UNIVERSITY COLLEGE LONDON

Faculty of Mathematics and Physical Sciences
Department of Physics & Astronomy

ENABLING SUPERNOVA COSMOLOGY
WITH LARGE TIME-DOMAIN SURVEYS

CATARINA SAMPAIO ALVES

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

SUPERVISORS:
prof. Hiranya Peiris
prof. Jason Mcewen

EXAMINERS:
prof. Richard Ellis
prof. Mark Sullivan

April 5, 2023
I, Catarina Sampaio Alves, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Date: April 5, 2023
Name: Catarina Sampaio Alves
Signature: _______________________________
If we had no winter, the spring would not be so pleasant: if we did not sometimes taste of adversity, prosperity would not be so welcome.

— Anne Bradstreet

“Space,” says the introduction to The Hitchhiker’s Guide, “is big. Really big. You just won’t believe how vastly, hugely, mind-bogglingly big it is.”

— Douglas Adams

Valeu a pena? Tudo vale a pena
Se a alma não é pequena.

Was it worth it? Everything is worthy
If the soul is not small.

— Fernando Pessoa
Abstract

It was recently discovered that the expansion of the Universe is accelerating. Type Ia supernovae (SN Ia) were crucial for this discovery and to constrain cosmological parameters. Current and upcoming large time-domain surveys will revolutionise the field by discovering at least one order of magnitude more SNe than the currently available datasets, which will lead to tighter cosmological parameter constraints. However, these surveys will also bring challenges due to the volume of data they observe. Thus, we require new methods to analyse, understand, and extract cosmological constraints from the data. In particular, the upcoming Rubin Observatory Legacy Survey of Space and Time (LSST) must rely on photometric classification to identify the observed SNe, instead of the traditional spectroscopic confirmation. In this thesis, we develop a methodology to perform this photometric classification based on dataset augmentation, wavelet features, and a machine learning classifier. Specifically, we find that augmenting the training set to make its features similar to the dataset we want to classify is crucial. Next, we use our methodology to measure the impact of the LSST observing strategy in photometric classification; this work contributes towards the community-focused optimisation of the observing strategy. Since the above work used simulated data, we next set a benchmark for the classification performance of our approach using the Zwicky Transient Facility real data; this is a precursor survey to LSST. Finally, we present a proof-of-concept of a neural network to predict lensed SN Ia parameters from light curves and images of those events.
Large time-domain surveys are a joint effort of many researchers and countries that have the potential to revolutionise the field in which they focus. Due to this significance it is important to maximise the scientific outputs of the surveys. This thesis focuses on enabling supernovae cosmology with large time-domain surveys through optimising survey strategy and developing machine learning approaches to classify supernovae and to estimate lensed supernovae parameters. Our work on the impact of the survey observing strategy for photometric classification of supernovae (Chapters 3 and 4) contributes towards the process of community-focused experimental design and optimisation of the Rubin Observatory Legacy Survey of Space and Time (LSST) observing strategy. This work also achieved academic impact through one published and one submitted scientific article, and over ten professional presentations in four continents, including invited talks and seminars. Moreover, the work earned the PhD candidate builder status in the LSST Dark Energy Science Collaboration, a recognition rarely accomplished by students and only awarded to 44 people in total so far. Additionally, the work lead to new international collaborations (such as Chapters 5 and 6) for the PhD candidate. Moreover, the data simulations developed for this thesis gathered significant interest in the community due to their broad applicability to other problems and for developing methodologies using the most up-to-date LSST simulations. Overall, the work developed in this thesis can be applied to future surveys to similarly optimise their scientific outputs.

The machine learning methodology developed in this thesis is applied to supernovae but it has broader applications. Since the methods are general, they can be adapted to other time-series data that requires classification. Thus, in aca-
In industrial and commercial settings, possible applications are identification of abnormalities on time-series medical data, identification of the best crop for an area based on historical satellite images, and prediction of the future direction of financial time-series data. Recent works on the medical field show the use of machine learning to assign patients to groups based on their predicted outcome and trajectory (e.g. Aguiar et al., 2022) and automated classification of electroencephalogram records for epilepsy diagnosis (e.g. Silva Lourenço et al., 2021). The UCR time series archive (Dau et al., 2019) stores datasets from several time-series classification problems where the machine learning methodology developed in this thesis could be used.

In addition to the above, the work presented on this thesis achieved impact through public outreach. It lead to two participations in the Three Minute Thesis competition, and over ten presentations about more general astronomy concepts during science festivals and public tours to the UCL Observatory.
Many people helped me through the PhD and made it an amazing experience. In this section, I express my gratitude and appreciation to them.

First, I would like to express my deepest gratitude to my amazing supervisors Hiranya Peiris and Jason McEwen. For these past years, you have been supportive, knowledgeable, and inspiring mentors. Thank you for your invaluable guidance, in particular during the pandemic, and for all the opportunities and insightful scientific discussions you provided me. I am extremely grateful for the amount of time you have invested to make me a better researcher. It has been a privilege to work with you and being your student.

Secondly, I would like to express my deepest appreciation to Michelle Lochner. Thank you for the engaging scientific and technical discussions and all the support through the PhD. I feel grateful for meeting and working with you. I would also like to recognise all the other brilliant scientists I was lucky to collaborate with - Rahul Biswas, Gautham Narayan, Richard Kessler, Ariel Goobar, and Brian Nord. Thank you all for the help and the opportunity to work in exciting projects. Additionally, I am grateful to everyone in the Rubin Observatory Legacy Survey of Space and Time Dark Energy Science Collaboration for an amazing, vibrant, and inclusive community that makes phenomenal cosmology research.

Many thanks to all the friends I made at UCL. In particular Francesca Gerardi for all the support and life conversations, Benjamin Stölzner for coffee knowledge and help during conferences and thesis-writing, Matt Scourfield for coming to my events and being the life of the party, Lillian Guo for understanding me, Pascal Förster for great get-togethers, Max von Wietersheim-Kramsta for your camaraderie, Tarek Allam Jr for help with snmachine, Davide Piras, Kiyam Lin, and Xie
Cheng for varied discussions at the Hub, Harry Johnston for random desk talks, Chris Pedersen for hanging out in New York, and Luisa Lucie-Smith, Constance Mahony, Martin Rey, Holly Andrews, and James Farr for showing me the ropes at UCL. I would also like to acknowledge all other people at group A and UCL that I had the chance to meet through societies, events, or friends and that made my PhD experience richer.

Special thanks to Carlos Martins for showing me links between astronomy and physics and kindling my interest for astronomy. Thank you for the research projects in cosmology we collaborated on and that led me to this PhD. I also appreciate your ongoing support.

A fond thanks to all my friends who supported me through these years. Thanks Nick, Tomáš, and Stefan for being such cool housemates and all the fun parties, Maria for being my fashionable best friend and checking up on me, Melo for our month-long trip and deep discussions, Zé for wholesome and silly conversations that make me smile, Ed for meeting at the start of our PhD and being close friends at the end, and Henrique for showing me different perspectives on a number of subjects. I also gratefully acknowledge Sofia, João, José, Ben, Deniz, and many more, for listening to my thoughts and rants about the PhD.

Thank you Daniel for always being there for me, and making my life happier. Our daily talks relaxed me and broadened my horizons. I appreciate all the time you spent helping me and trying to understand my research. You have been wonderful.

A big thanks to my family. Agradeço aos meus pais, Adélia e António, por todo o apoio ao longo dos anos e por me encorajarem a seguir as minhas paixões, mesmo que estas me levassem para longe. Mãe, a tua força e compaixão inspira-me a ser uma pessoa melhor todos os dias. Pai, obrigada por me introduzires à física. Obrigada André por seres um irmão fantástico e me acolheres sempre que volte a casa. Agradeço também a todos os meus primos, tios, e avós. Um obrigado especial para a avó Alzira que sempre me apoiou e incentivou a viver a minha vida. Obrigada por todos os momentos que passámos juntas e todas as histórias que me contaste. Gostava que estivesse aqui para me ver agora.
UCL Research Paper Declaration Form: referencing the doctoral candidate’s own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

1. For a research manuscript that has already been published (if not yet published, please skip to section 2):

| a) Where was the work published? (e.g. journal name) | The Astrophysical Journal Supplement Series |
| b) Who published the work? (e.g. Elsevier/Oxford University Press): | American Astronomical Society |
| c) When was the work published? | 18/01/2022 |
| d) Was the work subject to academic peer review? | Yes |
| e) Have you retained the copyright for the work? | Yes |

[If no, please seek permission from the relevant publisher and check the box next to the below statement]:

☐ I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.

2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3):

| a) Has the manuscript been uploaded to a preprint server? (e.g. medRxiv): | Please select. If yes, which server? Click or tap here to enter text. |
| b) Where is the work intended to be published? (e.g. names of journals that you are planning to submit to) | Click or tap here to enter text. |
| c) List the manuscript’s authors in the intended authorship order: | Click or tap here to enter text. |
| d) Stage of publication | Please select. |
3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4):

<table>
<thead>
<tr>
<th>Name</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catarina S. Alves</td>
<td>software, validation, formal analysis, investigation, data curation, writing (original draft), visualization</td>
</tr>
<tr>
<td>Hiranya V. Peiris</td>
<td>conceptualization, methodology, validation and interpretation, supervision, writing (original draft; review &amp; editing), funding acquisition</td>
</tr>
<tr>
<td>Michelle Lochner</td>
<td>conceptualization, methodology, software, validation and interpretation, writing (original draft; review &amp; editing)</td>
</tr>
<tr>
<td>Jason D. McEwen</td>
<td>conceptualization, methodology, validation and interpretation, supervision, writing (reviewing &amp; editing)</td>
</tr>
<tr>
<td>Tarek Allam Jr</td>
<td>software, data curation, writing (editing)</td>
</tr>
<tr>
<td>Rahul Biswas</td>
<td>software</td>
</tr>
</tbody>
</table>

4. In which chapter(s) of your thesis can this material be found?

Chapter 3

5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

<table>
<thead>
<tr>
<th>Role</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candidate</td>
<td>02/11/2022</td>
</tr>
<tr>
<td>Supervisor/ Senior Author (where appropriate)</td>
<td>2/11/2022</td>
</tr>
</tbody>
</table>
UCL Research Paper Declaration Form: referencing the doctoral candidate’s own published work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:
- have been uploaded to a preprint server;
- are in submission to a peer-reviewed publication;
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.

<table>
<thead>
<tr>
<th>1. For a research manuscript that has already been published (if not yet published, please skip to section 2):</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a)</strong> Where was the work published? (e.g. journal name)</td>
</tr>
<tr>
<td><strong>b)</strong> Who published the work? (e.g. Elsevier/Oxford University Press):</td>
</tr>
<tr>
<td><strong>c)</strong> When was the work published?</td>
</tr>
<tr>
<td><strong>d)</strong> Was the work subject to academic peer review?</td>
</tr>
<tr>
<td><strong>e)</strong> Have you retained the copyright for the work?</td>
</tr>
</tbody>
</table>

[If no, please seek permission from the relevant publisher and check the box next to the below statement]:

☐ I acknowledge permission of the publisher named under 1b to include in this thesis portions of the publication named as included in 1a.

<table>
<thead>
<tr>
<th>2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3):</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a)</strong> Has the manuscript been uploaded to a preprint server? (e.g. medRxiv):</td>
</tr>
<tr>
<td><strong>b)</strong> Where is the work intended to be published? (e.g. names of journals that you are planning to submit to)</td>
</tr>
<tr>
<td><strong>c)</strong> List the manuscript's authors in the intended authorship order:</td>
</tr>
<tr>
<td><strong>d)</strong> Stage of publication</td>
</tr>
</tbody>
</table>
### 3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4):

<table>
<thead>
<tr>
<th>Name</th>
<th>Contributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catarina S. Alves</td>
<td>software, validation, formal analysis, investigation, data curation, writing (original draft), visualization</td>
</tr>
<tr>
<td>Hiranya V. Peiris</td>
<td>conceptualization, methodology, validation and interpretation, supervision, writing (original draft; re-view &amp; editing), funding acquisition</td>
</tr>
<tr>
<td>Michelle Lochner</td>
<td>conceptualization, methodology, validation and interpretation, writing (original draft; review &amp; editing)</td>
</tr>
<tr>
<td>Jason D. McEwen</td>
<td>conceptualization, methodology, validation and interpretation, supervision, writing (review)</td>
</tr>
<tr>
<td>Richard Kessler</td>
<td>updating code for SNANA simulations, writing (review)</td>
</tr>
</tbody>
</table>

### 4. In which chapter(s) of your thesis can this material be found?

Chapter 4

### 5. e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/senior author unless this is not appropriate, e.g. if the paper was a single-author work):

<table>
<thead>
<tr>
<th>Candidate:</th>
<th>Date:</th>
<th>02/11/2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervisor/ Senior Author (where appropriate):</td>
<td>Date:</td>
<td>2/11/2022</td>
</tr>
</tbody>
</table>
## Contents

1 Introduction 
   1.1 The Universe ........................................ 33  
      1.1.1 General Relativity ................................. 34  
      1.1.2 Composition of the Universe ...................... 36  
      1.1.3 Expansion History ................................. 37  
      1.1.4 Distances ........................................ 38  
   1.2 Supernovae ............................................ 40  
      1.2.1 Types of Supernovae ................................. 41  
      1.2.2 SN Ia as standardisable candles .................. 46  
      1.2.3 Photometric classification ......................... 47  
      1.2.4 Lensed supernovae ................................ 49  
   1.3 Large Surveys ......................................... 50  
      1.3.1 Dark Energy Survey ................................. 50  
      1.3.2 Zwicky Transient Facility ......................... 51  
      1.3.3 Legacy Survey of Space and Time .................. 52  
      1.3.4 Time-Domain Extragalactic Survey ................ 57  
   1.4 Outline of the thesis ................................ 59  
      1.4.1 Candidate contributions ............................ 60  

2 Background methodology ................................. 63  
   2.1 Machine learning algorithms ............................ 63  
      2.1.1 Overview ........................................... 63  
      2.1.2 Gradient Boosting Decision Trees .................. 65  
      2.1.3 Gaussian Processes ................................ 67  

17
2.1.4 Neural Networks ........................................... 68
2.2 Wavelets ...................................................... 70
2.3 LSST Simulations ............................................ 71
  2.3.1 Observing Strategy Simulator ......................... 71
  2.3.2 Supernovae Simulations ................................. 72

3 Considerations for optimising photometric classification of supernovae from the Rubin Observatory 75
  3.1 Overview .................................................... 75
  3.2 Introduction ................................................. 76
  3.3 PLAsTiCC Dataset ........................................... 78
  3.4 Classification pipeline .................................... 79
    3.4.1 Light Curve Preprocessing ............................. 79
    3.4.2 Gaussian Process Modelling of Light Curves ........ 80
    3.4.3 Feature Extraction ....................................... 81
    3.4.4 Classification ............................................ 82
  3.5 Augmentation ............................................... 85
    3.5.1 Number and Class Balance of Synthetic Events ...... 87
    3.5.2 Redshift Augmentation .................................. 87
    3.5.3 Generating Realistic Synthetic Observations ......... 88
    3.5.4 Photometric Redshift .................................... 90
    3.5.5 Computational Resources ................................. 90
  3.6 Results and Implications for Observing Strategy .......... 90
    3.6.1 Survey Mode-specific Augmentation and its Effect on Performance ............................................ 91
    3.6.2 Light Curve Length ........................................ 92
    3.6.3 Inter-night Gaps .......................................... 93
    3.6.4 Observations Near Peak .................................. 95
  3.7 Discussion and Conclusions .................................. 96

4 Impact of Rubin Observatory cadence choices on supernovae photometric classification 99
  4.1 Overview .................................................... 99
6 Inference of lensed supernovae parameters 133
6.1 Overview .................................................. 133
6.2 Introduction .................................................. 134
6.3 Simulation of lensed supernovae .................. 136
6.4 Regression pipeline ....................................... 137
   6.4.1 Data Processing ..................................... 137
   6.4.2 Data Augmentation ................................. 138
   6.4.3 Regression ZipperNet ...................... 139
6.5 Results .................................................. 141
6.6 Discussion and Conclusion ......................... 143

7 Conclusions 145
7.1 Summary .................................................. 145
7.2 Outlook .................................................. 146

Appendices 149

A Appendix to Chapter 3 149
   A.1 Redshifting Implementation for Augmentation .... 149
   A.2 Comparison with other PLAsTiCC classifiers .... 151

B Appendix to Chapter 4 155
   B.1 Simulated CC SN rates .............................. 155
   B.2 Augmentation details and Classification Hyperparameters .... 155
   B.3 Computational Resources .......................... 158
   B.4 Well-measured Type Ia Supernovae .............. 160

Bibliography 162
1.1 Model light curves in the $ugrizy$ passbands of the Rubin Observatory Legacy Survey of Space and Time. Each row presents a model of a different extragalactic event; the first 6 rows show several types of SNe, the sixth shows a tidal disruption event (TDE), the seventh shows a kilonova (KN), and the eighth shows an active galactic nuclei. The image is from Kessler et al. (2019).

1.2 Illustration of the brighter-bluer (left panel) and brighter-broader (right panel) relationship of SNe Ia for nearby (blue open circles) and distant $(z < 0.8$, black filled circles) events. In both plots, brighter SNe Ia are toward the bottom and the behaviour is consistent between nearby and distant SNe Ia. The images are from Astier et al. (2006).

1.3 Example of a Type Ia SN light curve. Each colour represents observations in different passbands. The intensity of the light is measured in flux.

1.4 Forecast of the constraints on the dark energy equation of state $(w = w_0 + w_a (1 - a))$ from LSST probes after one year of data (Y1; left panel) and the full ten year survey (Y10; right panel), from each probe individually, and the joint forecast. 68% confidence intervals are shown in all cases and the plotted quantities $\Delta w_0$ and $\Delta w_a$ are the difference between $w_0$ and $w_a$ and their fiducial values of $-1$ and $0$, respectively (The LSST Dark Energy Science Collaboration et al., 2018, 2021).
2.1 Scheme of a GBDT. Decision trees are sequentially added to create an ensemble learning algorithm with a better performance than the individual decision trees.

2.2 Scheme of a neural network with an input layer $i$ with $n$ inputs, three hidden layers $h$, and an output layer with $m$ outputs. Figure from https://www.wandb.com/articles/fundamentals-of-neural-networks.

2.3 Three versions of the Symlet wavelet, one in each colour.

2.4 Scheme of the SNANA simulation models. This is the Figure 13 from Kessler et al. (2009b).

3.1 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 colours.

3.2 SN Ibc light curve, where the points show the observations, along with their errorbars, 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 the passbands (shown in different colours). The two-dimensional GP infers the light curve in passbands where there are no (or few) observations, unlike the one-dimensional GP.

3.3 Host galaxy photometric redshift distribution per supernova class, where SN Ia, SN Ibc, SN 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 3.5) is close to the desired test distribution.
3.4 The left panel shows a SN Ibc light curve where the points show the observations, along with their errorbars, 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 3.5. 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.

3.5 WFD test set normalised 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.

3.6 DDF test set normalised 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 customised for the specific survey mode characteristics.

3.7 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.

3.8 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 SN Ia, ~ 63% of SN Ibc, and ~ 66% of SN II from the test set events have median inter-night gap < 3.5 days and light curve lengths between 50 and 175 days.
3.9 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 SNe 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.

3.10 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.

4.1 Footprint of the baseline cadence with the number of WFD observations in the first three years (1.5 years of non-rolling followed by 1.5 years 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 to the ‘background’.

4.2 Host galaxy photometric redshift distribution per supernova class for Y0-3 baseline, where SN Ia, SN Ibc, 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.
4.3 Normalised test-set confusion matrix for the classifier trained on the augmented training set of the $\text{Y1.5-3 baseline}$ (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. ..

4.4 Normalised test-set confusion matrices for the classifier trained on the augmented training set of the $\text{Y1.5-3 baseline}$ 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. ..

4.5 Normalised test-set confusion matrix for the classifier trained on the augmented training set of the $\text{Y0-3 baseline}$ (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. ...............

4.6 Test-set recall (left panel) and precision (right panel) as a function of light curve length per SNe class for $\text{Y0-3 baseline}$. 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 only present the results for bins with more than 300 events. ...............

4.7 Test-set density of events as a function of light curve length for $\text{Y1.5-3 baseline vs no-roll (left panel), and Y0-3 baseline vs presto-color (right panel)}$. The dashed lines mark the high performance region between 50 and 175 days. ...............

25
4.8 Test-set recall (left panel) and precision (right panel) as a function of inter-night gap per SNe class for 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 only present the results for bins with more than 300 events.

4.9 Test-set density of events as a function of inter-night gap for Y1.5-3 baseline vs no-roll (left panel), and Y0-3 baseline vs presto-color (right panel).

4.10 Test-set recall (left panel) and precision (right panel) as a function of the length of the longest inter-night gap per SNe class for 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 only present the results for bins with more than 300 events. The reduced performance below 8 days corresponds to less than 5% of the events which tend to have very short light curves and therefore are not well classified.

4.11 Test-set recall (left panel) and precision (right panel) as a function of number of observations near peak per SNe class for 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 only present the results for bins with more than 300 events.

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

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

5.1 SN Ibc light curve, where the points show the observations. The errorbars are too small to be visible. This light curve illustrates the high ZTF cadence and that some of the events in the dataset lack observations in the i passband.
5.2 Test set normalised confusion matrix for the classifier trained on the balanced training set and its log loss performance. Each event is assigned to the class with the highest prediction.

6.1 Diagram of the ZipperNet architecture taken from Morgan et al. (2022).

6.2 Example of an image (top left) and its augmentations. Each column corresponds to a rotation of $0^\circ$, $90^\circ$, $180^\circ$, and $270^\circ$, respectively from the left to the right. The second row corresponds to an horizontal rotation of the first row.

6.3 Diagram of the regression ZipperNet architecture. The convolution branch is shown in blue and the LSTM branch in green. Once the branches join, the layers are red. The convolutional layer receives $griz$ images while the LSTM receive light curves. The final output is a 3-dimensional array, where each element corresponds to a different regression parameter. See Table 6.1 for the specifications of each layer. We made this visualisation with diagrams.net.

6.4 Learning curve of the regression DeepZipper. The lowest overall mean square error (MSE) of the test set is at epoch 39. We use the neural network at this point as our fully trained regression DeepZipper.

6.5 Regression ZipperNet predictions on the test set for lens mass ellipticity ($e_1, e_2$) and peak time of the SNe ($t_{\text{peak}}$) compared to their truth values. The grey line indicates a perfect prediction. While the regression ZipperNet has some success in predicting $t_{\text{peak}}$, it is unable to estimate the lens mass ellipticity parameters.

6.6 The right panel shows the learning curve of the modified regression DeepZipper that only predicts the lens ellipticity parameter $e_2$. The left panel shows the gradient flow of the neural network; each line represents the log average of the absolute value of the gradient of a training epoch per layer. If these lines were zero, the gradient would be null and the neural network would be unable to update its weights.
A.1 Distribution of the GP errors resulting from extrapolating GP fits to $u$ or $y$ passbands. $F_{\text{true}}$ is the true flux of an observation, and $F_{\text{predicted}}$ and $\sigma_{\text{predicted}}$ are the flux and its uncertainty predicted by a GP fit at the corresponding epoch and passband. An ideal error estimation results in a unit Gaussian (black line).

A.2 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 SNe class. In the first and third rows we reproduce the results of this work previously shown in Figures 3.7, 3.8, and 3.9. In the second and fourth rows we show the classification predictions obtained by Boone (2019).

A.3 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 SNe class. In the first and third rows we reproduce the results of this work previously shown in Figures 3.9 and 3.10. In the second and fourth rows we show the classification predictions obtained by Boone (2019).

B.1 Host galaxy photometric redshift distribution per supernova class, where SN Ia, SN Ibc, 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.
List of Tables

3.1 Breakdown of the number of SNe per class used in this work (see simulation details in Kessler et al., 2019). For each class, the number of events in the training and test set is shown. 79

3.2 Optimised hyperparameter values used for the LightGBM model. A description of the hyperparameters is given in the library documentation. 83

3.3 Confusion matrix for binary classification. 84

4.1 Breakdown of the number of SNe per class and observing strategy used in this work. (Left) events simulated between years 1.5 and 3 of the survey. (Right) events simulated between years 0 and 3 of the survey. 106

5.1 Breakdown of the number of SNe per class used in this work for the training and test sets. We show both the number of available training events for future work and the balanced subset we use in this work. 127

5.2 Optimised hyperparameter values used for the LightGBM model. 129
6.1 Layer specifications of the regression ZipperNet whose architecture is shown in Figure 6.3. The following parameters are shorthanded: kernel size \((k)\), padding \((p)\), stride \((s)\), value added to the denominator of the batch normalisation for numerical stability \((\text{eps})\), momentum \((m)\), and hidden units \((h)\). See the library documentation for a description of these hyperparameters and layers. Changes in size of the data representation passing through the layers are indicated with arrows. We indicate the Rectified Linear Unit (ReLU) activation functions with superscript ‘†’ in the name of the layers.

6.2 Mean square error of the test set for each individual parameter and overall. The mean square error is presented for the scaled values of regression parameters and for their true values. \(t_{\text{peak}}\) is the peak time of the SNe and \(e_1, e_2\) are the two lens mass ellipticity parameters of Equations 6.1 and 6.2.

B.1 Absolute and relative scale rate used to simulate core-collapse SNe in this work expressed in percentage. 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.

B.2 Parameters of Gaussian mixture models used to fit the number of observations of the test set light curves for each observing strategy. These values were later used to create an augmented training set (Section 4.4.3). 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.
B.3 Parameters of Gaussian mixture models used to fit the flux uncertainty distribution of the test set in each passband ($ugrizy$) and observing strategy. These values were later used to create an augmented training set (Section 4.4.3). 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. ........................................ 159

B.4 Values of the optimised LightGBM model hyper-parameters used in each observing strategy. The hyper-parameters are described in the library documentation. .................................................. 160
1.1 The Universe

Cosmology is the quantitative study of the Universe’s origin, evolution, and constituents. While existential questions about the Universe are ancient, modern cosmology started in the beginning 20th century with the discovery of the expansion of the Universe (Hubble, 1929). Cosmology is now an established field with many successes such as the development of a theoretical model that agrees very well quantitatively with the observations. The recent advent of large astronomical datasets make cosmology an active and exciting area of research. Moreover, the many fundamental questions still unanswered are alluring.

A fundamental assumption of modern cosmology is that on large scales the Universe appears the same to all observers regardless their individual location and direction in which they look. This idea can be more precisely postulated as the Cosmological Principle: the Universe is spatially homogeneous and isotropic on large scales. The statistical isotropy and homogeneity of the Cosmological Principle have been supported by a range of observations, such as measurements on the Cosmic Microwave Background (CMB) anisotropies (Penzias and Wilson, 1965; Planck Collaboration et al., 2020) and the large scale galaxy distribution (Davis et al., 1982).
### 1.1.1 General Relativity

In Newtonian theory, gravity is an attractive force between massive objects. However, in General Relativity\(^1\), gravity is a geometric property of the spacetime. This theory relates the spacetime geometry to the distribution of matter and energy of the Universe via the Einstein field equations,

\[
G_{\mu \nu} \equiv R_{\mu \nu} - \frac{1}{2} R g_{\mu \nu} = 8 \pi G T_{\mu \nu} + \Lambda g_{\mu \nu},
\]

where \(G_{\mu \nu}\) is the Einstein tensor, \(R_{\mu \nu}\) is the Ricci tensor, \(R\) is the Ricci scalar, \(g_{\mu \nu}\) is the metric tensor, \(G\) is the Newton’s constant, \(T_{\mu \nu}\) is the energy-momentum tensor, and \(\Lambda\) is the cosmological constant; the left-hand side of the equation is a function of the metric and the right-hand side of the energy. Initially, Einstein added the cosmological constant to reproduce a static Universe (Einstein, 1917). However, this constant was later dropped when Hubble discovered that the Universe is expanding. Nowadays, the cosmological constant was re-introduced to allow for an accelerated expansion of the Universe; \(\Lambda\) now represents the vacuum energy of the spacetime and contributes towards the total energy of the Universe.

Under General Relativity, a homogenous and isotropic Universe can be described by the Friedman-Lemaître-Robertson-Walker (FLRW) metric:

\[
ds^2 = -dt^2 + a(t)^2 \left[ \frac{dr^2}{1-k r^2} + r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \right],
\]

where \(t\) is the cosmic time, \((r, \theta, \phi)\) are the spatial coordinates in spherical coordinates, \(a(t)\) is the scale factor, and \(k\) is the spatial curvature constant that describes the Universe as open \((k = -1)\), flat \((k = 0)\), and closed \((k = 1)\). The scale factor \(a(t)\) is a dimensionless function of time that encodes the expansion (or contraction) of the Universe and it is defined to be one at the present time. To quantify the change in the scale factor, we define the Hubble parameter,

\[
H(t) = \frac{\dot{a}}{a},
\]

where \(\dot{a}\) is the derivative of \(a\) with respect to time. The current value of the Hubble

---

\(^1\) See Dodelson (2003) for a review of General Relativity from a cosmological perspective.
parameter, \( H_0 \equiv H(0) \), is called the Hubble constant.

Using the Einstein field equations (Equation 1.1), we can more generally describe the dynamics of the Universe through Friedman’s equations,

\[
\left( \frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \frac{\rho - \frac{k}{a^2}}{\rho} = H^2
\]

(1.4)

\[
\frac{\ddot{a}}{a} = -\frac{4\pi G}{3} \left( \frac{\rho + 3P}{\rho} \right) + \frac{\Lambda}{3},
\]

(1.5)

where \( \ddot{a} \) is the second derivative of \( a \) with respect to time, \( \rho \) is the energy density, and \( P \) is the pressure; these equations connect the content and the expansion of the Universe. Since the FLRW describes a homogeneous and isotropic Universe, the components of the Universe are perfect fluids, so we write their equation of state as \( P = w\rho \). Imposing the conservation of the energy-momentum, we obtain a third relation

\[
\rho + \frac{3\dot{a}}{a} (\rho + P) = 0,
\]

(1.6)

whose solution is \( \rho \propto a^{-3(1+w)} \). For a flat Universe, this solution allow us to relate the evolution of the scale factor with the equation of state:

\[
a(t) = \begin{cases} 
  t^{2/(1+w)} & w \neq -1 \\
  e^{Ht} & w = 1
\end{cases}
\]

(1.7)

We can then describe the expansion of the Universe accordingly to the fluid which dominates its total energy density at that time. Thus, it is useful to define the total energy density in a flat Universe, also called the critical density,

\[
\rho_c \equiv \frac{3H_0^2}{8\pi G}.
\]

(1.8)

For each component \( i \) of density \( \rho_i \), the normalised density parameter is

\[
\Omega_i \equiv \frac{\rho_i}{\rho_c}.
\]

(1.9)

Since each component evolves in time accordingly with their own equation of state, they affect the expansion of the Universe differently. In the following sections, we describe the components of the Universe and their effects on its evolution.
1.1.2 Composition of the Universe

The current concordance model of the Universe is referred to as the $\Lambda$CDM model or the 'standard model' of cosmology. The $\Lambda$ refers to the presence of dark energy in the form of a cosmological constant. The CDM alludes to the presence of cold dark matter. This model has been shown to describe the Universe well on large scales (e.g. Planck Collaboration et al., 2020). Accordingly with $\Lambda$CDM, the Universe is mainly composed of four ingredients:

- **Baryonic matter ($\Omega_b \approx 5\%$):** All atomic nuclei and electrons interacting through gravitational, electromagnetic, strong and weak forces. Thus, it includes galaxies, stars, planets, and living organisms; most of this matter is in diffuse hot gas around galaxies. Baryonic matter mostly contains hydrogen and light elements formed in the early Universe. Its equation of state parameter $w$ is 0 and a baryonic matter-dominated Universe has $a \propto t^{2/3}$.

- **Cold dark matter ($\Omega_{\text{CDM}} \approx 25\%$):** Predominant form of matter in the Universe, made of non-baryonic, pressureless and non-relativistic matter. It is called 'dark' due to its weak (or nonexistent) interaction through electromagnetic radiation. Currently its nature is unknown and it had not yet been directly observed or produced. Despite that, the presence of dark matter can be inferred from several observations: galaxy clusters (Oort, 1932; Zwicky, 1933), galaxy rotation curves (Freeman, 1970; Rubin and Ford, 1970), gravitational lensing (Clowe et al., 2006), and CMB (Planck Collaboration et al., 2020). This matter is thought to be 'cold', which means that the velocities of the particles are too small to erase structure formation in the early Universe. Over the years many dark matter candidates have been proposed, such as weakly interacting massive particles (Steigman and Turner, 1985), axions (Peccei and Quinn, 1977; Weinberg, 1978; Wilczek, 1978) and sterile neutrinos (Dodelson and Widrow, 1994). Similarly to baryonic matter, the dark matter equation of state parameter $w$ is 0 and a dark matter-dominated Universe has $a \propto t^{2/3}$.

The overall matter in the Universe is $\Omega_b + \Omega_{\text{CDM}} = \Omega_m \approx 30\%$.

- **Radiation ($\Omega_r < 1\%$):** Smaller fraction of the present-day Universe, made of

\footnote{See Workman et al. (2022) and Baudis (2018) for recent reviews.}
particles whose velocity is close to the speed of light. It mainly includes relic electromagnetic radiation and neutrinos produced in the early Universe in the form of the cosmic microwave and neutrino backgrounds. While radiation is currently subdominant, it had a major role in the formation of structure at small scales. Its equation of state parameter $w$ is $1/3$ and a radiation-dominated Universe has $a \propto t^{1/2}$.

- Dark energy ($\Omega_\Lambda \approx 70\%$): Largest component of the Universe and it is responsible for the current phase of accelerated expansion. The introduction of dark energy was motivated by observations of Type Ia supernovae (Perlmutter et al., 1999; Riess et al., 1998). However, its nature is even more obscure than that of the dark matter. Current observations hint that dark energy has the same properties of a cosmological constant (Planck Collaboration et al., 2020). In $\Lambda$CDM, its equation of state parameter $w$ is $-1$ and a dark-energy dominated Universe has $a \propto e^{Ht}$. Despite the observations being consistent with this model, if $\Lambda$ is considered the vacuum energy, the theoretical prediction underestimates the observed value by 120 orders of magnitude. Therefore, modified gravity theories and alternative dark energy models are active areas of research (e.g. Alves et al., 2019; Braglia et al., 2021).

### 1.1.3 Expansion History

In the beginning the Universe was small, hot and dense. However, we do not know the physics behind these early moments that are characterised by energy scales far above the reach of particle accelerators. The standard cosmological paradigm starts afterwards with a phase of quick expansion that increases the Universe’s size by 60-fold ($e^{60} \simeq 10^{26}$) in less than a second and it is known as ‘inflation’. During this expansion, causally connected patches are stretched, resulting in a flat, homogeneous and isotropic Universe. The exponential expansion was driven by the ‘inflaton’ scalar field; the small quantum fluctuations of the inflaton field later became density perturbations of cosmological scales and provided the seeds to form all the structures we observe in the Universe.

After the end of inflation, the inflaton decayed into the particle species in the Standard Model. This period is known as ‘reheating’ and radiation is the dominant
form of energy of the Universe, thus the scale factor grows as $a \propto t^{1/2}$. The Universe was filled with an ionised hot plasma of baryons and radiation in thermal equilibrium and coupled together. As the Universe expanded and cooled, neutrinos decoupled and formed a cosmic neutrino background yet to be detected. Next, the Big Bang nucleosynthesis begins where the protons and neutrons fuse to create light nuclei such as deuterium, helium and lithium.

At some point the Universe becomes dominated by matter and a new phase begins; the scale factor now grows as $a \propto t^{2/3}$. During this period, the Universe cools down to 3000 K and free protons and electrons can bind into neutral-hydrogen atoms without being immediately ionised by energetic photons; this is known as 'recombination'. The CMB was also emitted at this point. The Universe then remained dominated by matter for a long time, allowing the matter perturbations to grow during the dark ages. The radiation perturbations stay small because pressure acts opposite to gravity. In this period, the dark matter begins to collapse into halo-like structures through its own gravitational attraction. Eventually, the structures contained enough neutral hydrogen at high densities and the first stars and galaxies were born. When these objects were energetic enough to ionize the neutral hydrogen in the surrounding intergalactic medium, a period known as 'reionisation' started. As a result of reionisation, most of the Universe is not neutral. While this period of the Universe is still shrouded in mystery, future measurements of the 21-cm transition will map the distribution of neutral hydrogen at the epoch of reionisation.

Recently, after approximately 10 billion years, the expansion of the Universe started to accelerate. This acceleration is driven by the new dominant form of energy of the Universe, dark energy; the scale factor grows as $a \propto e^{Ht}$.

### 1.1.4 Distances

We can trace the expansion of the Universe through the redshift ($z$). When a photon is emitted and travels towards us, the Universe expands around it, making the photon lose energy; the photons are redshifted due to space dilation. Since their energy and momentum scale as $a^{-1}$, we define a cosmological redshift as

$$1 + z = \frac{a(t_{\text{obs}})}{a(t_{\text{em}})} = \frac{\lambda_{\text{obs}}}{\lambda_{\text{cm}}},$$

(1.10)
where $t_{\text{em}} (t_{\text{obs}})$ is the emitted (observed) time and $\lambda_{\text{em}} (\lambda_{\text{obs}})$ is the emitted (observed) wavelength. By convention, for observations today, $\alpha(t_{\text{obs}}) = 0$. However, while measuring redshifts can be quite straightforward, determining distances can be a challenge.

1.1.4.1 Standard Candles

We can probe cosmology using astrophysical objects with known intrinsic luminosities; these objects are referred to as ‘standard candles’. If the absolute luminosity $L$ is known, the observed brightness $F$ can be used to infer that the luminosity distance $D_L$ to the object is

$$D_L = \sqrt{\frac{L}{4\pi F}}. \tag{1.11}$$

The luminosity distance is also related to the comoving distance $\chi$, the distance between two objects that stays constant as the Universe expands:

$$D_L(z) = (1 + z) \chi(z), \tag{1.12}$$

where assuming a flat Universe, we have

$$\chi(z) = \int_0^z \frac{dz'}{H(z')} . \tag{1.13}$$

Thus we can rewrite Equation 1.12 as

$$D_L(z) = (1 + z) \int_0^z \frac{dz'}{H(z')} . \tag{1.14}$$

The relationships described in the above equations are useful to find the cosmology that best explains the observed data. An example of commonly used as standard candles are Type Ia supernovae (Riess et al., 1998; Perlmutter et al., 1999; Betoule et al., 2014), which we will discuss on Sections 1.2.1.1 and 1.2.2. Other standard candles are Cepheids and RR Lyrae.

1.1.4.2 Standard Rulers

Standard rulers are astronomical objects with known physical size. Using a reasoning analogous as Section 1.1.4.1, we can express the angular diameter distance $D_A$ of an astrophysical object as
By combining Equations 1.14 and 1.15 we obtain the relationship between luminosity and angular diameter distances, $D_L(z) = (1+z)^2 D_A(z)$.

Examples of standard rulers are the baryonic acoustic oscillations feature of the large-scale structure power spectrum (Seo and Eisenstein, 2003; Eisenstein et al., 2005, 1998) and the characteristic scale of transition to cosmic homogeneity of the universe (Ntelis et al., 2018).

1.1.4.3 Standard Sirens

In addition to standard candles, we can use gravitational waves (GW) to measure luminosity distances: when the members of a binary black hole are widely separated and slowly spiral together, the waveform of the resulting GW emission encodes the luminosity distance to a binary and its orientation (Holz and Hughes, 2005; Holz et al., 2018). For a wave propagating in the $z$-axis, one polarisation stretches and squeezes along the $x$- and $y$-directions ($h_+$) and the other stretches and squeezes along axes rotated by $45^\circ$ from the $x$- and $y$-axes ($h_\times$). The amplitude of these GW polarisations from a binary in a circular orbit with frequency $\Omega$, masses $m_1$ and $m_2$, and orientated such as the normal to its orbital plane makes an angle $i$ to our line of sight are

$$h_+ = \frac{2c}{D_L} \left( \frac{G M}{c^3} \right)^{5/3} \Omega^{2/3} \left[ 1 + \cos^2(i) \right] \cos(2\Phi(t)),$$

$$h_\times = \frac{4c}{D_L} \left( \frac{G M}{c^3} \right)^{5/3} \Omega^{2/3} \cos(i) \sin(2\Phi(t)),$$

(1.16)

where $\mathcal{M} = (m_1 m_2)^{3/5} (m_1 + m_2)^{-1/5}$ and $\Phi(t)$ is the integration of the orbital frequency $\Omega$ over the duration $t$ of the measurement. Once $\mathcal{M}$ and $i$ are estimated from the GW observation (Cutler and Flanagan, 1994), the distance $D_L$ is known (Holz and Hughes, 2005; Holz et al., 2018).

1.2 Supernovae

Supernovae (SNe) are luminous explosive events that mark the destruction of some stars. They can last several weeks and have long been observed and studied.
While the most recent directly observed SN in the Milky Way was in 1604, the earliest possible recorded SN dates to 4500±1000 BC (Joglekar et al., 2011). These explosive and high luminosity events are useful for modern cosmology because they can be detected in remote galaxies. Moreover, SNe are a major source of elements in the Universe. Not only do they eject elements synthesised in the core of the original star, but they also create new ones during the explosion itself. These elements enrich the molecular clouds where new stars form, changing the composition of each generation of stars from an almost pure mixture of hydrogen and helium to a more metal-rich composition. The difference in the abundance of the elements that form a star (metallicity) influences its life, such as its temperature, mass, the way it will die, the likelihood of having planets in its planetary system (e.g. Gonzalez, 1997; Baraffe et al., 1997; Tremonti et al., 2004; Fruchter et al., 2006; Wang and Fischer, 2014; Lodieu et al., 2019; Pakmor et al., 2022). Thus, understanding how metallicity increases with time is also a central astrophysics question in which SNe studies can help.

1.2.1 Types of Supernovae

Baade and Zwicky (1934) were the first to recognise SNe as a separate class of astrophysical events following previous work by Lundmark (1925). Later, Minkowski (1941, 1964) introduced the first scheme for SNe classification based on spectroscopic observations of 14 events.

We can classify SN based on their spectra around maximum light (Filippenko, 1997); some features are easier to observe in this period, and for distant SNe, the maximum light might be the only period in which the spectra has acceptable signal-over-noise ratio. The Type II SN (SN II) shows hydrogen lines, while the Type I lacks them. Different subtypes exist for each category, such as Type Ib SN that displays helium lines and Type Ic which does not show hydrogen nor helium. Type Ia SN (SN Ia) presents lines from higher-mass elements, like calcium, sulphur, silicon and iron, but lack hydrogen and helium. See Filippenko (1997) for more details in SN spectral classification.

1.2.1.1 SN Ia

It is currently thought that SN Ia is the thermonuclear explosion of a carbon-oxygen white dwarf. White dwarfs are supported by electron degeneracy pressure and are
usually stable. However, if they are in a binary system, they can start accreting mass from their companion star. Since white dwarfs use electron degeneracy pressure to counteract the force of gravity, their mass becomes inversely related to their volume and there is a finite mass that leads to a volume of zero; the Chandrasekhar mass limit \( (1.4M_\odot) \) is the highest mass a non-rotating white dwarf can obtain before gravity overcomes the electron degeneracy, and the white dwarf collapses (Carroll and Ostlie, 2017). As the white dwarf accretes mass, it gets hotter; it reaches the temperatures and densities needed for burning carbon shortly before the Chandrasekhar limit. At this stage, the burning creates a large instability, which turns into a runaway process (e.g. Hillebrandt and Niemeyer, 2000). The fusion results in a large release of energy in a few seconds and the white dwarf explodes (e.g. Röpke et al., 2007).

The energy of the explosion is transformed into kinetic energy of the ejecta\(^3\). While the ejecta cools rapidly, the radioactive elements recently created power it over a longer time scale. In particular, the energy released by the radioactive decay of \(^{56}\text{Ni}\) and \(^{56}\text{Co}\); the atomic nucleus of these elements are unstable and they lose energy by radiation\(^4\). The light curve is then powered by the diffusion of this radioactive decay energy into the expanding SN ejecta (Chatzopoulos et al., 2012; Bersten and Mazzali, 2017). Shortly after the explosion, the material is very dense, so the heat energy is unable to diffuse and becomes trapped until the ejecta expands. This heat-trapping releases light which has a peak luminosity at approximately three weeks after the explosion. It then fades out over the following months. Since SN Ia peak luminosity is \( \sim 10^{10} \) times brighter than the Sun, optical telescopes can find them in remote galaxies billions of light years away (Goobar and Leibundgut, 2011).

There are two popular types of progenitors models for SNe Ia. In the ‘single degenerate’, the mass of the white dwarf grows by accretion from an evolving binary companion (Whelan and Iben Jr, 1973). In the ‘double degenerate’ models, two white dwarfs merge (Iben Jr and Tutukov, 1984; Webbink, 1984). There are several evolutionary paths of binary systems that result in each type of SN Ia progenitor.

---

\(^3\)The explosion debris is called ejecta.

\(^4\)See Woosley (1988) and Goobar and Leibundgut (2011) for an overview of how radioactive decay powers SNe.
model (e.g. Yungelson, 2005; Pritchet et al., 2008; Sharon and Kushnir, 2022). Currently, the identity of the SN Ia progenitors is an open question, that when solved may help to constrain the binary–star evolution (Maoz and Mannucci, 2012; Livio and Mazzali, 2018; Ruiter, 2019).

Observationally, we find that SN Ia are very uniform in their appearance; they are ‘standardisable’ candles, and thus, useful to study the expansion of the Universe. We will detail this property in Section 1.2.2. The top row of Figure 1.1 shows a model of SN Ia light curves.

1.2.1.2 Peculiar SN Ia

As mentioned in Section 1.2.1.1, SN Ia have very homogeneous properties. However, they show some spread in their luminosity and spectral evolution. Both over-luminous and underluminous SNe Ia have been identified (Howell, 2011). Due to their different intrinsic luminosity, these peculiar events are usually a contaminant in SN Ia samples used to measure cosmological parameters. It has been argued that some of them should be separated into new SNe classes (e.g. Foley et al., 2013). Upcoming transient surveys will discover even wider diversity and possibly cement the class distinctions (Kessler et al., 2019).

The largest subclass of peculiar SN Ia is Type Iax SN (SN Iax; Foley et al., 2013; Jha, 2017). This class is based on SN 2002cx (Li et al., 2003). While similar to normal SN Ia, SN Iax have lower luminosity, lower ejecta velocity, and more variation on these parameters. Additionally, SN Iax typically have faster rises and their light curves show more variety (Magee et al., 2016; Jha, 2017; Kessler et al., 2019). The third row of Figure 1.1 shows a model of SN Iax light curves. The leading scenario for the creation of SN Iax is an explosion of a white dwarf, triggered by helium accretion to the Chandrasekhar mass, that does not necessarily fully disrupt the star (Jha, 2017).

Another large subclass of peculiar SN Ia is Type Ia-91bg SN (SN Ia-91bg; Filippenko et al., 1992; Leibundgut et al., 1993; Turatto et al., 1996). These events are redder, have lower peak luminosity, lower ejecta velocity and lack a secondary maximum in the near infrared (Benetti et al., 2005; Taubenberger et al., 2008; Wang et al., 2013; Taubenberger, 2017). SN Ia-91bg also decline faster and have a light curve rise time a few days shorter than normal SN Ia (Benetti et al., 2005; Tauben-
Figure 1.1: Model light curves in the $ugrizy$ passbands of the Rubin Observatory Legacy Survey of Space and Time. Each row presents a model of a different extragalactic event; the first 6 rows show several types of SNe, the sixth shows a tidal disruption event (TDE), the seventh shows a kilonova (KN), and the eighth shows an active galactic nuclei. The image is from Kessler et al. (2019).
berger et al., 2008; Wang et al., 2013). The second row of Figure 1.1 shows a model of SN Ia-91bg light curves. Several progenitor configurations and explosion mechanisms have been suggested to explain the creation of SN Ia-91bg, but without any direct observations, the exact formation remains unknown (Pakmor et al., 2013; Panther et al., 2019; Barkhudaryan et al., 2019).

1.2.1.3 SN II

Type II SNe are thought to be explosions of stars with masses between $8M_\odot$ and $18M_\odot$ (Carroll and Ostlie, 2017; Smartt et al., 2009). The explosion occurs when the star forms iron and thus can no longer fuse into more massive elements to produce further nuclear energy. When the fusion stops, the star can no longer counteract its own gravity, and collapses in on itself. This explosive ‘core collapse’ (CC) event happens in milliseconds and forms either a neutron star or a black hole, depending on the initial mass of the star. While most of the energy released goes to the emission of neutrinos that escape into space, a small fraction ($\sim 1\%$) is transferred to the surrounding material of the star. This causes the materials to unbind from the core and be expelled into space. Since some of the kinetic energy is thermalised as heat, the SN shines. For CC SN, the shock heat left over from the explosion is the main source of luminosity for the first few days until the radioactive decay of $^{56}$Ni and $^{56}$Co becomes dominant (Woosley, 1988). The timing for this transition depends on the parameters of the supernova, such as mass, kinetic energy, density profile, and composition. Other sources of energy are circumstellar interaction and pulsar emissions (Jeffery, 1999; Chatzopoulos et al., 2012). The total energy released by a CC SN is approximately 100 times that of a SN Ia, but they are optically fainter (Carroll and Ostlie, 2017; Kessler et al., 2009a).

In addition to the above, the stars that generate SNe II retain a significant amount of hydrogen in its outer layer at the time of explosion; this hydrogen is visible in the SN spectrum. SN II are then called hydrogen-rich CC SN. Accordingly with the amount hydrogen and the density structure of the outer layers, we can divide SN II into further subclasses (e.g. Schlegel, 1990; Filippenko et al., 1993; Kessler et al., 2019).

Observationally, SN II show a large luminosity variance, so they are poor distance indicators. Despite that, SN II are more abundant than SN Ia and there are
ongoing efforts to standardise their brightness (Hamuy and Pinto, 2002; Jaeger et al., 2015). The fourth row of Figure 1.1 shows a model of SN II light curves.

1.2.1.4 SN Ibc

Type Ib and Ic SN (SN Ibc) are CC SN. Their formation is then described in Section 1.2.1.3. However, these SNe lack the hydrogen features that SN II have, and SN Ic spectra further lacks both helium and silicon. The lack of these features implies that the progenitor star had been stripped of its hydrogen and helium envelope before the explosion; SN Ibc are then known as stripped-envelope CC SN (e.g. Kessler et al., 2019; Vincenzi et al., 2019).

In terms of photometric light curves, SN Ibc are similar to SN Ia but fainter and redder (Galbany et al., 2017). The fifth row of Figure 1.1 shows a model of SN Ibc light curves.

1.2.2 SN Ia as standardisable candles

As mentioned in Section 1.1.4.1, we use astronomical objects with known intrinsic luminosity to determine distances and study the expansion history of the Universe. However, the usefulness of any standard candle depends on how well we can distinguish between intrinsic luminosity differences and brightness differences due to distance. Baade (1938) was the first to suggest using SNe as standard candles and estimated that their intrinsic brightness dispersion was $\sim 1.1$ magnitudes. Afterwards, Kowal (1968) showed that the intrinsic peak brightness dispersion of the Type I SNe was only 0.61 magnitudes, and published the first Hubble diagram (distance modulus versus redshift) with SNe. Later, it was found that the subclass SN Ia was even more uniform, with an intrinsic dispersion of just $\sim 0.35$ magnitudes (Sandage and Tammann, 1993), close to the value accepted today. However, it took another couple of decades to obtain SN Ia at high redshifts and useful for cosmology (Branch and Tammann, 1992; Perlmutter et al., 1995; Branch, 1998). These led to the discovery of the accelerating expansion of the universe (Riess et al., 1998; Perlmutter et al., 1999). Subsequent studies have discovered more SNe Ia, which lead to increasingly strong constraints on cosmological parameters (e.g. Astier et al., 2006; Betoule et al., 2014; Scolnic et al., 2018a). The white dwarfs have the same fundamental mass limit (Chandrasekhar limit) so they explode into SN Ia with a similar energy. SN Ia then have very homogeneous spectral and photometric
properties. However, they are not standard candles because events at the same redshift show variability in brightness. This variability has been linked to other observables, and thus can be corrected for. SN Ia are then ‘standardisable’ candles.

To standardise SN Ia we use empirical calibrations between their peak brightness and secondary light curve properties. Phillips (1993) showed that brighter SN Ia take longer to fade (‘brighter-broader’ effect shown in the right panel of Figure 1.2). Later, Riess et al. (1996) and Tripp (1998) showed that brighter SN Ia also appear bluer, while fainter SNe appear redder (‘brighter-bluer’ effect shown in the left panel of Figure 1.2); some authors attribute part of this correlation to the reddening caused by dust in the host galaxies that can scatter and absorb SNe light (Tripp and Branch, 1999; Goobar, 2008). These two empirical calibrations are usually encoded through ‘stretch’ and ‘color’ parameters in light curve fitting models, respectively (e.g. Guy et al., 2007; Jha et al., 2007; Kenworthy et al., 2021). To account for them, the light curve fitters usually construct empirical models of SN Ia in which the distance modulus $\mu$ is

$$\mu = m - M + \alpha x_1 - \beta c,$$  

where $M$ and $m$ are the absolute and apparent magnitude, respectively, $x_1$ is the stretch parameter, and $c$ is the color parameter. The fit to each SN Ia light curve constrains $m$, $x_1$, and $c$. The remaining parameters ($M$, $\alpha$, $\beta$) can be constrained during cosmological fits.

Implementing the brighter-broader and brighter-bluer empirical correlations reduces the scatter in the peak brightness to $\sim 0.15$ magnitudes, which corresponds to an error in distance of only $\sim 7\%$. There is evidence that additional correlations could further reduce the scatter of the peak brightness distribution (e.g. Sullivan et al., 2006; Roman et al., 2018; Brout and Scolnic, 2021).

### 1.2.3 Photometric classification

Traditionally, the SNe used in cosmological analysis and astrophysical studies were identified based on their spectra. The resulting spectroscopically-classified SNe datasets are pure due to their low contamination of other classes. Since contamination bias the cosmological measurements, it is ideal to classify SNe and obtain
Figure 1.2: Illustration of the brighter-bluer (left panel) and brighter-broader (right panel) relationship of SNe Ia for nearby (blue open circles) and distant ($z < 0.8$, black filled circles) events. In both plots, brighter SNe Ia are toward the bottom and the behaviour is consistent between nearby and distant SNe Ia. The images are from Astier et al. (2006).

the more accurate redshift information using spectroscopy. However, with modern surveys such as the ones that will be described on Section 1.3, the SNe discovery rate is quickly outpacing the growth of the limited spectroscopic resources (e.g. Bernstein and Dark Energy Survey Collaboration, 2012; LSST Science Collaboration et al., 2009). Moreover, doing spectroscopy is expensive, in particular for the high-redshift SNe that play an important role in cosmological analyses. Thus, we must rely on the less accurate photometric classification, which we will explore in Chapters 3 and 4.

Photometric classification consists of identifying astronomical objects through their light curves, which are time-series of the observed time intensity; we can measure the intensity in several frequencies, as shown in Figure 1.3. Additional information about the event can also be used for classification, such as their redshift, sky position, host galaxy colour and mass. However, designing a photometric classification algorithm has major challenges. The light curves are usually sparsely and non-uniformly sampled. The observations might also occur in different photometric passbands of which only a subset is available each night. Moreover, the photometry uncertainties are heteroskedastic\(^5\), and the noise levels vary across passbands.

Typically, photometric classifiers are trained on subset of events from the survey that was spectroscopically classified. However, this biases the training set to-

\(^5\) The variance of heteroskedastic uncertainties is not constant across observations.
Figure 1.3: Example of a Type Ia SN light curve. Each colour represents observations in different passbands. The intensity of the light is measured in flux.

wards brighter events as they are easier to obtain spectra from; the training set is non-representative of the events we want to classify. Recent work has focused on overcoming the lack of representativeness (e.g. Muthukrishna et al., 2019; Revsbech et al., 2017; Pasquet et al., 2019; Boone, 2019; Carrick et al., 2021). We also present an approach in Chapter 3. Alternatively, photometric classifiers are trained on simulated representative training sets. While the previous approach was non-representative due to selection effects, this approach is non-representative if we have the wrong underlying SNe models or modelling of the observations. Since there is still much unknown about SNe, as shown in Section 1.2.1, we expect having to update their models in the coming years once we have more data from current and upcoming surveys.

See Jaschek and Jaschek (1987) for an introduction to stellar photometry and photometric classification, and the spreadsheet linked (compiled by Ragosta) for a list of more than 45 recent papers on photometric classification.

1.2.4 Lensed supernovae

Massive objects, such as galaxies or cluster of galaxies, bend the spacetime around them. Thus, the light from a background object (source) is deflected due to the gravitation potential of the foreground massive object (lens) on its way to the observer; this phenomena is known as gravitational lensing. In weak gravitational lensing the sources appear distorted, stretched or magnified, and it is typically measured by statistically analysing large numbers of sources. In strong lensing we study individual sources and the lensing can produce arcs, Einstein rings, and multiple images of the source (Einstein, 1936; Treu, 2010).

The observed positions of strongly lensed sources are widely used to map the
matter distributions of the lens. However, a time-variable source allows additional measurements of the lens, cosmic expansion rate, and dark energy models because we can measure the delay between each pair of images (Refsdal and Bondi, 1964; Holz, 2001; Kelly et al., 2015; Rodney et al., 2021a). The most common strongly lensed transients are quasars and SNe because they are sufficiently bright to be detected at cosmological distances (Hook et al., 1994; Helfand et al., 2001).

While quasars vary in brightness on the time scales of several years, SNe only last months. This faster timeframe allows SNe to be competitive in measuring time delays. Another advantage of lensed SNe (LSNe) is that we can use the common standardisable SN Ia (mentioned in Section 1.2.2) to improve the modelling of the lensing gravitational potential, and therefore yield a more precise $H_0$ measurement (Oguri and Kawano, 2003; Kolatt and Bartelmann, 1998; Foxley-Marrable et al., 2018; Birrer et al., 2022).

### 1.3 Large Surveys

In Section 1.2.2, we mentioned that we can use the standardisable candles SNe Ia to study the evolution of the Universe. To detect them and obtain the necessary information for photometric classification, we need to observe the sky. In this section, we will summarise the surveys considered in this thesis. We start by mentioning the science goals of the surveys. Next, we briefly describe the technical details of the telescopes and surveys. For the upcoming Rubin Observatory Legacy Survey of Space and Time, we also describe the process of observing strategy optimisation.

#### 1.3.1 Dark Energy Survey

1.3.1.1 Science Goals

The Dark Energy Survey (DES; The Dark Energy Survey Collaboration and Flaugher, 2005) was primarily designed to study the nature of dark energy. Using four complementary probes – galaxy clusters, weak lensing, Type Ia supernovae and baryon acoustic oscillations – DES provided data to constrain the properties of dark energy and other cosmological parameters, test $\Lambda$CDM and alternative models of gravity, and characterise dark matter (Dark Energy Survey Collaboration et
al., 2016; Abbott et al., 2018b, 2019a; Macaulay et al., 2019; Abbott et al., 2022). Additionally, Dark Energy Survey Collaboration et al. (2016) describes how DES data can also be used to study the solar system, the Milky Way, galaxy evolution, quasars, search for optical counterparts of gravitational waves, SNe and other transients.

1.3.1.2 Technical Details

DES was an optical/near-infrared survey that observed 5000 deg$^2$ of the Southern hemisphere sky between 2013 and 2019. The imaging data was acquired by the Dark Energy Camera (DECam; Flaugher et al., 2015), which was mounted on the Victor M. Blanco 4-meter Telescope at the Cerro Tololo Inter-American Observatory, high in the Chilean Andes. DECam has 570 megapixels and recorded images using five photometric passbands ($grizY$), spanning from 400 nm to 1080 nm (Flaugher et al., 2015). The typical exposure time was 90 s for $griz$ and 45 s for $Y$ (Flaugher et al., 2015; Dark Energy Survey Collaboration et al., 2016).

Over its lifetime, DES discovered and monitored more than 30000 optical transients. However, only $\sim 500$ SN Ia were spectroscopically-confirmed due to limited spectroscopic resources (The Dark Energy Survey Collaboration and Flaugher, 2005; Brout et al., 2019; Smith et al., 2020; Vincenzi et al., 2021). Despite that, DES discovered over 2500 SN-like events with host galaxy spectroscopic redshift; these events can be photometrically classified and used for cosmological analysis (Vincenzi et al., 2021).

1.3.2 Zwicky Transient Facility

1.3.2.1 Science Goals

The Zwicky Transient Facility (ZTF; Bellm et al., 2018, 2019; Graham et al., 2019a) is a 3-year time-domain survey designed to study supernova cosmology, the physics of supernovae and relativistic explosions, multi-messenger astrophysics, active galactic nuclei (AGN) and tidal disruption events, stellar variability, and Solar System objects. While DES focused on SN Ia samples at high redshifts $z \gtrsim 0.1$, the ZTF samples are mostly at $z < 0.1$. ZTF also acts as a precursor in terms of transient and variable object astronomy for the upcoming Rubin Observatory Legacy Survey of Space and Time (LSST Science Collaboration et al., 2009, 2017; Ivezić
et al., 2019a) that we will describe in Section 1.3.3.

To distinguish between dark energy models it is crucial to ‘anchor’ SNe Ia with a low redshift sample (Goliath et al., 2001; Astier et al., 2014). However, the current samples come from different photometric systems with hard to correlate systematic uncertainties, diverse filter sets and lack of precise calibration (Graham et al., 2019a; Dhawan et al., 2021). Thus the low redshift SN Ia sample is the main contributor to the systematic uncertainties of the estimates of the dark energy equation of state (Scolnic et al., 2018a; Foley et al., 2018). ZTF addresses this issues by providing a complete and unbiased SN Ia sample for $z < 0.1$. In addition to anchoring SN Ia, we can use this sample to study the large scale structure in the nearby universe and measure $H_0$ (Dhawan et al., 2021).

1.3.2.2 Technical Details

ZTF consists of a dedicated 576 megapixel camera (Dekany et al., 2016) on the Palomar 48-inch Schmidt telescope with a 47deg$^2$ field of view and an integral field unit spectrograph on the Palomar 60-inch telescope for spectral classification of explosive transients with $< 19$ magnitudes. Its survey time is divided into public surveys (40%), partnership surveys (40%), and Caltech programs (20%). During the public surveys, ZTF uses the passbands $g$ and $r$ to observe the entire northern visible sky every three nights and the Galactic Plane every night. In the first year, the bulk of the partnership observations was dedicated to a high-cadence survey of three visits per night in each of the passbands $g$ and $r$, and a slow, wide $i$-band survey (Bellm et al., 2019).

In the first $\sim 2.5$ years, ZTF has discovered and spectroscopically classified more than 3000 SNe Ia (Dhawan et al., 2021). While the sample size will increase as the survey progresses, it is already competitive to measure local large scale structure properties such as the growth rate (Graziani et al., 2020; Dhawan et al., 2021).

1.3.3 Legacy Survey of Space and Time

1.3.3.1 Science Goals

The upcoming Rubin Observatory Legacy Survey of Space and Time (LSST; LSST Science Collaboration et al., 2009, 2017; Ivezić et al., 2019a) was designed to
catalogue the Solar System, map the Milky Way, explore the transient optical sky and probe dark energy and dark matter. Since the bulk of this thesis focuses on LSST, we provide below an overview of each of the four main science goals.

• Making an Inventory of the Solar System: LSST will provide data (including the orbital parameters, light curves and colorimetry) for millions of small bodies in the Solar system. This will advance Solar system studies since understanding the distribution of these small bodies is a key element in testing theories of the formation and evolution of our planetary system. LSST will also find several trans-Neptunian objects and measure their properties, providing clues about the early environment in the outer Solar system and insights in the formation of the Solar system. Additionally, LSST will catalogue potentially hazardous asteroids (LSST Science Collaboration et al., 2009; Ivezić and The LSST Science Collaboration, 2011; LSST Science Collaboration et al., 2017).

• Mapping the Milky Way: LSST will study the Milky Way and its neighbours by producing a large and accurate photometric and astrometric dataset. With that dataset it is possible to study the distribution of main sequence stars beyond the presumed edge of the Galaxy’s halo, their metallicity distribution and their kinematics. By doing the above, LSST will determine what is the structure and accretion history of the Milky Way and the fundamental properties of all the stars within 300 pc of the Sun (LSST Science Collaboration et al., 2009; Ivezić and The LSST Science Collaboration, 2011; LSST Science Collaboration et al., 2017).

• Exploring the Transient Optical Sky: LSST’s capability of providing simultaneously large-area coverage, dense temporal coverage, accurate colour information and good image quality will enable the discovery and analysis of rare and exotic events. It will also facilitate new population and statistical studies of events due to the extension of the time–volume–colour space over current surveys. For example, LSST will be able to study known and unusual SN populations, parameterise their light curves (e.g. Hoeflich et al., 1998; Howell et al., 2007; Hicken et al., 2009; Bianco et al., 2014; Arcavi...
et al., 2017), and measure their intrinsic rates as a function of subtype and host environment properties (e.g. galaxy stellar mass; Graur et al., 2017). Goldstein and Nugent (2016) and Goldstein et al. (2018) predict LSST will discover 500 – 1000 multiply imaged SN Ia and Oguri and Marshall (2010) expect ∼ 80 lensed core-collapse SN. Scovacricchi et al. (2015) predicts LSST will detect a large and well-characterised sample of superluminous SN up to $z = 2.5$, which can be used to improve the cosmological constraints on $w$ and $\Omega_m$. Moreover, the telescope will characterise the variability of active galactic nuclei and monitor periodic variables such as RR Lyrae stars. LSST will also identify optical counterparts of transients and variables detected in other electromagnetic wavebands, such as radio transients associated with tidal disruption flares (Strubbe and Quataert, 2009; Giannios and Metzger, 2011) and in non-electromagnetic sources, such as gravitational waves and neutrino events (LIGO⁶, ICECUBE⁷) (LSST Science Collaboration et al., 2009; Ivezić and The LSST Science Collaboration, 2011; LSST Science Collaboration et al., 2017; Cowperthwaite et al., 2019). In particular, it is expected that LSST will find serendipitously ≥ 80 kilonovae if it follows the survey baseline cadence (Scolnic et al., 2017; Scolnic et al., 2018b; Setzer et al., 2019). Other than that, LSST will also find kilonovae as targets-of-opportunity triggered by detections of GW signals (Scolnic et al., 2018b; Cowperthwaite et al., 2019).

• Probing Dark Energy and Dark Matter: As mentioned in Section 1.1, dark energy manifests itself in the expansion rate of the universe. LSST will study dark energy using weak lensing, baryon acoustic oscillations, SN, cluster counts, and other probes of geometry and growth of structure (LSST Science Collaboration et al., 2009). In connection to rare transients, we can measure the Hubble constant by obtaining redshifts from standard sirens with electromagnetic detections of kilonovae (see Section 1.1.4.3) and use their GW counterparts to measure the luminosity distances (Schutz, 1986; Holz and Hughes, 2005; Nissanke et al., 2013; Vitale and Chen, 2018). As this measure is independent of the measurements based on SNe (Riess et al., 2016).

⁶ http://www.ligo.caltech.edu
⁷ http://icecube.wisc.edu
or on the CMB (Planck Collaboration, 2016; Planck Collaboration et al., 2020), it can shed some light in the $\sim 3\sigma$ disagreement between the aforementioned methods (Freedman, 2017; Vitale and Chen, 2018; Feeney et al., 2019). Due to LSST’s ability to produce a large and uniform data set with high quality for multiple techniques, it is possible to cross check the result from each technique, detect unknown systematics and cross-calibrate known systematics. Additionally, as each probe constraints different cosmological parameters, they can break degeneracies between the parameters (e.g. Fig. 1.4).

1.3.3.2 Technical Details
LSST is a ground-based optical telescope facility being built in Cerro Pachón, which will repeatedly survey the Southern half of the sky every few nights for ten years (LSST Science Collaboration et al., 2009, 2017; Ivezić et al., 2019a). The telescope will have a 8.4 m primary mirror, a 9.6 deg$^2$ field of view, a 3.2 Gpx camera, and will survey $\sim 18000$ deg$^2$ of sky in six photometric bands ($ugrizy$) covering the wavelength range 320 – 1050 nm (LSST Science Collaboration et al., 2009; Ivezić and The LSST Science Collaboration, 2011; LSST Science Collaboration et al., 2017; Ivezić et al., 2019a). Around 90% of the observing time will be dedicated to the main survey mode: wide-fast-deep (WFD). During this mode, the telescope will survey $\sim 18000$ deg$^2$ of sky, visiting each location around 825 times (summed over all six bands) over the ten year period. It will reach magnitudes of 24.7 in single exposure and 27.5 co-added in the $r$ band (Ivezić and The LSST Science Collaboration, 2011; Ivezić et al., 2019a). The remainder of the observing time will be dedicated to mini-surveys, such as the Deep-Drilling-Fields (DDF) survey which will observe small patches of sky with a higher cadence and depth (Ivezić and The LSST Science Collaboration, 2011; LSST Science Collaboration et al., 2017; Jones et al., 2020). Other mini-surveys will cover the ecliptic plane, Galactic plane, and the Large and Small Magellanic Clouds. LSST will be roughly operating at ten times the scale of ZTF, yielding $\sim 10$ million alerts per night (Graham et al., 2019b).

1.3.3.3 Observing Strategy Optimization
Ivezić and The LSST Science Collaboration (2011) describes the minimum requirements to fulfil LSST science goals. However, this description is not complete, leaving flexibility for the choice and implementation of the survey observing strategy.
Figure 1.4: Forecast of the constraints on the dark energy equation of state $(w = w_0 + w_a(1 - a))$ from LSST probes after one year of data (Y1; left panel) and the full ten year survey (Y10; right panel), from each probe individually, and the joint forecast. 68% confidence intervals are shown in all cases and the plotted quantities $\Delta w_0$ and $\Delta w_a$ are the difference between $w_0$ and $w_a$ and their fiducial values of −1 and 0, respectively (The LSST Dark Energy Science Collaboration et al., 2018, 2021).
The strategy encompasses a number of aspects such as the survey footprint, time allocation to the different survey modes and targets-of-opportunity, season length, inter- and intra-night gaps, exposure time per visit, cadence of repeat visits and depth of the different passbands. Changes in the observing strategy influence the scientific output of LSST but the optimisation is challenging due to the diverse goals that often have conflicting requirements (LSST Science Collaboration et al., 2009; Ivezić et al., 2019a). Thus, the Rubin Observatory Construction and Early Operations teams involved the broad scientific community in the process of setting and refining the observing strategy (Bianco et al., 2021). This collaboration resulted in the Community Observing Strategy Evaluation Paper (LSST Science Collaboration et al., 2017), Cadence White Papers (e.g. Scolnic et al., 2018b; Lochner et al., 2018; Olsen et al., 2018; Feigelson et al., 2019; Smith et al., 2019), Cadence Notes (e.g. Lochner et al., 2021; Frohmaier et al., 2021), metrics to compare simulated observing strategies, and guidance from the LSST Science Advisory Committee and the Survey Cadence Optimization Committee (SCOC). Recently the Survey Cadence Optimization Committee (2022) Phase 1 report summarised the findings from the latest Cadence Notes and provided recommendations. They also asked for metrics and input on a range of topics, such as passband pairing and alternation, detailed inter- and intra-night cadence, footprint, DDFs strategies, and early science. The SCOC aims to deliver the recommendation which will define the baseline strategy for starting LSST in December 2022 (Survey Cadence Optimization Committee, 2022). Thus, the observing strategy optimisation work that we describe in Chapters 3 and 4 is critical and timely.

1.3.4 Time-Domain Extragalactic Survey

1.3.4.1 Science Goals

The 4-metre Multi-Object Spectroscopic Telescope (4MOST; De Jong et al., 2019) was developed to address several questions in cosmology, high-energy astrophysics, galaxy evolution, and Galactic Archaeology. It will perform these tasks by providing spectroscopic complements to large-area surveys such as Gaia (T. Prusti 8).

8 See a comprehensive list of Cadence White Papers in https://www.lsst.org/submitted-whitepaper-2018
9 See a comprehensive list of Cadence Notes in https://www.lsst.org/content/survey-cadence-notes-2021

57
et al., 2016), Euclid (Laureijs et al., 2011), VLT Survey Telescope (VLT; Capaccioli and Schipani, 2011), DES, and LSST.

In this section we focus on the campaign for spectroscopic follow-up of 4MOST, the Time-Domain Extragalactic Survey (TiDES; Swann et al., 2019). This survey aims to study the nature of dark energy, the extragalactic transient universe, and cosmology and galaxy evolution with AGN. More concretely, TiDES will spectroscopically classify live SNe, build an optimised training set for photometric classification and obtain spectroscopic redshifts for galaxies which had SNe. In addition to SNe science, TiDES will use AGN to build a Hubble diagram up to $z \approx 2.5$, and to constrain the cosmological equation of state. While observing AGNs, the survey will also measure the mass of their associated supermassive black holes; this is a crucial parameter in galaxy evolution (Swann et al., 2019). The events for follow-up will be selected from large sky surveys such as LSST (Swann et al., 2019). In the following paragraphs we expand upon the link between the TiDES SNe science goals and LSST.

The number of LSST-observed SNe that will be spectroscopically-confirmed through TiDES depends on several aspects, such as its overlap with LSST both in footprint and time, and their cadence (none of which is already decided); at the moment TiDES is set to start in early 2024 and LSST later in the same year.

As previously mentioned, in addition to using the spectroscopically-confirmed SNe for cosmology, TiDES will use its $\sim 30000$ live transient observations to construct an optimised training set for photometric classifiers (Swann et al., 2019; Carrick et al., 2021). However, Swann et al. (2019) and Carrick et al. (2021) estimate that TiDES will only be able to classify transients down to $r = 22.5$ magnitudes, which is brighter than the $r \approx 24$ magnitudes LSST will be able to obtain in a single visit. Consequently, the performance of photometric classifiers trained on this magnitude-limited dataset depends on solving the lack of representativity of the dataset. Carrick et al. (2021) explored this problem in detail and concluded that introducing faint spectroscopic-confirmed events from other sources (such as the VLT or the Extremely Large Telescope (Gilmozzi and Spyromilio, 2007)), and augmenting the training data leads to a more representative training set, and thus to a higher classification performance.
TiDES is expected to obtain \( \sim 50000 \) host galaxy spectroscopic redshifts (Swann et al., 2019). These will be used for LSST SN Ia cosmology (The LSST Dark Energy Science Collaboration et al., 2018; Swann et al., 2019) and to improve the performance of photometric classifiers via the use of spectroscopic redshift instead of photometric redshifts (Swann et al., 2019). The latter point is crucial since several photometric classifiers use the redshift information when classifying transients (e.g. Lochner et al., 2016; Muthukrishna et al., 2019; Boone, 2019; Carrick et al., 2021; Alves et al., 2022). Furthermore, Mitra and Linder (2021) found that using photometric redshifts for SNe at \( z \lesssim 0.2 \) bias the dark energy cosmology inference, and thus spectroscopic follow-up should be used for all SNe at \( z \lesssim 0.2 – 0.3 \).

1.3.4.2 Technical Details

TiDES makes use of the 4MOST wide-field spectroscopic survey facility that will be mounted in the four-metre-class Visible and Infrared Survey Telescope for Astronomy at Paranal, Chile. The instrument will have a 4.2 deg\(^2\) field of view, 1624 fibres leading into two low-resolution spectrographs, and 812 fibres feeding light a high-resolution spectrograph. Over its five-year survey, 4MOST will cover \( > 17000 \text{deg}^2 \) of sky twice (De Jong et al., 2019; Guiglion et al., 2019). Swann et al. (2019) anticipates that most of TiDES targets will come from LSST. Since wherever 4MOST points there will be LSST live transients and their density is too low for efficient stand-alone observations (TiDES uses \( \sim 2\% \) of 4MOST fibres), TiDES will ‘piggy-back’ on the other 4MOST surveys (Swann et al., 2019; Carrick et al., 2021).

TiDES contains three sub-surveys: TiDES-SN will target all bright (\( r < 22.5 \) magnitudes) live transients in a 4MOST field, which results in the \( \sim 30000 \) transients mentioned in Section 1.3.4.1. TiDES-Hosts will mostly select SNe with a complete light curve from LSST observations. The last sub-survey, TiDES-RM, will select AGN targets from the LSST DDF survey based on their variability history (Swann et al., 2019).

1.4 Outline of the thesis

In this section we outline the thesis structure and the contribution of the PhD candidate to the work undertaken. Chapter 2 describes the background for the methods
used in the following chapters. We introduce machine learning algorithms, such as gradient boosted decision trees and neural networks, and the data processing techniques gaussian processes and wavelets. Additionally, we briefly describe the methodology behind LSST simulations. In Chapter 3 we present the first exploration of the impact of observing strategy on photometric SNe classification with LSST. We also present the classification methodology used to obtain the results. The following Chapter 4 applies the same methodology to SNe simulated under more recent LSST observing strategies. Our results contribute towards the LSST observing strategy optimisation described in Section 1.3.3.3. In Chapter 5, we adapt the photometric SNe classification methodology for ZTF. We also show early results of this application on real ZTF SNe. Chapter 6 shows a proof-of-concept for using machine learning methods on lensed SN Ia to obtain their lens parameters and peak time. In Chapter 7, we describe future work and outline the conclusions of this thesis.

The thesis contains material from the following papers, together with ongoing work:

- Chapter 3: Considerations for optimising photometric classification of supernovae from the Rubin Observatory. This work was published as Catarina S. Alves, Hiranya V. Peiris, Michelle Lochner, Jason D. McEwen, Tarek Allam Jr., Rahul Biswas, and The LSST Dark Energy Science Collaboration, 2022, The Astrophysical Journal Supplement Series, 258, 23, and was carried out in collaboration with the named co-authors.

- Chapter 4: Impact of Rubin Observatory cadence choices on supernovae photometric classification. This work was published as Catarina S. Alves, Hiranya V. Peiris, Michelle Lochner, Jason D. McEwen, Richard Kessler, and The LSST Dark Energy Science Collaboration, 2023, The Astrophysical Journal Supplement Series, 265 (2), 43, and was carried out in collaboration with the named co-authors.

1.4.1 Candidate contributions

The PhD candidate searched and wrote the background information presented in the introductory Chapters 1 and 2.
In Chapter 3, the candidate updated most of the existing codebase to produce the results and prepare it to be publicly released. They also implemented new functionalities, in particular, a module to augment the training set (Section 3.5), the ability to use two-dimensional Gaussian process fits (Section 3.4.2), and a new machine learning model for photometric classification (Section 3.4.4). Other co-authors contributed to the codebase by discussing implementation, reviewing, and writing code that the candidate reviewed. Additionally, the candidate dealt with all the work necessary to publicly release the codebase. More broadly, the candidate processed and analysed all the data, performed experiments, and produced all the visualisations of Chapter 3. The results and interpretation were discussed among all co-authors.

In Chapter 4, the candidate acquired the input files necessary to simulate the SNe under recent LSST observing strategies. They later adapted an existing file to produce such simulations. One co-author helped to acquire the files and updated the SuperNova ANAlysis (Kessler et al., 2009b) code to solve the issues mentioned in Section 4.3.4; the candidate uncovered those implementation issues. Similarly to Chapter 3, all the processing, analysis, experiments, and visualisations were done by the candidate and the results and interpretation were discussed among co-authors.

In Chapter 5, the candidate did everything except the data acquisition. The results and conceptualisation were discussed with the co-authors.

In Chapter 6, the candidate simulated the lensed SNe by adapting the input files of the existing deeplenstronomy (Morgan et al., 2021) codebase (Section 6.3). Additionally, they significantly modified the neural network of ZipperNet Morgan et al. (2022); these modifications were discussed with the co-author. The candidate did all the processing, analysis, experiments, and visualisations of Chapter 6 except for Figure 6.1. The conceptualisation and results were discussed with the co-author.
In this chapter we present the methodologies we use throughout the thesis. In Section 2.1 we provide an overview of machine learning (Section 2.1.1) followed by details of Gradient Boosting Decision Trees (Section 2.1.2) and Gaussian processes (Section 2.1.4), which we extensively use to build photometric classifiers in Chapters 3, 4, and 5. In Section 2.1.3 we describe several neural networks architectures that we later use in Chapter 6 for estimating the properties of lensed SNe. Section 2.2 focuses on wavelets, which we use as features for our previously mentioned photometric classifiers (Chapters 3, 4, and 5). Lastly, in Section 2.3, we describe the key components to generate LSST simulations; we later use these simulations in Chapters 3 and 4.

2.1 Machine learning algorithms

2.1.1 Overview

Machine learning allows systems to learn how to solve problem from data without being explicitly programmed for that purpose. While some of the techniques have been known for more than fifty years (e.g. Samuel, 1959), the recent advent of large datasets led to the popularisation of machine learning and to a rapid increase of its applications in many fields, including Astrophysics (e.g. Lahav et al., 1995; Cranmer et al., 2021; Lochner and Bassett, 2021).

Often machine learning methods are categorised into supervised, unsupervised, and semi-supervised methods (e.g. Hodge and Austin, 2004; Chandola et
Supervised methods require a training set comprised of input features and their corresponding labelled output. These methods are traditionally used for classification tasks, where the dataset include objects from several classes, and regression tasks, where we want to predict continuous variables.

Unsupervised methods do not require a training set but rather only unlabelled input features. These methods are often used for clustering tasks where we want to group objects with similar properties.

Semi-supervised methods fall in between; during their training, the algorithms combine a small amount of labelled data with a larger amount of unlabelled input features. They are often used when labelled data is scare. Semi-supervised algorithms are used for classification, regression, and clustering (Xiaojin, 2008).

Although the categories described above are the most common, there are other machine learning categories, such as reinforcement learning (Kaelbling et al., 1996) and dimensionality reduction (Hotelling, 1933).

In this thesis we use supervised methods both for photometric classification of SNe (Chapters 3, 4, and 5) and regression of lensed SNe parameters (Chapter 6); see these chapters for a brief mention of other machine learning algorithms used for similar tasks. The following sections describe the algorithms we used.

As previously mentioned, a supervised machine learning algorithm requires a training set. This dataset is used to fit the parameters of the algorithm during the training procedure. Once the training is complete, we assess the performance of the algorithm on a new dataset that was not used in the training; we call it test set. Prior to training the algorithm, we must select some parameters; these are called hyperparameters and must be optimised outside the training algorithm. Generally, they describe high-level characteristics of the algorithm, such as when to stop the training. To optimise the hyperparameters, we can use a metric to evaluate the performance of the algorithm for all possible combinations of hyperparameter values on an independent validation set (grid-search); unlike the training set that is used to fit the model, the validation set is used to select the hyperparameters. We chose the
hyperparameter combination with the best performance. Since this choice is based on the validation set, it should prevent the model from overfitting to the training set.

To address overfitting problems, one way to optimise the hyperparameters is with a $k$-fold cross-validation (Kohavi et al., 1995; Rao et al., 2008). In this approach, the training set is divided into $k$ subsets, where $k$ is typically 5 or 10. Then, we use $k - 1$ sets to train the learning algorithm and the remainder set for validation. We repeat the procedure $k$ times to use each set as validation once. Finally, we chose the hyperparameter combination with the best performance given by the average score over the $k$ validation sets. Afterwards, we can verify if the algorithm is overfitting by comparing its performance on the validation set with an independent test set never used in the training and optimisation procedure. If the validation performance is better, the algorithm is overfitting. The division of a dataset into training, validation, and test sets depends on the problem and the amount of data available.

2.1.2 Gradient Boosting Decision Trees

A decision tree is a supervised learning method that maps input features to output classes by breaking up this complex decision problem into a union of simpler decisions. Those decisions are based on the value of the features, thus decision tree classifiers are easy to interpret when they are small (for example, the events that last less than a week and are only observed in the green passband are from class A). The tree starts with a node that includes all the training set (‘root node’). Next, the data is divided into the following nodes accordingly to the binary splits. This process is repeated for each node of the tree until we reach a ‘leaf node’ (or ‘child’), where no more splits are made. The leaf node yields the prediction of the decision tree for all the objects there. When the decision tree is used for classification, the final prediction of each leaf node corresponds to the probability of belonging to each class; this value comes from the fraction of training objects in each class at that node. When we are in a regression task, the leaf node predicts the single value of the target variable; it is the average target value of the training objects in that node. However, this classifier algorithm is sensitive to minor changes in the data and can easily became complex. Thus, in general, decision trees do not generalise well beyond the training data due to overfitting (Safavian and Landgrebe, 1991; James...
et al., 2013).

Aggregating many decision trees leads to a higher predictive performance. The two main approaches to create these ensembles are bagging (Breiman, 1996) and boosting (Freund and Schapire, 1996). These approaches differ in the trade-off between minimising the difference between the average prediction of the learning algorithm and the true value (bias), and the variability of the algorithms predictions for a given true value (variance):

- **Bagging**: In this approach we create several subsets of the data from the training sample chosen randomly with replacement. Next, each subset is independently used to train decision trees. At the end, the prediction of the ensemble is the average of the predictions from the independent individual trees. Bagging decreases the variance in the predictions but has no effect on the bias. Random forests is an example of this approach to ensemble learning algorithms.

- **Boosting**: In this approach, we sequentially add new decision trees that prioritise difficult-to-classify objects; the weight of an object is based on the last iteration. The ensemble prediction is the sum of the predictions from all the trees. Boosting decreases both the bias and the variance in the predictions (Bartlett et al., 1998). Boosted decision trees is an example of this approach.

In Chapters 3, 4, and 5 we use **Gradient Boosting Decision Trees** (GBDT) for classification. This learning algorithm combines boosting with gradient descent optimisation to construct an ensemble of decision trees. At each training step, the algorithm computes the gradient of a given loss function with respect to the value predicted by the ensemble and adds a tree in the direction of the gradient. In other words, additive models are sequentially fitted to the residuals by least squares at each iteration (Friedman, 2001, 2002). Figure 2.1 shows a graphical example. In this thesis we use the GBDT implementation of the python package `LightGBM`\(^1\) (Barbier et al., 2016; Ke et al., 2017; Zhang et al., 2017), which has a number of hyperparameters such as

\(^1\)lightgbm.readthedocs.io
2.1.3 Gaussian Processes

Gaussian processes (GPs) are a supervised learning algorithm that can be used for regression and classification tasks (Rasmussen and Williams, 2005). They are a generalisation of the Gaussian probability distribution; the probability distribution that describes random variables is replaced by a stochastic process that describes properties of functions. We can then think of a GP as a prior over a set of functions. Thus, when we constrain it with observations, we obtain a posterior containing the functions that are consistent with the provided observations. Formally, a GP is a collection of random variables such that any finite set of them have a joint Gaussian

A GP is fully specified by a mean function $\mu$ and a covariance function (also known as ‘kernel’) $K$. Their choice determines the weight of the diverse functions in the prior of the GP. Typically the mean function is null. However this is a loose limitation because the mean of the posterior process is not restricted to zero. The major issue is then finding $K$ as it encodes our beliefs about the function which we wish to learn (e.g. smoothness, length scale) (Rasmussen and Williams, 2005). For guidance on common kernels and their properties see Chapter 2 of Duvenaud (2014). Often the squared-exponential kernel (Kim et al., 2013; Fakhouri et al., 2015; Pessi et al., 2019) and the Matérn–3/2 kernel (Boone, 2019) are adapted, respectively

$$K(x_1,x_2;\sigma,l) = \sigma^2 \exp\left(\frac{(x_1-x_2)^2}{2l^2}\right)$$  \hspace{1cm} (2.1)

and

$$K(x_1,x_2;\sigma,l) = \sigma^2 \left(1 + \sqrt{3\frac{(x_1-x_2)^2}{l^2}}\right) \exp\left(-\sqrt{3\frac{(x_1-x_2)^2}{l^2}}\right),$$  \hspace{1cm} (2.2)

where $x_1, x_2$ are the input points, $\sigma^2$ is the amplitude scale of the functions produced by the GP, and $l$ is the length scale over which the functions vary. While there are numerous implementations of GP, the python packages George (Ambikasaran et al., 2014), GPy (authors, 2016), and sklearn (Pedregosa et al., 2011) are popular.

### 2.1.4 Neural Networks

Neural networks are machine learning algorithms inspired by information processing in biological systems (McCulloch and Pitts, 1943; Nielsen, 2015). The basic unit of these algorithms is a neuron; several neurons together form a layer, and combining multiple layers result on a neural network. Figure 2.2 shows a graphical representation of a neural network. In feedforward neural networks, neurons take inputs, perform a linear transformation followed by a non-linear one, and output a scalar number. Given an input $\vec{x}$, the output from the linear transformation of the $i$th neuron with associated weights $\vec{w}_i$ and bias $b_i$ is

$$z_i = \vec{w}_i^T \cdot \vec{x} + b_i.$$  \hspace{1cm} (2.3)
The non-linear transformation is then applied to the result: \( a_i = \sigma(z_i) \), where \( \sigma \) represents the non-linear function known as activation function. The appropriate activation function depends on the task, but common ones are step-functions, sigmoid, hyperbolic tangent, rectified linear units (ReLUs), leaky rectified linear units, and exponential linear units (Nwankpa et al., 2018). The operation described is then repeated for every neuron, using the \( a_i \) outputs of a layer as inputs to the following one. If all neurons of neighbouring layers are connected to each other, the layers are known as ‘fully-connected layers’.

Neural networks with a single hidden layer (layer that does not correspond to the input nor output) can approximate any continuous function with arbitrary accuracy (Hornik et al., 1989; Cybenko, 1989; Nielsen, 2015). However, there is no prescription on how to find the weights and bias values to construct the appropriate neural network. Hence, in practice, more hidden nodes and layers are typically used; more complicated functions require more hidden nodes to approximate it to the desired accuracy. In particular, multiple layers allow neural networks to learn features at various levels of abstraction. In practice, the applicability of neural networks depends on computational power, training data availability, task complexity, and other factors.

**Convolutional neural networks** (CNN; LeCun et al., 1989) are a type of feed-forward neural network with at least one convolutional layer prior to fully-connected
layers. In this context, a convolution can be represented by a matrix multiplication between the weights and the input data. The bias and activation function are also applied, as described before. While the weights in traditional neural networks are one-dimensional, in CNNs, both the input data and the weights are multi-dimensional. We can describe a convolution layer as sliding kernel over an image and calculating the sum of all products between the kernel parameters and the corresponding pixels. CNNs are often used for image classification and recognition. However for sequential data, other types of neural networks are strongly preferred.

Recurrent neural networks (RNNs; Elman, 1990) were designed to model sequences of data, such as sentences and time-series. They handle sequences of variable length by having a hidden state whose activation at each time is dependent on that of the previous time. **Long Short-Term Memory** (LSTM Hochreiter and Schmidhuber, 1997) networks are a type of RNNs capable of learning order dependence in sequence prediction problems. They introduce a ‘memory cell’ that can retain information for long periods of time, allowing LSTM to learn longer-term dependencies. Consequently LSTM have a higher performance than traditional RNNs (e.g. Chung et al., 2014).

In Chapter 6 we will use LSTM and CNN to construct neural networks.

### 2.2 Wavelets

We can represent any function as an infinite sum of sines and cosines. The discovery of this transformation of functions to a different basis led to the discovery and development of Fourier analyses. However, in Fourier analyses the frequency resolution is known at the expense of no time resolution. While one would like to have infinite resolution on both dimensions, the Heisenberg uncertainty inequality states that it is impossible to know both the exact frequency and exact time of occurrence of this frequency:

\[ \Delta t \Delta f \geq \frac{1}{4\pi}, \]  

(2.4)

where \( \Delta t \) and \( \Delta f \) are the resolutions in time and frequency, respectively (Riou and Vetterli, 1991).

Wavelet analysis allow us to process functions at different resolutions and
Figure 2.3: Three versions of the Symlet wavelet, one in each colour.

hence obtain information on time and frequency at the same time. We can use a large ‘window’ to notice gross features in the function and at a small ‘window’ to notice details (Rioul and Vetterli, 1991; Graps, 1995). The basis of this transformation are wavelets, wave-like oscillations with a finite domain (they are non-zero in a finite region); Figure 2.3 shows an example of wavelets (Nievergelt and Nievergelt, 1999). See Rioul and Vetterli (1991) and Graps (1995) for more details on wavelet analysis.

In this thesis we use wavelet transforms to obtain features for classification of transient events. This was previously done in (Varughese et al., 2015; Lochner et al., 2016; Sooknunan et al., 2021; Narayan et al., 2018). Due to the properties of wavelet transforms, these features are include both time and frequency information and they do not assume any physical knowledge about the events.

2.3 LSST Simulations

2.3.1 Observing Strategy Simulator

As mentioned in Section 1.3.3.3, the requirements to fulfil LSST science goals described in Ivezić and The LSST Science Collaboration (2011) leave flexibility in the implementation of the survey observing strategy. Thus, to empower the community and make science-driven optimisations, the Rubin Observatory released numerous simulated sequences of LSST pointings using different survey strategies (Bianco et al., 2021).

The software used to make the simulations changed over the years (Delgado et al., 2014; Delgado and Reuter, 2016; Naghib et al., 2019) but the current algo-
algorithm to decide which passband to use and which direction to point to at each moment (scheduler) is the LSST Feature-Based Scheduler (FBS; Naghib et al., 2019). This scheduler is automated and it does not use hand-crafted sequences of astronomical observations (proposals); it is instead based on a Markovian model. The decision to make a specific observation depends on the time it takes to redirect the telescope and dome from the current position to the next one, short-term science-driven requirements (e.g. whether to revisit a field in the same night), long-term requirements (e.g. maintain uniform coverage of the sky), the relative quality of the possible observations, and the general preference for observing fields around the meridian (Naghib et al., 2019). Moreover, the Markovian Decision Process used is memoryless (the next state is independent of the history), which makes the scheduler adaptable as any decision only requires the current state of the system. This is particularly important to deal with interruptions, such as maintenance downtime. See Naghib et al. (2019) for a detailed explanation of FBS.

2.3.2 Supernovae Simulations

In this thesis (Chapters 3 and 4) we use simulated light curves of SNe. These are built on top of the simulated sequences of LSST pointings mentioned on the previous section. The LSST schedulers (both previous and current) output a record of each simulated pointing of the survey. However, for generating light curves it is more convenient to reorganise the output such that all the observations of the same event are together. Then, one can simply iterate over them to generate light curves for events located at different pointings.

The python package OpSimSummary\(^2\) (Biswas et al., 2020) provides the link between the LSST schedulers output and the commonly used light curve simulation code SuperNova ANALysis (SNANA; Kessler et al., 2009b); it reorders the observations and translates the output cadence file into a format that SNANA can use. SNANA computes light-curve fluxes and uncertainties from image properties, rather than using actual images. It requires rest-frame source models and volumetric rate versus redshift for the classes of events we want to simulate, cosmological parameters (e.g. \(\Omega_\Lambda, w_0\)), telescope transmission in each passband, calibration reference, observing and image properties, and random numbers to generate Poisson fluctu-

\(^2\) https://github.com/LSSTDESC/OpSimSummary
Figure 2.4: Scheme of the \texttt{SNANA} simulation models. This is the Figure 13 from Kessler et al. (2009b).

...ations (Kessler et al., 2009b); some of these quantities are specified in the cadence file. Figure 2.4 illustrates the \texttt{SNANA} simulation models. In addition to light curves, \texttt{SNANA} also generates metadata associated with the simulated events (e.g. host galaxy photometric redshift and its uncertainty). Refer to Kessler et al. (2009b) for a detailed description of the simulation code \texttt{SNANA}. 
Considerations for optimising photometric classification of supernovae from the Rubin Observatory

This chapter is based on the paper Considerations for optimizing photometric classification of supernovae from the Rubin Observatory by Catarina S. Alves, Hiranya V. Peiris, Michelle Lochner, Jason D. McEwen, Tarek Allam Jr., Rahul Biswas, and The LSST Dark Energy Science Collaboration. The work was performed in collaboration with the named co-authors and the version presented here contains minor modifications to suit the thesis format.

3.1 Overview

The Vera C. Rubin Observatory will increase the number of observed SNe (supernovae) by an order of magnitude; however, it is impossible to spectroscopically confirm the class for all the SNe discovered. Thus, photometric classification is crucial but its accuracy depends on the not-yet-finalised observing strategy of LSST ( Rubin Observatory’s Legacy Survey of Space and Time). We quantitatively analyse the impact of the LSST observing strategy on SNe classification using simulated multi-band 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 supernovae 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 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 supernova classification with LSST.

3.2 Introduction

The upcoming LSST (LSST Science Collaboration et al., 2009, 2017; Ivezić et al., 2019a) is expected to discover, during its ten-year duration, at least one order of magnitude more SNe than the current available SNe samples (Guillochon et al., 2017). As mentioned in Section 1.2.3, 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., 2009a). 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 (SN PhotCC; Kessler et al., 2010a) in preparation for DES (Dark Energy Survey; The Dark Energy Survey Collaboration and Flaugher, 2005). Recently, PLAsTiCC\(^1\) (The PLAsTiCC team et al., 2018; Kessler et al., 2019) was launched in preparation for LSST, which will reach fainter magnitudes and have a ~ 4 times larger survey area compared to DES. The classifiers applied to the datasets from these challenges employed parametric fits, template fits, and machine learning models such as neural networks, boosted decision trees, support vector machine,

\(^1\)https://www.kaggle.com/c/PLAsTiCC-2018/
and gradient boosting (e.g. Kessler et al., 2010b; Lochner et al., 2016; Charnock and Moss, 2017; Pasquet et al., 2019; Muthukrishna 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 non-representative spectroscopically-confirmed events that are biased towards lower redshifts. Thus, recent work has focused on overcoming the lack of representativeness (Muthukrishna et al., 2019; Pasquet et al., 2019; Revsbech et al., 2017; Boone, 2019; Carrick et al., 2021). Photometric classification performance also depends on the survey observing strategy; however, this dependence has not yet been explored.

As mentioned in Section 1.3.3.3, 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 optimised (LSST Science Collaboration et al., 2017; Ivezić et al., 2018; Lochner et al., 2018; Gonzalez et al., 2018; Laine 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., 2019a). 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 optimisation (Scolnic et al., 2018b; Lochner et al., 2018, 2022). 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 smmachine² (Lochner et al., 2016) for use with LSST data and build a classifier based on wavelet features obtained from GP (Gaussian process) 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

---

² https://github.com/LSSTDESC/snmachine
the classification performance for different observing strategies with fixed total exposure time, where a reduced season length could be compensated by a higher cadence.

In Sections 3.3 and 3.4 we summarise the PLAsTiCC dataset and describe the classification pipeline, respectively. Section 3.5 focuses on the augmentation methodology. Our results and their implications for observing strategy are described in Section 3.6. We conclude in Section 3.7.

### 3.3 PLAsTiCC Dataset

The PLAsTiCC (The PLAsTiCC team et al., 2018; PLAsTiCC Team and PLAsTiCC Modelers, 2019) dataset 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\(^3\) is now outdated (Jones et al., 2020). PLAsTiCC mimicked future LSST observations in two survey modes: the WFD (Wide-Fast-Deep) survey, which covers almost half the sky and was used for 99% of the events, and the DDF (Deep-Drilling-Fields) survey, small patches of the sky with more frequent and deeper observations that have smaller flux uncertainties.

The simulations were divided into a non-representative spectroscopically-confirmed training set biased towards brighter events, and 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 modelled on the magnitude-limited TiDES (4-metre Multi-Object Spectroscopic Telescope Time Domain Extragalactic Survey, described in Section 1.3.4; Swann et al., 2019), resulting in a sample with a mean redshift \(\sim 0.3\). The unblinded dataset is available in PLAsTiCC Team and 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

\(^3\)Simulation minion\_1016: https://docushare.lsst.org/docushare/dsweb/View/Collection-4604
Table 3.1: Breakdown of the number of SNe per class used in this work (see simulation details in Kessler et al., 2019). For each class, the number of events in the training and test set is shown.

<table>
<thead>
<tr>
<th>SN class</th>
<th>(N_{\text{training}}) (%)</th>
<th>(N_{\text{test}}) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN Ia</td>
<td>2313 (58%)</td>
<td>1659831 (59%)</td>
</tr>
<tr>
<td>SN Ibc</td>
<td>484 (12%)</td>
<td>175094 (6%)</td>
</tr>
<tr>
<td>SN II</td>
<td>1193 (30%)</td>
<td>1000150 (35%)</td>
</tr>
<tr>
<td>Total</td>
<td>3990 (100%)</td>
<td>2835075 (100%)</td>
</tr>
</tbody>
</table>

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 3.1 shows a breakdown of the numbers of SNe in each class.

### 3.4 Classification pipeline

In this section we describe how we upgraded the photometric classification pipeline \textit{snmachine} for use with PLAsTiCC data. Augmentation was a crucial step in this process, and it is discussed in greater detail in Section 3.5.

#### 3.4.1 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 which contains the observations flagged as detected, and which has no inter-night gaps larger than 50 days. To introduce uniformity in the dataset, 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
Figure 3.1: 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 colours.

did for the test set. Figure 3.1 shows an example of light curve preprocessing.

3.4.2 Gaussian Process Modelling of Light Curves

We modelled each light curve with a GP regression (e.g. Section 2.1.3; MacKay, 2003; Rasmussen and Williams, 2005), following previous works that successfully used GP-modelled light curves in their classification pipelines (e.g. Lochner et al., 2016; 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 3.2 shows that such a two-dimensional GP fit infers the SNe light curve even in passbands where there are none 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, modelling 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 modelling explo-
Figure 3.2: SN Ibc light curve, where the points show the observations, along with their errorbars, 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 the passbands (shown in different colours). The two-dimensional GP infers the light curve in passbands where there are no (or few) observations, unlike the one-dimensional GP.

sive transients with sudden changes in their flux. The time dimension length-scale and amplitude were optimised 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 George⁴ (Ambikasaran et al., 2014).

3.4.3 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; Sooknunan et al., 2021; Narayan et al., 2018). 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., 2019). To obtain the wavelet decomposition,

⁴george.readthedocs.io/
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. Chen et al., 2013; Istas, 1992; Pope, 2019, and references therein), we note that our method of modelling the sparse light curves with GP fits and then using wavelet decomposition to obtain classification features is innovative. This approach was briefly mentioned in Varughese et al. (2015), and firstly 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).

Commonly, SALT2 models (Guy et al., 2007) are used to fit SNe and extract their light curve parameters (e.g. Jones et al., 2018). However, these models are geared towards SN Ia and therefore, they would mostly help in distinguishing between SN Ia and non-SN Ia, while in this work we focus on classifying SN Ia, SN Ibc, and SN II. Moreover, this approach goes against our aim of using a model-independent approach to feature classification and it is computationally intensive. Thus, we do not use SALT2 features.

3.4.4 Classification

We augmented the training set as described in Section 3.5, prior to training a classifier. We used the Gradient Boosting Model implementation of the package
Table 3.2: Optimised hyperparameter values used for the LightGBM model. A description of the hyperparameters is given in the library documentation.

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>WFD setting</th>
<th>DDF setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>boosting_type</td>
<td>gbdt</td>
<td>gbdt</td>
</tr>
<tr>
<td>learning_rate</td>
<td>0.24</td>
<td>0.24</td>
</tr>
<tr>
<td>max_depth</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>min_child_samples</td>
<td>25</td>
<td>70</td>
</tr>
<tr>
<td>min_split_gain</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>n_estimators</td>
<td>115</td>
<td>45</td>
</tr>
<tr>
<td>num_leaves</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

LightGBM\(^5\) (Ke et al., 2017), in particular the GBDT (Gradient Boosting Decision Tree; Friedman, 2001). As described in Section 2.1.2, 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 prioritise 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 optimised the GBDT hyperparameters (parameters of the model that must be set before the learning process starts) by maximising the performance of a 5-fold cross-validated grid-search on the augmented training set. First, each hyperparameter was optimised 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 optimisation. Finally we optimised this six-dimensional grid through a standard grid search. The resulting hyperparameter values are shown in Table 3.2. Since training and testing on the same events leads to overfitting, we placed in the same cross-validation fold all synthetic events that were derived from the same original event. While alternative hyperparameter optimisation 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.

\(^5\)lightgbm.readthedocs.io
3.4.4.1 Performance Evaluation

For a given event, our classifier outputs the probability of it belonging to each SN class. The higher the probability, the more confident the classifier is in its prediction. In order to evaluate the classification performance in this work, we used a metric that takes into account the probabilistic predictions of the classifier while rewarding confident predictions. Following The PLAsTiCC team et al., 2018, we used the PLAsTiCC weighted log-loss metric (Malz et al., 2019). This metric is related to the notion of entropy and can be interpreted as a measure of information. The PLAsTiCC weighted log-loss is given by

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

(3.1)

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. This is particularly relevant when we are more interested in a few classes, for example, rare events or SN Ia. The classifier will then focus more on correctly and confidently predicting the up-weighted classes. Since in this work, we investigated the impact of the LSST observing strategy in the general SNe classification, we gave the same weight to every SNe class, following the PLAsTiCC challenge.

We used confusion matrices to visualise the mislabelled classes; Table 3.3 shows the confusion matrix for a binary classification. For ease of comparison, we normalised 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

\[
\text{recall} = \frac{TP}{TP + FN}
\]  
(3.2)

and

\[
\text{precision} = \frac{TP}{TP + FP}.
\]  
(3.3)

A perfect classifier has both recall and precision equal to one. However, in
practical cases, these quantities have an inverse relationship, where we can in-
crease the value of one at the cost of decreasing the other. For example, assigning
all the events predicted with more than 33% probability as SN Ia to that class will
likely correctly assign most of the true SN Ia (high recall). However, it will also
include more non-SN Ia events than if we had a higher threshold (low precision).
This contamination can bias the estimation of cosmological parameters. We briefly
discuss contamination and ways to mitigate its impact on cosmology in Section 3.7.

3.5 Augmentation

As previously outlined, the PLAsTiCC training set is non-representative of the test
set in redshift (see Figure 3.3) and also imbalanced: the most common SNe class
has \(\sim 4.8\) 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 3.5.1).

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
Redshift distribution per class

Figure 3.3: Host galaxy photometric redshift distribution per supernova class, where SN Ia, SN Ibc, SN 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 3.5) is close to the desired test distribution.

Figure 3.4: The left panel shows a SN Ibc light curve where the points show the observations, along with their errorbars, 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 3.5. 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.

We adapted the approach used in Boone (2019) for our training set augmentation. Figure 3.4 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 3.5.1).
2. Model the original light curve with a two-dimensional GP fit in time and wavelength (as described in Section 3.4.2).
3. Choose a redshift for the synthetic event (Section 3.5.2).
4. Create synthetic observations at the new redshift, making use of the GP fit to
5. Generate a photometric redshift (Section 3.5.4).

The WFD and DDF surveys have very different characteristics and enable qualitatively different science goals. Hence we found it is important to use customised 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 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 customised for the two survey modes as necessary.

### 3.5.1 Number and Class Balance of Synthetic Events

As we wish to optimise 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 ($\sim 59\%$ SN Ia, $\sim 6\%$ SN Ib/c, $\sim 35\%$ SN II). However this gave worse performance, biasing the predictions toward the most common class.

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

### 3.5.2 Redshift Augmentation

As previously outlined, redshift augmentation was found to be critical for the PLAsTiCC dataset (Boone, 2019). Figure 3.3 shows the bias of the training set towards low-redshift events in comparison to the test set. We augmented each training set event of the WFD survey between

$$z_{\text{min}} \approx \max \{0, 0.90z_{\text{ori}} - 0.10\} \quad \text{and}$$

$$z_{\text{max}} \approx 1.43z_{\text{ori}} + 0.43,$$

(3.4)
where \( z_{\text{ori}} \) 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 \( \log(z_{\text{min}}) \), and maximum value \( \log(z_{\text{max}}) \). Then, we calculated the redshift of the new augmented event \( z_{\text{aug}} \),

\[
  z_{\text{aug}}(z^*) = -z^* + z_{\text{min}} + z_{\text{max}}.
\]  

(3.5)

For the deeper DDF survey, the corresponding \( z_{\text{max}} \) 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 Boone (2019), which also uses a class-agnostic augmentation. We derive the aforementioned redshift limits for augmentation in Appendix A.1.

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

### 3.5.3 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 summarise 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\(^6\). 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 \( (1 + z_{\text{aug}})/(1 + z_{\text{ori}}) \). 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

\(^6\) 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.
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 $z_{\text{aug}}$. Further details of this redshifting implementation are given in Appendix A.1.

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 2-detection trigger based on PLAsTiCC (Kessler et al., 2019). Boone (2019) fitted an error function to the observations from the full dataset to predict the probability of detection as a function of signal-to-noise ratio (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 constraining 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.

3.5.4 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.

3.5.5 Computational Resources

We performed our computations on an Intel(R) Xeon(R) 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 $4 \times 10^4$ events takes ~9 hrs. Reducing the dimensionality using PCA takes ~30 min for an augmented training set of $4 \times 10^4$ events and optimising the LightGBM classifier on the same training set takes ~8 hrs. After we computed the test set features, generating predictions with the trained classifier takes ~10 min. Overall, the entire classification pipeline takes ~70 core hours of computing time for WFD and 12 for DDF in this setting.

3.6 Results and Implications for Observing Strategy

We now turn to our results on the PLAsTiCC dataset 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 3.6.1. We then explore the performance as a function of light curve length (Section 3.6.2), median inter-night gap (Section 3.6.3), number of gaps > 10 days (Section 3.6.3), and number of observations near the peak (Section 3.6.4).
3.6.1 Survey Mode-specific Augmentation and its Effect on Performance

Figure 3.5 shows the confusion matrix for the classifier trained on an augmented WFD training set as described in Section 3.5. Despite the use of general wavelet features which 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 similar to other classifiers, the performance is weakest for SN 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 3.5.

We now compare the DDF test set classification performance when using a classifier which 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 3.6 shows that the classifier optimised for the WFD test set obtains a worse performance on the DDF test set, with a higher log-loss (0.570 vs 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.
Figure 3.6: DDF test set normalised 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 customised for the specific survey mode characteristics.

Figure 3.7: 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.

3.6.2 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 3.7 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 the range the results are dominated by small-number effects. As expected, events observed for longer are better characterised 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 by a higher cadence.
Figure 3.8: 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; $\sim 64\%$ of SN Ia, $\sim 63\%$ of SN Ibc, and $\sim 66\%$ of SN II from the test set events have median inter-night gap $< 3.5$ days and light curve lengths between 50 and 175 days.

Our findings support the minimum five-month season length recommendation in Lochner et al. (2018, 2022).

3.6.3 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 gap. The left panel of Figure 3.8 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, 2022) 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 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.
Figure 3.9: 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 SNe 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.

The upper left panels of Figure 3.9 show the recall and precision are broadly independent of the number of large gaps in a light curve\(^7\). 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 3.2, 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 3.9 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 recommended a regular cadence without inter-night gaps larger than 10 – 15 days (Lochner et al., 2018, 2022), we find that requiring a median inter-night gap of $< 3.5$ days has little impact on the classification performance.

\(^7\) SN Ibc and SN II have a small recall increase for 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 SN Ibc and SN II. Note that uncertainties are also larger for cases with a greater number of large gaps due to small-number statistics.
Figure 3.10: 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 $\geq 9$ 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.

...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 independent of the inter-night gap.

### 3.6.4 Observations Near Peak

Obtaining observations near the peak of a SN Ia light curve is generally considered critical to obtain a reliable cosmological distance modulus. To investigate whether SNe classification has a similar requirement, we analysed 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 maximises the GP fit predicted flux. Figure 3.10 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 characterisation of light curve shape. However, since we cannot predict when a 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.
3.7 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 specialised features constructed using domain knowledge about SNe). We find that we can augment the original training set in a number of aspects (the photometric redshift distribution per supernovae class, the distribution of the observing cadence, and the flux uncertainty distribution) to construct a representative training set for machine learning classification. This is crucial as we demonstrated by the difference in the classification performance for the DDF survey with and without representative training sets in Section 3.6.1. While previous studies using the PLAsTiCC data had indicated augmentation is important, we extend this conclusion for a new framework and go into more detail by studying the augmentation for the WFD and DDF surveys individually.

Our classifier yields similar performance to the top PLAsTiCC submission (Boone, 2019; Hložek et al., 2020) and competitive results in core-collapse SN contamination (see below; Kessler and Scolnic, 2017; Jones et al., 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 SN Ia with > 50\% probability) which is comparable to the \( \sim 5\% \) contamination obtained in Jones et al. (2018) with Pan-STARRS SNe. This could be further improved by optimising 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 marginalise 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 customised 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 five-month season length recommendation in Lochner et al. (2018, 2022). 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 Gaussian process methods we use to interpolate such gaps effectively. Finally, a regular cadence which achieves $>9$ observations near the peak of the light curve provides effective classification performance. In Appendix A.2 we show that these results also hold if we replace our classification predictions with the predictions obtained by Boone (2019), who use 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.

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. In Section 4, we will use our augmentation and classification pipeline to study the SNe classification performance of more recent observing strategy simulations.

We publicly released the photometric transient classification library snmachine8 with the paper associated to this work (Alves et al., 2022). The library also contains some example Jupyter notebooks which can be used to reproduce this work. In the future, the snmachine pipeline can be extended to facilitate the classification of other transient classes.

---

8 https://github.com/LSSTDESC/snmachine
4

Impact of Rubin Observatory cadence choices on supernovae photometric classification

This chapter is based on the paper *Impact of Rubin Observatory cadence choices on supernovae photometric classification* by Catarina S. Alves, Hiranya V. Peiris, Michelle Lochner, Jason D. McEwen, Richard Kessler, and The LSST Dark Energy Science Collaboration. The work was performed in collaboration with the named co-authors and the version presented here contains minor modifications to suit the thesis format.

4.1 Overview

LSST (Vera C. Rubin Observatory’s Legacy Survey of Space and Time) will discover an unprecedented number of SNe (supernovae), making spectroscopic classification for all the events infeasible. LSST will thus rely on photometric classification, whose accuracy depends on the not-yet-finalised LSST observing strategy. In this work, we analyse the impact of cadence choices on classification performance using simulated multi-band light curves. First, we simulate SNe with an LSST baseline cadence, a non-rolling cadence, and a presto-color cadence which observes each sky location three times per night instead of twice. Each simulated dataset 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 supernovae relative to the background region of the same observing strategy. 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.

4.2 Introduction

As mentioned in Section 1.2, SNe are used for diverse astrophysical and cosmological studies, such as measurements of the Universe’s accelerated expansion (e.g. Riess et al., 1998; Perlmutter et al., 1995; Astier et al., 2006; Kessler et al., 2009a; Betoule et al., 2014; Scolnic et al., 2018a; Abbott et al., 2019b; 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 LSST (LSST Science Collaboration et al., 2009, 2017; Ivezić et al., 2019a). Thus, as referred in Section 3.2, LSST will rely on photometric classification, utilising spectroscopically-confirmed SNe samples to train classifiers.

SNPhotCC (Supernova Photometric Classification Challenge; Kessler et al., 2010a) and PLAsTiCC† (Photometric LSST Astronomical Time-Series Classification Challenge; The PLAsTiCC team et al., 2018; Kessler et al., 2019) catalysed the development of photometric classifiers in preparation for the DES (Dark Energy Survey; The Dark Energy Survey Collaboration and Flaugher, 2005) and LSST (LSST Science Collaboration et al., 2009, 2017; Ivezić et al., 2019a), respectively. Many of the resulting recent classifiers rely on machine learning methods, such as neural networks (Charnock and Moss, 2017; Muthukrishna et al., 2019; Möller and Boissière, 2019; Villar et al., 2020; Boone, 2021; Qu and Sako, 2022), boosted decision trees (Boone, 2019; Alves et al., 2022), and self-attention mechanisms.

†https://www.kaggle.com/c/PLAsTiCC-2018/
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. Chapter 3; Lochner et al., 2016). However, classifiers are usually trained with either simulated datasets, which may suffer from model misspecification, or spectroscopically-confirmed events, which are non-representative 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; Pasquet et al., 2019; Boone, 2019; Carrick et al., 2021). This previous work has demonstrated that the bias introduced by non-representative training sets can be corrected.

As shown in Chapter 3, another crucial factor which impacts the accuracy of photometric classification is the survey observing strategy. Over the course of ten years, LSST will 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 intra-night 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 optimisation is challenging due to the diverse goals of LSST (LSST Science Collaboration et al., 2009; Ivezic et al., 2019a, see Section 1.3.3.1 for a description of the science goals).

Recently, the Survey Cadence Optimization Committee (2022) Phase 1 report (hereafter: Ph1R) narrowed down the choice of possible observing strategies and recommended new simulations\(^2\) to respond to the findings of the previous optimisation work (e.g. LSST Science Collaboration et al., 2017; Lochner et al., 2018; Scolnic et al., 2018b; Gonzalez et al., 2018; Olsen et al., 2018; Laine et al., 2018; Jones et al., 2020; Bianco et al., 2019; Alves et al., 2022; Lochner et al., 2022) and enable further optimisation. In particular, it is not yet decided whether LSST will use a rolling cadence\(^3\), and whether it will visit each sky pointing two or three times.


\(^3\) In a rolling cadence strategy, LSST observes a part of the sky at a higher cadence than the rest. After a fixed period, usually one year, the rolling moves to a different part of the sky.
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 intra-night observing strategy, since we expect these factors to have the greatest impact on the efficacy of light-curve classification.

Our work builds upon Chapter 3 by studying the performance of photometric SN classification for light curves simulated with different LSST observing strategies for the first three years 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 multi-band light curves using the SNANA package\(^4\) (SuperNova ANAlysis; Kessler et al., 2009b). These simulated datasets included a non-representative spectroscopically confirmed training set, biased towards brighter events. Next, we followed the classification approach of Section 4.4, using the photometric transient classification library snmachine\(^5\) (Lochner et al., 2016; Alves et al., 2022) to build a classifier based on wavelet features obtained from GP (Gaussian process) 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 4.3 we describe the LSST observing strategies and the framework that we used to generate our SNe datasets. Our classification and augmentation methodologies that relied on snmachine are presented in Section 4.4. Section 4.5 focuses on our results and their implications for observing strategy. We conclude in Section 4.6.

### 4.3 Simulation of LSST Supernovae

#### 4.3.1 Overview

In this work we simulated LSST-like SN light curves for the first three years of the survey using three observing strategies: baseline_v2.0, noroll_v2.0, and

\(^4\)https://snana.uchicago.edu/

\(^5\)https://github.com/LSSTDESC/snmachine
These observing strategies were created with the FBS (Feature-Based Scheduler; 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 4.3.2, the SNe models in Section 4.3.3, and the simulations infrastructure in Section 4.3.4.

Following The PLAsTiCC team et al. (2018) and Kessler et al. (2019), we simulated the WFD (Wide-Fast-Deep) survey, which is the main survey of LSST (containing \( \sim 98\% \) of our simulated events), and the DDF (Deep-Drilling-Fields) 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 4.4.3) but are not included in the test set.

We used 0.2% of the simulations to construct a non-representative spectroscopically-confirmed training set. The training set was biased towards brighter events, with a median redshift \( \sim 0.3 \). 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 TiDES (4-metre Multi-Object Spectroscopic Telescope Time Domain Extragalactic Survey; Swann et al., 2019).

### 4.3.2 Observing Strategies

The SCOC\(^6\) (Rubin’s Survey Cadence Optimization Committee) 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

---

\(^6\) For further details see https://www.lsst.org/content/charge-survey-cadence-optimization-committee-scoc.
optimisations (Ph1R), including an update of LSST baseline observing strategy: baseline_v2.0. Under the updated baseline, the telescope observes each field twice with a gap of approximately 15 min during twilight and 33 min 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 which is observed more often (rolling at 90% strength) and a ‘background’ area. This two-band rolling cadence is defined by declination and shown in Figure 4.1. 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 three years 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 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 (noroll_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 dataset as Y1.5-3 baseline; we used Y0-3 baseline when considering the entire three years. The simulations are otherwise identical.
Another aim of the new observing strategies is to investigate modifications of
the intra-night 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 tran-
sient detection. The presto-color family (Bianco et al., 2018, 2019) encompasses
a number of variations of the third visit inclusion, such as different intra-night gaps
between the observations (e.g. 1.5 hrs to 4 hrs between the first pair of observa-
tions and the third), whether the initial pair of visits is in consecutive passbands
\((g + r, r + i, r + z, i + z)\) or mixed passbands \((g + i, r + z, i + y)\), 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 intra-night gaps between 1.5 hrs and 4 hrs
has minimal impact. We thus chose a presto-color cadence whose third visit has
an intermediate value of 2.5 hrs for the intra-night gap (presto_gap2.5_mix_v2.0,
hereafter referred to as presto-color). Since the total number of visits per point-
ing is fixed, adopting a presto-color cadence results in longer inter-night 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 three years of LSST. For more details on the
simulations, see Ph1R and the descriptions in the associated Jupyter Notebook.

### 4.3.3 Supernovae Models

Following Chapter 3, 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 Guy et al. (2010), Kessler et al. (2010b, 2013), Villar et al. (2017), Pierel et al.
(2018), and Guillochon 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 B.1 of Appendix B.1 for the
relative rates used to simulate CC SNe in this work. Table 4.1 shows the resulting
number of SNe per class for each observing strategy.
Table 4.1: Breakdown of the number of SNe per class and observing strategy used in this work. (Left) events simulated between years 1.5 and 3 of the survey. (Right) events simulated between years 0 and 3 of the survey.

<table>
<thead>
<tr>
<th></th>
<th>Y1.5-3 baseline</th>
<th>Y0-3 baseline</th>
<th>No-roll</th>
<th>Presto-color</th>
</tr>
</thead>
<tbody>
<tr>
<td>SN class</td>
<td>N&lt;sub&gt;training&lt;/sub&gt; (%)</td>
<td>N&lt;sub&gt;test&lt;/sub&gt; (%)</td>
<td>N&lt;sub&gt;training&lt;/sub&gt; (%)</td>
<td>N&lt;sub&gt;test&lt;/sub&gt; (%)</td>
</tr>
<tr>
<td>SN Ia</td>
<td>1738 (64%)</td>
<td>755330 (62%)</td>
<td>3421 (65%)</td>
<td>1563427 (62%)</td>
</tr>
<tr>
<td>SN Ibc</td>
<td>231 (8%)</td>
<td>57854 (5%)</td>
<td>460 (9%)</td>
<td>119703 (5%)</td>
</tr>
<tr>
<td>SN II</td>
<td>759 (28%)</td>
<td>405139 (33%)</td>
<td>1392 (26%)</td>
<td>818241 (33%)</td>
</tr>
<tr>
<td>Total</td>
<td>2728 (100%)</td>
<td>1218323 (100%)</td>
<td>5273 (100%)</td>
<td>2501371 (100%)</td>
</tr>
</tbody>
</table>

4.3.4 Framework for Generating Simulations

Our SNe simulations were built on top of the observing strategy cadences produced by FBS previously discussed in Section 4.3.2 (Naghib et al., 2019); as mentioned in Section 2.3.1, 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. 

---

7 The observing strategies are hosted in [https://epyc.astro.washington.edu/~lynnej/opsim_downloads/fbs_2.0/](https://epyc.astro.washington.edu/~lynnej/opsim_downloads/fbs_2.0/).

8 [https://github.com/LSSTDESC/OpSimSummary](https://github.com/LSSTDESC/OpSimSummary)
(Biswas et al., 2020, 2022) to reorder of 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 datasets 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 only keep events with at least two detections in our datasets; 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-year survey to the first three years, 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\(^9\) 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, we used the more realistic value of 100,000 photoelectrons/pixel. We also corrected the code to ensure that any observations in the same band in a given night are co-added and count as a single observation. We provide our SNANA input files for each observing strategy simulation on zenodo.

### 4.4 Photometric Classification

We followed the approach of Chapter 3 to photometrically classify the SNe simulated from each observing strategy. In that chapter, we benchmarked our classification approach against the winning PLAsTiCC entry (Boone, 2019) and showed that our classification results were generalisable; they hold if we replace our classification...\(^9\) In this work we used SNANA version v11_04i.

---

\(^9\) In this work we used SNANA version v11_04i.
tion predictions with the predictions of Boone (2019). Here we used the photometric transient classification library \textit{snmachine} (Lochner et al., 2016; Alves et al., 2022) and updated it to handle the output files of \textit{SNANA} (FITS files). In the sections below, we describe the main steps of the approach and any modifications relative to Chapter 3.

4.4.1 Light Curve Preprocessing

Following Section 3.4.1, 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 inter-night gaps $> 50$ days; we selected our preprocessed light curve as the sequence of observations which 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.

4.4.2 Gaussian Process Modelling of Light Curves

We used GP regression (e.g. MacKay, 2003; Rasmussen and Williams, 2005) to model each light curve. Following Boone (2019) and Section , 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 $6000\,\text{Å}$ and used maximum likelihood estimation to optimise the time dimension length-scale and amplitude per event. We implemented the GPs with the python package \texttt{George}\footnote{\texttt{george.readthedocs.io/}} (Ambikasaran et al., 2014). We note that Stevance and Lee (2022) investigated possible improvements to using GPs for SNe light curve fitting. We leave these extensions to future work on SNe classification.

4.4.3 Augmentation

We applied the methodology developed in Chapter 3 to augment the training set of each simulated observing strategy to be representative of their respective test set in terms of the photometric redshift distribution per SNe class, the cadence of ob-
Figure 4.2: Host galaxy photometric redshift distribution per supernova class for Y0-3 baseline, where SN Ia, SN Ibc, 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.

servations, and the flux uncertainty distribution. We delineate below the departures from the augmentation procedure described in Section 3.5 and refer the details to Appendix B.2.

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 a SN model (as mentioned in Section 4.3.3), the redshift distribution of the events changed with respect to Chapter 3. Consequently, we used a different distribution to produce the augmented training sets, as detailed in Appendix B.2.

Following our previous work, we generated 15440 WFD synthetic events for each SNe class. Figure 4.2 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. Comparable figures for the other observing strategies are shown in Figure B.1 of Appendix B.2.

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 the test set; Table B.2 of Appendix B.2 shows the parameters used for each observing strategy. For the flux uncertainty, we followed Boone (2019) and Section 3.5.3 by combining the uncertainty predicted by the GP in quadrature with a value drawn from the flux uncertainty distribution of the test set. Table B.3 of Appendix B.2 shows the parameters of the Gaussian mixture model used to fit the flux uncertainty distribution of each passband and observing strategy.
4.4.4 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 Section 3.4.3, which we briefly summarise 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 Section 3.4.3 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\(^\text{11}\). These decomposition choices resulted in 7008 redundant wavelet coefficients per event. Following Lochner et al. (2016) and Section 3.4.3, we reduced the dimensionality of the wavelet space to 40 components using PCA (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.

4.4.5 Classification

We used \texttt{snmachine} to build a photometric classifier trained on the augmented training set. We used GBDT (Gradient Boosting Decision Trees; Friedman, 2002), classifiers whose predictions are based on ensembles of decision trees. We trained the classifier for each observing strategy separately, using dedicated augmented training sets (Section 4.4.3) and features (Section 4.4.4). The GBDT classifier hyperparameters were optimised following the procedure described in Section 3.4.4; Table 3.2 of Appendix B.2 shows the values of the hyperparameters per observing strategy.

\(^{11}\)Using the \texttt{PyWavelets} (Lee et al., 2019) package as part of \texttt{snmachine}.
4.4.5.1 Performance Evaluation

Similarly to Chapter 3, we used the PLAsTiCC weighted log-loss metric (The PLAsTiCC team et al., 2018; Malz et al., 2019) to optimise 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 normalised 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 and precision of each SNe class. These are defined as in Equations 3.2 and 3.2, respectively.

The computational performance of this procedure and an estimate of the resources needed for reproducing this analysis are discussed in Appendix B.3.

4.5 Results and Implications for Observing Strategy

Here we present our results on the impact of rolling cadence and the intra-night gap on SNe classification performance. We perform a comparative analysis for these two cases relative to the baseline strategy in Section 4.5.1. In Chapter 3 we found that light curve length (time difference between the first and last observation after the light curve preprocessing described in Section 4.4.1) and inter-night gap (time difference between consecutive observations which are more than 12 hrs apart) were the key properties of observing strategies affecting classification performance. We thus investigate how the no-roll and preto-color families affect the recall and precision as a function of several factors including the above.

4.5.1 Overall Classification Performance

Figure 4.3 shows the confusion matrices for classifiers trained on the augmented training set of Y1.5-3 baseline and no-roll. The Y1.5-3 baseline classifier yields a slightly higher performance for SN Ibc, 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 to 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 4.1; 65% of the test set events) and the background region (35% of the test set events). The confusion matrices in Figure 4.4 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 as compared to the background of the same observing strategy.

Figure 4.5 shows the confusion matrices for Y0-3 baseline and presto-color. The baseline cadence outperforms presto-color for SN Ia, SN II; the PLAsTiCC log-loss metric degrades by ~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 (Chapter 3 reported a log-loss metric of 0.550 for this case). This indicates substantial performance gains achieved by recent updates to the FBS scheduler.
Figure 4.4: Normalised 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.

Figure 4.5: Normalised 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 ~10% higher SNe classification performance.
4.5.2 Light Curve Length

We found in Chapter 3 that the light curve length of an event has a significant impact on classification performance 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 4.6, our results for the Y0-3 baseline show a similar performance behaviour. We also find that these conclusions generalise to the other observing strategies analysed here, and hence the conclusions of Chapter 3 carry over to these new cadence simulations. We note that the recall and precision figures (Figure 4.6 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 4.7 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.
**Figure 4.7**: Test-set density of events as a function of light curve length for Y1.5-3 baseline vs no-roll (left panel), and Y0-3 baseline vs presto-color (right panel). The dashed lines mark the high performance region between 50 and 175 days.

**Figure 4.8**: Test-set recall (left panel) and precision (right panel) as a function of inter-night gap per SNe class for 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 only present the results for bins with more than 300 events.

### 4.5.3 Median Inter-night Gaps

The observing strategies proposed for LSST have different intra- and inter-night gaps distributions. Given the finite total number of observations available, these distributions are intrinsically linked. In Section 3.6.3 we demonstrated that the median inter-night gap was a crucial factor in photometric SNe classification, and Ph1R has highlighted the necessity for science-motivated metrics to measure the impact of intra-night 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 towards characterisation of the light curve. Here we firstly investigate the impact of a higher intra-night cadence on the median inter-night gap, and hence classification performance.

Figure 4.8 shows that, in accordance with our expectations, a lower median
inter-night gap leads to higher precision and recall for Y0-3 baseline. SN II show a larger sensitivity to cadence as compared to the other types because SN Ia with high median inter-night gaps are misclassified as SN II, driving down the precision of the latter. Overall, we find similar results for the other observing strategies analysed.

While this overall conclusion still holds, our results show that the classification performance depends less on the median inter-night gap for this set of observing strategy simulations compared to Section 3.6.3, where we recommended a median inter-night 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 inter-night 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 4.9 shows that the peak of the median inter-night gap for 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 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 4.9. However, the small shift in the median inter-night gap distribution does not explain the degradation in performance seen for presto-color. We now
Figure 4.10: Test-set recall (left panel) and precision (right panel) as a function of the length of the longest inter-night gap per SNe class for 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 only present the results for bins with more than 300 events. The reduced performance below 8 days corresponds to less than 5% of the events which tend to have very short light curves and therefore are not well classified.

turn to other cadence properties in order to understand this result.

4.5.4 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 which are potentially important for classification include large gaps in the light curve and the number of observations near the peak.

In Section 3.6.3 we found that the GPs successfully interpolate between large gaps (> 10 days) so the classifier is still able to identify the SNe. Figure 4.10 confirms that the recall and precision of SNe either slowly decrease or remain constant with the increase of the length of longest inter-night gap. These conclusions also generalise to the other cadences we study. This indicates that the GP step is generally able to interpolate large gaps.

A related consideration for characterising 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 facilitates accurate photometric classification. In this work, we estimate the SNe peak as the moment that maximises 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.
Figure 4.11: Test-set recall (left panel) and precision (right panel) as a function of number of observations near peak per SNe class for 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 only present the results for bins with more than 300 events.

in all passbands to calculate this quantity. Similarly to Section 3.6.4, we find that the classification performance generally increases with the number of observations near the peak for all the new observing strategies. Figure 4.11 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 4.11 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 4.3 and 4.5 for the observing strategies considered.

The distribution of the longest inter-night gap in Y1.5-3 baseline exhibits two peaks, as shown in the left panel of Figure 4.12. This is due to the fact that different areas of the sky start rolling at different times, and Y1.5-3 baseline therefore includes some events in areas of the sky which have not yet started rolling.
Thus, we see a second peak at higher values of longest inter-night gap for 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 towards longer gaps. By contrast, the right panel of Figure 4.12 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 which have a long gap of 20 days or more.

Figure 4.13 compares the distributions of the number of observations near the SNe peak for the various cadences. The left panel of Figure 4.13 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 4.13 shows that presto-color events have \(~\) 10\% more events with < 15 observations near peak compared to Y0-3 baseline. While this difference does not make a large visual impact, it is nevertheless in a regime which strongly affects classification performance.

Figure 4.5 shows that presto-color mainly impacts classification of type II SNe. In turn, Figure 4.11 shows that type II SNe classification performance is a strong function of number of observations near peak as compared to the other classes, for < 15 observations near peak. Since presto-color has more events in this regime, one may therefore expect that SNII 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.

### 4.6 Discussion and Conclusions

We have presented the impact of LSST cadence choices on the performance of SNe photometric classification, using simulated multi-band light curves from the LSST baseline cadence, the non-rolling cadence, and a presto-color cadence. For each dataset considered, we augmented the non-representative 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 inter-night 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 per-
formance 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 SN Ia which form the basis of seven cosmology metrics; in Appendix B.4 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 (dark and light bands of Figure 4.1). We found that ~ 50% of the SN Ia in the active region fulfilled the light curve requirements, compared with only ~ 20% of the SN 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 SN 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 4.7 is skewed towards 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 (Kessler and Scolnic, 2017; Jones et al., 2017). While a Bayesian methodology can marginalise over the contamination, minimising such contamination reduces
systematic uncertainties in cosmological constraints (Kunz et al., 2007; Lochner 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 Appendix A.2. In the future, we could 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-focussed experimental design and optimisation of the LSST observing strategy.
Photometric classification of supernovae from Zwicky Transient Facility

This chapter is part of an ongoing work in collaboration with Prof Ariel Goobar.

5.1 Overview

The upcoming LSST (Rubin Observatory’s Legacy Survey of Space and Time) will rely on photometric classification to identify most of the discovered SNe (supernovae). Previously, we built classifiers with the photometric transient classification library snmachine and used them on simulated LSST datasets. However, the SNe models are continuously evolving as we learn more about these events so we expect the current simulations to differ from real data. Thus, it is crucial to test our approach on real data to further improve the algorithm, and have a more accurate understanding of its real performance. In this work we classify real SN-like events from the Zwicky Transient Facility Bright Transient Survey. First, we use the NASA Extragalactic Database to match the observed SNe with their nearest galaxy and obtain an estimation of the photometric redshift of the event. Next, we divide the SNe into a representative training set and a small test set. Afterwards, we train a snmachine classifier to obtain a benchmark performance for future studies on non-representative training sets. Our results indicate that the spectroscopically-confirmed training set is too small for the classifier to generalise and we require data augmentation even in this representative setting.
5.2 Introduction

The number of SNe discovered by current and upcoming astronomical surveys such as ZTF (Zwicky Transient Facility; Bellm et al., 2018, 2019; Graham et al., 2019a) and LSST (Rubin Observatory Legacy Survey of Space and Time; LSST Science Collaboration et al., 2009, 2017; Ivezić et al., 2019a) far surpasses the available spectroscopic resources for follow-up. Thus, as mentioned in Section 1.2.3, we must rely on photometric classification to identify the events. The smaller subset of spectroscopically-confirmed SNe will be used to train and evaluate the photometric classifiers.

As mentioned in Chapters 3 and 4, the SNPhotCC (Supernova Photometric Classification Challenge; Kessler et al., 2010a) and PLAsTiCC (Photometric LSST Astronomical Time-Series Classification Challenge; The PLAsTiCC team et al., 2018; Kessler et al., 2019) challenges promoted the development and understanding of the strengths and limitations of photometric classifiers. These challenges relied on realistic simulated datasets but we nevertheless expect real data to be different; our knowledge of SNe, atmosphere and instruments is incomplete. Thus, it is important to also test our approaches and pipelines with real data.

As described in Sections 1.3.2.2 and 1.3.3.2, ZTF has a larger field of view (47deg²) than LSST (9.6deg²). Moreover, ZTF has half the number of passbands and can only reach magnitudes of ~20.5 while LSST can reach ~24.7 magnitudes. Despite that, with ~10% of the LSST nightly alert rate, ZTF is the closest survey to LSST. Thus, it acts as a precursor for the latter survey (Graham et al., 2019a), and we can use the SNe observed by ZTF to test LSST pipelines. However, to assess the photometric classification performance we require events whose class is known. Thus, Sánchez-Sáez et al. (2021), Carrasco-Davis et al. (2021), and Pimentel et al. (2022) constructed labelled catalogs of ZTF data by cross-matching several catalogs (see Förster et al. (2021) for a description) and spectroscopically classified events from the Transient Name Server database¹. Alternatively, we can use the events spectroscopically-classified by the ZTF Bright Transient Survey (BTS; Fremling et al., 2020; Perley et al., 2020).

¹https://wis-tns.weizmann.ac.il/
The BTS is a large spectroscopic campaign with the aim of classifying all ZTF extragalactic transient events brighter than 18.5 magnitudes on either $g$ or $r$ passbands, except AGNs (active galactic nuclei) (Fremling et al., 2020). BTS will also target transients up to 19 magnitudes if there are available spectroscopic resources. Perley et al. (2020) showed that with some cuts to remove variables and poorly-observed transients, the BTS dataset is spectroscopically 97% complete at < 18 magnitudes, 93% complete at < 18.5 magnitudes, and 75% complete at < 19 magnitudes. Thus, BTS provides a complete and unbiased view of the transient optical sky at these magnitudes.

In this work we use the photometric transient classification library snmachine (Lochner et al., 2016; Alves et al., 2022) to build a classifier and assess its performance on BTS data. In this first study we use a small representative and balanced training set to obtain a benchmark performance for our approach. We leave a more realistic and unrepresentative scenario which requires dataset augmentation, such as the one in Chapters 3 and 4, for future work. First we use the NASA Extragalactic Database (NED) to obtain the photometric redshift and its uncertainty from the closest galaxy to each event. Next, following the approach of Chapter 3, we use snmachine to build a photometric classifier whose features are galaxy photometric redshifts and their uncertainties, as well as wavelet features obtained from Gaussian process (GP) fits. We then assess the classification performance to set a benchmark performance for our approach.

In Section 5.3 we describe the dataset used. The data processing and classification pipeline are described in Section 5.4. Our results and benchmark performance are presented in Section 5.5. We conclude in Section 5.6.

5.3 Dataset

In this work we used the ZTF light curves associated with the BTS spectroscopically-confirmed SNe events. The ZTF light curves came from the partnership survey mentioned in Section 1.3.2.2 and were observed the passbands $g$, $r$, and $i$. While the BTS dataset we used has 2726 SNe events, a similar number as two of the
training sets of the Chapter 3.3, the events here are subdivided into 35 classes instead of three. Since some classes have fewer than five events and there are too many classes to predict without data augmentation, we grouped the supernovae for classification into the following four classes:

- **SN Ia**: events BTS classified as SN Ia, SN Ia-norm, SN Ia-91T
- **SN Ibc**: events BTS classified as SN Ibc, SN Ib, SN Ic, SN Ibn, SN Icn, SN Ib-pec, SN Ic-BL, SN Ib/c?, SN Ibc, SN Ibn?, SN Ic?
- **SN II**: events BTS classified as SN II, SN IIb, SN IIn, SN IIp, SN II-pec, SN II?, SN IIb-pec, SN IIb?, SN IIn?
- **Other**: events BTS classified as SN Ia-pec, SN Ia-91bg, SN Iax, SN Ia-CSM, SLSN-I, SLSN-II, SLSN-I.5, SLSN-I?, SN Ia-03fg, SN Ia-91bg?, SN Ia?

The question mark in the class indicates uncertainty in the spectroscopic classification of the underlying event.

We used the BTS Sample Explorer\(^3\) \cite{Perley2020} to select the spectroscopically-confirmed SNe for our analysis. While this dataset contains the spectroscopic redshift of the events, this quantity will be missing from most of the ZTF and LSST events. Thus, we used NED to match each SNe to the nearest galaxy within a radius of 30\(\,\text{arcsec}\); we assigned the photometric redshift and uncertainty of the galaxy to the SNe. However, we only successfully matched 37% of the events with a galaxy; if we had access and used other galaxy databases such as the Set of Identifications, Measurements and Bibliography for Astronomical Data (SIMBAD; Wenger et al., 2000), we would expect a higher number of matches. Due to the low fraction of SNe with identifiable host galaxies, we applied a methodology similar to Section 3.5.4 to construct a photometric redshift for each of the other events: first we selected a random matched event and calculated the difference between its spectroscopic and photometric redshifts. The constructed redshift of the unmatched event is the sum of its spectroscopic redshift and this difference. We remove from our analysis the 17 unmatched SNe without spectroscopic redshift. We also remove 95 events with null flux in all passbands for the entire duration of the light curve.

\(^3\)https://sites.astro.caltech.edu/ztf/bts/explorer.php
Table 5.1: Breakdown of the number of SNe per class used in this work for the training and test sets. We show both the number of available training events for future work and the balanced subset we use in this work.

<table>
<thead>
<tr>
<th>SN class</th>
<th>$N_{\text{training}}$</th>
<th>$N_{\text{test}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Available (%)</td>
<td>Balanced (%)</td>
</tr>
<tr>
<td>SN Ia</td>
<td>1513 (72%)</td>
<td>125 (28%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>374 (72%)</td>
</tr>
<tr>
<td>SN Ibc</td>
<td>125 (6%)</td>
<td>125 (28%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>38 (7%)</td>
</tr>
<tr>
<td>SN II</td>
<td>385 (19%)</td>
<td>125 (28%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>93 (18%)</td>
</tr>
<tr>
<td>Other</td>
<td>68 (3%)</td>
<td>68 (16%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18 (3%)</td>
</tr>
<tr>
<td>Total</td>
<td>2091 (100%)</td>
<td>443 (100%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>518 (100%)</td>
</tr>
</tbody>
</table>

We divided the dataset into a training and test set while preserving the percentage of events from each subclass. Due to the small size of the original dataset, we only kept 20% of the events (523 SNe) for testing. Since class balance affects the performance of classifiers, in this work to benchmark the classification performance we subsampled the training set to ensure the same number of SN Ia, SN Ibc, and SN II. We ignored the class ‘Other’ in this subsampling because there are too few events in this class; we expect a poor performance for the class ‘Other’. We will use the remaining training set events in future work that involves dataset augmentation such as the one described in Sections 3.5 and 4.4.3. This future work will be evaluated in the same test set to compare the performance with the benchmark. Table 5.1 shows the resulting number of SNe per class.

5.4 Photometric Classification

We broadly followed the approach of Chapter 3 to classify the ZTF light curves. However, we left the augmentation procedure (described in Section 3.5) for future work where we will compare the resulting classification performance with the benchmark obtained in this work. The sections below describe the main steps of our approach and our upgrades to `snmachine` for use with ZTF and BTS data.

5.4.1 Light Curve Preprocessing

The ZTF SNe used in this work have low redshift with a median of $\sim 0.06$, hence the time dilation of their light curves is small. While the SNe are usually detected for a few months, most of the information used to distinguish the classes is near the
Figure 5.1: SN Ibc light curve, where the points show the observations. The errorbars are too small to be visible. This light curve illustrates the high ZTF cadence and that some of the events in the dataset lack observations in the $i$ passband.

peak. Since the ZTF light curves have high cadence, as exemplified in Figure 5.1, and low redshift, we selected our preprocessed light curve as the sequence of observations between 30 days before the peak and 50 days afterwards; we used the peak estimation provided by the BTS dataset. We also investigated wider light curve cuts but these resulted in a worse classification performance because the classifier became more tuned to SNe with long detection details; however, these SNe are non representative of the rest of the dataset as they are closer and rare.

5.4.2 Feature Extraction

This work aims to provide a benchmark of applying the methodology described in Sections 3.4 and 4.4 on real data. Thus, our feature extraction approach is based on wavelets and similar to the one described in Sections 3.4.3 and 4.4.4. We first used two-dimensional GPs (e.g. MacKay, 2003; Rasmussen and Williams, 2005) in time and wavelength to model the light curve of each SNe; we fitted them with a null mean function and a Matérn 3/2 kernel for the covariance. Then, we used the GP models to interpolate between the light curve observations and construct a uniform time grid. Following Chapter 3, we sampled the SNe once per day, resulting in 80 grid points. Afterwards, we decomposed the sampled light curve using Stationary Wavelet Transform and the symlet family of wavelets; this resulted in 960 redundant wavelet coefficients per event. Following Lochner et al. (2016) and Alves et al. (2022), we used PCA (Principal Component Analysis Pearson, 1901; Hotelling, 1933) to reduce the dimensionality of the wavelet space to 40 components, which
Table 5.2: Optimised hyperparameter values used for the LightGBM model.

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>boosting_type</td>
<td>gbdt</td>
</tr>
<tr>
<td>learning_rate</td>
<td>0.03</td>
</tr>
<tr>
<td>max_depth</td>
<td>7</td>
</tr>
<tr>
<td>min_child_samples</td>
<td>20</td>
</tr>
<tr>
<td>min_split_gain</td>
<td>1.1</td>
</tr>
<tr>
<td>n_estimators</td>
<td>115</td>
</tr>
<tr>
<td>num_leaves</td>
<td>10</td>
</tr>
</tbody>
</table>

we used for photometric classification.

As mentioned in Section 5.3, the ZTF SNe used in this work have spectroscopic redshift instead of the less precise photometric redshift. To increase the similarity of our classification setting with the one we will encounter for most of the ZTF and LSST events, we estimated the photometric redshift and its uncertainty as described in Section 5.3, and included them as classification features. Thus, our classifier had 42 features: two redshift features (photometric redshift and its uncertainty) and 40 wavelet features.

5.4.3 Classification

Similarly to Chapters 3 and 4, we used snmachine to build a GBDT (Gradient Boosting Decision Trees; Friedman, 2002) classifier with the 42 features described in the previous section. We optimised the classifier hyperparameters following the procedure described in Section 3.4.4; in this work we used the balanced training set summarised in Table 5.1. The Table 5.2 shows the optimised hyperparameter values.

5.4.4 Performance Evaluation

Following Sections 3.4.4.1 and 4.4.5.1, we used the PLAsTiCC weighted log-loss metric (The PLAsTiCC team et al., 2018; Malz et al., 2019) with the same weight for each class to optimise the photometric classifier and to evaluate its performance. We complemented this metric with a confusion matrix which summarises the performance of the photometric classifier across the multiple classes.
Figure 5.2 shows the confusion matrix for the classifier trained on the balanced training set summarised in Table 5.1. While most of the SNe Ia, SNe Ibc and SNe II are correctly classified, the events in the remaining class (Other) are poorly predicted. This result was expected as this class encompasses many different types of SN-like events, some of which are closer to SN Ia, SN Ibc, or SN II than to the other events in the same class. The classifier is then unable to generalise for that class. The classification performance for SN Ia, SN Ibc, and SN II is also worse than the obtained in Chapters 3 and 4. Despite that, in line with the previous studies, SN Ia has the best performance followed by SN Ibc and SN II with similar performances among them. We expect the classification performance and the log-loss metric to improve for a larger training set, such as the one used in Chapters 3 and 4. Since the training set is too small, we require data augmentation to improve the classification performance.

Since we have the spectroscopically-confirmed redshifts of the BTS events, we tested that using them as classification features leads to a better performance than using the estimated photometric redshift and uncertainty; we found an improvement of 1 – 10% in precision, recall and log loss. In this case, SN Ia was also the easiest
class to predict. However, as described in Section 5.3, the spectroscopic redshift will be missing from most of the ZTF and LSST events; this result validated that the higher accuracy of the spectroscopic redshift leads to higher performance.

5.6 Discussion and Conclusion

We have presented a benchmark for photometric classification of SN using real data, in particular ZTF SNe. To construct this benchmark we selected a balanced and representative training set from the ZTF SN-like events that were spectroscopically-confirmed with the BTS. Then, we built a classifier using the photometric transient classification library snmachine based on model-independent wavelet features and estimated photometric redshift and uncertainty. We also verified that using spectroscopic redshift leads to higher performance. In line with previous studies using LSST-like datasets, we find that SN Ia is the easiest class to identify.

Our benchmark classifier yielded a significantly worse performance than we found in Chapters 3 and 4 for LSST-like simulated dataset, despite the higher cadence of observations and the representative setting. However, the ZTF training dataset was much smaller and we did not perform any dataset augmentation, a step we found to be crucial on the previous analyses. Thus, in future work, we will compare this benchmark performance with a classifier trained on the entire available training set where we will use augmentation to balance the SNe classes; we expect the new classifier to yield a higher performance. Afterwards, we will select brighter events in the training set, augment them, and train a new classifier. Based on Chapters 3 and 4, we expect that by augmenting the non-representative training set by a factor of 11.6, the new classifier will achieve a competitive performance compared with the classifiers from those chapters. These future step-wise analyses and comparisons will aid in identifying and understanding any difference between a real-data analysis and the simulation-based analysis we did in Chapters 3 and 4.

The BTS dataset used in this work has 2726 spectroscopically-classified SNe. We expect to have more for the upcoming LSST because, as mentioned in Section 1.3.4, TiDES (Time-Domain Extragalactic Survey; Swann et al., 2019) will
spectroscopically classify $\sim 30000$ LSST transients. However, despite the larger number, these events will be magnitude-limited. Thus, as shown in Chapters 3 and 4, we still require augmentation to make a representative training set. Testing both our classification and augmentation methodologies in ZTF will then improve our understanding of the relative importance of representativeness and augmentation for our classification approach on real SNe.

If further time and effort are available, we could use the methodology described in this chapter to classify the ZTF SNe that were not spectroscopically confirmed. We would use all the identified SNe as our training set. Then, similarly to Chapters 3 and 4, we would augment the training set to be representative of the events we want to classify. After training the classifier, we would use it to classify the events. We expect this classifier to have a similar performance to the one trained with the brighter events of the spectroscopically-confirmed BTS dataset, which we described above as future work. Thus, we could extrapolate our conclusions and the cosmological impact of contamination from the BTS dataset to all ZTF SNe.

With this chapter, we provide a benchmark for ZTF light curves, which is the first step for using snmachine classifiers to identify ZTF SNe. Our results highlight the need for data augmentation in representative settings if the training set is small.
6

Inference of lensed supernovae parameters

This chapter is part of an ongoing work in collaboration with Dr Brian Nord. It is based upon work that was supported by the Visiting Scholars Award Program of the Universities Research Association. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Universities Research Association, Inc.

6.1 Overview

The recent DES (Dark Energy Survey) and the upcoming LSST (Vera C. Rubin Observatory Legacy Survey of Space and Time) aim to better understand dark energy. An important parameter for this understanding is the Hubble constant $H_0$. Currently, there is a strong disagreement on its value, but gravitationally LSNe (lensed supernovae) can provide an independent measurement. Traditionally, this requires modelling of the lens system to measure LSNe properties. However, due to the data deluge LSST will bring, it is crucial to construct a faster method. Here we present a machine learning pipeline to predict LSNe parameters from simulated time-series images for DES; we modify an existing multi-branch deep neural network that combines LSTM (long short-term memory) layers with convolutional layers for this use. In this proof-of-concept demonstration, we predict the peak night of SN Ia (type Ia SN) and the ellipticity of the associated lens. We find that while the neural network can predict the SN Ia peak time with a root mean square error of 9.7 days, it is unable to accurately estimate the ellipticity parameters of the lens. This work demonstrates that further work is needed to predict LSNe parameters with this neu-
6.2 Introduction

Since the discovery of the expansion of the Universe, astronomers have tried to determine its expansion rate. In particular, the present cosmic expansion rate, Hubble constant $H_0$, is critical for understanding the evolution of the Universe (e.g. Hubble, 1929; Sandage, 1958). There are multiple ways to measure $H_0$, such as the Cepheid-supernova distance ladder, extrapolation from the cosmic microwave background anisotropies, and strongly lensed variable sources, like SNe and quasars. However, in recent years, measurements from early-time probes (Abbott et al., 2018a; Addison et al., 2018; Philcox et al., 2020; Planck Collaboration et al., 2020) have been in tension with the ones from late-time probes (Birrer et al., 2019; Wong et al., 2019; Riess et al., 2021) (see Bernal et al. (2016), Verde et al. (2019), Bernal et al. (2021), and Shah et al. (2021) for a summary). This disagreement strongly motivates the need for new, independent, probes of $H_0$. Time-delay cosmography is promising in this regard since it performs a geometric measurement of $H_0$ (Refsdal, 1964).

Time-delay cosmography (TDC) is a technique that measures the time delays between multiple images of a distant source that have undergone strong gravitational lensing. The time delays depend on a number of aspects such as the matter distribution in the lens, the overall matter distribution along the line-of-sight, and the cosmological parameters; TDC thus requires the modelling of the strong lensing system in addition to the time delay measurements. This approach has been used with gravitationally lensed time-varying quasars to measure $H_0$ to a precision greater than five per cent (Wong et al., 2019; Shajib et al., 2020).

In addition to quasars, SNe are important time-varying sources that are sufficiently bright to be detected at cosmological distances. Since SNe only last months, they require fast identification and analysis to allow us to obtain time delay measurements. This faster timeframe compared with quasars allows SNe to be competitive in measuring time delays. Another advantage of LSNe is that a common subtype of SNe, SNe Ia, has a standardisable brightness (Tripp, 1998); as discussed in
Section 1.2.2. The standardisability can improve the modelling of the lensing gravitational potential, and therefore yield a more precise \( H_0 \) measurement (Oguri and Kawano, 2003; Kolatt and Bartelmann, 1998; Foxley-Marrable et al., 2018; Birrer et al., 2022). While only a few LSNe have been detected (Kelly et al., 2015; Goobar et al., 2017; Rodney et al., 2021b), we expect the number to increase by more than two orders of magnitude in the coming years with recent and upcoming surveys such as DES (Gruendl et al., 2018), ZTF (Zwicky Transient Facility; Bellm et al., 2018, 2019; Graham et al., 2019a), Hyper Suprime-Cam (Aihara et al., 2017), and LSST (Ivezic et al., 2019b). To measure the properties of LSNe, it is common to model the lens system using the SNe location in each image (Kelly et al., 2015; Goobar et al., 2017; Rodney et al., 2021b) and software such as *glafic* (Oguri, 2010, 2015) and *LENSTOOL* (Jullo et al., 2007; Kneib et al., 2011). However, we need new, faster, approaches to deal with the increased number of LSNe that future surveys will discover; Wojtak et al. (2019) estimates that LSST will detect 89 type Ia LSNe and 254 CC LSNe per year. Deep neural networks can bypass the traditional lens modelling that relies on spectroscopic follow up and directly infer parameters. Moreover, deep neural networks can estimate uncertainties and are faster. Thus we can apply them as the observations arrive instead of waiting days to obtain these results from the traditional approach, and possibly missing the optimal follow up moment for additional analysis or studies.

Recently, Morgan et al. (2022) constructed a deep neural network classifier to identify LSNe in optical survey datasets. The neural network, ZipperNet, is innovative in its combined use of convolution layers for images and LSTM (long short-term memory) layers for light curves; these two layers are combined into one coherent deep learning architecture, so the training of the network optimises weights in both types of layers simultaneously. The architecture is illustrated in Figure 6.1. The success of ZipperNet motivates the use of a similar network architecture for regression of LSNe parameters.

In this work we demonstrate the predictive ability of a modified ZipperNet for LSNe parameters, hereby referred as ‘regression ZipperNet’. First, we simulated DES images of Type Ia LSNe with the software package *deeplenstronomy* (Morgan et al., 2021). Next, we processed the data to construct the input to the neural
network. Afterwards, we trained the regression ZipperNet to predict the peak night of SN Ia and the ellipticity of the associated lens and assessed its performance.

Section 6.3 describes the simulations used for training and testing the regression ZipperNet, the data processing methodology, and the architecture of the network. In Section 6.5, we present the regression results. We conclude in Section 6.6.

6.3 Simulation of lensed supernovae

In this work we simulated LSNe observed by the DECam (Dark Energy Camera; Flaugher et al., 2015) as astronomical systems with one galaxy with one SN Ia behind a foreground galaxy. To simulate the events we followed one approach of Morgan et al. (2022) by using a three-day cadence and properties similar to the Dark Energy Spectroscopic Instrument DECam Observation of Transients (DESI-DOT) program (Palmese & Wang; DECam Proposal 2021A-0148).

We simulated the events in the $griz$ passbands using the open-source software package, deeplenstronomy (Morgan et al., 2021), which is built around the commonly used package lenstronomy (Birrer and Amara, 2018; Birrer et al., 2021); lenstronomy performs gravitational lensing calculations, modelling, and simulations. deeplenstronomy provides additional features such as image and SN in-
jection, probability distribution sampling, and survey observational characteristics. Following Morgan et al. (2022), we simulated the DESI-DOT dataset using the real observing conditions (seeing and sky brightness) from the DES wide-field survey (Abbott et al., 2018a) and an exposure time of 60 seconds. For each event we obtained 15 images of 45 by 45 pixels in each passband, covering 42 days; this results in a total of 60 images per event. To generate our dataset, we adapted the deeplenstronomy input files from Morgan et al. (2022), which are provided in Morgan (2021)\(^1\); we simulated 624 LSNe. Following Morgan et al. (2022), we did not simulate microlensing and we leave this improvement for further work. See Morgan et al. (2022) for more details on the simulations.

We divided the dataset such that 90\% of the events are in the training set and the remainder in the test set, respectively 561 and 63 events. While realistically the test set would be larger than the training set, in this work we want to present a proof-of-concept. We thus leave the more realistic setting for further work.

### 6.4 Regression pipeline

#### 6.4.1 Data Processing

Over the 42 days of our observations, the main differences in the images are the presence of one or more point sources, and small effects due to the observing conditions. Following Morgan et al. (2022), we condensed the image information of each SNe into a single-image per passband; we averaged all images of each passband on a pixel-by-pixel basis. Then, we scaled the pixels values to range from 0 to 1 on a per-event basis to preserve colour relationships. The resulting images have a higher overall signal-to-noise ratio and the faint objects are more visible. Additionally, we constructed light curves from the time-series of images. Following Morgan et al. (2022), we measured a background-subtracted brightness from each image with a predefined circular aperture of 20-pixel radius (corresponds to 5.26 arcseconds for DECam) at the centre of each image. Similarly to the condensed images, we scaled the extracted brightness to range from 0 to 1 on a per-event basis.

\(^1\) We simulated only the configuration 3 from the input file high_cad_data.yaml.
Figure 6.2: Example of an image (top left) and its augmentations. Each column corresponds to a rotation of 0°, 90°, 180°, and 270°, respectively from the left to the right. The second row corresponds to a horizontal rotation of the first row.

The inputs for the regression DeepZipper are the 45 by 45 pixels condensed image in each passband (four images in total) and the constructed light curve with 15 observations in each passband. The outputs are the peak time of the SNe and two parameters of the lens mass ellipticity,

$$e_1 = \frac{1 - q_{\text{lens}}}{1 + q_{\text{lens}}} \cos(2\phi_{\text{lens}})$$  \hspace{1cm} (6.1)

and

$$e_2 = \frac{1 - q_{\text{lens}}}{1 + q_{\text{lens}}} \sin(2\phi_{\text{lens}}),$$  \hspace{1cm} (6.2)

where $q_{\text{lens}}$ is the ratio of the projected axis of the lens and $\phi_{\text{lens}}$ is the lens orientation angle. We obtained these parameters (among others such as the velocity dispersion of the lens) directly from the deeplenstronomy (Morgan et al., 2021) simulation. We scale the regression parameters individually over the entire dataset so they range from 0 to 1.

6.4.2 Data Augmentation

We applied the same methodology as Morgan et al. (2022) to augment the dataset eight-fold; we mirrored the images and rotated them three times. Figure 6.2 shows
an example of the augmentation. The light curves for each set of eight events is the same since the circular aperture extraction method (mentioned in Section 6.4.1 and detailed in Morgan et al., 2022) is invariant to rotations and reflections. The meta-data associated with the events (e.g. peak time of the SNe, lens mass ellipticity) is also the same. Since training and testing on the same events leads to overfitting, we rotate and mirror the images in the training and test set independently.

### 6.4.3 Regression ZipperNet

Following the original ZipperNet Morgan et al. (2022), our modified version for regression treats similarly the image-based information and the light curve-based information. The architecture is shown in Figure 6.3 with the specifications of each layer in Table 6.1.

Figure 6.3 shows that the images pass through convolutional layers, are flattened into one dimensional arrays, and are condensed in size. The light curves pass through recurrent layers with LSTM cells, are flattened into one dimensional arrays, and condensed. Then, the flatten output of each branch is concatenated
Table 6.1: Layer specifications of the regression ZipperNet whose architecture is shown in Figure 6.3. The following parameters are shorthanded: kernel size \((k)\), padding \((p)\), stride \((s)\), value added to the denominator of the batch normalisation for numerical stability \((\text{eps})\), momentum \((m)\), and hidden units \((h)\). See the library documentation for a description of these hyperparameters and layers. Changes in size of the data representation passing through the layers are indicated with arrows. We indicate the Rectified Linear Unit (ReLU) activation functions with superscript \(^\dagger\) in the name of the layers.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>conv</td>
<td>Conv2D — ((k: 3, p: 2, s: 3)) — ((4,45,45 \rightarrow 48,16,16))</td>
</tr>
<tr>
<td>batchnorm(^\dagger)</td>
<td>BatchNorm2D — ((\text{eps}: 10^{-5}, m: 0.1))</td>
</tr>
<tr>
<td>maxpool</td>
<td>MaxPool2D — ((k: 2)) — ((48,16,16 \rightarrow 48,8,8))</td>
</tr>
<tr>
<td>flatten</td>
<td>Flatten — ((48,8,8 \rightarrow 3072))</td>
</tr>
<tr>
<td>fc1(^\dagger)</td>
<td>Fully-Connected — ((3072 \rightarrow 408))</td>
</tr>
<tr>
<td>fc2</td>
<td>Fully-Connected — ((408 \rightarrow 25))</td>
</tr>
<tr>
<td>lstm</td>
<td>LSTM — ((h: 32)) — ((15,4 \rightarrow 32))</td>
</tr>
<tr>
<td>fc3</td>
<td>Fully-Connected — ((32 \rightarrow 25))</td>
</tr>
<tr>
<td>fc4(^\dagger)</td>
<td>Fully-Connected — ((50 \rightarrow 8))</td>
</tr>
<tr>
<td>fc5</td>
<td>Fully-Connected — ((8 \rightarrow 3))</td>
</tr>
</tbody>
</table>

and mapped to the three output features we want to predict: two parameters for ellipticity and the SNe peak time. This last layer can be changed to accommodate different number of parameters.

Since the output of the image and light curve branches is concatenated, the weights on both layers were optimised simultaneously during the training of the network. For the loss function we used the mean square error of the output, where all the parameters have the same weight. Following Morgan et al. (2022), we used Adam (Kingma and Ba, 2014) as the optimiser algorithm\(^2\) with a learning rate of 0.001, \(\beta_1 = 0.9\), \(\beta_2 = 0.999\), and \(\varepsilon = 10^{-8}\). However, we found that a batch size of 32 for training performed better than the batch size of 5 used by Morgan et al. (2022); the gradient flow approached zero for the latter batch size, which prevented the proper training of the network. Figure 6.4 shows that the regression ZipperNet achieved the best test set performance for 39 epochs (i.e. passes over the training data). Further iterations decrease the test set performance as the network starts to overfit to the training data. Thus, we choose the fully trained network as the point during training with the lowest testing mean square error (39 epochs). We

\(^2\) See the documentation for a brief description of the hyperparameters.
Figure 6.4: Learning curve of the regression DeepZipper. The lowest overall mean square error (MSE) of the test set is at epoch 39. We use the neural network at this point as our fully trained regression DeepZipper.

implemented the regression DeepZipper with the package PyTorch (Paszke et al., 2019).

6.5 Results

We find that the current version of the regression ZipperNet is unable to accurately estimate all the parameters used. Figure 6.5 and Table 6.2 show that while the network has some success in estimating the peak time of the SNe, it fails at predicting the lens mass ellipticity parameters.

We now focus on the parameter with the worse performance, $e_2$. We first modify the last layer of the network (fc5) to return a single output. Next, we train the modified regression ZipperNet as described in Section 6.4.3. The right panel of Figure 6.6 shows that the neural network is unable to learn this parameter as the test set performance remains constant regardless of the number of training epochs. However, the gradient flow is non-null (see left panel of Figure 6.6), which means the neural network weights are updated in each iteration. These results indicate that the current implementation of the regression ZipperNet is unable to predict the lens ellipticity parameter $e_2$ and that this is not related with the optimisation procedure. We find similar results for $e_1$. 
Figure 6.5: Regression ZipperNet predictions on the test set for lens mass ellipticity ($e_1$, $e_2$) and peak time of the SNe ($t_{peak}$) compared to their truth values. The grey line indicates a perfect prediction. While the regression ZipperNet has some success in predicting $t_{peak}$, it is unable to estimate the lens mass ellipticity parameters.

Table 6.2: Mean square error of the test set for each individual parameter and overall. The mean square error is presented for the scaled values of regression parameters and for their true values. $t_{peak}$ is the peak time of the SNe and $e_1$, $e_2$ are the two lens mass ellipticity parameters of Equations 6.1 and 6.2.

<table>
<thead>
<tr>
<th></th>
<th>$t_{peak}$</th>
<th>$e_1$</th>
<th>$e_2$</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scaled values</td>
<td>0.164</td>
<td>0.895</td>
<td>1.050</td>
<td>0.703</td>
</tr>
<tr>
<td>True values</td>
<td>94.135</td>
<td>0.014</td>
<td>0.017</td>
<td>31.389</td>
</tr>
</tbody>
</table>
Figure 6.6: The right panel shows the learning curve of the modified regression DeepZipper that only predicts the lens ellipticity parameter $e_2$. The left panel shows the gradient flow of the neural network; each line represents the log average of the absolute value of the gradient of a training epoch per layer. If these lines were zero, the gradient would be null and the neural network would be unable to update its weights.

6.6 Discussion and Conclusion

We have presented an implementation of a neural network similar to ZipperNet Morgan et al. (2022) can be used to predict LSNe parameters. We used the package deepenstronomy (Morgan et al., 2021) to simulate LSNe images with properties similar to DESI-DOT. Next, we changed ZipperNet from classification to regression, and trained it in the generated dataset. We found that with minor modifications, our regression ZipperNet can predict the scaled SNe peak time with a mean square error of 0.164; this corresponds to a standard error of 9.7 days in the true SNe peak time. While this error is large, the improvement in performance shown in Figure 6.4 indicates that this type of neural network has the ability to learn this parameter. However, the regression ZipperNet is unable to estimate the ellipticity parameters of the lens. This proof-of-concept shows that we need additional changes before using this type of neural network architecture for estimating LSNe parameters, and eventually $H_0$.

To improve the SN Ia peak prediction, we need to further optimise the regres-
sion ZipperNet architecture. This involves more experimentation with the architecture and the hyperparameters. For the ellipticity parameters, other approach is necessary. Since Figure 6.6 shows that the neural network training is working, the problem might be in the dataset. Thus, we need to verify the assumptions of the our deeplenstronomy simulations (Section 6.3) and augmentation (Section 6.4.2). While this approach worked for classifying LSNe (Morgan et al., 2022), it might not be valid or require further changes for predicting the LSNe parameters.

In this work, we used a dataset with properties like DESI-DOT. However, the original ZipperNet Morgan et al. (2022) was also successful on classifying LSST simulations. Thus, we will apply the regression ZipperNet to LSST data in future work. At this point we will also include microlensing in the simulations; while Foxley-Marrable et al. (2018) expects its effects to be small, Goldstein et al. (2018) found that microlensing can lead to systematic errors of $\sim 4\%$ in time delays for LSST.

If further time and effort are available, we would extend the network output to obtain other lensed SNe parameters, time-delay, and $H_0$ measurements; the current codebase$^3$ allows an easy implementation of this extension and any further change to the network. Afterwards, we would compare our $H_0$ results with Arendse et al. (in prep) who successfully implemented a neural network with a different architecture and inputs to infer $H_0$ from LSST LSNe. Our approach is complementary to Arendse et al. (in prep) because the different methodologies and simulations allow us to independently validate the results. We would also study the prediction differences to understand what drives them and use that information to further improve both methods. Once our approach is validated and improved from the previous step, we would incorporate it in the strong lensing and time-domain analysis pipelines in LSST.

We release the regression ZipperNet code and Jupyter notebooks which can be used to reproduce this work$^4$.

---

$^3$The regression ZipperNet code is in https://github.com/Catarina-Alves/DeepZipper. This repository also includes demo Jupyter notebooks to perform the analysis and change the regression parameters.

$^4$See footnote 3.
7.1 Summary

In this PhD thesis we focused on the preparation for the upcoming LSST. We developed and implemented an approach for photometric classification of SNe. At the same time, we measured the impact of the LSST observing strategy survey design on classification performance using simulated LSST-like SNe. Then, we set a benchmark performance for our approach on real ZTF data, which is a precursor to LSST. Additionally, we performed a proof-of-concept demonstration of the ability of multi-branch deep neural networks to measure LSNe parameters from photometric surveys.

We extended an existing photometric classification pipeline and associated software for use with LSST (Chapter 3). Crucially we implemented a methodology to increase the number and representativeness of training set events. This is critical for LSST as the spectroscopically-confirmed SNe available for the training of photometric classifiers will be brighter than most of the observed SNe. Next, we used our approach to study the classification performance for simulated SNe with different properties. We found that the survey season length and cadence significantly impact the performance, with longer events and shorter median inter-night gap $< 3.5$ days yielding a more accurate classification.

In Chapter 4 we explored in more detail the impact of LSST cadence choices on photometric classification of SNe. We simulated LSST-like SNe for the current baseline observing strategy, a version without rolling cadence, and a version with
an additional visit per night. We confirmed that the light curve length, median inter-night gap and number of observations near the SNe peak, which differ between the observing strategies, affect the photometric classification. While a moderate rolling cadence mildly improves the overall performance, the region of the sky observed at a higher cadence has a significantly higher classification performance and up to a factor of 2.7 more cosmologically-useful SN Ia than the rest of the sky; these results motivate the implementation of a rolling cadence within the baseline observing strategy. However, the additional visit per night leads to more irregularly sampled light curves which degrades the classification performance. The results of Chapters 3 and 4 contribute towards the optimisation of the LSST observing strategy and are the first works to explore impact of the observing strategy on photometric classification of SNe.

In Chapter 5 we applied the previously developed methodology to real SNe events from ZTF. We set an initial benchmark performance for photometric classification in a representative setting without training set augmentation. However, our results indicate that even in this representative setting, augmentation is likely needed for a high classification performance. With this work we started an important validation of our approach with real SNe.

The work presented in Chapter 6 expands our preparation for LSST to LSNe. We modified an existing multi-branch deep neural network with previous success in LSNe classification to predict LSNe parameters. Our proof-of-concept showed that the considered neural network architecture is challenging to adapt to a regression problem. Thus, additional changes are needed before using for LSST and comparing with different approaches currently under development. This work contributes to developing a faster alternative to the traditional methodology to measure LSNe parameters which requires modelling of the lens system; such approach is too time-consuming to use for all LSNe from LSST.

7.2 Outlook

While the initial survey strategy for LSST will be decided in the coming months, further optimisation will be performed throughout the survey. Thus, we require a
set of metrics to measure the impact of observing strategy on LSST science that can be calculated frequently. Since there are numerous possibilities for any change in the observing strategy, it is crucial for these metrics to be fast and have a low computational cost. DESC (LSST Dark Energy Science Collaboration) has an ongoing effort to develop and improve these metrics for cosmology (e.g. Lochner et al., 2022). As mentioned in Chapter 4, one possible extension of our work is to develop a photometric classification fast metric that could be incorporated in the DESC metrics. This fast metric would work directly on the observing strategy cadences produced by the LSST official scheduler (Naghib et al., 2019), avoiding the time-consuming simulations and classification of Chapters 3 and 4. Despite that, the results of Chapter 4 could aid in the development of the metric and to ensure it is indeed a proxy for photometric classification performance.

Another extension of Chapters 3 and 4 is to study the impact of photometric classification on other transient classes besides SNe. To do this, we could update the photometric transient classification library snmachine. Since other transients have different light curve shapes, the update would require modifications to the GP modelling step, such as the possibility for events to use different kernels. We would also verify if the augmentation step works for the new classes and perform any required adjustments.

As mentioned in Chapter 5, it is crucial to validate our photometric classification approach on real data to improve it. Thus, on that chapter we set an initial benchmark performance to later compare for future studies on non-representative training sets. First, our augmentation procedure should be implemented on the representative setting. This should improve the classification performance by increasing the available dataset to train the classifier on. In further future work one would select the brightest events to construct a non-representative training set, augment it and train another photometric classifier. Comparing this last performance with the previous two would provide a precise understanding of the relative importance of representativeness and augmentation for our classification approach on real SNe.

In Chapter 6 we presented a proof-of-concept for a neural network to predict LSNe parameters. This neural network can be further optimised by modifying the architecture. A straightforward future extension could be predicting more SNe pa-
rameters, in particular predicting $H_0$ directly. Next the approach could be applied to LSST simulated LSNe. At this point one could compare the performance of our method with Arendse et al. (in prep), which is currently developing a neural network with a different architecture for the same purpose. Since these methods are different they can be used to validate each other results and hedge the weaknesses before applying the to real LSST LSNe.

The work presented in this thesis focused on enabling SNe cosmology with large time-domain surveys, in particular preparing for the upcoming LSST. This survey will revolutionise Astronomy with the huge number of new and fainter events. Thus, we must prepare now so we can obtain the most cosmology out of it. Moreover, our work will also aid in the continuous optimisation of LSST observing strategy and it can be applied to other surveys.
Appendix to Chapter 3

A.1 Redshifting Implementation for Augmentation

In this appendix we provide further details of the augmentation procedure described in Section 3.5. In particular, we present the technique for redshifting a light curve, and derive the redshift limits for augmentation shown in Equation (3.4).

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

$$
f_{\lambda;\text{aug}}(\lambda) = \frac{1 + z_{\text{ori}}}{1 + z_{\text{aug}}} \left( \frac{d_L(z_{\text{ori}})}{d_L(z_{\text{aug}})} \right)^2 f_{\lambda} \left( \frac{1 + z_{\text{ori}}}{1 + z_{\text{aug}}} \lambda \right), \quad (A.1)
$$

where $\lambda$ is the observed wavelength, $d_L$ is the luminosity distance, and $f_{\lambda}$ 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 3.4.2 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

$$
F_{\text{ori}b} = \mathcal{GP} \left( \frac{1 + z_{\text{ori}}}{1 + z_{\text{aug}}} \lambda_b \right), \quad (A.2)
$$

where $\mathcal{GP}$ represents the mean of the GP fit used to model the flux observations of the original SNe, and $\lambda_b$ is the central wavelength of passband $b$. We calculated
Figure A.1: Distribution of the GP errors resulting from extrapolating GP fits to $u$ or $y$ passbands. $F_{\text{true}}$ is the true flux of an observation, and $F_{\text{predicted}}$ and $\sigma_{\text{predicted}}$ are the flux and its uncertainty predicted by a GP fit at the corresponding epoch and passband. An ideal error estimation results in a unit Gaussian (black line).

these central wavelengths using the LSST throughputs\(^1\). 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 $z_{\text{aug}}$. We assumed a flat $\Lambda$CDM cosmology with $H_0 = 70$ km/s/Mpc and $\Omega_m = 0.3$. We estimated the flux of the synthetic event in each passband $b$ as

$$F_{\text{aug}b} = \frac{1 + z_{\text{ori}}}{1 + z_{\text{aug}}} \left[ \frac{d_L(z_{\text{ori}})}{d_L(z_{\text{aug}})} \right]^2 F_{\text{ori}b}, \quad (A.3)$$

and estimated its uncertainty similarly.

As previously discussed in Section 3.5.2, the GP fit is more reliable close to 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 A.1 shows 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 ($\lambda_y - \lambda_u$) and maximum ($\lambda_y - \lambda_u$) wavelength ranges. Thus, when generating a synthetic SN at higher redshifts, we have that $\lambda_u - (1 + z_{\text{ori}})/(1 + z_{\text{aug}}) \lambda_u \leq \lambda_y - \lambda_u$. Similarly, for events generated at lower redshifts, we obtain the redshift limits for

\(^1\)https://github.com/lsst/throughputs
augmentation presented in Section 3.5.2:

\[
\begin{align*}
    z_{\text{min}} &= \max \left\{ 0, (1 + z_{\text{ori}}) \left( 2 - \frac{\lambda_z}{\lambda_y} \right)^{-1} - 1 \right\} \\
    &\approx \max \left\{ 0, 0.90 z_{\text{ori}} - 0.10 \right\}, \\
    z_{\text{max}} &= (1 + z_{\text{ori}}) \left( 2 - \frac{\lambda_z}{\lambda_y} \right)^{-1} - 1 \\
    &\approx 1.43 z_{\text{ori}} + 0.43.
\end{align*}
\]  

(A.4)

### A.2 Comparison with other PLAsTiCC classifiers

Section 3.6 presented the results of our classifier on the impact of observing strategy on photometric classification. In this appendix, we show that our results are generalisable beyond our classification pipeline, by replacing the our classification predictions with those obtained by Boone (2019). We use the publicly available predictions for SN Ia, SN Ibc, and SN II in the test set\(^2\); we choose the predictions obtained with a classifier optimized on the log-loss metric, which equally weights all the PLAsTiCC classes \(w_j = 1\) in Equation 3.1. The choice of this flat-weighted metric reduces the impact of additional classes upweighted in the original challenge, but unused in the present work. Figures A.2 and A.3 show that the classifier used in Boone (2019) has the same performance behaviour as ours. This further indicates that our conclusions are general and not an artifact of our classification architecture.

---

\(^2\) [http://supernova.lbl.gov/avocado_plasticc/predictions/predictions_plasticc_test_flat_weight.csv](http://supernova.lbl.gov/avocado_plasticc/predictions/predictions_plasticc_test_flat_weight.csv)
Figure A.2: 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 SNe class. In the first and third rows we reproduce the results of this work previously shown in Figures 3.7, 3.8, and 3.9. In the second and fourth rows we show the classification predictions obtained by Boone (2019).
Figure A.3: 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 SNe class. In the first and third rows we reproduce the results of this work previously shown in Figures 3.9 and 3.10. In the second and fourth rows we show the classification predictions obtained by Boone (2019).
B.1 Simulated CC SN rates

In this appendix we present the absolute and relative rates used to simulate CC SNe in Chapter 4. These rates follow Shivvers et al. (2017) with the adjustments described in Section 4.3.4. Table B.1 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 4.1.

B.2 Augmentation details and Classification Hyperparameters

Section 4.4.3 described the differences between the augmentation procedure used in this work and the one in the 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 4.3.3) altered the redshift distribution of the events. In this work, we augmented each event between the redshift limits $z_{\text{min}}$ and $z_{\text{max}}$ from Equation 3.4. However, we used a different class-agnostic target distribution. In particular, we drew an auxiliary value $\zeta^*$ from a log-trapezoidal distribution; the probability density function of the trapezoid distribution is
Table B.1: Absolute and relative scale rate used to simulate core-collapse SNe in this work expressed in percentage. 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.

<table>
<thead>
<tr>
<th>Subtypes</th>
<th>Model name</th>
<th>Abs. scale rate (rel. scale rate) %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core-collapse</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>see Hydrogen Rich</td>
<td>69.6 (100)</td>
</tr>
<tr>
<td>Ib+Ic</td>
<td>see Stripped Envelope</td>
<td>30.4 (100)</td>
</tr>
<tr>
<td>Hydrogen Rich - class SN II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II (IIP, IIL)</td>
<td>SNII-NMF</td>
<td>32.45 (23.325)</td>
</tr>
<tr>
<td></td>
<td>SNII-Templates</td>
<td>32.45 (23.325)</td>
</tr>
<tr>
<td>IIn</td>
<td>SNII-n-MOSFiT</td>
<td>4.7 (3.35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.7 (6.7)</td>
</tr>
<tr>
<td>Stripped Envelope - class SN Ibc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ib</td>
<td>SNIb-Templates</td>
<td>5.4 (17.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.8 (35.6)</td>
</tr>
<tr>
<td>Ic</td>
<td>SNIc-Templates</td>
<td>3.75 (12.35)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.5 (24.7)</td>
</tr>
</tbody>
</table>

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

(B.1)

where \( x_{\text{min}} = \log(z_{\text{min}}) \), \( x_{\text{max}} = \log(z_{\text{max}}) \), and \( \Delta x = x_{\text{max}} - x_{\text{min}} \). Then, we calculated the redshift of the new augmented event following Equation 3.5.

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 B.2) and the flux uncertainty distribution of each passband (Table B.3). 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 B.1.

In this section we also show the hyperparameters values of the GBDT classifier used for each observing strategy (Table B.4).
Figure B.1: Host galaxy photometric redshift distribution per supernova class, where SN Ia, SN Ibc, 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 B.2: Parameters of Gaussian mixture models used to fit the number of observations of the test set light curves for each observing strategy. These values were later used to create an augmented training set (Section 4.4.3). 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>
<thead>
<tr>
<th></th>
<th>Y1.5-3 baseline</th>
<th>No-roll</th>
</tr>
</thead>
<tbody>
<tr>
<td>weights</td>
<td>[0.322, 0.678]</td>
<td>[0.289, 0.215, 0.181, 0.315]</td>
</tr>
<tr>
<td>means</td>
<td>[16.86, 61.15]</td>
<td>[31.93, 18.07, 59.65, 45.35]</td>
</tr>
<tr>
<td>variances</td>
<td>[46.62, 561.06]</td>
<td>[30.23, 42.76, 75.83, 34.96]</td>
</tr>
</tbody>
</table>

B.3 Computational Resources

We simulated the observing strategy datasets on a Intel E5-2680v4 @ 2.4 GHz. Each dataset with $2.5 \times 10^6$ events takes $\sim 200$ core hours to simulate. The data processing, classification and analysis was performed on an Intel(R) Xeon(R) CPU E5-2697 v2 (2.70GHz). Using a single core, the pipeline takes $\sim 5.6$ hrs to preprocess these events. Modelling them with GPs and performing their wavelet decomposition takes $\sim 44.4$ hrs. Generating a augmented training set with 15440 events takes $\sim 9$ hrs, and reducing the dimensionality of their wavelet features using PCA takes $\sim 45$ min. Optimising the GBDT classifier takes $\sim 5.5$ hrs. Obtaining the test set predictions on the pre-computed test set features with the trained classifier takes 5 min. Overall, the entire classification pipeline takes $\sim 200 + 70$ core hours of computing time for each observing strategy.
Table B.3: Parameters of Gaussian mixture models used to fit the flux uncertainty distribution of the test set in each passband (ugrizy) and observing strategy. These values were later used to create an augmented training set (Section 4.4.3). 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>
<thead>
<tr>
<th></th>
<th>Y1.5-3 baseline</th>
<th>No-roll</th>
<th>Y0-3 baseline</th>
<th>Presto-color</th>
</tr>
</thead>
<tbody>
<tr>
<td>weights</td>
<td>[0.24, 0.76]</td>
<td>[0.25, 0.75]</td>
<td>[0.29, 0.71]</td>
<td>[0.64, 0.36]</td>
</tr>
<tr>
<td>u means</td>
<td>[2.34, 1.92]</td>
<td>[2.29, 1.95]</td>
<td>[2.24, 1.93]</td>
<td>[1.96, 2.35]</td>
</tr>
<tr>
<td>variances</td>
<td>[0.46, 0.10]</td>
<td>[0.45, 0.10]</td>
<td>[0.38, 0.09]</td>
<td>[0.09, 0.28]</td>
</tr>
<tr>
<td>weights</td>
<td>[0.88, 0.12]</td>
<td>[0.11, 0.89]</td>
<td>[0.88, 0.12]</td>
<td>[0.10, 0.90]</td>
</tr>
<tr>
<td>g means</td>
<td>[1.26, 2.06]</td>
<td>[2.13, 1.27]</td>
<td>[1.26, 1.98]</td>
<td>[2.12, 1.41]</td>
</tr>
<tr>
<td>variances</td>
<td>[0.11, 0.93]</td>
<td>[1.08, 0.12]</td>
<td>[0.11, 0.85]</td>
<td>[0.98, 0.14]</td>
</tr>
<tr>
<td>weights</td>
<td>[0.27, 0.73]</td>
<td>[0.76, 0.24]</td>
<td>[0.29, 0.71]</td>
<td>[0.23, 0.45, 0.32]</td>
</tr>
<tr>
<td>r means</td>
<td>[1.92, 1.57]</td>
<td>[1.59, 1.92]</td>
<td>[1.89, 1.58]</td>
<td>[2.82, 1.60, 2.20]</td>
</tr>
<tr>
<td>variances</td>
<td>[0.38, 0.09]</td>
<td>[0.09, 0.40]</td>
<td>[0.34, 0.08]</td>
<td>[0.30, 0.08, 0.12]</td>
</tr>
<tr>
<td>weights</td>
<td>[0.62, 0.38]</td>
<td>[0.63, 0.37]</td>
<td>[0.63, 0.37]</td>
<td>[0.66, 0.34]</td>
</tr>
<tr>
<td>i means</td>
<td>[1.97, 2.45]</td>
<td>[1.98, 2.44]</td>
<td>[1.99, 2.45]</td>
<td>[1.96, 2.48]</td>
</tr>
<tr>
<td>variances</td>
<td>[0.08, 0.22]</td>
<td>[0.08, 0.23]</td>
<td>[0.08, 0.20]</td>
<td>[0.10, 0.26]</td>
</tr>
<tr>
<td>weights</td>
<td>[1.]</td>
<td>[1.]</td>
<td>[1.]</td>
<td>[1.]</td>
</tr>
<tr>
<td>z means</td>
<td>[2.65]</td>
<td>[2.65]</td>
<td>[2.67]</td>
<td>[2.68]</td>
</tr>
<tr>
<td>variances</td>
<td>[0.14]</td>
<td>[0.14]</td>
<td>[0.14]</td>
<td>[0.17]</td>
</tr>
<tr>
<td>weights</td>
<td>[1.]</td>
<td>[1.]</td>
<td>[1.]</td>
<td>[0.62, 0.38]</td>
</tr>
<tr>
<td>y means</td>
<td>[3.23]</td>
<td>[3.22]</td>
<td>[3.23]</td>
<td>[3.07, 3.52]</td>
</tr>
<tr>
<td>variances</td>
<td>[0.15]</td>
<td>[0.14]</td>
<td>[0.14]</td>
<td>[0.07, 0.13]</td>
</tr>
</tbody>
</table>
Table B.4: Values of the optimised LightGBM model hyper-parameters used in each observing strategy. The hyper-parameters are described in the library documentation.

<table>
<thead>
<tr>
<th>Hyper-parameter name</th>
<th>Y1.5-3 baseline</th>
<th>No-roll</th>
<th>Y0-3 baseline</th>
<th>Presto-color</th>
</tr>
</thead>
<tbody>
<tr>
<td>boosting_type</td>
<td>gbdt</td>
<td>gbdt</td>
<td>gbdt</td>
<td>gbdt</td>
</tr>
<tr>
<td>learning_rate</td>
<td>0.14</td>
<td>0.14</td>
<td>0.16</td>
<td>0.18</td>
</tr>
<tr>
<td>max_depth</td>
<td>19</td>
<td>19</td>
<td>13</td>
<td>19</td>
</tr>
<tr>
<td>min_child_samples</td>
<td>70</td>
<td>70</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>min_split_gain</td>
<td>0.6</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>n_estimators</td>
<td>115</td>
<td>115</td>
<td>115</td>
<td>115</td>
</tr>
<tr>
<td>num_leaves</td>
<td>50</td>
<td>50</td>
<td>55</td>
<td>50</td>
</tr>
</tbody>
</table>

B.4 Well-measured Type Ia Supernovae

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

$$380 \text{ nm} < \frac{\bar{\lambda}_{\text{obs}}}{1+z} < 700 \text{ nm}, \quad (B.2)$$

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

$$\frac{t - t_{\text{peak}}}{1+z}, \quad (B.3)$$

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

- at least 3 observations before peak with phase $>-20$
- at least 8 observations after peak with phase $<60$
- at least 1 observation with phase $\leq 10$
- at least 1 observation with phase $\geq 20$
• $\sigma_C < 0.04$, where $\sigma_C$ 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 3.4.1 rather than the three-year-long light curves.


Arendse, N. et al. (in prep). Inferring the Hubble constant from strongly lensed supernovae in LSST with spatio-temporal neural networks (pages 144, 148).


Jeffery, D. J. (1999). Radioactive decay energy deposition in supernovae and the exponential/quasi-


Pan-STARRS Supernovae. I. Systematic Uncertainty from Core-collapse Supernova Contamination.

1710.00846, p. 51. DOI: 10.3847/1538-4357/aab6b1 (pages 82, 96, 122).

Cadence Choices for the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST).


of artificial intelligence research 4, pp. 237–285 (page 64).


Kelly, P. L. et al. (2015). Multiple images of a highly magnified supernova formed by an early-type clus-

Kenworthy, W. et al. (2021). SALT3: An Improved Type Ia Supernova Model for Measuring Cosmic

Kessler, R. et al. (2015). The difference imaging pipeline for the transient search in the Dark En-

Classification Challenge (PLAsTiCC). Publications of the Astronomical Society of the Pacific


