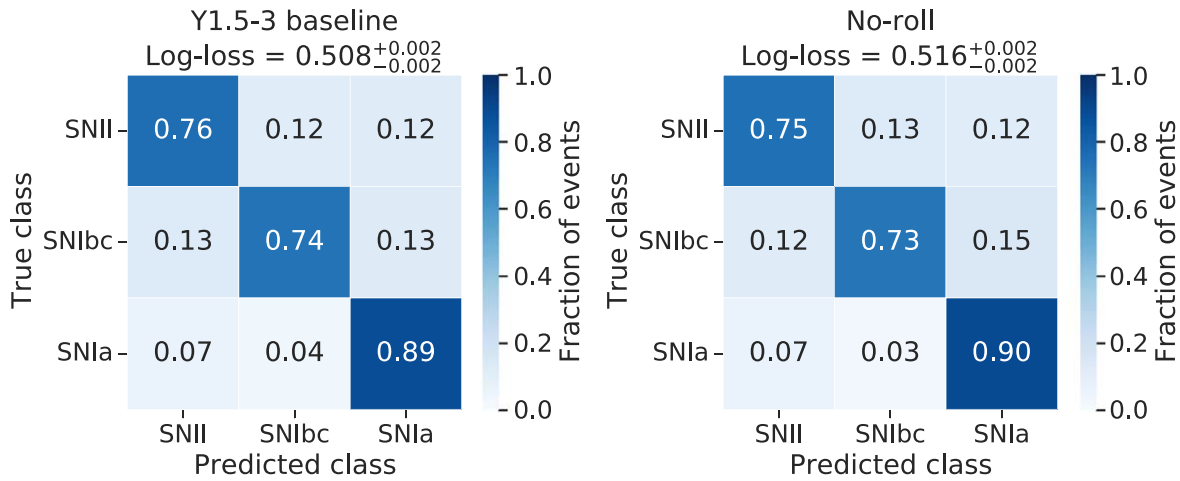


Figure 3. Normalized test-set confusion matrix for the classifier trained on the augmented training set of the Y1.5-3 baseline (left panel) and the no-roll cadences (right panel). The uncertainty in the log-loss corresponds to the 95% confidence intervals obtained by bootstrapping. The results show a slightly higher SNe classification performance when rolling is implemented at the level in the baseline cadence.

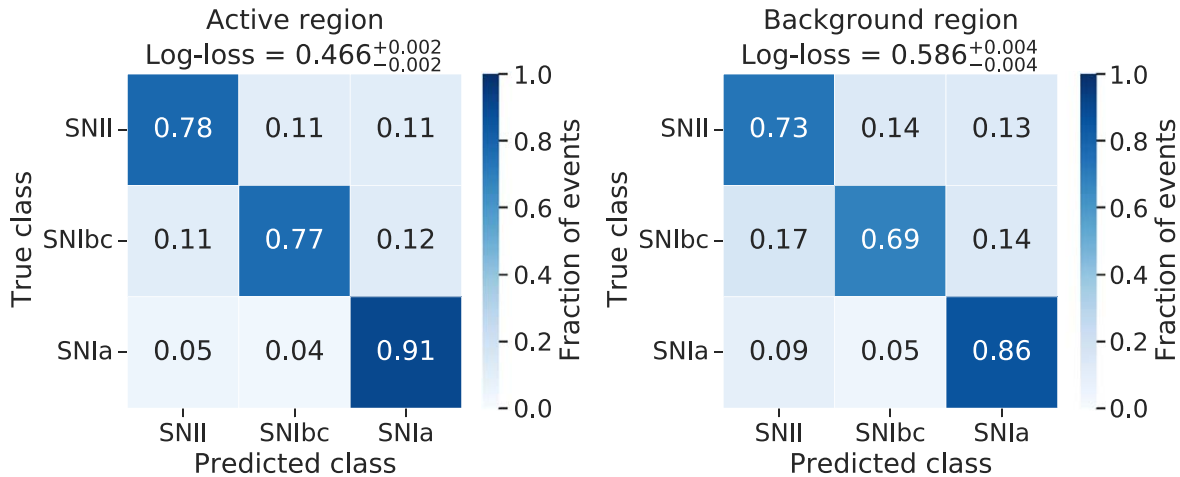


Figure 4. Normalized test-set confusion matrices for the classifier trained on the augmented training set of the Y1.5-3 baseline cadence. The left panel shows the results for the active region of the rolling cadence, and the right panel, for the background region. The uncertainty in the log-loss corresponds to the 95% confidence intervals obtained by bootstrapping. The results show a significantly higher SNe classification performance for the active region of the rolling cadence.

gap (time difference between consecutive observations that are more than 12 hr apart) were the key properties of observing strategies affecting classification performance. We thus investigate how the no-roll and presto-color families affect the recall and precision as a function of several factors including the above.

4.1. Overall Classification Performance

Figure 3 shows the confusion matrices for classifiers trained on the augmented training set of the Y1.5-3 baseline and no-roll. The Y1.5-3 baseline classifier yields a slightly higher performance for SN Ibc and SN II and a percent-level improvement in the PLAsTiCC log-loss metric. This small difference indicates that rolling at this level makes a negligible difference in the overall efficacy of SNe photometric classification. However, this result masks a significant difference between the classification efficacy between the active and background regions due to an averaging effect. Therefore, we also investigated the difference in performance between the active region (which we visually identified as the dark bands in Figure 1; 65% of the test-set events) and the background region (35% of the test-set events). The confusion

matrices in Figure 4 show that the classification performance of the active region is higher than of the background region for all SNe classes. Indeed the log-loss metric improves by 25% for events in the active region compared with the background.

Figure 5 shows the confusion matrices for Y0-3 baseline and presto-color. The baseline cadence outperforms presto-color for SN Ia and SN II; the PLAsTiCC log-loss metric degrades by $\sim 10\%$ for presto-color. While adding a third visit per night is expected to improve performance for early classification and for fast transient detection, our results indicate that this choice moderately degrades classification performance for long-lived transients.

All the observing strategies considered in this work yield a higher performance in terms of the log-loss metric compared with the observing strategy used for PLAsTiCC (Alves et al. 2022 reported a log-loss metric of 0.550 for this case). This indicates substantial performance gains achieved by recent updates to the FBS scheduler.

4.2. Light-curve Length

We found in Alves et al. (2022) that the light-curve length of an event has a significant impact on classification performance

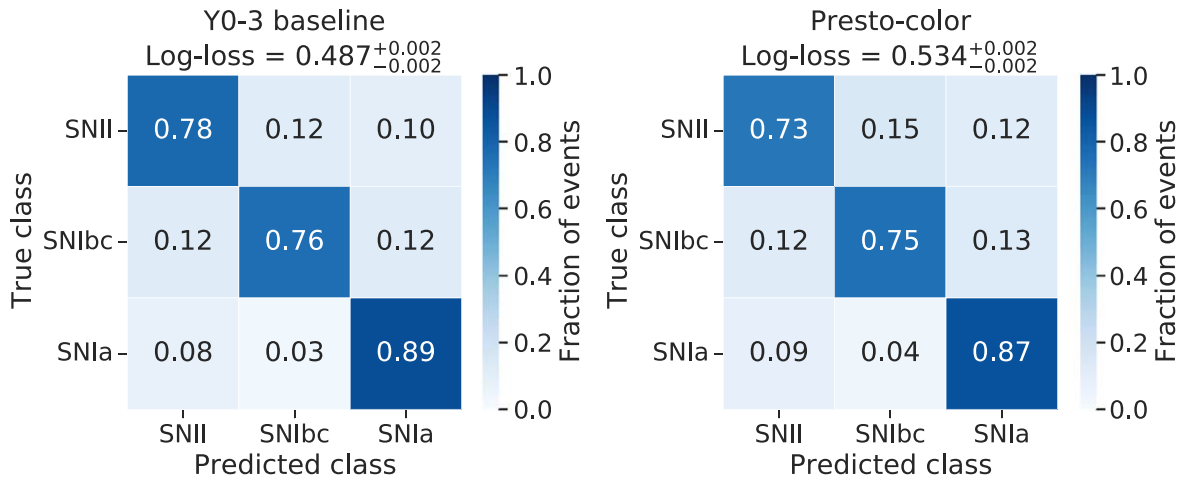


Figure 5. Normalized test-set confusion matrix for the classifier trained on the augmented training set of the Y0-3 baseline (left panel) and the presto-color cadences (right panel). The uncertainty in the log-loss corresponds to the 95% confidence intervals obtained by bootstrapping. The results show that visiting each event twice per night (baseline cadence) instead of three times yields a $\sim 10\%$ higher SNe classification performance.

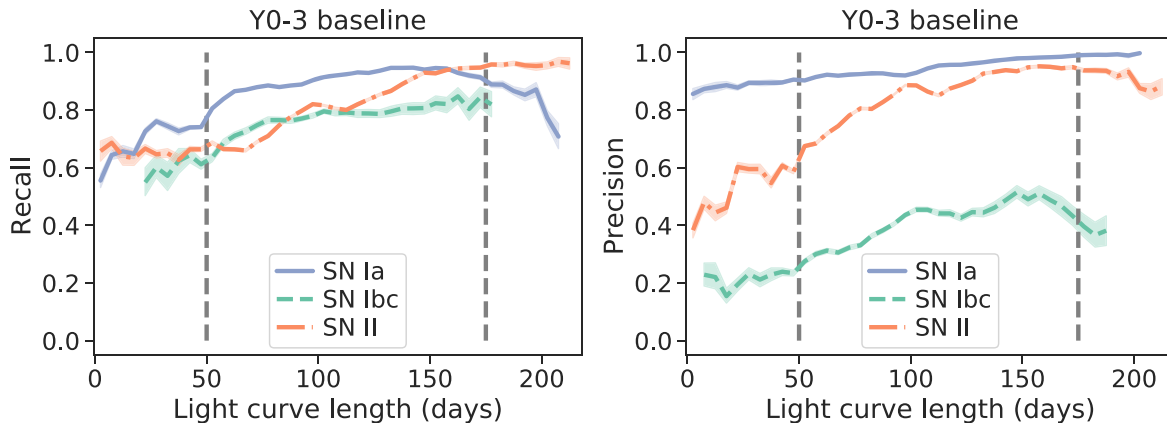


Figure 6. Test-set recall (left panel) and precision (right panel) as a function of light-curve length per SNe class for the Y0-3 baseline. The shaded areas correspond to the 95% confidence limits obtained by bootstrapping the recall and precision values for each bin. The dashed lines mark the high performance region between 50 and 175 days. To remove small-number effects we present only the results for bins with more than 300 events.

for long-lived transients such as SNe. In particular, longer light curves are easier to distinguish within a classifier since they incorporate more information about the time evolution of the event. In that work, we focused on events with light-curve length between 50 and 175 days due to their higher performance. As shown in Figure 6, our results for the Y0-3 baseline show a similar performance behavior. We also find that these conclusions generalize to the other observing strategies analyzed here, and hence the conclusions of Alves et al. (2022) carry over to these new cadence simulations. We note that the recall and precision figures (Figure 6 and subsequent figures) show a small scatter above our statistical uncertainties, likely arising from the limited diversity of the simulations in those particular bins. Figure 7 shows that the distribution of light-curve lengths is similar for all the cadences. Indeed, the cadence choices currently under consideration (v2.0) have similar distributions of gaps larger than 50 days, so the light-curve length distribution correlates more with the intrinsic duration of the events and our preprocessing of the light curves than with the cadences. Therefore, even though this is a very important factor for

overall classification performance, it is not strongly affected by observing strategy choices.

4.3. Median Internight Gaps

The observing strategies proposed for LSST have different intra- and internight gaps distributions. Given the finite total number of observations available, these distributions are intrinsically linked. In Alves et al. (2022), we demonstrated that the median internight gap was a crucial factor in photometric SNe classification, and Ph1R has highlighted the necessity for science-motivated metrics to measure the impact of intranight gaps. Since the timescale for changes in SN light curves is in days, multiple observations in a single night in the same filter do not contribute toward characterization of the light curve. Here we first investigate the impact of a higher intranight cadence on the median internight gap and hence classification performance.

Figure 8 shows that, in accordance with our expectations, a lower median internight gap leads to higher precision and recall for the Y0-3 baseline. SNe II show a larger sensitivity to cadence as compared with the other types because SNe Ia with

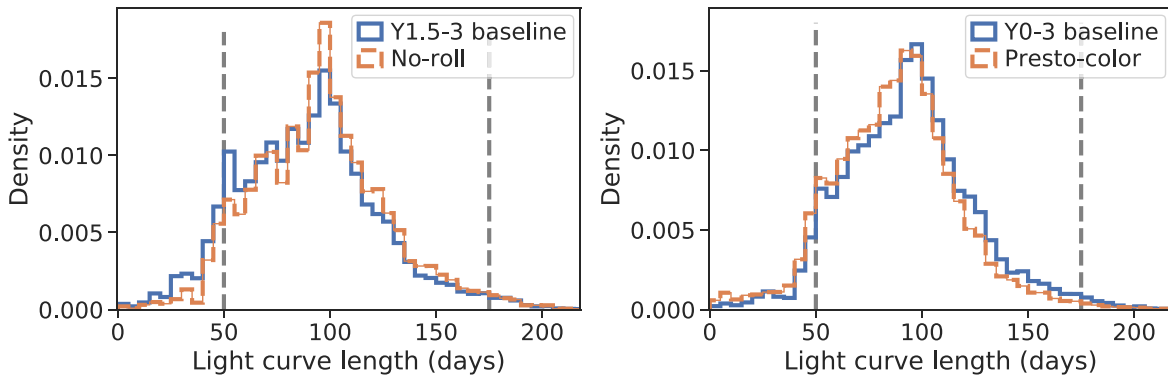


Figure 7. Test-set density of events as a function of light-curve length for the Y1.5-3 baseline versus no-roll (left panel) and Y0-3 baseline versus presto-color (right panel). The dashed lines mark the high performance region between 50 and 175 days.

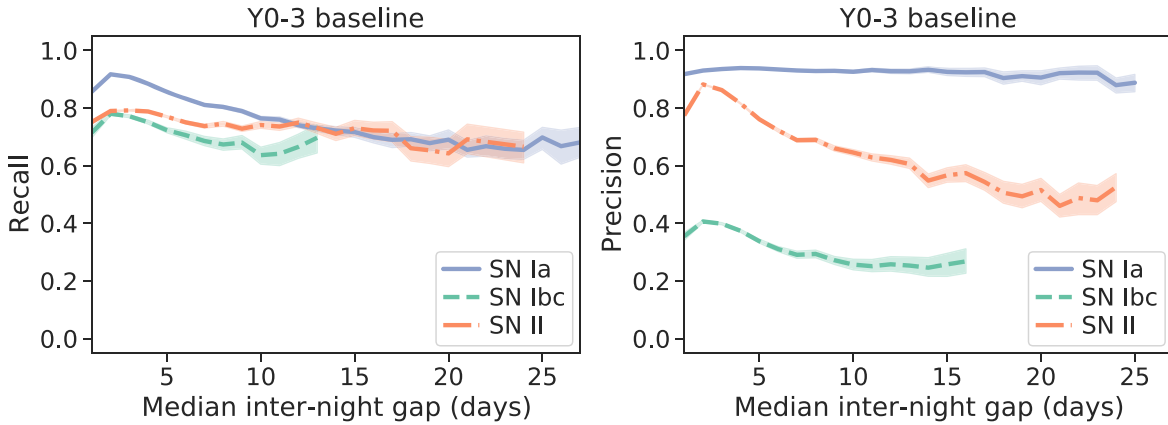


Figure 8. Test-set recall (left panel) and precision (right panel) as a function of internight gap per SN class for the Y0-3 baseline. The shaded areas correspond to the 95% confidence limits obtained by bootstrapping the recall and precision values for each bin. To remove small-number effects, we present only the results for bins with more than 300 events.

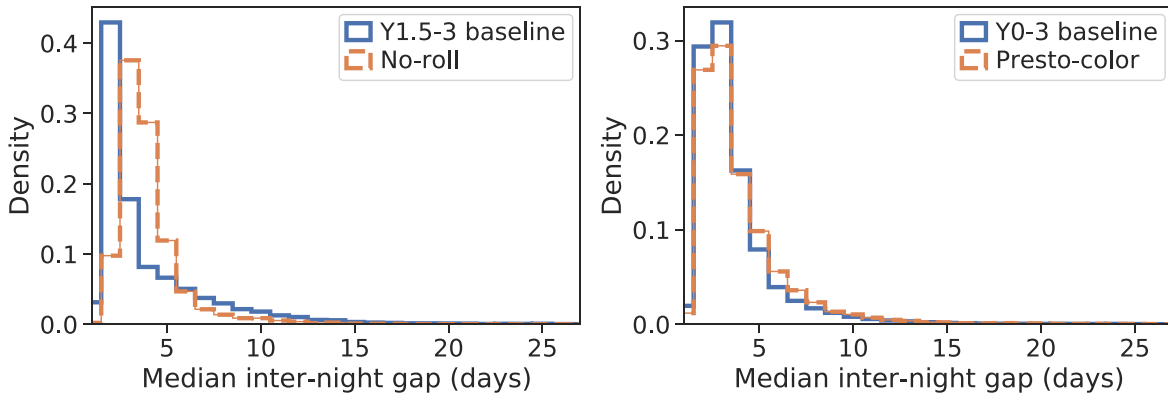


Figure 9. Test-set density of events as a function of internight gap for the Y1.5-3 baseline versus no-roll (left panel) and Y0-3 baseline versus presto-color (right panel).

high median internight gaps are misclassified as SNe II, driving down the precision of the latter. Overall, we find similar results for the other observing strategies analyzed.

While this overall conclusion still holds, our results show that the classification performance depends less on the median internight gap for this set of observing strategy simulations compared with Alves et al. (2022), where we recommended a median internight gap of $\lesssim 3.5$ days. Our new results suggest a cut of $\lesssim 5.5$ days; however, all the current observing strategies aim for a lower median internight gap, making such a recommendation redundant. We attribute this reduced

sensitivity of classification results to cadence to the recent improvements made to the FBS scheduler.

The left panel of Figure 9 shows that the peak of the median internight gap for the Y1.5-3 baseline is lower than for no-roll. However, while rolling improves the cadence of the events in the active region, the events in the background region are less regularly sampled than no-roll, which leads to the heavier tail of Y1.5-3 baseline. Thus, overall, the classification performance is not significantly improved by rolling.

Since the total exposure time is fixed, the addition of a third visit each night leads to the presto-color events being visited

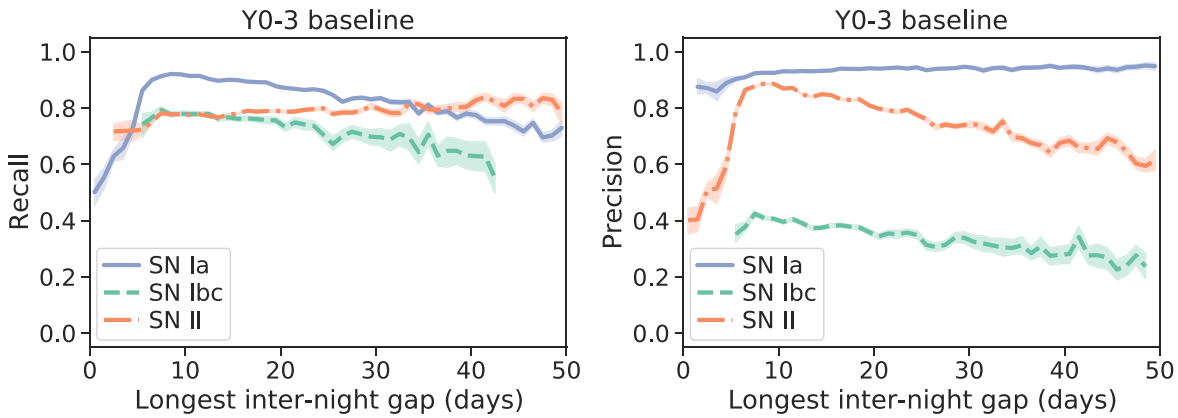


Figure 10. Test-set recall (left panel) and precision (right panel) as a function of the length of the longest inter-night gap per SN class for the Y0-3 baseline. The shaded areas correspond to the 95% confidence limits obtained by bootstrapping the recall and precision values for each bin. To remove small-number effects we present only the results for bins with more than 300 events. The reduced performance below 8 days corresponds to less than 5% of the events that tend to have very short light curves and therefore are not well classified.

fewer nights. Consequently, this cadence has sparser observations than the Y0-3 baseline. This is reflected in a slightly higher median inter-night gap for `presto-color`, as shown in the right panel of Figure 9. However, the small shift in the median inter-night gap distribution does not explain the degradation in performance seen for `presto-color`. We now turn to other cadence properties in order to understand this result.

4.4. Regularity of Sampling

In this section we investigate whether the performance differences seen for the cadences considered arise from the (ir) regularity of sampling. Characteristics of the regularity of sampling that are potentially important for classification include large gaps in the light curve and the number of observations near the peak.

In Alves et al. (2022) we found that the GPs successfully interpolate between large gaps (>10 days) so the classifier is still able to identify the SNe. Figure 10 confirms that the recall and precision of SNe either slowly decrease or remain constant with the increase of the length of the longest inter-night gap. These conclusions also generalize to the other cadences we study. This indicates that the GP step is generally able to interpolate large gaps.

A related consideration for characterizing the regularity of light-curve sampling is observing SNe near peak brightness, where the shape of the light curve changes rapidly. These observations are critical for obtaining a reliable cosmological distance modulus and facilitate accurate photometric classification. In this work, we estimate the SN peak as the moment that maximizes the GP fit predicted flux in any passband. Then we define the number of observations near the peak as those 10 days before and 30 days after peak brightness; we sum the observations in all passbands to calculate this quantity. Similarly to Alves et al. (2022), we find that the classification performance generally increases with the number of observations near the peak for all the new observing strategies. Figure 11 shows the results for Y0-3 baseline as a representative example. For Type Ia SNe, this performance levels off around 15 observations near the peak. This is comparable to the SNe cosmology metric used in Lochner et al. (2022), which requires 5 observations before peak and 10 observations after.

Figure 11 also shows a drop in performance for ~ 2 observations near the peak. We find that such events generally only contain the latter part of the transient, and their light curves tend to be flat. The classifier predicts events with flat GP fits as SNe II: the latter tend to have long light curves so it is likely that a flat part of the light curve will be observed. Once there are more observations near the peak, the light curves are not as flat so the SN II recall decreases until there is sufficient information in the light curve for the classifier to correctly identify the transient shape.

Having established the influence of these characteristics on classification performance, we now consider how they impact the relative classification performance seen in Figures 3 and 5 for the observing strategies considered.

The distribution of the longest inter-night gap in the Y1.5-3 baseline exhibits two peaks, as shown in the left panel of Figure 12. This is due to the fact that different areas of the sky start rolling at different times, and the Y1.5-3 baseline therefore includes some events in areas of the sky that have not yet started rolling. Thus, we see a second peak at higher values of the longest inter-night gap for the Y1.5-3 baseline. The peak in the distribution corresponding to the rolling region is at shorter timescales than in `no-roll`. Overall, these differences do not result in a significant change in performance, because there is no significant tail produced toward longer gaps. By contrast, the right panel of Figure 12 shows that, as expected, the distribution of the longest gaps for `presto-color` does exhibit a broad tail, due to the more irregular sampling: `presto-color` has 15% more events that have a long gap of 20 days or more.

Figure 13 compares the distributions of the number of observations near the SNe peak for the various cadences. The left panel of Figure 13 shows the impact of rolling (Y1.5-3 baseline), with a bimodal distribution corresponding to the “background” and “active” areas. While the difference between this distribution and that of `no-roll` may appear visually large, the latter only has 5% more events in the poorly classified region (<15 observations near peak). This again results in very little difference in classification performance due to rolling. The right panel of Figure 13 shows that `presto-color` events have $\sim 10\%$ more events with <15 observations near peak compared with the Y0-3 baseline. While this difference does not make a large visual impact, it is

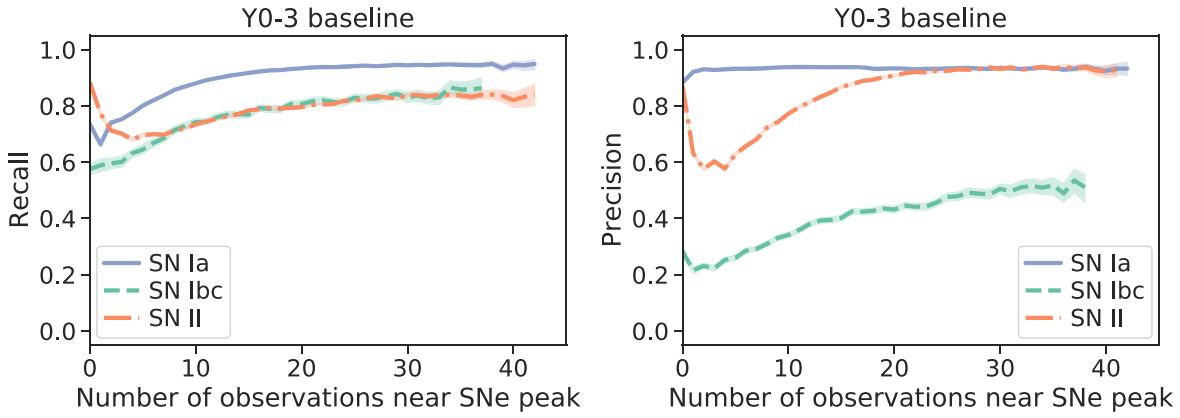


Figure 11. Test-set recall (left panel) and precision (right panel) as a function of the number of observations near peak per SN class for the Y0-3 baseline. The shaded areas correspond to the 95% confidence limits obtained by bootstrapping the recall and precision values for each bin. To remove small-number effects we present only the results for bins with more than 300 events.

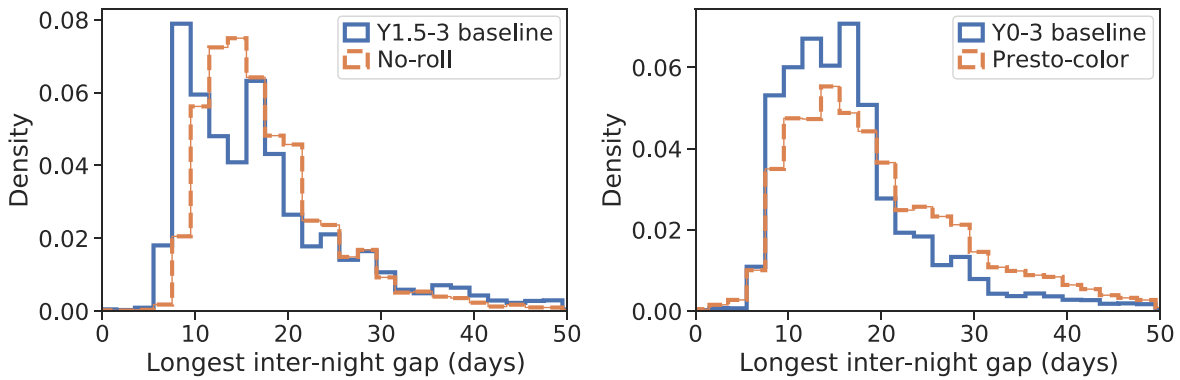


Figure 12. Test-set density of events as a function of the length of the longest inter-night gap for the Y1.5-3 baseline versus no-roll (left panel) and Y0-3 baseline versus presto-color (right panel).

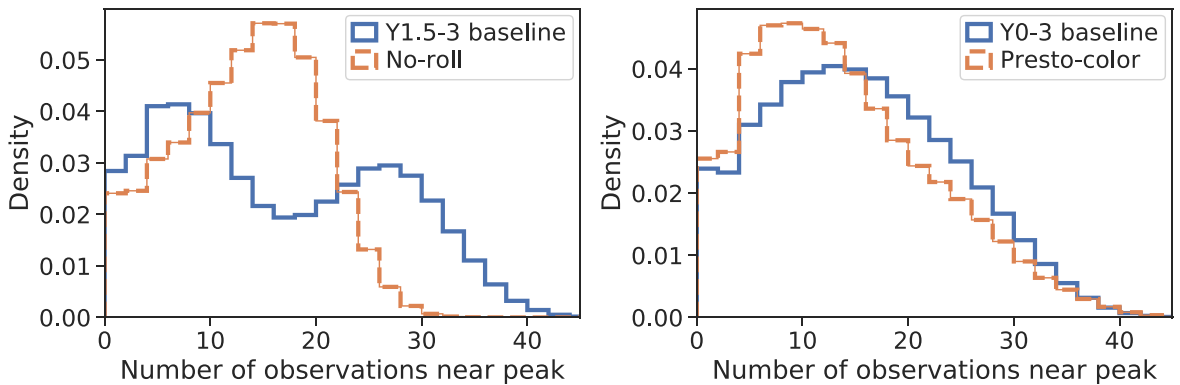


Figure 13. Test-set density of events as a function of the number of observations near peak for the Y1.5-3 baseline versus no-roll (left panel) and Y0-3 baseline versus presto-color (right panel).

nevertheless in a regime that strongly affects classification performance.

Figure 5 shows that `presto-color` mainly impacts classification of Type II SNe. In turn, Figure 11 shows that Type II SNe classification performance is a strong function of the number of observations near peak compared with the other classes, for <15 observations near peak. Since `presto-color` has more events in this regime, one may therefore expect that SN II classification is particularly degraded for this cadence, and this expectation is confirmed by our results.

Overall, `presto-color` exhibits small but significant changes in the distribution of the longest inter-night gap and the number of observations near the peak. These combine to result in irregularly sampled light curves, which in turn leads to degraded classification performance.

5. Discussion and Conclusions

We have presented the impact of LSST cadence choices on the performance of SNe photometric classification, using

simulated multiband light curves from the LSST baseline cadence, the nonrolling cadence, and a presto-color cadence. For each data set considered, we augmented the nonrepresentative training set to be representative of the test set and built a classifier using the photometric transient classification library `snmachine`. In line with previous studies, we confirmed that the light-curve length, median internight gap, and number of observations near the SNe peak, which differ between the cadences, affect the photometric classification.

Previous works argued that a rolling cadence benefits SNe science due to the improved sampling but that more in-depth simulations and studies were needed (LSST Science Collaboration et al. 2017; Lochner et al. 2018). We find that the considered rolling cadence (which increases by 90% the footprint weight of the active region) only mildly improves the overall classification performance. However, crucially, our results show that the active region of the rolling cadence as implemented in the current baseline strategy has a significantly higher classification performance than the background region. This in turn suggests that the SN Ia light curves in the active region could be better measured, and hence more useful for cosmological analyses. We now investigate this point.

Lochner et al. (2022) defines a set of light-curve requirements for well-measured SNe Ia that form the basis of seven cosmology metrics; in Appendix D we present the updated version of these requirements currently being used by LSST DESC. Considering all but one of the updated requirements (ignoring a color-related requirement due to its computationally intensive nature), we compared the SN Ia light curves in the active and background regions of the Y1.5-3 baseline. We found that $\sim 50\%$ of the SNe Ia in the active region fulfilled the light-curve requirements, compared with only $\sim 20\%$ of the SNe Ia in the background region. While these results are indicative rather than definitive due to ignoring the color requirement, they suggest that the 25% improvement in the classification performance log-loss metric in the active region is also associated with an increase of up to a factor of 2.7 in the number of cosmologically useful SNe Ia in the active region. These results taken together strongly motivate the implementation of a rolling cadence within the baseline observing strategy.

We also found that the `presto-color` cadence led to shorter and sparser light curves: the light-curve length distribution of this cadence on Figure 7 is skewed toward lower values. Additionally, there are more `presto-color` events with large gaps and fewer observations near the SNe peak. These results indicate that the events simulated under this cadence have a more heterogeneous sampling than the baseline events. Irregular sampling, especially around the peak where the light curve varies more rapidly, results in worse constraints on its shape; therefore the classifier is less able to distinguish between the SNe classes.

Since the third visit per night implemented in `presto-color` is in part motivated by facilitating early transient classification, our results imply that there is a trade-off in the observing strategy requirements of early and full light-curve classification.

The accuracy of SN Ia photometric classification and core-collapse contamination affect the measurements of the dark energy equation-of-state parameter (Jones et al. 2017; Kessler & Scolnic 2017). While a Bayesian methodology can marginalize over the contamination, minimizing such contamination reduces systematic uncertainties in cosmological constraints

(Kunz et al. 2007; Knights et al. 2013; Roberts et al. 2017; Jones et al. 2018). Since SNe cosmology with LSST is expected to be limited by systematic uncertainties, the relationship between the efficacy of photometric classification and cosmological constraints is of crucial importance.

We expect these conclusions to be general and hold for different photometric classifiers that rely on the full light curve, as shown in Alves et al. (2022). In the future, we plan to develop a fast proxy metric to evaluate the impact of cadence choices on photometric classification directly on the observing strategy cadences produced by the FBS (Naghib et al. 2019), avoiding the time-consuming SNe simulations and classification steps that were necessary in this work. More broadly, our results contribute to the pioneering process of community-focused experimental design and optimization of the LSST observing strategy.

This paper has undergone internal review in the LSST Dark Energy Science Collaboration. The authors would like to thank Mi Dai, V. Ashley Villar, and Ayan Mitra for their helpful comments and reviews.

Author contributions are listed below.

C.S.A.: software, validation, formal analysis, investigation, data curation, writing (original draft), visualization.

H.V.P.: conceptualization, methodology, validation and interpretation, supervision, writing (original draft; review and editing), funding acquisition.

M.L.: conceptualization, methodology, validation and interpretation, writing (original draft; review and editing).

J.D.M.: conceptualization, methodology, validation and interpretation, supervision, writing (review).

R.K.: updating code for `SNANA` simulations, writing (review).

We thank R. Lynne Jones, Rahul Biswas, and Gautham Narayan for helpful discussions. We also thank Philippe Gris for the updated requirements for well-measured Type Ia SNe. This work was partially enabled by funding from the UCL Cosmoparticle Initiative. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101018897 CosmicExplorer). This work has also been enabled by support from the research project grant Understanding the Dynamic Universe funded by the Knut and Alice Wallenberg Foundation under Dnr KAW 2018.0067. The work of H.V.P. was partially supported by the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine. M.L. acknowledges support from South African Radio Astronomy Observatory and the National Research Foundation (NRF) toward this research. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF. This work used facilities provided by the UCL Cosmoparticle Initiative. This work was completed in part with resources provided by the University of Chicago's Research Computing Center.

The DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des Particules in France; the Science & Technology Facilities Council in the United Kingdom; and the Department of Energy, the National Science Foundation, and the LSST Corporation in the United States. DESC uses resources of the IN2P3 Computing Center (CC-IN2P3—Lyon/Villeurbanne—France) funded by the

Centre National de la Recherche Scientifique; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under contract No. DE-AC02-05CH11231; STFC DiRAC HPC Facilities, funded by UK BEIS National E-infrastructure capital grants; and the UK particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE Contract DE-AC02-76SF00515.

Software: Astropy (Astropy Collaboration et al. 2013, 2018), George (Ambikasaran et al. 2014), Jupyter (Kluyver et al. 2016), LightGBM (Barbier et al. 2016; Ke et al. 2017; Zhang et al. 2017), Matplotlib (Hunter 2007; Caswell et al. 2020), NumPy (Harris et al. 2020), pandas (McKinney 2010; pandas development team 2020), pickle (Van Rossum 2020), pytest (<https://github.com/pytest-dev/pytest>), pywt (Lee et al. 2019a, 2019b), scikit-learn (Pedregosa et al. 2011), SciPy (Virtanen et al. 2020), seaborn (Waskom et al. 2020), smmachine (Lochner et al. 2016; Alves et al. 2022).

Appendix A Simulated CC SN Rates

In this appendix we present the absolute and relative rates used to simulate CC SNe in this work. These rates follow Shivvers et al. (2017) with the adjustments described in Section 2.4. Table 2 includes both the rates of each CC SNe class and the models used (see Kessler et al. 2019 for further details). The resulting number of SNe for each class is shown in Table 1.

Table 2

Absolute and Relative Scale Rate Used to Simulate Core-collapse SNe in This Work Expressed in Percentages

| Subtypes | Model Name | Abs. Scale Rate (Rel. Scale Rate) % | |
|--------------------------------|-----------------------|-------------------------------------|-------------|
| Core-collapse | | | |
| II | see Hydrogen Rich | 69.6 (100) | |
| Ib+Ic | see Stripped Envelope | 30.4 (100) | |
| Hydrogen Rich—class SN II | | | |
| II (IIP, IIL) | SNII-NMF | 32.45 (23.325) | 64.9 (93.3) |
| | SNII-Templates | 32.45 (23.325) | |
| IIIn | SNIIIn-MOSFiT | 4.7 (3.35) | 4.7 (6.7) |
| Stripped Envelope—class SN Ibc | | | |
| Ib | SNIb-Templates | 5.4 (17.8) | 10.8 (35.6) |
| Ic | SNIc-Templates | 3.75 (12.35) | 7.5 (24.7) |

Note. The rates follow Shivvers et al. (2017), and the SNe models are described in Kessler et al. (2019); SNIb-Templates and SNIc-Templates are both described together as SNIbc-Templates.

Appendix B

Augmentation Details and Classification Hyperparameters

Section 3.3 described the differences between the augmentation procedure used in this work and the one in Section 4 of Alves et al. (2022). In particular, we changed the distribution used to create the augmented training sets because the removal of the SNIbc-MOSFiT model (mentioned in Section 2.3) altered the redshift distribution of the events. In this work, we augmented each event between the redshift limits z_{\min} and z_{\max} from Section 4.2 of Alves et al. (2022):

$$\begin{aligned} z_{\min} &\approx \max\{0, 0.90 z_{\text{ori}} - 0.10\} \text{ and} \\ z_{\max} &\approx 1.43 z_{\text{ori}} + 0.43, \end{aligned} \quad (\text{B1})$$

where z_{ori} is the spectroscopic redshift of the original event. However, we used a different class-agnostic target distribution. In particular, we drew an auxiliary value z' from a log-trapezoidal distribution; the probability density function of the trapezoid distribution is

$$f(x) = \begin{cases} \frac{2}{\Delta x} \left[\frac{x_{\max} - 0.8x}{\Delta x} - 0.1 \right] & x \in [x_{\min}, x_{\max}] \\ 0 & \text{otherwise} \end{cases} \quad (\text{B2})$$

where $x_{\min} = \log(z_{\min})$, $x_{\max} = \log(z_{\max})$, and $\Delta x = x_{\max} - x_{\min}$. Then, we calculated the redshift of the new augmented event following Alves et al. (2022), $z_{\text{aug}}(z^*) = -z^* + z_{\min} + z_{\max}$.

For each observing strategy, we also adjusted the parameters of the Gaussian mixture models used to fit the number of observations per light curve (Table 3) and the flux uncertainty distribution of each passband (Table 4). We fitted Gaussian mixture models to the test set and used visual inspection to select the number components. The resulting photometric distributions are shown in Figure 14.

Table 3

Parameters of Gaussian Mixture Models Used to Fit the Number of Observations of the Test-set Light Curves for Each Observing Strategy

| | Y1.5-3 baseline | No-roll |
|-----------|------------------|------------------------------|
| weights | [0.322, 0.678] | [0.289, 0.215, 0.181, 0.315] |
| means | [16.86, 61.15] | [31.93, 18.07, 59.65, 45.35] |
| variances | [46.62, 561.06] | [30.23, 42.76, 75.83, 34.96] |
| | Y0-3 baseline | Presto-color |
| weights | [0.679, 0.321] | [0.626, 0.374] |
| means | [33.6, 65.1] | [27.11, 53.72] |
| variances | [199.09, 453.67] | [125.04, 252.05] |

Note. These values were later used to create an augmented training set (Section 3.3). We used visual inspection to select the number components of the Gaussian mixture models; that number is indicated through the number of weights provided for each observing strategy. The weight, mean, and variance of each component are displayed in the same order.

Table 4
Parameters of Gaussian Mixture Models Used to Fit the Flux Uncertainty Distribution of the Test Set in Each Passband (*ugrizy*) and Observing Strategy

| | | Y1.5-3 baseline | No-roll | Y0-3 baseline | Presto-color |
|----------|-----------|--------------------|--------------|------------------|--------------------|
| <i>u</i> | weights | [0.24, 0.76] | [0.25, 0.75] | [0.29, 0.71] | [0.64, 0.36] |
| | means | [2.34, 1.92] | [2.29, 1.95] | [2.24, 1.93] | [1.96, 2.35] |
| | variances | [0.46, 0.10] | [0.45, 0.10] | [0.38, 0.09] | [0.09, 0.28] |
| <i>g</i> | weights | [0.88, 0.12] | [0.11, 0.89] | [0.88, 0.12] | [0.10, 0.90] |
| | means | [1.26, 2.06] | [2.13, 1.27] | [1.26, 1.98] | [2.12, 1.41] |
| | variances | [0.11, 0.93] | [1.08, 0.12] | [0.11, 0.85] | [0.98, 0.14] |
| <i>r</i> | weights | [0.27, 0.73] | [0.76, 0.24] | [0.29, 0.71] | [0.23, 0.45, 0.32] |
| | means | [1.92, 1.57] | [1.59, 1.92] | [1.89, 1.58] | [2.82, 1.60, 2.20] |
| | variances | [0.38, 0.09] | [0.09, 0.40] | [0.34, 0.08] | [0.30, 0.08, 0.12] |
| <i>i</i> | weights | [0.62, 0.38] | [0.63, 0.37] | [0.63, 0.37] | [0.66, 0.34] |
| | means | [1.97, 2.45] | [1.98, 2.44] | [1.99, 2.45] | [1.96, 2.48] |
| | variances | [0.08, 0.22] | [0.08, 0.23] | [0.08, 0.20] | [0.10, 0.26] |
| <i>z</i> | weights | [1.] | [1.] | [1.] | [1.] |
| | means | [2.65] | [2.65] | [2.67] | [2.68] |
| | variances | [0.14] | [0.14] | [0.14] | [0.17] |
| <i>y</i> | weights | [1.] | [1.] | [1.] | [0.62, 0.38] |
| | means | [3.23] | [3.22] | [3.23] | [3.07, 3.52] |
| | variances | [0.15] | [0.14] | [0.14] | [0.07, 0.13] |

Note. These values were later used to create an augmented training set (Section 3.3). We used visual inspection to select the number components of the Gaussian mixture models; that number is indicated through the number of weights provided for each observing strategy. The weight, mean, and variance of each component are displayed in the same order.

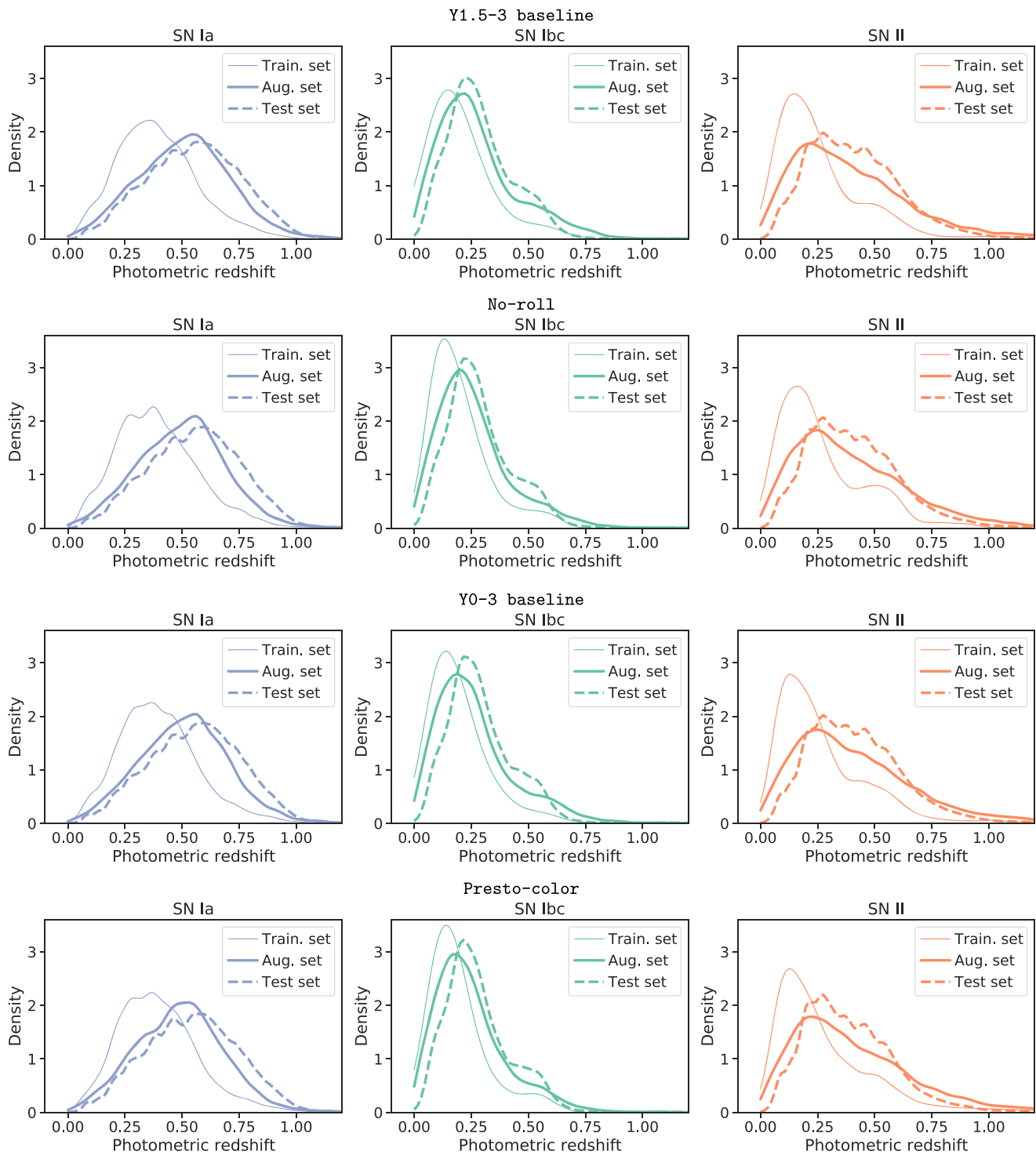


Figure 14. Host-galaxy photometric redshift distribution per supernova class, where SN Ia, SN Ibc, and SN II are shown, respectively, on the left, middle, and right panels. Each row shows the training (solid line), augmented training (bold solid line), and test-set distributions (dashed line) for each observing strategy. For all the observing strategies, the augmented training-set distribution is closer to the test set than the original training set.

Table 5
Values of the Optimized LightGBM Model Hyperparameters Used in Each Observing Strategy

| Hyperparameter name | Y1.5-3 baseline | No-roll | Y0-3 baseline | Presto-color |
|---------------------|-----------------|---------|---------------|--------------|
| boosting_type | gbdt | gbdt | gbdt | gbdt |
| learning_rate | 0.14 | 0.14 | 0.16 | 0.18 |
| max_depth | 19 | 19 | 13 | 19 |
| min_child_samples | 70 | 70 | 70 | 70 |
| min_split_gain | 0.6 | 0.0 | 0.1 | 0.0 |
| n_estimators | 115 | 115 | 115 | 115 |
| num_leaves | 50 | 50 | 55 | 50 |

Note. The hyperparameters are described in the library documentation (<https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMClassifier.html#lightgbm.LGBMClassifier>).

In this section we also show the hyperparameters values of the GBDT classifier used for each observing strategy (Table 5).

Appendix C Computational Resources

We simulated the observing strategy data sets on an Intel E5-2680v4 @ 2.4 GHz. Each data set with 2.5×10^6 events takes ~ 200 core hours to simulate. The data processing, classification, and analysis were performed on an Intel(R) Xeon(R) CPU E5-2697 v2 (2.70GHz). Using a single core, the pipeline takes ~ 5.6 hr to preprocess these events. Modeling them with GPs and performing their wavelet decomposition takes ~ 44.4 hr. Generating an augmented training set with 15,440 events takes ~ 9 hr, and reducing the dimensionality of their wavelet features using principal component analysis takes ~ 45 minutes. Optimizing the GBDT classifier takes ~ 5.5 hr. Obtaining the test-set predictions on the precomputed test-set features with the trained classifier takes 5 minutes. Overall, the entire classification pipeline takes $\sim 200 + 70$ core hours of computing time for each observing strategy.

Appendix D Well-measured Type Ia SNe

To measure cosmological parameters accurately, it is crucial to obtain a large sample of well-measured SN Ia. Lochner et al. (2022) presented a set of requirements to denote a SN Ia light curve as well measured; these requirements have recently been updated and refined. The updated requirements only use the light-curve observations in the *grizy* passbands, which have signal-to-noise ratio > 1 , and which satisfy

$$380 \text{ nm} < \frac{\bar{\lambda}_{\text{obs}}}{1+z} < 700 \text{ nm}, \quad (\text{D1})$$

where $\bar{\lambda}_{\text{obs}}$ is the mean wavelength of the telescope in the passband of the considered observation. They also limit the light curves to the observations with phases between 20 days before and 60 days after peak; the phase of the light curve is given by

$$\frac{t - t_{\text{peak}}}{1+z}, \quad (\text{D2})$$




where t is the time of the observation and t_{peak} is the time of the SNe peak brightness. The requirements are

1. At least three observations before peak with phase > -20 .
2. At least eight observations after peak with phase < 60 .
3. At least one observation with phase ≤ 10 .

4. At least one observation with phase ≥ 20 .
5. $\sigma_C < 0.04$, where σ_C is color uncertainty obtained when fitting the light curve with the SALT2 package (Guy et al. 2007).

In this work we ignore the last requirement because SALT2 fits are computationally intensive. We also use the light curves preprocessed as described in Section 3.1 rather than the 3-yr-long light curves.

ORCID iDs

Catarina S. Alves  <https://orcid.org/0000-0002-6164-9044>
Hiranya V. Peiris  <https://orcid.org/0000-0002-2519-584X>
Michelle Lochner  <https://orcid.org/0000-0003-2221-8281>
Richard Kessler  <https://orcid.org/0000-0003-3221-0419>

References

- Abbott, T. M. C., Allam, S., Andersen, P., et al. 2019, *ApJL*, 872, L30
Allam, T., Jr., & McEwen, J. D. 2021, arXiv:2105.06178
Alves, C. S., Peiris, H. V., Lochner, M., et al. 2022, *ApJS*, 258, 23
Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O’Neil, M. 2014, arXiv:1403.6015v2
Astier, P., Guy, J., Regnault, N., et al. 2006, *A&A*, 447, 31
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
Barbier, J., Dia, M., Macris, N., et al. 2016, Advances in Neural Information Processing Systems 29, ed. M. Lee et al. <https://proceedings.neurips.cc/paper/2016/hash/621bf66ddb7c962aa0d22ac97d69b793-Abstract.html>
Betoule, M., Kessler, R., Guy, J., et al. 2014, *A&A*, 568, A22
Bianco, F. B., Drout, M. R., Graham, M. L., et al. 2019, *PASP*, 131, 068002
Bianco, F. B., Graham, M., Drout, M. R., et al. 2018, Presto-Color: An LSST Cadence for Explosive Physics & Fast Transients, White Paper <https://authors.library.caltech.edu/94963/2/1812.03146.pdf>
Biswas, R., Alves, C., Setzer, C., Frohmaier, C., & Azfar, F. 2022, LSSTDESC/OpSimSummary: v3.0.0, Zenodo, doi:10.5281/ZENODO.6350796
Biswas, R., Daniel, S. F., Hložek, R., Kim, A. G., & , P. Y. 2020, *ApJS*, 247, 60
Boone, K. 2019, *AJ*, 158, 257
Boone, K. 2021, *AJ*, 162, 275
Brout, D., Scolnic, D., Popovic, B., et al. 2022, *ApJ*, 938, 110
Carrick, J. E., Hook, I. M., Swann, E., et al. 2021, *MNRAS*, 508, 1
Caswell, T. A., Droettboom, M., Lee, A., et al. 2020, matplotlib/matplotlib: REL: v3.3.2, Zenodo, doi:10.5281/ZENODO.4030140
Charnock, T., & Moss, A. 2017, *ApJL*, 837, L28
Friedman, J. H. 2002, *Comput. Stat. Data Anal.*, 38, 367
Gonzalez, O. A., Clarkson, W., Debattista, V. P., et al. 2018, arXiv:1812.08670
Guillochon, J., Nicholl, M., Villar, V. A., et al. 2018, *ApJS*, 236, 6
Guy, J., Astier, P., Baumont, S., et al. 2007, *A&A*, 466, 11
Guy, J., Sullivan, M., Conley, A., et al. 2010, *A&A*, 523, A7
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, 585, 357
Hložek, R., Ponder, K. A., Malz, A. I., et al. 2020, arXiv:2012.12392

- Hotelling, H. 1933, *J. Educ. Psychol.*, 24, 417
- Hunter, J. D. 2007, *CSE*, 9, 90
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
- Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2017, *ApJ*, 843, 6
- Jones, D. O., Scolnic, D. M., Riess, A. G., et al. 2018, *ApJ*, 857, 51
- Jones, R. L., Yoachim, P., Ivezić, Z., Neilsen, E. H., & Ribeiro, T. 2020, Survey Strategy and Cadence Choices for the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST) v1.2, Zenodo, doi:10.5281/ZENODO.4048838
- Ke, G., Meng, Q., Finley, T., et al. 2017, Advances in Neural Information Processing Systems 30, ed. I. Guyon et al., 3146, <http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf>
- Kessler, R., Bassett, B., Belov, P., et al. 2010b, *PASP*, 122, 1415
- Kessler, R., Becker, A. C., Cinabro, D., et al. 2009a, *ApJS*, 185, 32
- Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009b, *PASP*, 121, 1028
- Kessler, R., Conley, A., Jha, S., & Kuhlmann, S. 2010a, arXiv:1001.5210
- Kessler, R., Guy, J., Marriner, J., et al. 2013, *ApJ*, 764, 48
- Kessler, R., Marriner, J., Childress, M., et al. 2015, *AJ*, 150, 172
- Kessler, R., Narayan, G., Avelino, A., et al. 2019, *PASP*, 131, 094501
- Kessler, R., & Scolnic, D. 2017, *ApJ*, 836, 56
- Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, in Positioning and Power in Academic Publishing: Players, Agents and Agendas, ed. F. Loizides & B. Schmidt (Amsterdam: IOS Press), 87
- Knights, M., Bassett, B. A., Varughese, M., et al. 2013, *JCAP*, 2013, 039
- Kunz, M., Bassett, B. A., & Hlozek, R. A. 2007, *PhRvD*, 75, 103508
- Laine, S., Martínez-Delgado, D., Trujillo, I., et al. 2018, arXiv:1812.04897
- Lee, G., Gommers, R., Waselewski, F., Wohlfahrt, K., & O’Leary, A. 2019a, *JOSS*, 4, 1237
- Lee, G. R., Gommers, R., Wohlfahrt, K., et al. 2019b, PyWavelets/pywt: PyWavelets v1.1.1, Zenodo, doi:10.5281/ZENODO.3510098
- Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, *ApJS*, 225, 31
- Lochner, M., Scolnic, D., Almoubayyed, H., et al. 2022, *ApJS*, 259, 58
- Lochner, M., Scolnic, D. M., Awan, H., et al. 2018, arXiv:1812.00515
- Lokken, M., Gagliano, A., Narayan, G., et al. 2023, *MNRAS*, 520, 2887
- LSST Science Collaboration, Abell, P. A., Allison, J., et al. 2009, arXiv:0912.0201
- LSST Science Collaboration, Marshall, P., Anguita, T., et al. 2017, arXiv:1708.04058
- MacKay, D. J. C. 2003, Information Theory, Inference and Learning Algorithms (Cambridge: Cambridge Univ. Press)
- Malz, A. I., Hlozek, R., Allam, T., et al. 2019, *AJ*, 158, 171
- McKinney, W. 2010, in Proc. of the 9th Python in Science Conf., ed. S. van der Walt & J. Millman, 56
- Möller, A., & de Boissière, T. 2020, *MNRAS*, 491, 4277
- Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hlozek, R. 2019, *PASP*, 131, 118002
- Naghieb, E., Yoachim, P., Vanderbei, R. J., Connolly, A. J., & Jones, R. L. 2019, *AJ*, 157, 151
- Olsen, K., Di Criscienzo, M., Jones, R. L., et al. 2018, arXiv:1812.02204
- pandas development team, T 2020, pandas-dev/pandas: Pandas, v2.0.0rc1, Zenodo, doi:10.5281/zenodo.3509134
- Pasquet, J., Pasquet, J., Chaumont, M., & Fouchez, D. 2019, *A&A*, 627, A21
- Pearson, K. 1901, *London, Edinburgh, Dublin Philos. Mag. J. Sci.*, 2, 559
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, 12, 2825
- Perlmutter, S., Deustua, S., Gabi, S., et al. 1995, AAS Meeting, 187, 86
- Pierel, J. D. R., Rodney, S., Avelino, A., et al. 2018, *PASP*, 130, 114504
- Pimentel, Ó., Estévez, P. A., & Förster, F. 2022, *AJ*, 165, 18
- Qu, H., & Sako, M. 2022, *AJ*, 163, 57
- Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning (Cambridge, MA: MIT Press)
- Revsbech, E. A., Trotta, R., & van Dyk, D. A. 2017, *MNRAS*, 473, 3969
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, *AJ*, 116, 1009
- Roberts, E., Lochner, M., Fonseca, J., et al. 2017, *JCAP*, 2017, 036
- Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018a, *ApJ*, 859, 101
- Scolnic, D. M., Lochner, M., Gris, P., et al. 2018b, arXiv:1812.00516
- Shivvers, I., Modjaz, M., Zheng, W., et al. 2017, *PASP*, 129, 054201
- Stevance, H. F., & Lee, A. 2022, *MNRAS*, 518, 5741
- Survey Cadence Optimization Committee 2022, Survey Cadence Optimization Committee’s Phase I Recommendations, <https://pstn-053.lsst.io/>
- Swann, E., Sullivan, M., Carrick, J., et al. 2019, *Msng*, 175, 58
- The Dark Energy Survey Collaboration, Flaugher, B. 2005, *IJMPA*, 20, 3121
- The PLASTiCC team, Allam, T., Jr., Bahmanyar, A., et al. 2018, arXiv:1810.00001
- Van Rossum, G. 2020, The Python Library Reference, release 3.8.2 (Wilmington, DE: Python Software Foundation)
- Villar, V. A., Berger, E., Metzger, B. D., & Guillochon, J. 2017, *ApJ*, 849, 70
- Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020, *ApJ*, 905, 94
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *NatMe*, 17, 261
- Waskom, M., Botvinnik, O., Gelbart, M., et al. 2020, mwaskom/seaborn: v0.11.0 (September 2020), Zenodo, doi:10.5281/ZENODO.4019146
- Zhang, H., Si, S., & Hsieh, C.-J. 2017, arXiv:1706.08359