The Astrophysical Journal Supplement Series, 258:23 (14pp), 2022 Febuary https://doi.org/10.3847/1538-4365/ac3479

Considerations for Optimizing the Photometric Classification of Supernovae from the Rubin Observatory

Catarina S. Alves[1], Hiranya V. Peiris[1,2], Michelle Lochner[3,4,5], Jason D. McEwen[6], Tarek Allam, Jr.[6], and Rahul Biswas[2]
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 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 (SARAO), The Park, Park Road, Pinelands, Cape Town 7405, South Africa
[5] African Institute for Mathematical Sciences, 6 Melrose Road, Muizenberg, 7945, South Africa
[6] Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK
Received 2021 July 26; revised 2021 October 26; accepted 2021 October 28; published 2022 January 18

Abstract

The Vera C. Rubin Observatory will increase the number of observed supernovae (SNe) by an order of magnitude; however, it is impossible to spectroscopically confirm the class for all SNe discovered. Thus, photometric classification is crucial, but its accuracy depends on the not-yet-finalized observing strategy of Rubin Observatory's Legacy Survey of Space and Time (LSST). We quantitatively analyze the impact of the LSST observing strategy on SNe classification using simulated multiband light curves from the Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC). First, we augment the simulated training set to be representative of the photometric redshift distribution per SNe class, the cadence of observations, and the flux uncertainty distribution of the test set. Then we build a classifier using the photometric transient classification library `snmachine`, based on wavelet features obtained from Gaussian process fits, yielding a similar performance to the winning PLAsTiCC entry. We study the classification performance for SNe with different properties within a single simulated observing strategy. We find that season length is important, with light curves of 150 days yielding the highest performance. Cadence also has an important impact on SNe classification; events with median inter-night gap <3.5 days yield higher classification performance. Interestingly, we find that large gaps (>10 days) in light-curve observations do not impact performance if sufficient observations are available on either side, due to the effectiveness of the Gaussian process interpolation. This analysis is the first exploration of the impact of observing strategy on photometric SN classification with LSST.

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

1. Introduction

The upcoming Rubin Observatory Legacy Survey of Space and Time (LSST; LSST Science Collaboration et al. 2009; Collaboration et al. 2017; Ivezić et al. 2019) is expected to discover, during its 10 yr duration, at least one order-of-magnitude more supernovae (SNe) than the currently available SNe samples (Guillochon et al. 2017). Traditionally, SNe that are used in astrophysical and cosmological studies need to be spectroscopically classified (e.g., Riess et al. 1998; Astier et al. 2006; Kessler et al. 2009). However, this will be impossible for most events detected by LSST due to the limited spectroscopic resources; thus, LSST will rely on photometric classification, using the events that will be spectroscopically classified as its training set.

Previous efforts to understand the strengths and limitations of photometric classification algorithms resulted in the Supernova Photometric Classification Challenge (`SNPhotCC`; Kessler et al.2010a) in preparation for the Dark Energy Survey (DES; The Dark Energy Survey Collaboration & Flaugher 2005). Recently, the Photometric LSST Astronomical Time-Series Classification Challenge[7] (PLAsTiCC; The PLAsTiCC Team et al. 2018; Kessler et al. 2019) was launched in preparation for LSST, which will reach fainter magnitudes and have a $\sim$4 times larger survey area compared to DES. The classifiers applied to the data sets from these challenges employed parametric fits, template fits, and machine-learning models such as neural networks, boosted decision trees, a support vector machine, and gradient boosting (e.g., Kessler et al. 2010; Lochner et al. 2016; Charnock & Moss 2017; Muthukrishna et al. 2019; Pasquet et al. 2019; Villar et al. 2020).

To obtain accurate classification, the training set must be representative of the test set (e.g., Lochner et al. 2016). However, photometric classifiers are typically trained with nonrepresentative spectroscopically confirmed events that are biased toward lower redshifts. Thus, recent work has focused on overcoming the lack of representativeness (Revsbech et al. 2017; Boone 2019; Muthukrishna et al. 2019; Pasquet et al. 2019; Carrick et al. 2021). Photometric classification performance also depends on the survey observing strategy; however, this dependence has not yet been explored.

[7] https://www.kaggle.com/c/PLAsTiCC-2018/

The LSST observing strategy encompasses diverse considerations such as season length, survey footprint, single-visit exposure time, inter-night gaps, and cadence of repeat visits in different passbands. The observing strategy is currently being optimized (Collaboration et al. 2017; Gonzalez et al. 2018; Ivezić et al. 2018; Laine et al. 2018; Lochner et al. 2018; Jones et al. 2020), a challenging task since the survey has diverse goals (LSST Science Collaboration et al. 2009; Ivezić et al. 2019). Recently, the Rubin Observatory LSST Dark Energy Science Collaboration (DESC) Observing Strategy Working Group investigated the impact of observing strategy on cosmology and made recommendations for its optimization (Lochner et al. 2018, 2021; Scolnic et al. 2018). In particular, SNe cosmology requires a high and regular cadence with long season lengths (how long a field is observable in a year).

In this work, we upgrade the photometric transient classification library `snmachine`[8] (Lochner et al. 2016) for use with LSST data and build a classifier based on wavelet features obtained from Gaussian process (GP) fits. We also include the host-galaxy photometric redshifts and their uncertainties as features. We make several other improvements to deal with the greater realism of the PLAsTiCC data, including training set augmentation. Using this improved classifier, we study the performance of photometric SNe classification for subsets of light curves with different cadence properties, using the single observing strategy simulated for the PLAsTiCC challenge. We note that this approach is different from studying the classification performance for different observing strategies with fixed total exposure time, where a reduced season length could be compensated for with a higher cadence.

In Sections 2 and 3 we summarize the PLAsTiCC data set and describe the classification pipeline, respectively. Section 4 focuses on the augmentation methodology. Our results and their implications for observing strategy are described in Section 5. We conclude in Section 6.

2. PLAsTiCC Data Set

The PLAsTiCC (The PLAsTiCC Team et al. 2018; PLAsTiCC Team & PLAsTiCC Modelers 2019) data set consists of simulations of 18 different classes of transients and variable stars. It contains three-year-long light curves of 3.5 millions events observed in the LSST *ugrizy* passbands, as well as their host-galaxy photometric redshifts and uncertainties. Although the simulations included realistic observing conditions, the observing strategy used[9] is now outdated (Jones et al. 2020). PLAsTiCC mimicked future LSST observations in two survey modes: the Wide-Fast-Deep (WFD) survey, which covers almost half the sky and was used for 99% of the events, and the Deep-Drilling-Fields (DDF) survey, small patches of the sky with more frequent and deeper observations that have smaller flux uncertainties.

The simulations were divided into a nonrepresentative spectroscopically confirmed training set biased toward brighter events, which is 0.2% of the size of the test set. The training set was much smaller, to mimic the data that will be available at the start of LSST science operations from current and near-term spectroscopic surveys. In particular, the training set was loosely modeled on the magnitude-limited 4 m Multi-Object Spectroscopic Telescope Time Domain Extragalactic Survey (Swann et al. 2019), resulting in a sample with a mean redshift ∼0.3. The unblinded data set is available in PLAsTiCC Team & PLAsTiCC Modelers (2019), the model libraries are presented in PLAsTiCC Modelers (2019), and more details about the models and simulations, including the description of the training set, host-galaxy photometric redshifts, and their uncertainties, are given in Kessler et al. (2019). In this work we provide observing strategy recommendations to improve photometric classification of SNe in particular, so we restrict ourselves to the PLAsTiCC classes SN Ia, SN Ibc, and SN II; Table 1 shows a breakdown of the numbers of SNe in each class.

Table 1
Breakdown of the Number of SNe per Class Used in This Work (See Simulation Details in Kessler et al. 2019)

SN Class	N_{training} (%)	N_{test} (%)
SN Ia	2313 (58%)	1,659,831 (59%)
SN Ibc	484 (12%)	175,094 (6%)
SN II	1193 (30%)	1,000,150 (35%)
Total	3990 (100%)	2,835,075 (100%)

Note. For each class, the number of events in the training and test sets is shown.

3. Classification Pipeline

In this section we describe how we upgraded the photometric classification pipeline `snmachine` for use with PLAsTiCC data. Augmentation was a crucial step in this process, and it is discussed in greater detail in Section 4.

3.1. Light-curve Preprocessing

PLAsTiCC light curves have long gaps (>50 days) in the observations because any given sky location is not visible from the Vera C. Rubin Observatory site for several months of the year. Additionally, the SNe are only detected for a few months, so including the entire three-year-long light curve provides irrelevant information to the classifier, which in turn degrades its performance. In order to isolate the observing season that contains the SNe, we selected the season that contains the observations flagged as detected, and that has no inter-night gaps larger than 50 days. To introduce uniformity in the data set, we translated the resulting light curves so their first observation is at time zero. However, this results in light curves that peak at different times, so we explored additionally shifting all training set light curves randomly in time to capture a larger variability of peak times. We found that augmenting with this random shift led to a less representative training set, and thus to a worse classification performance. Therefore, in this work, we simply aligned the first observation of the training events at time zero, such as we did for the test set. Figure 1 shows an example of light-curve preprocessing.

3.2. Gaussian Process Modeling of Light Curves

We modeled each light curve with a GP regression (e.g., MacKay 2003; Rasmussen & Williams 2005), following previous works that successfully used GP-modeled light curves in their classification pipelines (e.g., Lochner et al. 2016;

[8] https://github.com/LSSTDESC/snmachine

[9] Simulation `minion_1016`: https://docushare.lsst.org/docushare/dsweb/View/Collection-4604.

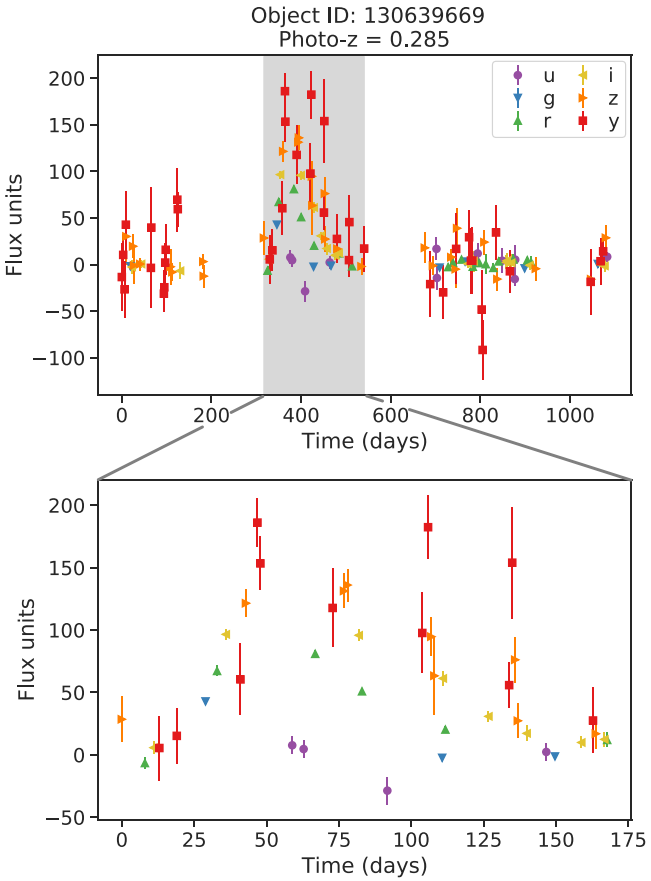

Figure 1. Example of a simulated LSST type Ibc supernova light curve from PLAsTiCC, showing how we preprocess light curves to remove season gaps. The original and processed light curves are shown in the top and bottom panels, respectively. The processed light curve corresponds to the shaded region on the original light curve with the first observation translated to time zero. The observations in different passbands are shown in different colors.

Revsbech et al. 2017). Unlike the previous examples that fitted separate GPs to each passband, Boone (2019) fitted GPs both in time and wavelength, thus allowing the GPs to incorporate cross-band information. Figure 2 shows that such a two-dimensional GP fit infers the SNe light curve even in passbands where there are no or only a few observations, in contrast to the one-dimensional GP fit. Thus, we used two-dimensional GPs to fit light curves both in time and wavelength.

We chose a null mean function for the GP, modeling the events as perturbations to a flat background. Following Boone (2019), we used the once-differentiable Matérn $3/2$ kernel for the GP covariance, which is appropriate for modeling explosive transients with sudden changes in their flux. The time dimension length-scale and amplitude were optimized per event, using maximum likelihood estimation. We fixed the length-scale of the wavelength dimension to 6000 Å as in Boone (2019), since they found that this value produces reasonable models for all classes in PLAsTiCC. The GPs were implemented with the package `George`[10] (Ambikasaran et al. 2015).

3.3. Feature Extraction

In this work we followed the wavelet decomposition approach of Lochner et al. (2016) to extract features. Since this is a model-independent approach to feature extraction, it does not assume any physical knowledge about the observed phenomena; hence, it is applicable to any time-series data. Moreover, recent results showed wavelet decomposition was successful for general transient classification (Varughese et al. 2015; Lochner et al. 2016; Narayan et al. 2018; Sooknunan et al. 2021). This model-independent approach had not been used previously by the winning PLAsTiCC entries.

Following Lochner et al. (2016), we used a stationary wavelet transform and the `symlet` family of wavelets; the wavelet decomposition was implemented with the package `PyWavelets` (Lee et al. 2019a). To obtain the wavelet decomposition, we first used the GPs to interpolate all light curves onto the same time grid of 277 days (maximum light-curve length of the events); we chose approximately one grid point per day and used a two-level wavelet decomposition, following Lochner et al. (2016). These choices resulted in 6624 (highly redundant) wavelet coefficients per event. While it is common to combine GP fits and wavelet analysis (e.g., Istas 1992; Chen et al. 2013; Pope 2019, and references therein), we note that our method of modeling the sparse light curves with GP fits and then using wavelet decomposition to obtain classification features is unusual. This approach was briefly mentioned in Varughese et al. (2015), and first implemented in Lochner et al. (2016).

Following Lochner et al. (2016), we reduced the dimensionality of this wavelet space using principal component analysis (PCA; Pearson 1901; Hotelling 1933) on the wavelet coefficients of the augmented training set. After comparing the classifier performance on a validation set (we set aside 5% of the test set for validation) with different numbers of PCA components, we found that 20–50 components were the best to distinguish different types of SNe (their log-loss differs by around 2%). We chose 40 components (99.995% of the total variance) due to its slightly better performance.

Finally, we also include the photometric redshift and its uncertainty as classification features. Unlike our previous results on the SNPhotCC challenge (Lochner et al. 2016), we find that these features are crucial for solving the more realistic classification challenge presented by the PLAsTiCC data. This is also confirmed by other PLAsTiCC analyses (Boone 2019; Hložek et al. 2020).

3.4. Classification

We augmented the training set as described in Section 4, prior to training a classifier. We used the gradient boosting model implementation of the package `LightGBM`[11] (Ke et al. 2017), in particular the gradient boosting decision tree (GBDT; Friedman 2001). These are ensemble classifiers that produce predictions using ensembles of decision trees. The boosting improves the ensemble prediction by sequentially adding new decision trees that prioritize difficult-to-classify events. Boosted decision trees are commonly used in machine-learning pipelines, including most of the top solutions to PLAsTiCC challenge (Hložek et al. 2020), due to their robust predictions, capacity for handling missing data, and flexibility (Friedman 2001; Ke et al. 2017).

We optimized the GBDT hyperparameters (parameters of the model that must be set before the learning process starts) by maximizing the performance of a five-fold cross-validated grid

[10] george.readthedocs.io/

[11] lightgbm.readthedocs.io

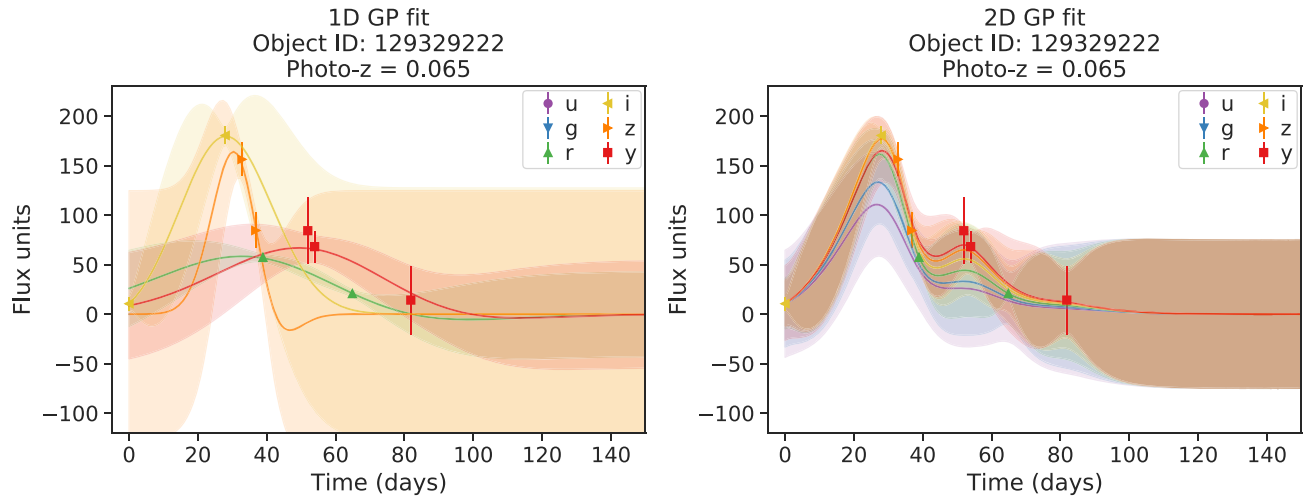

Figure 2. SN Ibc light curve, where the points show the observations, along with their error bars, and the lines and the shaded regions show the mean and standard deviation of the GP fit, respectively. The left panel shows the one-dimensional GP fit to each available passband, and the right panel shows the two-dimensional GP fit to all of the passbands (shown in different colors). The two-dimensional GP infers the light curve in passbands where there are no (or few) observations, unlike the one-dimensional GP.

Table 2
Optimized Hyperparameter Values Used for the `LightGBM` Model

Hyperparameter	WFD Setting	DDF Setting
`boosting_type`	`gbdt`	`gbdt`
`learning_rate`	0.24	0.24
`max_depth`	16	19
`min_child_samples`	25	70
`min_split_gain`	0.3	0.3
`n_estimators`	115	45
`num_leaves`	50	50

Note. A description of the hyperparameters is given in the library documentation at https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMClassifier.html#lightgbm.LGBMClassifier.

search on the augmented training set. First, each hyperparameter was optimized individually using a one-dimensional grid, keeping the other hyperparameters at default values. Then, we constructed a six-dimensional grid with three possible values for each hyperparameter informed by the earlier one-dimensional optimization. Finally we optimized this six-dimensional grid through a standard grid search. The resulting hyperparameter values are shown in Table 2. Since training and testing on the same events lead to overfitting, we placed in the same cross-validation fold all synthetic events that were derived from the same original event. While alternative hyperparameter optimization techniques can be considered (e.g., Bayesian optimization; Mockus et al. 1978; Snoek et al. 2012), a simple grid search strategy as described above proved to be effective.

3.4.1. Performance Evaluation

In order to evaluate the classification performance, we used the PLAsTiCC weighted log-loss metric (The PLAsTiCC Team et al. 2018; Malz et al. 2019) given by

$$\text{Log-loss} = -\left(\frac{\sum_{i=1}^{M} w_i \cdot \sum_{j=1}^{N_i} \frac{y_{ij}^*}{N_i} \cdot \ln p_{ij}}{\sum_{i=1}^{M} w_i} \right), \quad (1)$$

where M is the total number of classes, N_i is the number of events in class i, y_{ij}^* is 1 if observation j belongs to type i and 0 otherwise, p_{ij} is the predicted probability that event j belongs to class i, and w_i is the weight of the class i. The weights can be changed to give different importances to different classes; however, following the PLAsTiCC challenge, we gave the same weight to every SNe class.

Table 3
Confusion Matrix for Binary Classification

		True Class	
		Positive (P)	Negative (N)
Predicted	P	True positive (TP)	False positive (FP)
Class	N	False negative (FN)	True negative (TN)

We used confusion matrices to visualize the mislabeled classes; Table 3 shows the confusion matrix for a binary classification. For ease of comparison, we normalized the confusion matrices by dividing each entry by the true number of each SNe class; hence, the identity matrix represents a perfect classification.

For a single SNe class, it is also common to use the recall (also called completeness/sensitivity) to measure the fraction of correctly classified SNe, and the precision to measure the fraction of SNe assigned to the considered class that are indeed from that class. These are defined as

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

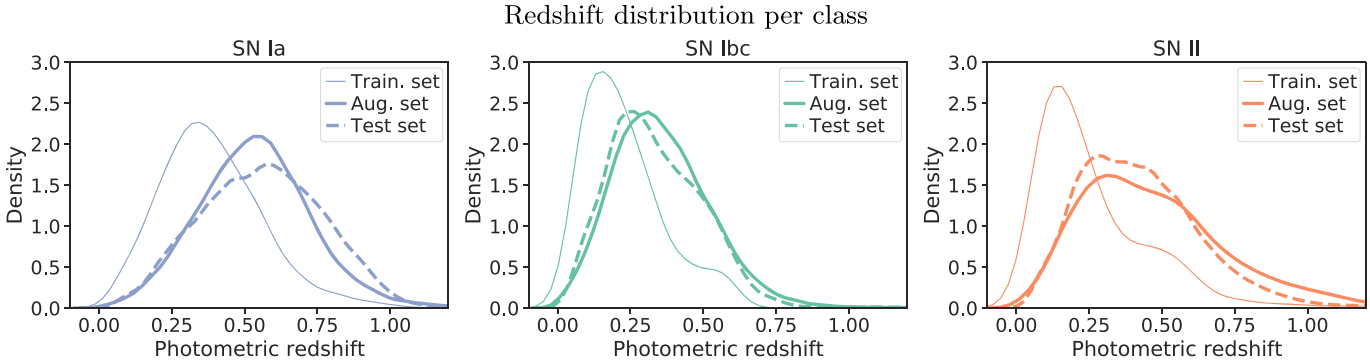

Figure 3. Host-galaxy photometric redshift distribution per supernova class, where SNe Ia, SNe Ibc, and SNe II are shown, respectively, on the left, middle, and right panels. The distribution of the training, augmented training, and test sets are shown as fine solid, bold solid, and dashed lines, respectively. Although the training set distribution is not representative of the test set, the augmented training set (Section 4) is close to the desired test distribution.

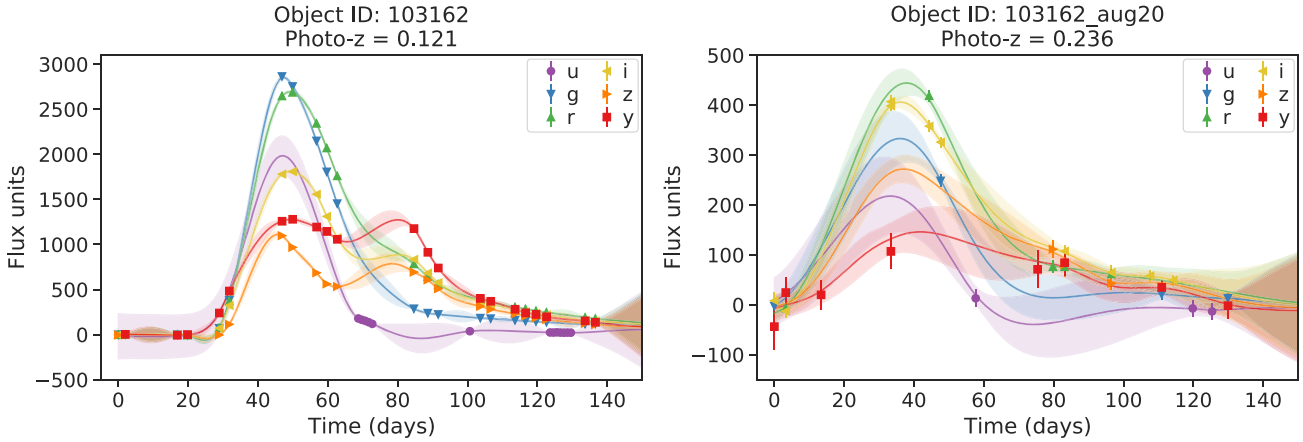

Figure 4. The left panel shows an SN Ibc light curve where the points show the observations, along with their error bars, and the lines and the shaded regions show the mean and standard deviation of the GP fit, respectively. The right panel shows a synthetic event at a different redshift generated from the original event using the procedure described in Section 4. Note that, in this example, the original event was simulated in the higher-cadence DDF survey, and we generated the synthetic event in the WFD survey.

and

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{3}$$

4. Augmentation

As previously outlined, the PLAsTiCC training set is nonrepresentative of the test set in redshift (see Figure 3) and also imbalanced: the most common SNe class has ∼4.8 times more events than the least common. However, to obtain accurate classification, the training set must be representative (Lochner et al. 2016) and balanced (as later discussed in Section 4.1).

Recent augmentation approaches rely on generating synthetic light curves from the GPs fitted to training set events (Revsbech et al. 2017; Boone 2019). In particular, Boone (2019) simulated new sets of observations for each object such that they match the cadence, depth, and uncertainty of observations of the test set, which ensured the representativity of these properties irrespective of the quality of the original event. The augmented observations were drawn from the mean prediction of the GP, and blocks of observations were dropped to simulate season boundaries. Additionally, Boone (2019) introduced redshift augmentation, where the observations of a new synthetic event are simulated at a different redshift from the original.

We adapted the approach used in Boone (2019) for our training set augmentation. Figure 4 shows a synthetic light curve generated using our augmentation procedure, which can be summarized as follows:

1. Choose the number of synthetic events to create (Section 4.1).
2. Model the original light curve with a two-dimensional GP fit in time and wavelength (as described in Section 3.2).
3. Choose a redshift for the synthetic event (Section 4.2).
4. Create synthetic observations at the new redshift, making use of the GP fit to the original event (Section 4.3).
5. Generate a photometric redshift (Section 4.4).

The WFD and DDF surveys have very different characteristics and enable qualitatively different science goals. Hence we found that it is important to use customized augmentation for the two survey modes, in contrast to the approach of the winning PLAsTiCC entries. Since the DDF survey has a different redshift distribution, higher cadence, and higher signal-to-noise ratio (S/N) than the WFD survey, we must use a different augmentation and, consequently, a different classifier.

We now describe the augmentation procedure in detail. The reader should keep in mind, where relevant, that the

augmentation procedure was customized for the two survey modes as necessary.

4.1. Number and Class Balance of Synthetic Events

As we wish to optimize classification performance for all SNe classes, we generated an augmented training set with the same number of events per class (i.e., a balanced training set). We also investigated an augmentation of the training set to resemble the class proportions of the test set (∼59% SNe Ia, ∼6% SNe Ibc, ∼35% SNe II). However, this gave worse performance, biasing the predictions toward the most common class.

We determined that the performance of the classifier stabilized when the size of the WFD augmented training set was around 4×10^4. In this final configuration, each training set SN was augmented up to 140 times. The DDF augmented training set stabilized around 8×10^3, and each DDF training set SN was augmented up to 70 times.

4.2. Redshift Augmentation

As previously outlined, redshift augmentation was found to be critical for the PLAsTiCC data set (Boone 2019). Figure 3 shows the bias of the training set toward low-redshift events in comparison to the test set. We augmented each training set event of the WFD survey between

$$z_{\min} \approx \max\{0,\ 0.90\, z_{\rm ori} - 0.10\} \text{ and}$$
$$z_{\max} \approx 1.43\, z_{\rm ori} + 0.43, \quad (4)$$

where $z_{\rm ori}$ is the spectroscopic redshift of the original event. For the augmentation, we used a target distribution that is *class-agnostic*. First, we drew an auxiliary value z^* from a log-triangular distribution with minimum value and mode $\log(z_{\min})$, and maximum value $\log(z_{\max})$. Then, we calculated the redshift of the new augmented event $z_{\rm aug}$,

$$z_{\rm aug}(z^*) = -z^* + z_{\min} + z_{\max}. \quad (5)$$

For the deeper DDF survey, the corresponding $z_{\max}$ was increased by 40%; otherwise, the same procedure was followed. These limits arise due to the fact that for a given event in the original training set, its GP fit is more reliable close to the observations; hence, we limit the GP extrapolation in wavelength when generating synthetic events, which translates into the above redshift constraint. This distribution differs slightly from that of Boone (2019), which also used a class-agnostic augmentation. We derive the aforementioned redshift limits for augmentation in Appendix A.

The process of actually redshifting the light curve after choosing the new redshift is discussed below.

4.3. Generating Realistic Synthetic Observations

The first step in generating the synthetic light curves is selecting the epochs at which mock observations will be made. Our implementation proceeded as in Boone (2019); we summarize the approach as follows. First, we stretched the observed epochs of the original event to account for the time dilation due to the difference between the original and augmented redshifts. We also removed any observations that fell outside the observing window as a consequence. Then we randomly picked a target number of observations from a Gaussian mixture model based on the test set.[12] However, this fails to account for the change in the cadence due to redshift augmentation; events shifted to higher redshifts have a lower density of observations than the events observed at those redshifts. In order to account for this, we multiplied this target number by $(1 + z_{\rm aug})/(1 + z_{\rm ori})$. We then generated additional observations at the same epochs as existing observations in randomly selected passbands, associating each synthetic observation with an observed epoch in the original light curve. Further, to avoid creating synthetic light curves where most of the observations are obtained through this procedure, we capped the number of additional observations generated to be less than 50% of the total number of observations in the original light curve. If this procedure resulted in more observations than the original target number drawn from the Gaussian mixture model, we then randomly dropped observations (original or new) until the target number was reached. Otherwise, to introduce additional variability, we randomly dropped 10% of the synthetic observations.

Once we determined the epochs at which new observations would be generated, we redshifted the light curve as follows. We first computed the central wavelengths of the *ugrizy* passbands of the synthetic event as seen at the redshift of the original event. Then, we computed the mean and uncertainty of the GP fit to the original event, at the observed epochs of the original event associated with the synthetic observations but at the redshifted wavelengths. The steps so far dealt with the time dilation but not with the cosmological dimming of the synthetic event. Assuming a standard cosmological model, we redshifted the flux of the synthetic event and its uncertainty, such that it is observed at $z_{\rm aug}$. Further details of this redshifting implementation are given in Appendix A.

Following Boone (2019), we then combined the flux uncertainty of the augmented events predicted by the GP in quadrature with a value drawn from the flux uncertainty distribution of the test set, in order to achieve a more representative flux uncertainty distribution for the augmented training set. We also drew noise to add to the flux of the augmented events from a Gaussian with standard deviation of the aforementioned value from the flux uncertainty distribution.

Finally, we imposed quality cuts on the synthetic events in order to decide whether to add them to the augmented training set. To make the synthetic events as similar as possible to the test-set events, we use a two-detection trigger based on PLAsTiCC (Kessler et al. 2019). Boone (2019) fitted an error function to the observations from the full data set to predict the probability of detection as a function of S/N, and applied this probabilistic threshold to all observations. Boone (2019) then accepted an event if at least two of its observations were predicted as detected. However, we find that this was insufficient to constrain a GP; thus, we required an additional observation, without requiring it to be predicted as detected by the chosen probabilistic detection threshold. Note that all synthetic light curves in the augmented DDF training set meet this quality cut as they are generated with a higher number of observations.

[12] The model used contained one component with mean 24.5 and standard deviation 8.5 for WFD, and two components with probabilities for each component of [0.34, 0.66], means of [57.4, 92.8] and standard deviations of [16.5, 18.4] for DDF. While a mixture model was fitted for both WFD and DDF, a single component was found to be the best fit for WFD.

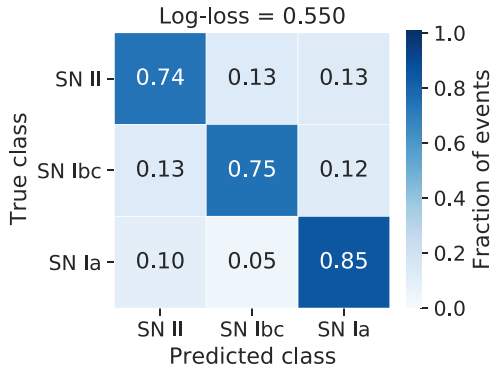

Figure 5. WFD test-set normalized confusion matrix for the classifier trained on the augmented training set and its log-loss performance. Each event is assigned to the class with the highest prediction.

4.4. Photometric Redshift

In order to simulate realistic photometric redshifts for the synthetic events, following Boone (2019), we chose a random event from the ∼4% of test-set events that had a spectroscopic redshift measurement, and calculated the difference between its spectroscopic and photometric redshifts. We then added this difference to the true redshift of the augmented event to generate a photometric redshift.

4.5. Computational Resources

We performed our computations on an Intel® Xeon® CPU E5-2697 v2 (2.70GHz). Using a single core, the pipeline takes ∼1 min to fit GPs to 1000 events, and to perform their wavelet decomposition. Generating a balanced augmented training set with 4×10^4 events takes ∼9 hr. Reducing the dimensionality using PCA takes ∼30 minutes for an augmented training set of 4×10^4 events, and optimizing the LightGBM classifier on the same training set takes ∼8 hr. After we computed the test-set features, generating predictions with the trained classifier takes ∼10 minutes. Overall, the entire classification pipeline takes ∼70 core hours of computing time for WFD and 12 for DDF in this setting.

5. Results and Implications for Observing Strategy

We now turn to our results on the PLAsTiCC data set and consider in detail their implications for various aspects of the LSST observing strategy. We study classification performance for SNe with different properties within the single simulated observing strategy that is available in PLAsTiCC. We present results related to classification performance for the two different survey modes (WFD and DDF) in Section 5.1. We then explore the performance as a function of light-curve length (Section 5.2), median inter-night gap (Section 5.3), number of gaps > 10 days (Section 5.3), and number of observations near the peak (Section 5.4).

5.1. Survey Mode-specific Augmentation and Its Effect on Performance

Figure 5 shows the confusion matrix for the classifier trained on an augmented WFD training set as described in Section 4. Despite the use of general wavelet features that were not specifically designed for SNe classification, the classifier obtains a log-loss of 0.55. This performance is comparable to that obtained by the top three submissions to PLAsTiCC for these SN classes (Boone 2019; Hložek et al. 2020). We note that similarly to other classifiers, the performance is weakest for SNe Ibc (75% recall but 39% precision).

The DDF survey contains fainter events with higher cadence, as well as lower flux uncertainty compared to the WFD survey. Unlike the PLAsTiCC submissions, we therefore carried out a separate augmentation for this survey mode and built a custom classifier for it, as discussed in Section 4.

We now compare the DDF test-set classification performance when using a classifier that is based on the augmented PLAsTiCC training set (which mixes WFD and DDF events) versus one trained on an augmented DDF-only training set. Figure 6 shows that the classifier optimized for the WFD test set obtains a worse performance on the DDF test set, with a higher log-loss (0.570 versus 0.384) and a lower recall for SNe II and SNe Ia. These results illustrate the vital need for matching augmented training sets to the characteristics of the different survey modes. It also strongly highlights the better classification performance that can be obtained for SNe in the DDF survey compared to the WFD survey.

5.2. Light-curve Length

The season length is an important factor for observing strategy, which can be tuned by taking additional observations in suboptimal conditions (such as at high airmass). We compared the classification performance of light curves of different lengths, as a proxy for season length. The right panel of Figure 7 shows that 94% of events in the test set have light-curve lengths between 50–175 days. We focus on this interval in the recall (left panel) and precision (middle panel) plots, as outside that range, the results are dominated by small number effects. As expected, events observed for longer are better characterized by the feature extraction step, and hence yield higher recall and precision. Again, we note that for a fixed total exposure time, a reduced season length could be compensated for with a higher cadence. Our findings support the minimum 5 month season length recommendation in Lochner et al. (2018, 2021).

5.3. Inter-night Gaps

The cadence of observation, as quantified by the inter-night gap when no observations are taken in any passband, is a critical factor in LSST observing strategy that impacts all transient science goals. To investigate this effect, we compared the performance of SNe with different median inter-night gaps. The left panel of Figure 8 shows that cadence has an important impact on SNe classification; events whose median inter-night gap is <3.5 days yield higher recall and precision. Such events comprise nearly 70% of the entire test set. These events are better sampled and thus have a higher light-curve quality. Moreover, for a fixed SN Ia recall of 80%, the core-collapse SN contamination is 6.8% for events whose median inter-night gap is <3.5 days, and 8.0% otherwise. These results support previous works such as Lochner et al. (2018, 2021) that call for SN Ia light curves to have frequent observations in order to reduce the uncertainty on the cosmological distance modulus.

However, the median inter-night gap does not fully capture the impact of gaps in the light curve. A 3.5 day median inter-night gap does not imply a uniform cadence; it is entirely possible that such light curves contain much larger gaps. To investigate the impact of such "gappy" light curves, we studied

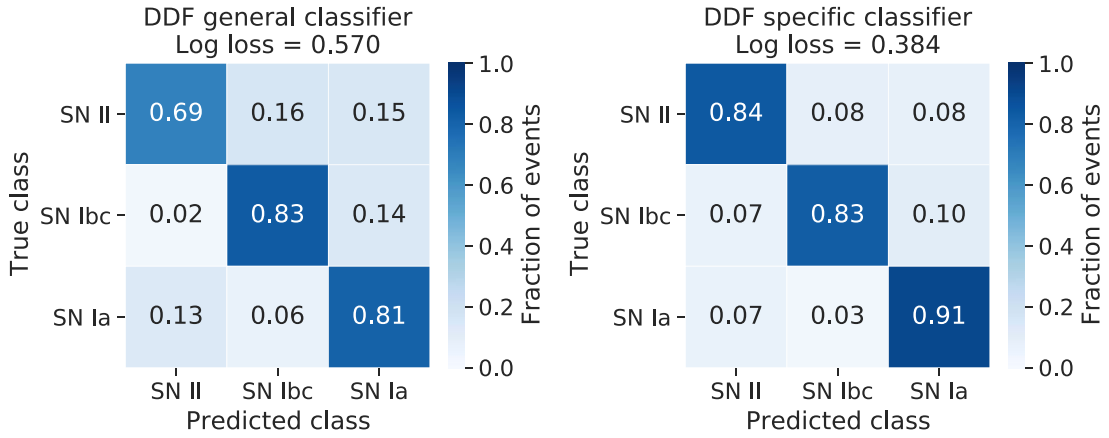

Figure 6. DDF test-set normalized confusion matrix for the classifier trained with (left panel) the general WFD+DDF augmented training set and (right panel) with the DDF-only augmented training set. The results show the importance of using an augmented training set customized for the specific survey mode characteristics.

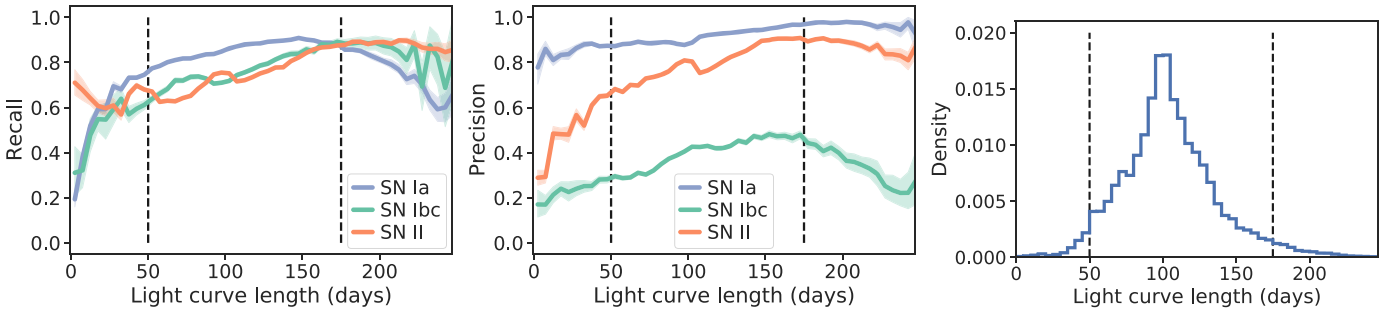

Figure 7. WFD test-set recall (left panel) and precision (middle panel) as a function of light-curve length per SNe class. The right panel shows the density of events as a function of light-curve length. Because of the low number of events in the tails of the distribution, we restrict our analysis to between 50 and 175 days (comprising 94% of the events). Recall and precision increase for longer light curves.

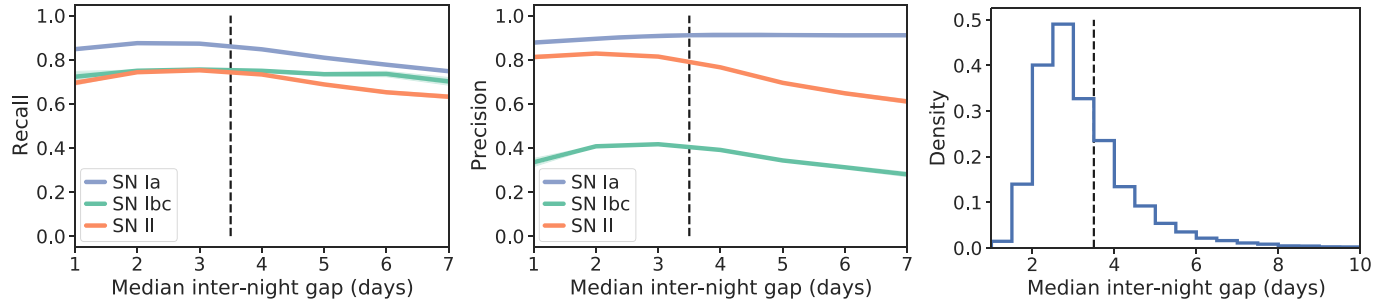

Figure 8. WFD test-set recall (left panel) and precision (middle panel) as a function of median inter-night gap per SNe class for events with light curves between 50 and 175 days long. In general, the recall and precision are higher for events whose median inter-night gap is <3.5 days (left side of the black line). The right panel shows the density of events as a function of median inter-night gap; ∼64% of SNe Ia, ∼63% of SNe Ibc, and ∼66% of SNe II from the test-set events have median inter-night gap <3.5 days and light-curve lengths between 50 and 175 days.

the classification performance as a function of the number of large gaps (<10 days) in a subsample of events with a median inter-night gap <3.5 days.

The upper-left panels of Figure 9 show that the recall and precision are broadly independent of the number of large gaps in a light curve.[13] We tested this with <20 day gaps and found similar results.

We expect that the reason for these surprising findings is that the GP fits can still constrain a light-curve fit sufficiently well if there are enough points on either side of large gaps. This is demonstrated in Figure 2, which shows an example of a GP fit to an event with four gaps >10 days, one of which is >20 days. We then compare the classification performance as a function of the length of the longest inter-night gap per light curve, to investigate at which point the performance degrades due to inability of GP fits to constrain a light-curve fit. The bottom panels of Figure 9 show that the recall and precision of SNe either slowly decrease or remain constant with the increase of the length of longest inter-night gap. While previous works

[13] SNe Ibc and SNe II have a small recall increase for a higher number of large gaps; we find that these events correspond to longer light curves at lower redshifts, which tend to have a higher recall for SNe Ibc and SNe II. Note that uncertainties are also larger for cases with a greater number of large gaps due to small number statistics.

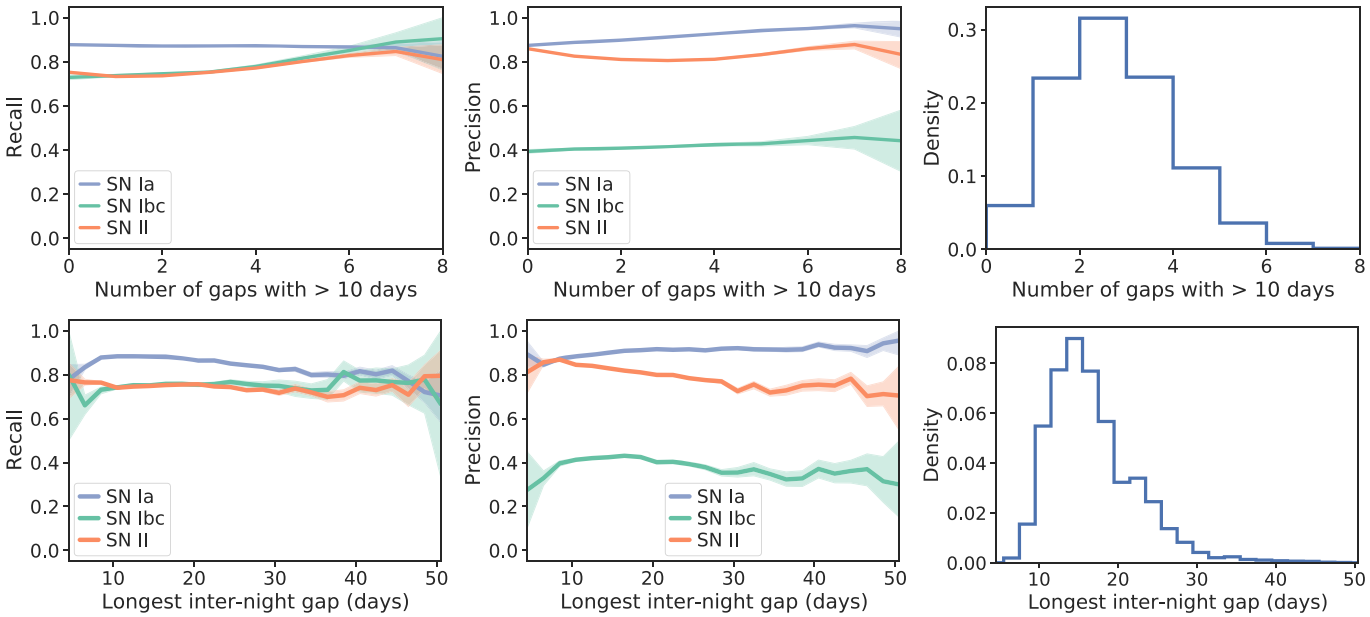

Figure 9. WFD test-set recall (left panels) and precision (middle panels) as a function of the number of gaps longer than 10 days (top row) and the length of the longest inter-night gap (bottom row), per SN class. We only included events with median inter-night gap <3.5 days and light curves between 50 and 175 days long. These results show that large gaps do not significantly impact the SNe classification for this subset. The right panel shows the density of events as a function of the number of large gaps (top row) and the length of the longest inter-night gap (bottom row). We note that the results in the tails of the distribution are dominated by small number effects.

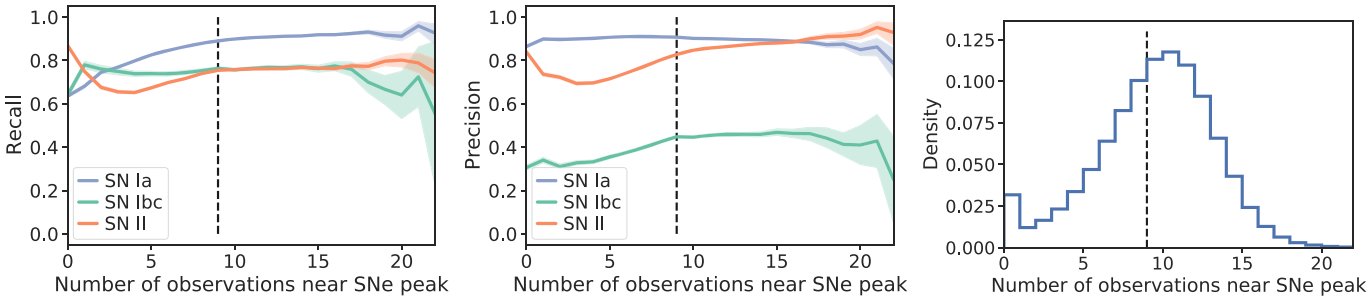

Figure 10. WFD test-set recall (left panel) and precision (middle panel) as a function of the number of observation between 10 days before and 30 days after the peak per SNe class. We only included events with median inter-night gap <3.5 days and light curves between 50 and 175 days long. In general, the recall and precision increase with the number of observations near the peak, until they reach an approximately constant value for events with ⩾9 observations. The right panel shows the density of events as a function of number of observations near peak. We note that, as with previous plots, the results in the tails of the distribution are dominated by small number effects.

recommended a regular cadence without inter-night gaps larger than 10–15 days (Lochner et al. 2018, 2021), we find that requiring a median inter-night gap of <3.5 days is sufficient for photometric classification methods using GPs that incorporate cross-band information to model the light curves and generate features.

We also find that 98% of DDF events have a median inter-night gap of <3.5 days, and hence the DDF sample performs uniformly well independently of the inter-night gap.

5.4. Observations near the Peak

Obtaining observations near the peak of an SN Ia light curve is generally considered critical to obtaining a reliable cosmological distance modulus. To investigate whether SNe classification has a similar requirement, we analyzed the classification performance as a function of the number of observations near the peak (defined as 10 days before and 30 days after the peak). We estimated the peak time as the moment that maximizes the GP fit predicted flux. Figure 10 shows that the recall and precision generally increase with the number of observations near the peak, reaching a constant value for events with more than nine observations. This improvement in performance is likely due to better characterization of light-curve shape. However, since we cannot predict when an SN will be observed, this result only further demonstrates the importance of frequent observations to increase the likelihood of obtaining observations near the peak. These results agree with Takahashi et al. (2020), who found that SNe light curves without observations near the peak were more often misclassified.

6. Discussion and Conclusions

We have presented a quantitative analysis of the impact of various factors related to the LSST observing strategy on the performance of SNe photometric classification, using the PLAsTiCC simulation. We use the photometric transient classification library `snmachine`, based on model-independent

wavelet features (instead of specialized features constructed using domain knowledge about SNe). In line with previous studies using the PLAsTiCC data, we confirm that augmentation for a number of aspects (the photometric redshift distribution per supernovae class, the distribution of the observing cadence, and the flux uncertainty distribution) is crucial for obtaining a representative training set for machine-learning classification.

Our classifier yields a similar performance to the top PLAsTiCC submission (Boone 2019; Hložek et al. 2020) and competitive results in core-collapse SN contamination (see below; Jones et al. 2017; Kessler & Scolnic 2017), which is essential for measurements of the dark energy equation-of-state parameter. We obtain a core-collapse SN contamination of 8.3% (for SNe predicted to be SNe Ia with >50% probability), which is comparable to the ∼5% contamination obtained in Jones et al. (2018) with Pan-STARRS SNe. This could be further improved by optimizing the classifier for SN Ia classification rather than overall classification performance, as was done in PLAsTiCC. Jones et al. (2018) demonstrated that this level of contamination provides competitive cosmological constraints when using a Bayesian methodology to marginalize over the contamination. Hence, we expect our contamination levels to also be acceptable for cosmology when used along with a Bayesian methodology such as Bayesian Estimation Applied to Multiple Species (Kunz et al. 2007; Lochner et al. 2013; Roberts et al. 2017; Jones et al. 2018).

Turning to the question of how observing strategy impacts classification, our results demonstrate the importance of customized training set augmentation for each LSST survey mode (WFD and DDF). We find that the season length is important—in general, better classification performance is obtained for longer light curves. This supports the minimum 5 month season length recommendation in Lochner et al. (2018, 2021). Further, we show that good classification performance requires a cadence with a median inter-night gap of <3.5 days. Surprisingly, however, we find that large gaps of >10 days do not impact the classification performance for events exhibiting such a cadence, due to the ability of the GP methods we use to interpolate such gaps effectively. Finally, a regular cadence that achieves >9 observations near the peak of the light curve provides effective classification performance. In Appendix B we show that these results also hold if we replace our classification predictions with the predictions obtained by Boone (2019), who used a different feature set and an independent classification framework with somewhat different augmentation choices.

These results provide guidance for further refinement of the LSST observing strategy on the question of SNe photometric classification. While the PLAsTiCC simulation used in this analysis has an outdated cadence, we expect our general conclusions to hold for any reasonable variation currently under consideration. Our augmentation and classification pipeline will be used in the future to study the SNe classification performance of more recent observing strategy simulations in detail.

Since the release of PLAsTiCC, new and more realistic observing strategy simulations have been released. These simulations include improvements to the scheduler, more realistic weather, and changes to the cadence in different bands. While new transient simulations using the more recent baseline observing strategy may result in different classification performance, we still expect our broad conclusions to remain unchanged. Future work will include investigating the dependence of classification performance on different observing strategy simulations.

With this paper, we publicly release the photometric transient classification library `snmachine`.[14] The library also contains some example Jupyter notebooks, which can be used to reproduce this work. In the future, the `snmachine` pipeline will be extended to facilitate the classification of other transient classes.

This paper has undergone internal review in the LSST Dark Energy Science Collaboration. The authors would like to thank Kyle Boone, Patrick D. Aleo, and Philippe Gris for their helpful comments and reviews.

Author contributions are listed below.

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

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

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

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

T.A.: software, data curation, and writing (editing).

R.B.: software.

We thank Christian N. Setzer for useful discussions and for setting up the initial code framework to handle the PLAsTiCC light curves and metadata. We also thank Gautham Narayan for useful discussions and for contributing to reviewing the updated `snmachine` software. This work was partially enabled by funding from the UCL Cosmoparticle Initiative. This work used facilities provided by the UCL Cosmoparticle Initiative; and we thank the HPC systems manager Edd Edmondson for his support. H.V.P. and R.B. were partially supported by the research environment grant "Gravitational Radiation and Electromagnetic Astrophysical Transients (GREAT)" funded by the Swedish Research council (VR) under Dnr 2016-06012, and the research project grant "Understanding the Dynamic Universe" funded by the Knut and Alice Wallenberg Foundation under Dnr KAW 2018.0067. 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. T.A. is supported in part by STFC.

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 BIS National E-infrastructure capital grants; and the UK particle

[14] https://github.com/LSSTDESC/snmachine

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. 2015), 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), pytest (Krekel et al. 2004), pywt (Lee et al. 2019a, 2019b), scikit-learn (Pedregosa et al. 2011), SciPy (Virtanen et al. 2020), seaborn (Waskom et al. 2020), snmachine (Lochner et al. 2016).

Appendix A
Redshifting Implementation for Augmentation

In this Appendix we provide further details of the augmentation procedure described in Section 4. In particular, we present the technique for redshifting a light curve, and derive the redshift limits for augmentation shown in Equation (4).

Consider a multiband SN light curve at redshift $z_{\rm ori}$, from which we want to create a synthetic multiband light curve at redshift $z_{\rm aug}$. For each epoch, the spectrum of the new synthetic SN is

$$f_{\lambda;\rm aug}(\lambda) = \frac{1+z_{\rm ori}}{1+z_{\rm aug}}\left[\frac{d_{\rm L}(z_{\rm ori})}{d_{\rm L}(z_{\rm aug})}\right]^2 \times f_\lambda\left(\frac{1+z_{\rm ori}}{1+z_{\rm aug}}\lambda\right), \quad \text{(A1)}$$

where λ is the observed wavelength, $d_{\rm L}$ is the luminosity distance, and f_λ is the spectrum of the original event. Note that the spectrum of the synthetic SN depends on the original spectrum evaluated at redshifted wavelengths. The two-dimensional GP fit described in Section 3.2 then models the convolution of the original spectrum with the *ugrizy* passbands to predict the measured flux. Thus, for each epoch of the synthetic SN, we estimated the flux in the original event at each redshifted passband b (where b = *u*, *g*, *r*, *i*, *z*, *y*) as

$$F_{\rm ori\,b} = \mathcal{GP}\left(\frac{1+z_{\rm ori}}{1+z_{\rm aug}}\lambda_{\rm b}\right), \quad \text{(A2)}$$

where $\mathcal{GP}$ represents the mean of the GP fit used to model the flux observations of the original SNe, and $\lambda_{\rm b}$ is the central wavelength of passband b. We calculated these central wavelengths using the LSST throughputs.[15] Similarly, we estimated the flux uncertainty in each passband as the uncertainty of the GP fit.

Finally, we adjusted the fluxes of the synthetic event and their uncertainties to the desired redshift $z_{\rm aug}$. We assumed a flat Λ cold dark matter cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1} and $\Omega_{\rm m} = 0.3$. We estimated the flux of the synthetic event in each passband b as

$$F_{\rm aug\,b} = \frac{1+z_{\rm ori}}{1+z_{\rm aug}}\left[\frac{d_{\rm L}(z_{\rm ori})}{d_{\rm L}(z_{\rm aug})}\right]^2 F_{\rm ori\,b}, \quad \text{(A3)}$$

and estimated its uncertainty similarly.

[15] https://github.com/lsst/throughputs

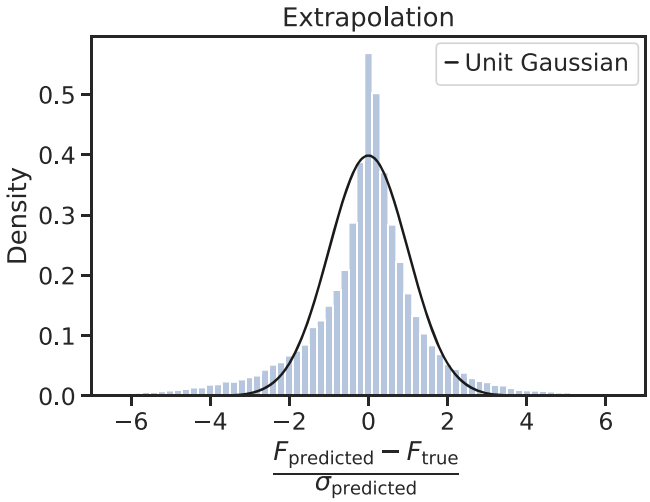

Figure 11. Distribution of the GP errors resulting from extrapolating GP fits to *u* or *y* passbands. $F_{\rm true}$ is the true flux of an observation, and $F_{\rm predicted}$ and $\sigma_{\rm predicted}$ are the flux and its uncertainty, respectively, predicted by a GP fit at the corresponding epoch and passband. An ideal error estimation results in a unit Gaussian (black line).

As previously discussed in Section 4.2, the GP fit is more reliable close to the observations. To test the GP extrapolation, for every SNe in the training set, we fitted a GP with the observations in the *ugriz* passbands. Then, we compared the observed flux in the *y* passband with the flux predictions of the GP fit at the same epochs. Additionally, we repeated this procedure to test the GP extrapolation in the *u* passband using the observations in the *grizy* passbands. Figure 11 shows that the GP is reliable despite underestimating some flux errors. Since the GP errors increase at wavelengths far from the original observation, we restricted our extrapolation to minimum $(\lambda_g - \lambda_u)$ and maximum $(\lambda_y - \lambda_z)$ wavelength ranges. Thus, when generating a synthetic SN at higher redshifts, we have that $\lambda_u - (1+z_{\rm ori})/(1+z_{\rm aug})\lambda_u \leqslant \lambda_g - \lambda_u$. Similarly, for events generated at lower redshifts, we obtain the redshift limits for augmentation presented in Section 4.2:

$$\begin{aligned} z_{\rm min} &= \max\left\{0, (1+z_{\rm ori})\left(2 - \frac{\lambda_{\rm z}}{\lambda_{\rm y}}\right)^{-1} - 1\right\} \\ &\approx \max\{0, 0.90\, z_{\rm ori} - 0.10\}, \\ z_{\rm max} &= (1+z_{\rm ori})\left(2 - \frac{\lambda_{\rm g}}{\lambda_{\rm u}}\right)^{-1} - 1 \\ &\approx 1.43\, z_{\rm ori} + 0.43. \end{aligned} \quad \text{(A4)}$$

Appendix B
Comparison with Other PLAsTiCC Classifiers

Section 5 presented the results of our classifier on the impact of observing strategy on photometric classification. In this Appendix, we show that our results are generalizable beyond our classification pipeline, by replacing the our classification predictions with those obtained by Boone (2019). We use the publicly available predictions for SNe Ia, SNe Ibc, and SNe II in the test set;[16] we choose the predictions obtained with a

[16] http://supernova.lbl.gov/avocado_plasticc/predictions/predictions_plasticc_test_flat_weight.csv

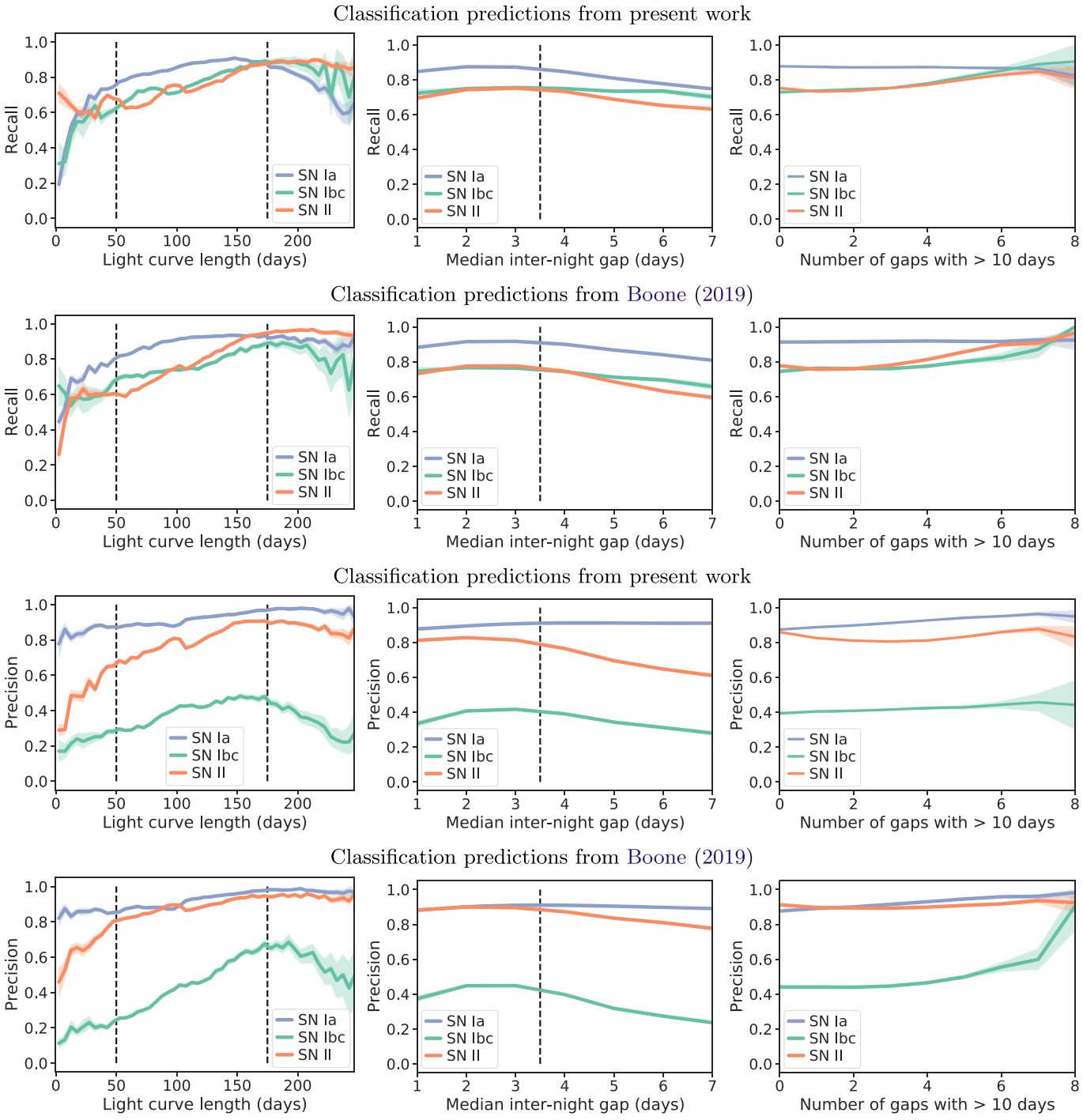

Figure 12. WFD test-set recall (top two rows) and precision (bottom two rows) as a function of light-curve length (left panels), median inter-night gap (middle panels), and number of gaps larger than 10 days (right panels), per SN class. In the first and third rows, we reproduce the results of this work previously shown in Figures 7, 8, and 9. In the second and fourth rows, we show the classification predictions obtained by Boone (2019).

classifier optimized on the log-loss metric, which equally weights all of the PLAsTiCC classes ($w_i = 1$ in Equation (1)). The choice of this flat-weighted metric reduces the impact of additional classes up-weighted in the original challenge, but unused in the present work. Figures 12 and 13 show that the classifier used in Boone (2019) has the same performance behavior as ours. This further indicates that our conclusions are general and not an artifact of our classification architecture.

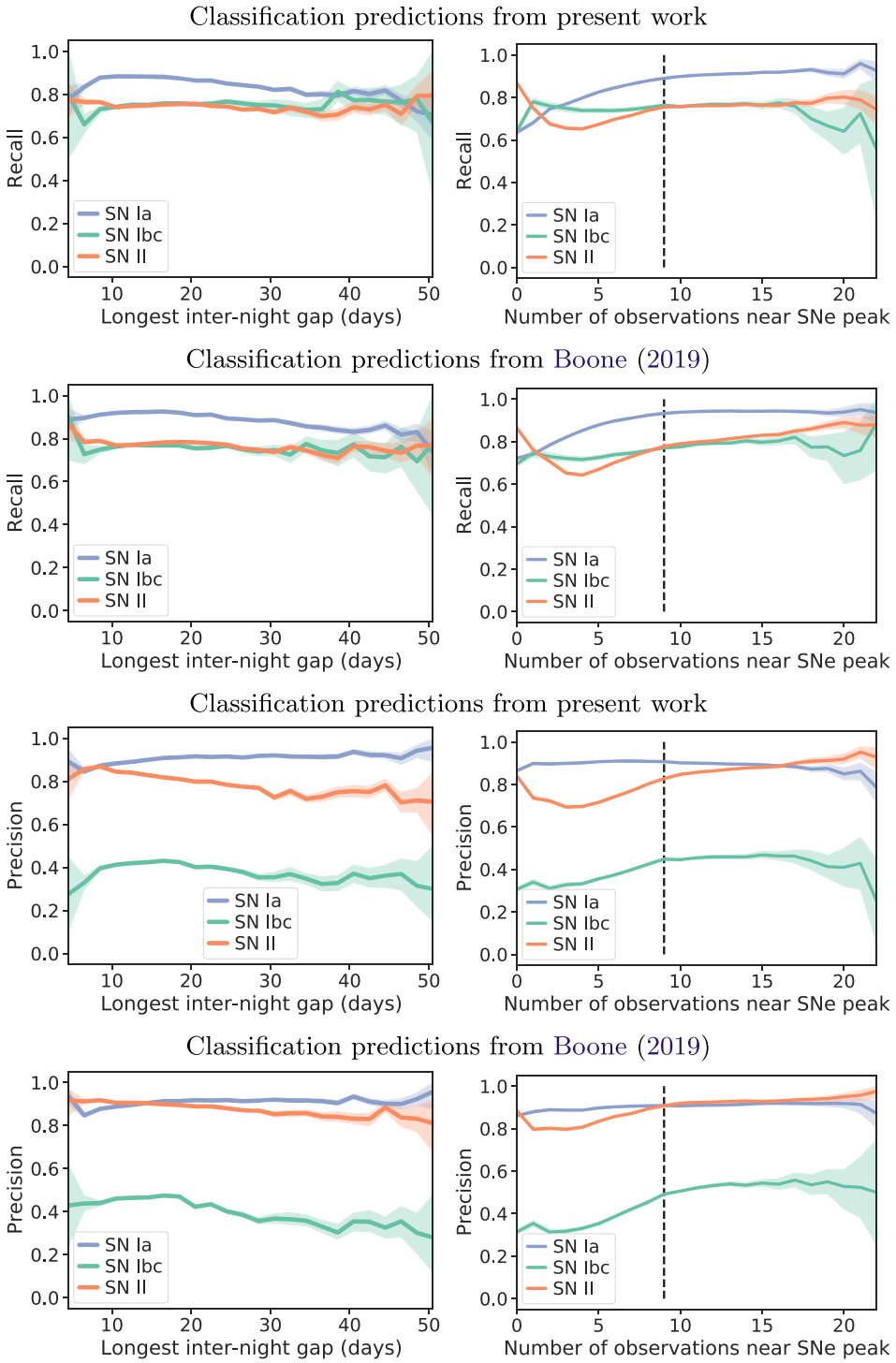

Figure 13. WFD test-set recall (top two rows) and precision (bottom two rows) as a function of the length of the longest inter-night gap (left panels) and the number of observations between 10 days before and 30 days after the peak length (right panels), per SN class. In the first and third rows, we reproduce the results of this work previously shown in Figures 9 and 10. In the second and fourth rows, we show the classification predictions obtained by Boone (2019).

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
Tarek Allam, Jr. https://orcid.org/0000-0002-8117-9254
Rahul Biswas https://orcid.org/0000-0002-5741-7195

References

Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2015, ITPAM, 38, 252
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. D. Lee, https://proceedings.neurips.cc/paper/2016/file/621bf66ddb7c962aa0d22ac97d69b793-Paper.pdf
Boone, K. 2019, AJ, 158, 257
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
Chen, N., Qian, Z., & Meng, X. 2013, Math. Probl. Eng., 2013, 461983
Friedman, J. H. 2001, Ann. Stat., 29, 1189
Gonzalez, O. A., Clarkson, W., Debattista, V. P., et al. 2018, arXiv:1812.08670
Guillochon, J., Parrent, J., Kelley, L. Z., & Margutti, R. 2017, ApJ, 835, 64
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
Istas, J. 1992, AIHPB, 28, 537
Ivezić, Ž., Jones, L., & Ribeiro, T. 2018, Call for White Papers on LSST Cadence Optimization LSST Document 28382, https://docushare.lsst.org/docushare/dsweb/Get/Document-28382
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., Ivezic, 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,, 3146
Kessler, R., Bassett, B., Belov, P., et al. 2010, PASP, 122, 1415
Kessler, R., Becker, A. C., Cinabro, D., et al. 2009, ApJS, 185, 32
Kessler, R., Conley, A., Jha, S., & Kuhlmann, S. 2010a, arXiv:1001.5210
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. Scmidt (Amsterdam: IOS Press), 87
Krekel, H., Oliveira, B., Pfannschmidt, R., et al. 2004, pytest v6.2.2, https://github.com/pytest-dev/pytestpytest v6.2.2
Kunz, M., Bassett, B. A., & Hlozek, R. 2007, PhRvD, 75, 103508
Laine, S., Martinez-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 1.1.1, Zenodo, doi:10.5281/ZENODO.3510098
Lochner, M., Bassett, B. A., Varughese, M., et al. 2013, JCAP, 2013, 039
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. 2021, arXiv:2104.05676
Lochner, M., Scolnic, D. M., Awan, H., et al. 2018, arXiv:1812.00515
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)
McKinney, W. 2010, in Proceedings of the 9th Python in Science Conf., ed. S. van der Walt & J. Millman, 56
Malz, A. I., Hložek, R., Allam, T., et al. 2019, AJ, 158, 171
Mockus, J., Tiesis, V., & Zilinskas, A. 1978, Towards Glob. Optim., 2, 117
Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, PASP, 131, 118002
Narayan, G., Zaidi, T., Soraisam, M. D., et al. 2018, ApJS, 236, 9
pandas development team 2020, pandas-dev/pandas: Pandas, 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 Phil. Mag. J. Sci., 2, 559
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825
PLAsTiCC Modelers 2019, Libraries & Recommended Citations for using PLAsTiCC Models, Zenodo, doi:10.5281/ZENODO.2612896
PLAsTiCC TeamPLAsTiCC Modelers 2019, Unblinded Data for PLAsTiCC Classification Challenge, Zenodo, doi:10.5281/ZENODO.2535746
Pope, C. A. 2019, PhD thesis, Univ. Leeds, https://etheses.whiterose.ac.uk/25066/
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., Lochner, M., Gris, P., et al. 2018, arXiv:1812.00516
Snoek, J., Larochelle, H., & Adams, R. P. 2012, arXiv:1206.2944
Sooknunan, K., Lochner, M., Bassett, B. A., et al. 2021, MNRAS, 502, 206
Swann, E., Sullivan, M., Carrick, J., et al. 2019, Msngr, 175, 58
Takahashi, I., Suzuki, N., Yasuda, N., et al. 2020, PASJ, 72, 89
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
Varughese, M. M., von Sachs, R., Stephanou, M., & Bassett, B. A. 2015, MNRAS, 453, 2849
Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020, ApJ, 905, 94
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261
Waskom, M., Botvinnik, O., Gelbart, M., et al. 2020, mwaskom/seaborn: v0.11.0 (Sepetmber 2020), Zenodo, doi:10.5281/ZENODO.4019146
Zhang, H., Si, S., & Hsieh, C.-J. 2017, arXiv:1706.08359