Considerations for optimizing photometric classification of supernovae from the Rubin Observatory

Catarina Alves

Supervised by:
Hiranya Peiris and Jason McEwen

University College London

CosKASI ECR seminar

Collaborators

Catarina Alves
(UCL)

Hiranya Peiris
(UCL / OKC)

Michelle Lochner
(UWC / SARAO)

Jason McEwen
(UCL / MSSL)

Tarek Allam Jr
(UCL / MSSL)

Rahul Biswas
(OKC)

LSST

The Rubin Observatory **L**egacy **S**urvey of **S**pace and **T**ime

LSST Telescope Mount Assembly Group Photo.
Credit: Asturfeito

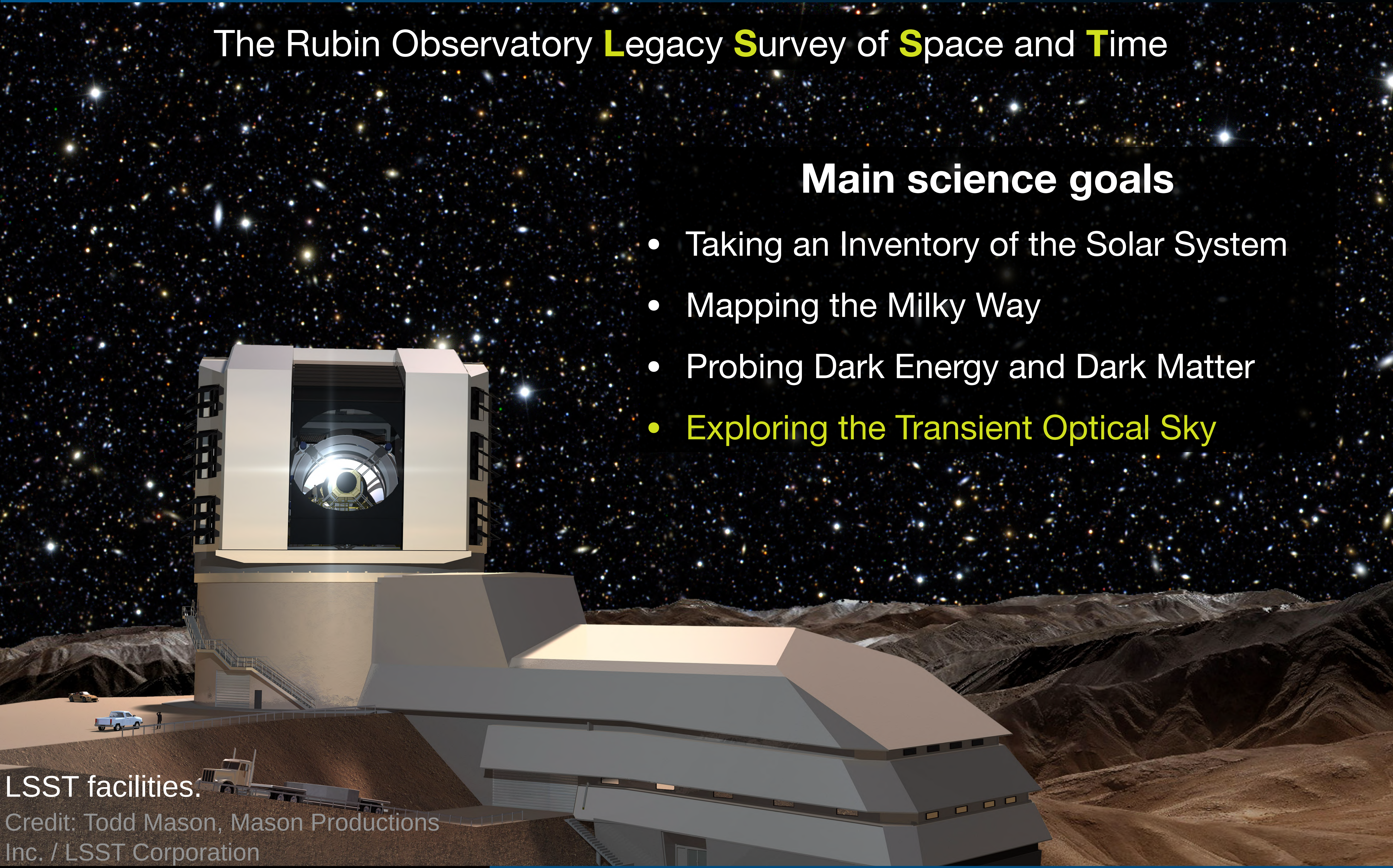
- Cerro Pachón, Chile
- 8.4m wide-field telescope
- 3.2 Gpx camera
  - world's largest digital camera
- 3.5-degree field of view
- Each image the size of 40 moons

LSST

The Rubin Observatory **L**egacy **S**urvey of **S**pace and **T**ime

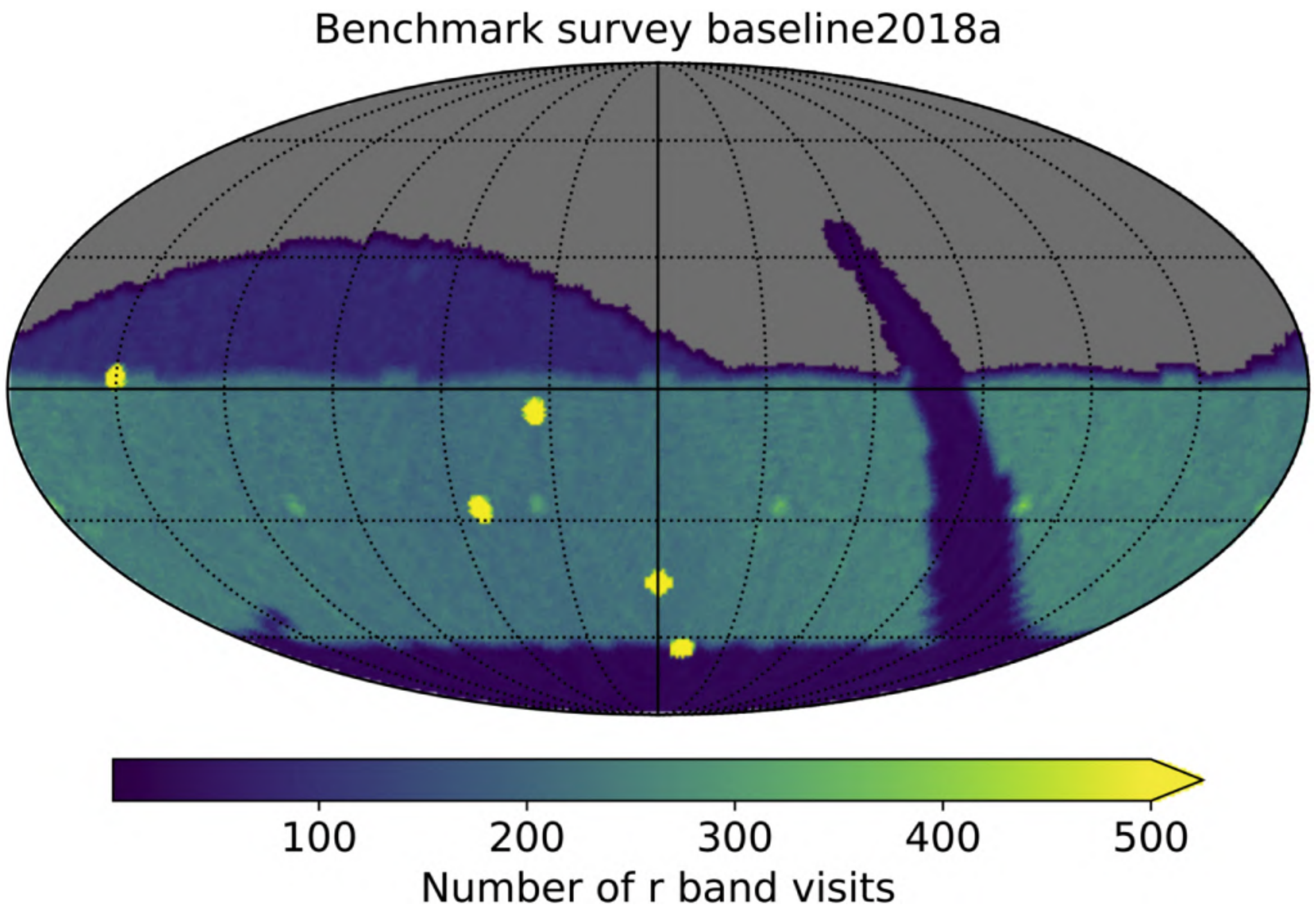
Main science goals

- Taking an Inventory of the Solar System
- Mapping the Milky Way
- Probing Dark Energy and Dark Matter
- Exploring the Transient Optical Sky



LSST facilities.
Credit: Todd Mason, Mason Productions Inc. / LSST Corporation

LSST and Transients



Distribution of the r-band visits on the sky for a simulated realization of the baseline cadence.
Credit: Ivezic et al. (ArXiv: 0805.2366v5, living reference document)

LSST key numbers

- Wide (18000+ deg2)
- Fast (~3 days)
- Deep (25-28 mag)
- 10 years
- 6 filters (320-1050 nm)
- Specialised surveys, such as Deep-Drilling-Fields (DDF) → more frequent and deeper observations
- **~10 million alerts per night**

Motivation

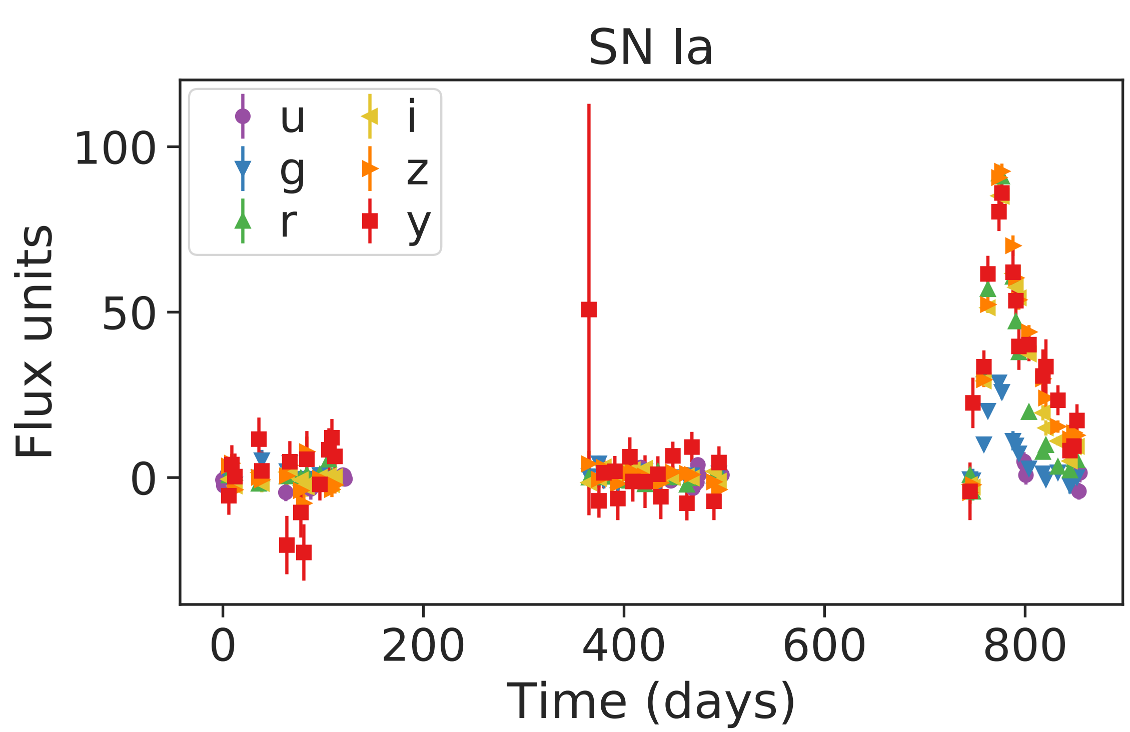
- Supernovae (SNe) are used for astrophysical and cosmological studies
- LSST will discover at least one order of magnitude more SNe than the current available SNe samples
- Limited spectroscopic resources → photometric classification
- Photometric classification performance depends on the survey observing strategy
- **First study to analyze the impact of the LSST observing strategy on SNe classification**

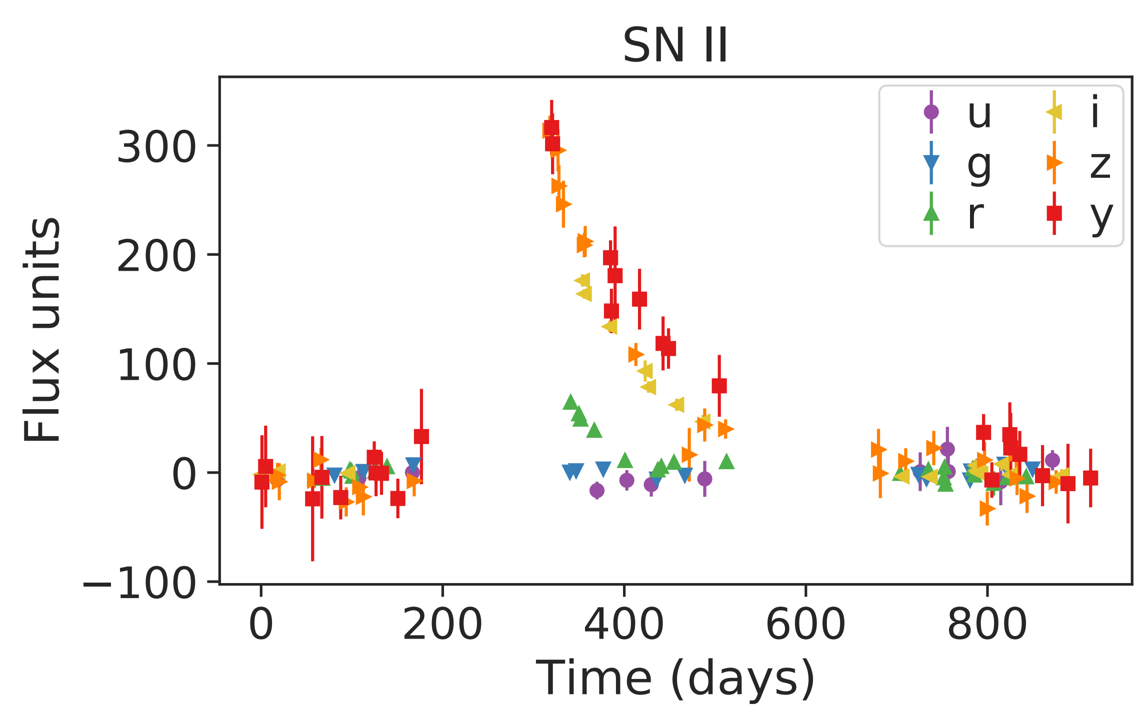
PLAsTiCC

- **P**hotometric **L**SST **As**tronomical **Ti**me-Series **C**lassification **C**hallenge
- Simulated multi-band light curves for 3 years of LSST
- Simulated host-galaxy photometric redshifts and uncertainties
- Realistic observing conditions but outdated observing strategy
- Simulations in two survey modes:
  - Wide-Fast-Deep (WFD) → 99% of the events
  - Deep-Drilling-Fields (DDF) → 1% of the events

PLAsTiCC

- 3.5 millions events → 18 different classes of transients and variable stars
- This work focuses on classifying SN Ia, SN Ibc, SN II
- Simulated spectroscopically-confirmed training set biased towards nearby, brighter events → non-representative

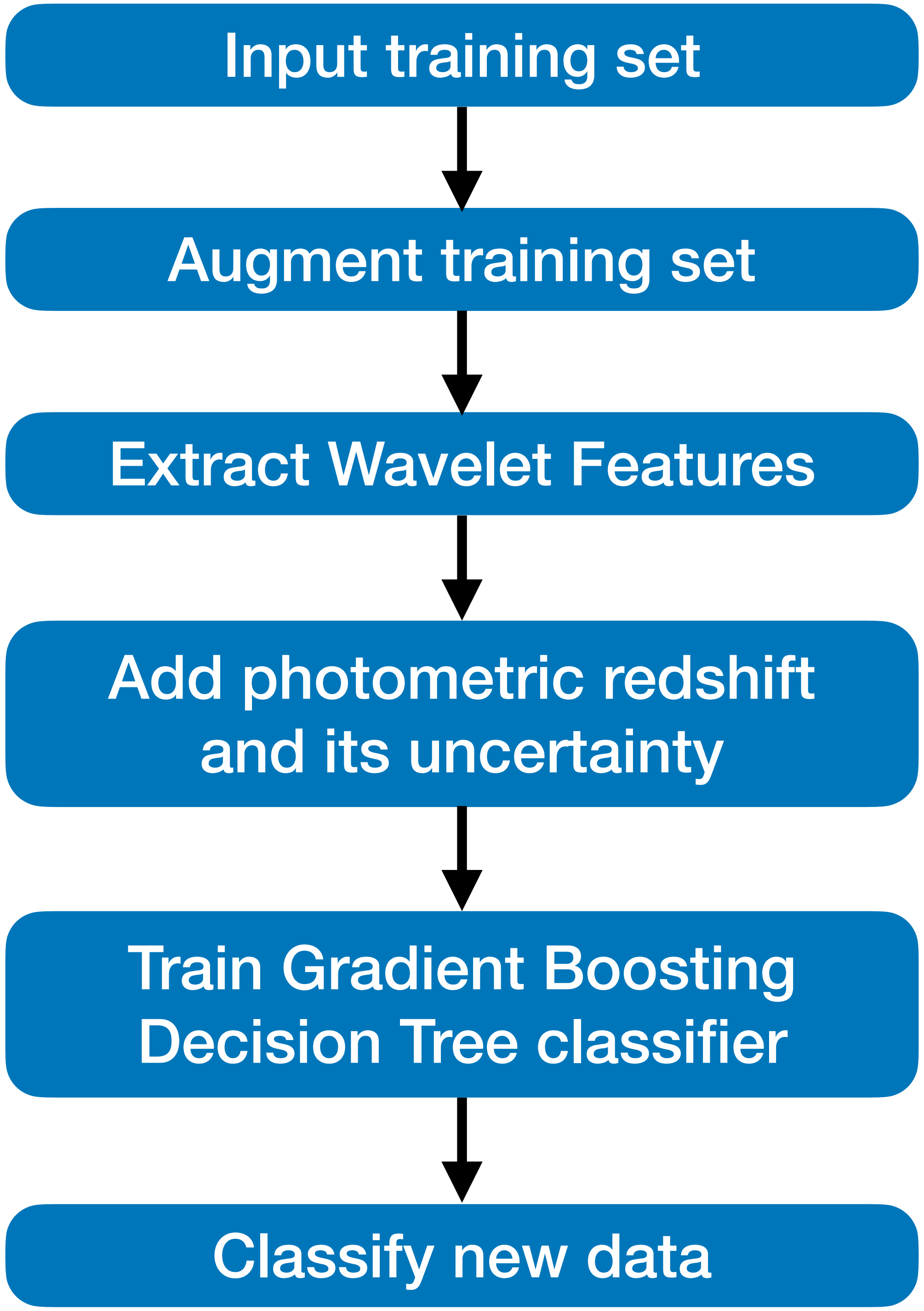




The PLAsTiCC team et al. 2018; PLASTICC Team & PLASTICC Modelers 2019

snmachine pipeline

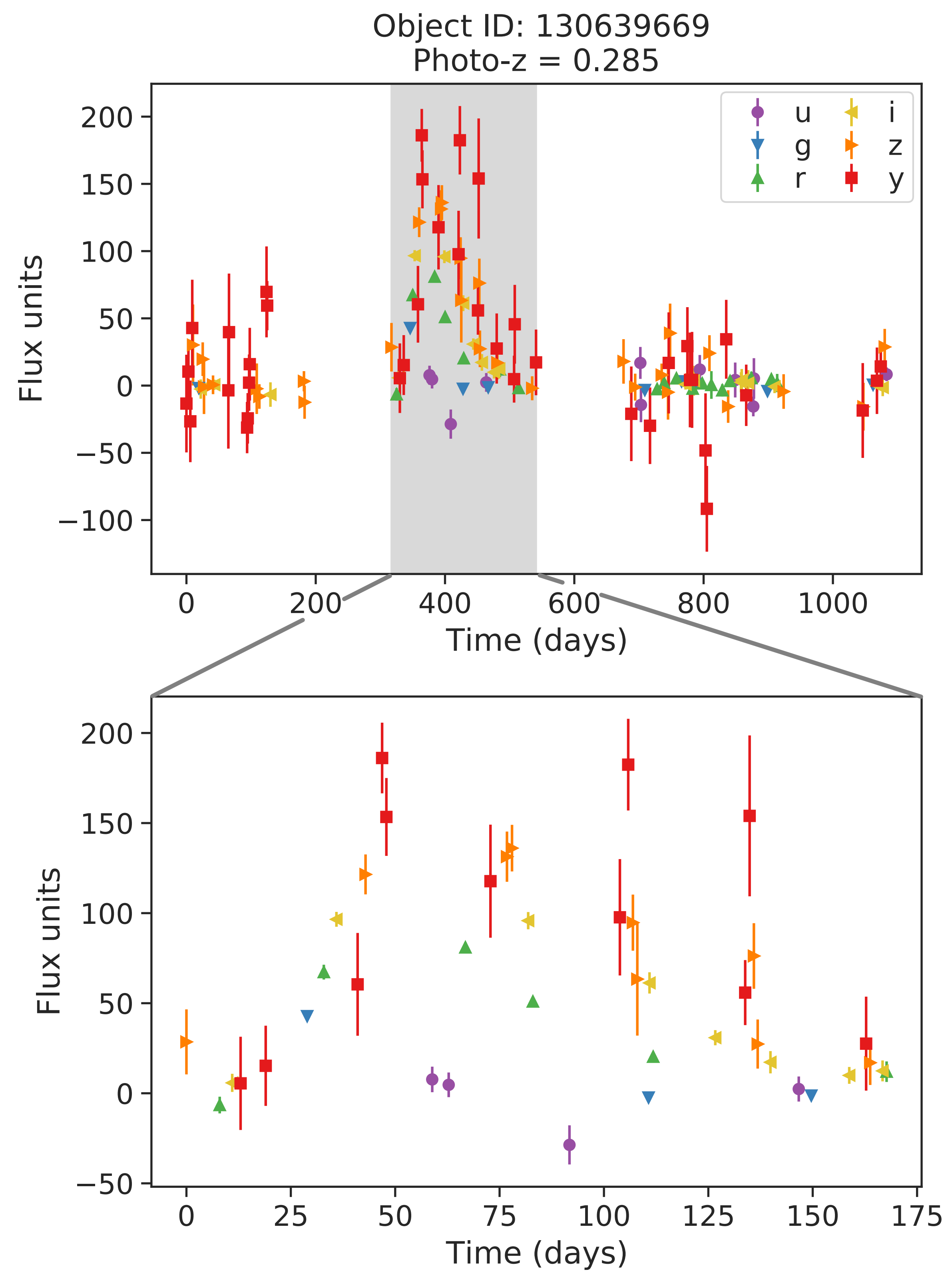
- Build a classifier using the photometric transient classification library snmachine (Lochner et al. 2016)
- Original version of snmachine used in Lochner et al. (2016), Narayan et al. (2018), Malz et al. (2019), Carrick et al. (2020), Sooknunan et al. (2021)
- snmachine upgraded for use with LSST data
- Public release with the accompanying paper



Light curve preprocessing

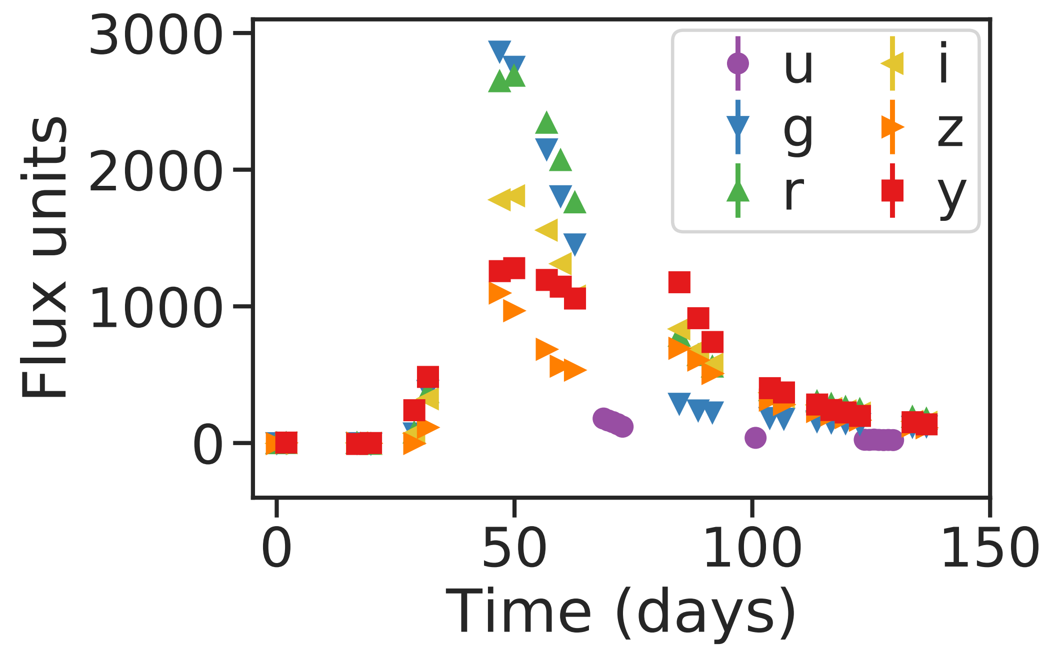
- Isolate the observing season that contains the SNe
  - season which contains the observations flagged as detected
  - no inter-night gaps larger than 50 days

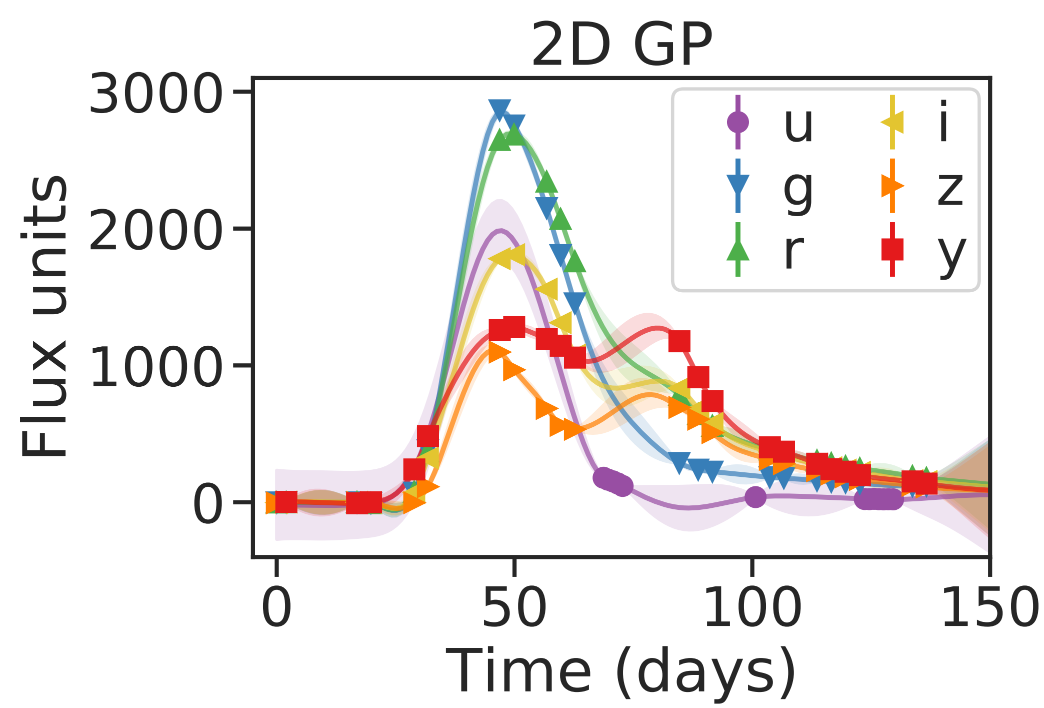
- Introduce uniformity in the dataset → translate the light curves so their first observation is at time zero



Gaussian process modeling

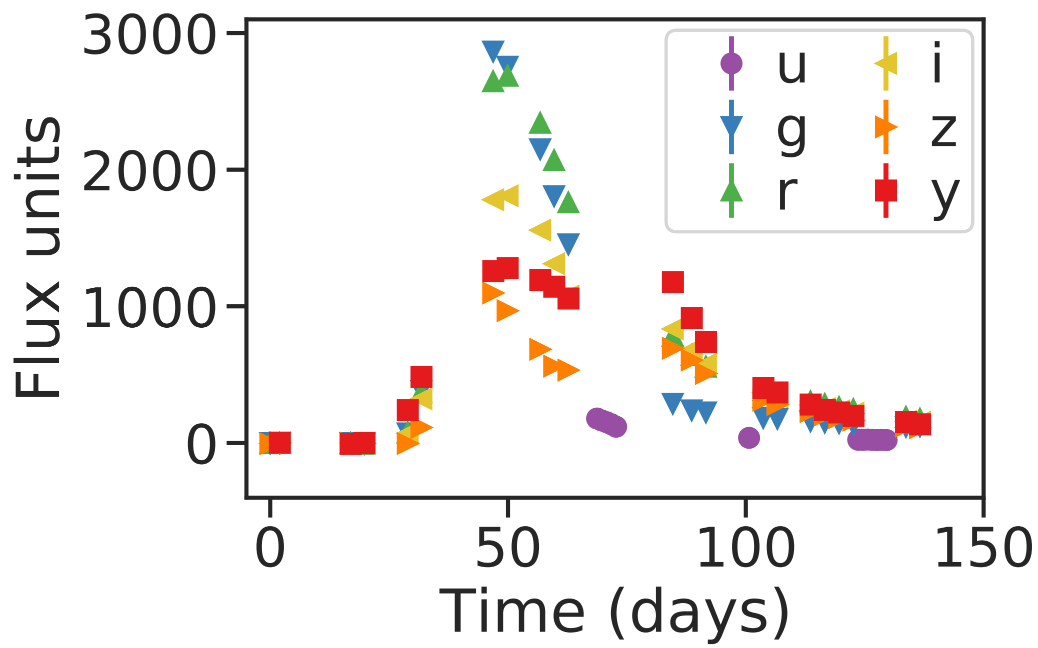
- Model each light curve with a 2D Gaussian process (GP)

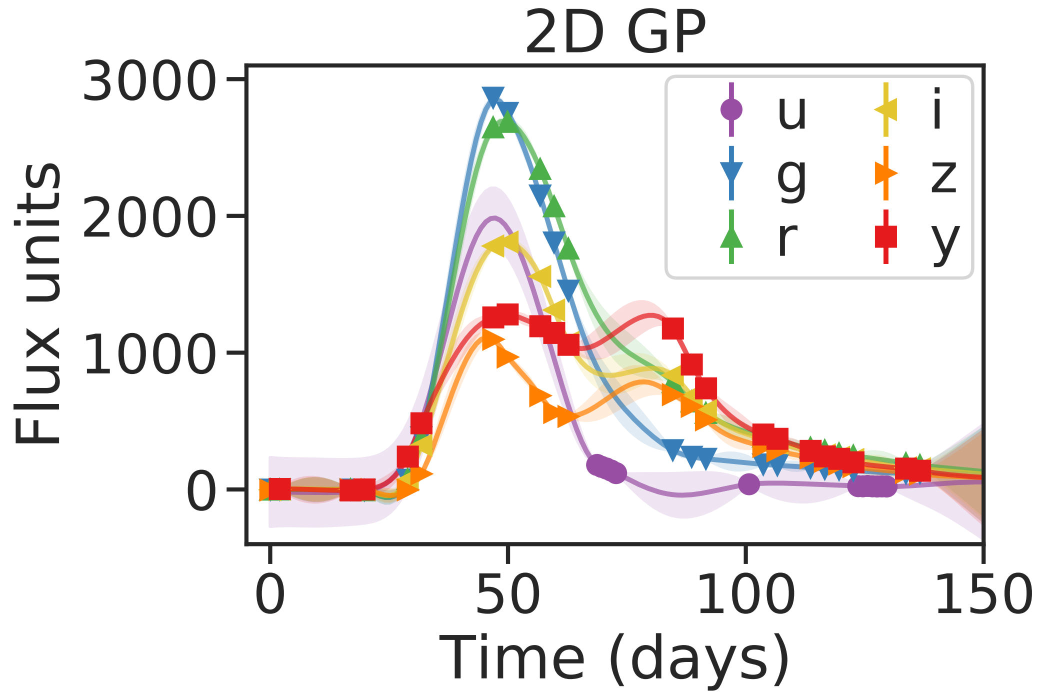




Gaussian process modeling

- Model each light curve with a 2D Gaussian process (GP)
- **What is a Gaussian process?**
- A GP is a probability distribution over possible functions that are consistent with a set of observations
- Characterised by its mean function and its covariance function/kernel
- Predicts the flux at new times



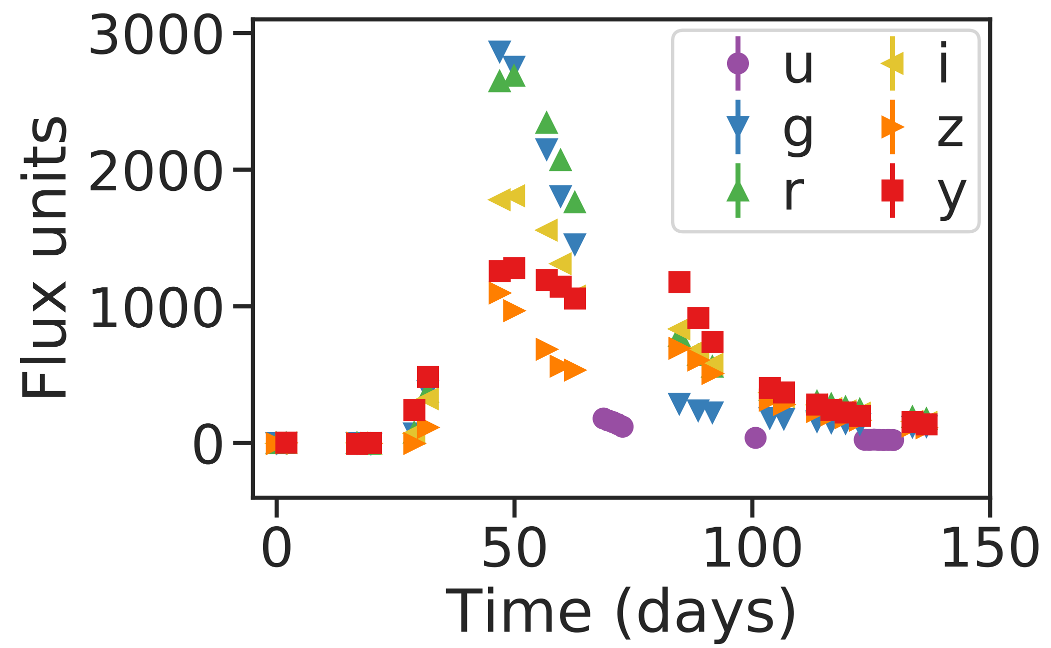


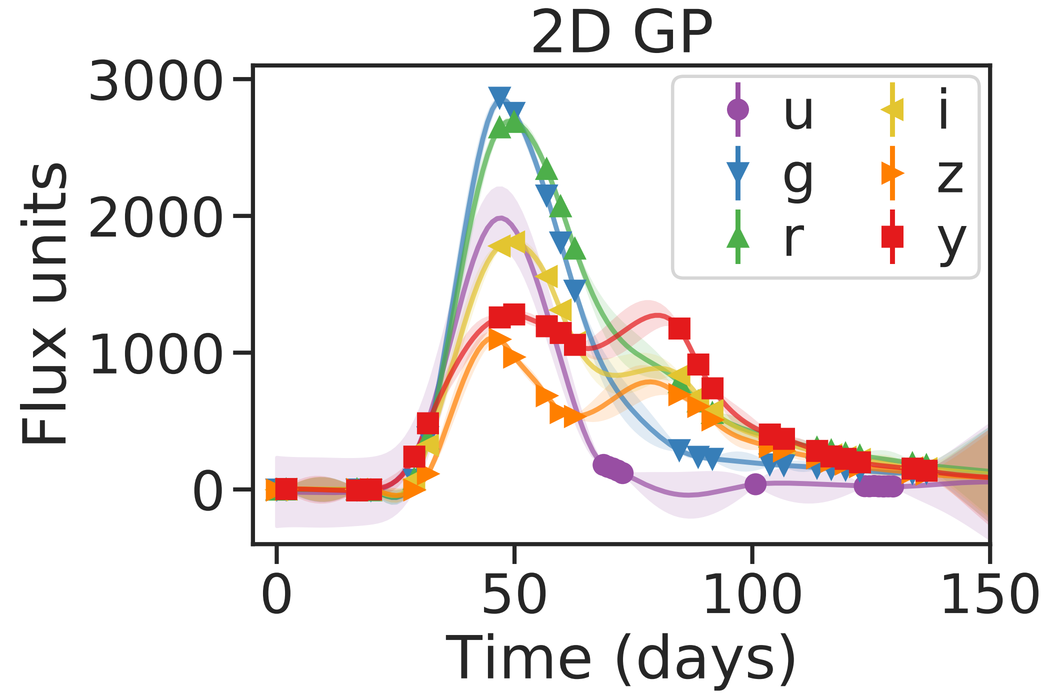
Gaussian process modeling

- Model each light curve with a 2D Gaussian process (GP)

- 2D GPs are fitted both in time and wavelength
  → incorporate cross-band information
  → infers the flux in passbands where there are no observations

- GPs fitted with
  - null mean function
  - Matern-3/2 kernel





Training set augmentation

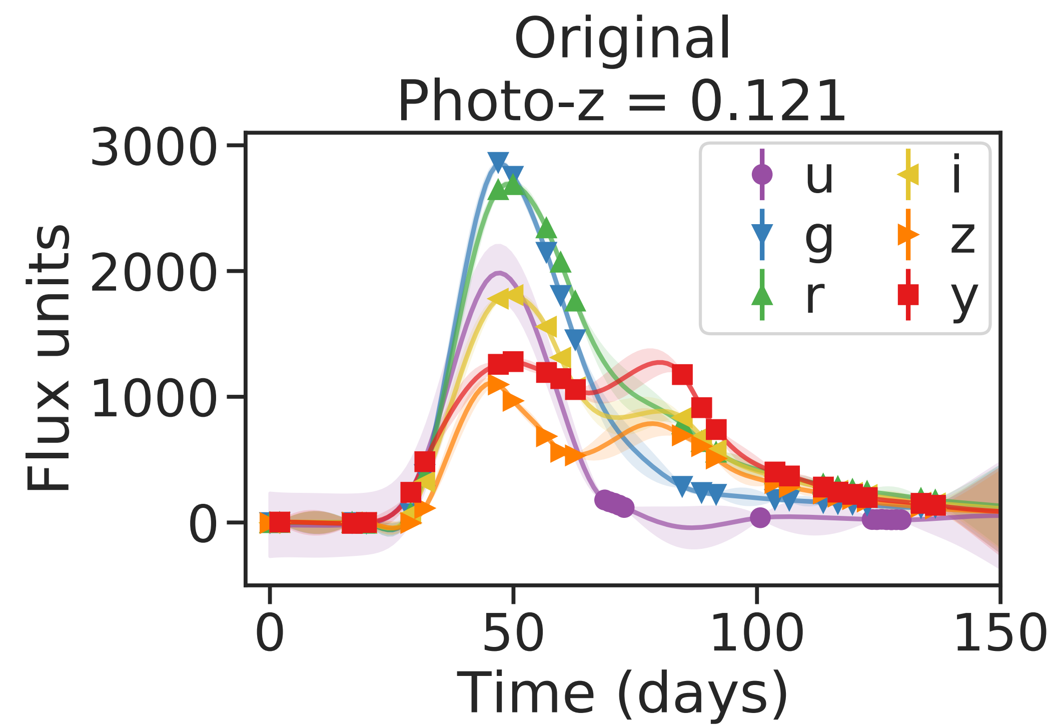
- The training set is:
  - non-representative of the test set
  - imbalanced: ~4.8 times more SN Ia than SN Ibc
- Accurate classification → training set must be representative and balanced
- **Solution:** Augment the simulated training set to be representative of
  - the photometric redshift distribution per SNe class,
  - the cadence of observations,
  - and the flux uncertainty distribution of the test set

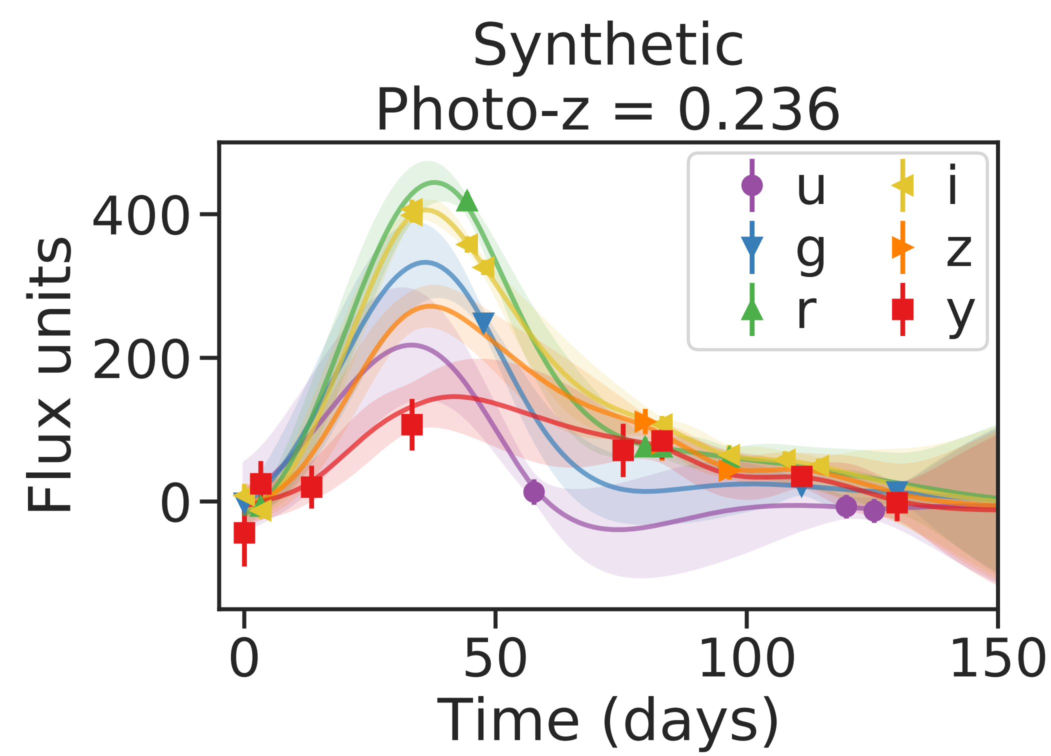
  (based on Boone, 2019)

Augmentation approach

1. Choose the number of new events to create
2. Model the original light curve with a 2D GP fit in time and wavelength
3. Choose a redshift for the new event
4. Create synthetic observations at the new redshift, making use of the GP fit to the original event
5. Generate a photometric redshift and its uncertainty

- Same number of augmented events from **each SN class**





Wavelet features

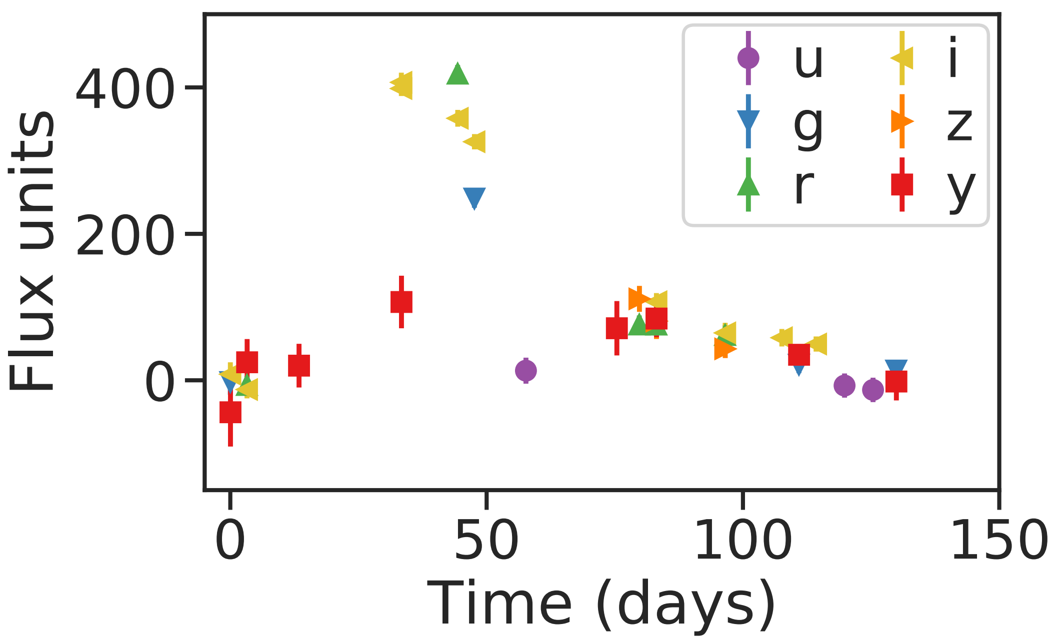
- Wavelet features are model independent
- Localised both in time and frequency
- General features → can characterise many classes of transients
- Successful for general transient classification (e.g. Varughese et al. 2015; Lochner et al. 2016; Gautham Narayan et al. 2018; Sooknunan et al. 2021)
- Approach not previously used by the winning PLAsTiCC entries

Wavelet features

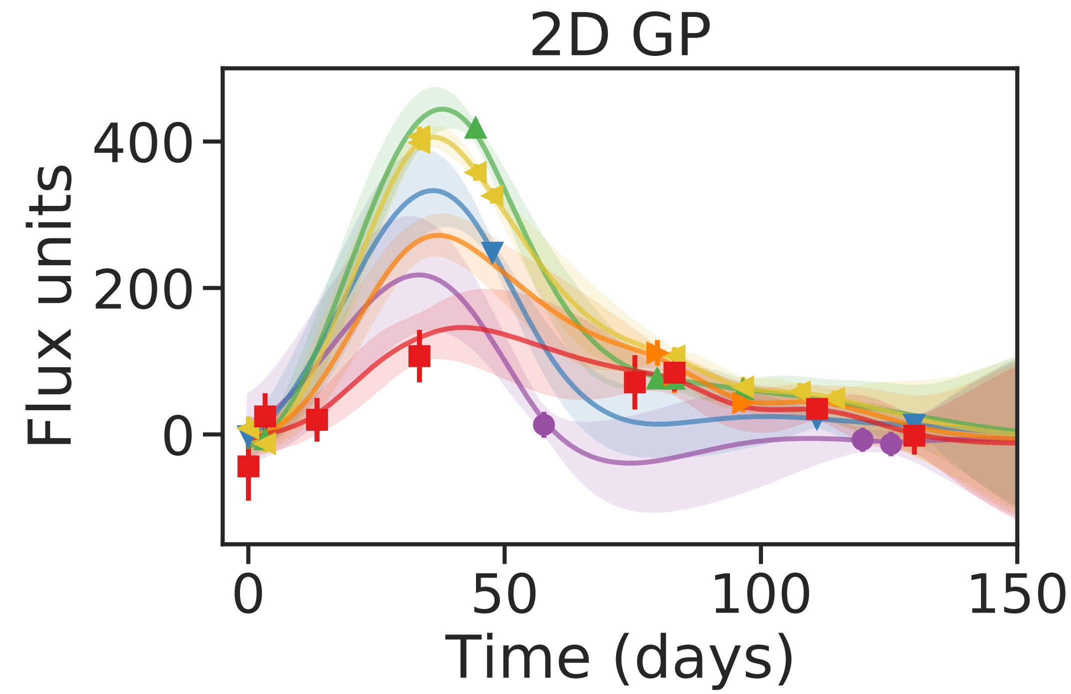
- Wavelet features are model independent
- Localised both in time and frequency
- General features → can characterise many classes of transients
- Successful for general transient classification (e.g. Varughese et al. 2015; Lochner et al. 2016; Gautham Narayan et al. 2018; Sooknunan et al. 2021)
- Approach not previously used by the winning PLAsTiCC entries

How do we extract wavelet features?

Wavelet features extraction



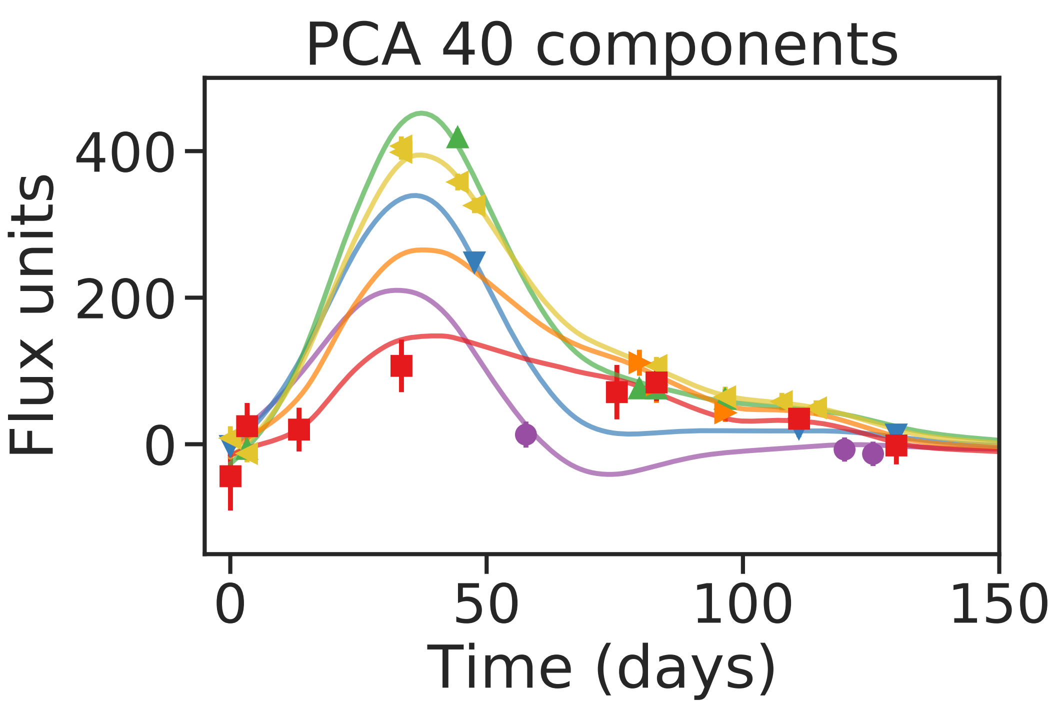
Gaussian process fit



Wavelet decomposition

Wavelet Components

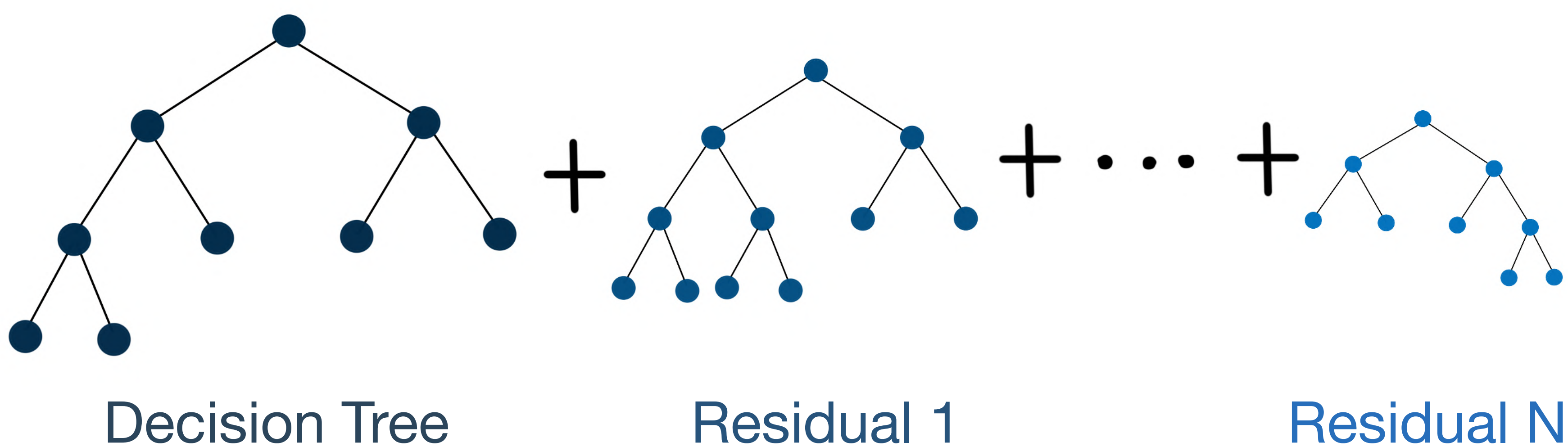
PCA



Classification

- Used Gradient Boosting Decision Tree (GBDT) to classify the events
  - Ensemble classifier → 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.

- Optimised the GBDT hyperparameters by maximising the performance of a 5-fold cross-validated grid-search on the augmented training set.



Performance evaluation

- To evaluate the classification performance, we used the PLAsTiCC weighted log-loss metric (The PLAsTiCC team et al. 2018; Malz et al. 2019):

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

  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.

- Weights can be changed to give different importances to different classes.

Performance evaluation

- To evaluate the classification performance, we used the PLAsTiCC weighted log-loss metric (The PLAsTiCC team et al. 2018; Malz et al. 2019):

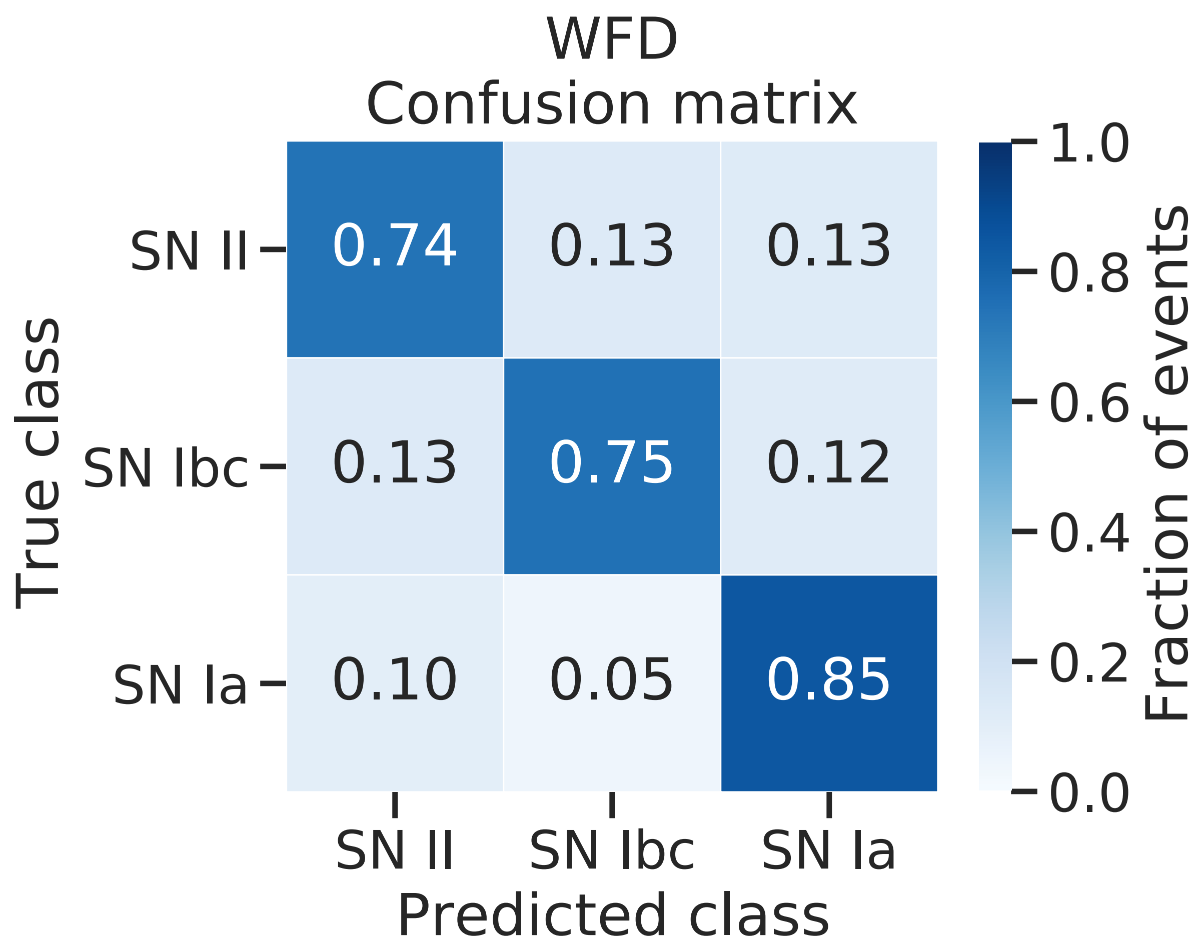
$$\textbf{Log-loss} = -\left(\frac{\sum_{i=1}^{M} w_i \cdot \sum_{j=1}^{N_i} \frac{y_{ij}^{*}}{N_i} \cdot \ln p_{ij}}{\sum_{i=1}^{M} w_i}\right)$$

  - Probabilistic metric → includes classification uncertainty
  - Disfavours classifiers that neglected any classes

- Following the PLAsTiCC challenge, we gave the same weight to every SNe class.

Classification and performance

- Use the augmented training set to train a classifier with the features:
    - wavelet features (40 PCA components)
    - photometric redshift + its uncertainty
- Use the PLAsTiCC weighted log-loss metric (Malz et al. 2019)
- Performance comparable to that obtained by the top three submissions to PLAsTiCC

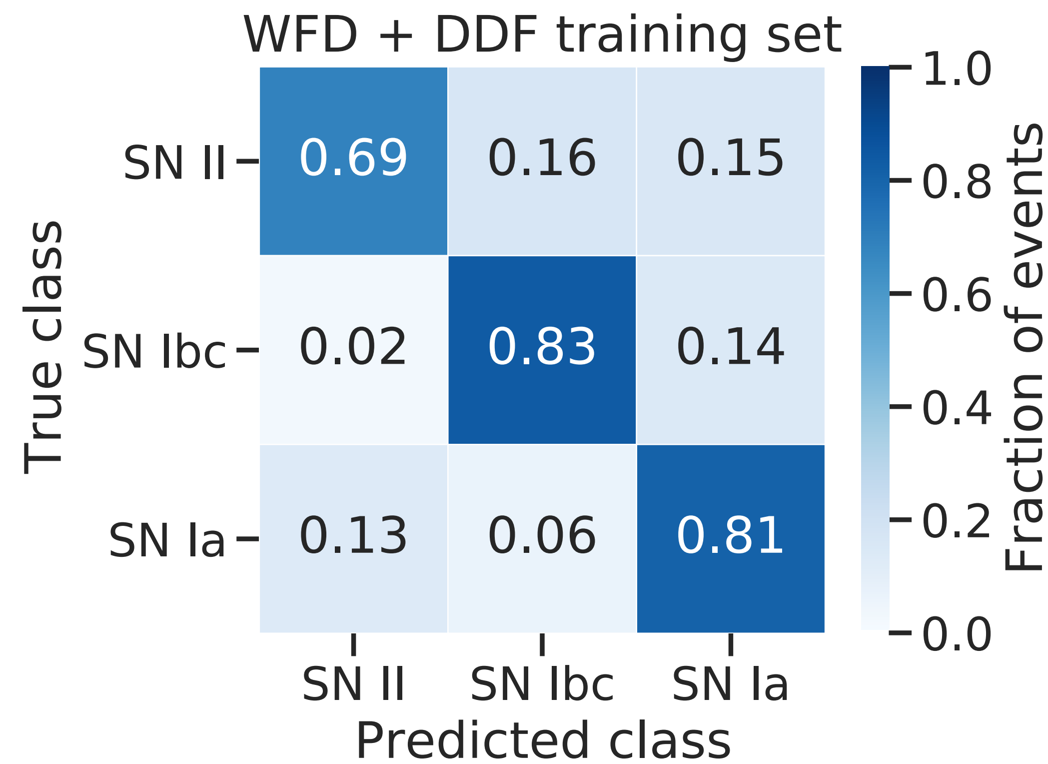


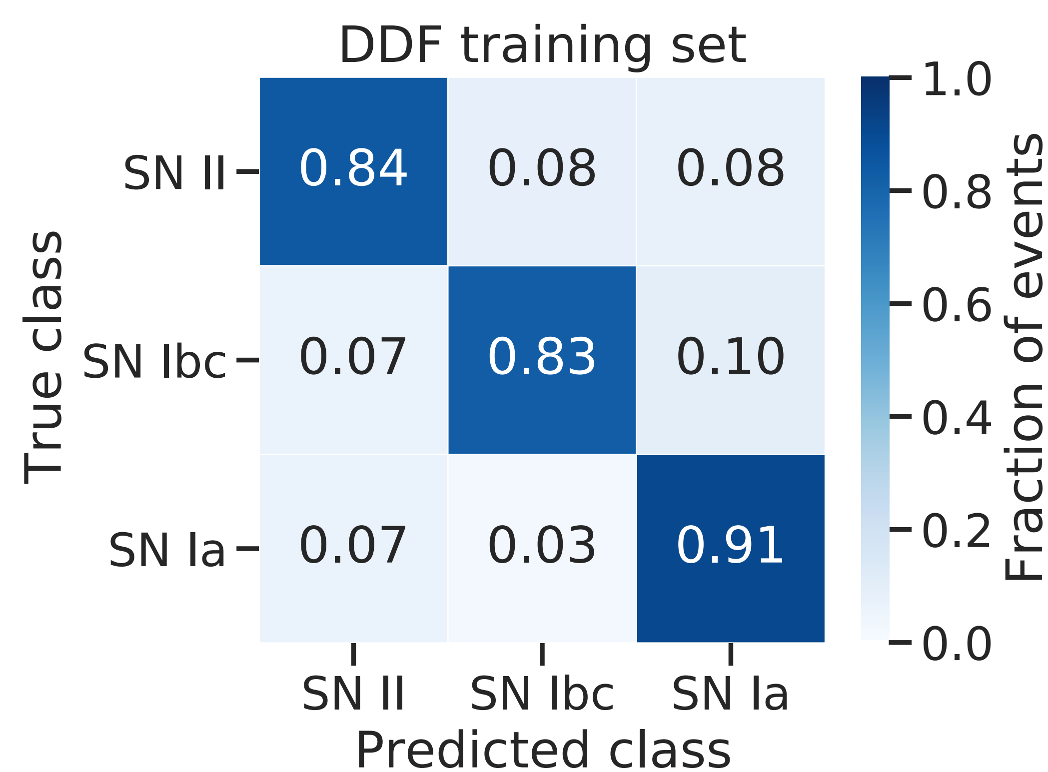
DDF classification performance

- Compare the DDF classification performance using classifiers based on the augmented
    - WFD + DDF training set
    - DDF training set

- Results show that it is crucial to match the augmented training sets to the characteristics of the different survey modes

- DDF survey yields higher classification performance than WFD survey





Observing strategy

- **What are the implications for observing strategy?**
- We study classification performance for SNe with different properties within the single simulated observing strategy that is available in PLAsTiCC
- Measure the performance using:
  - $\text{Recall} = \dfrac{\text{TP}}{\text{TP} + \text{FN}}$
  - $\text{Precision} = \dfrac{\text{TP}}{\text{TP} + \text{FP}}$

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

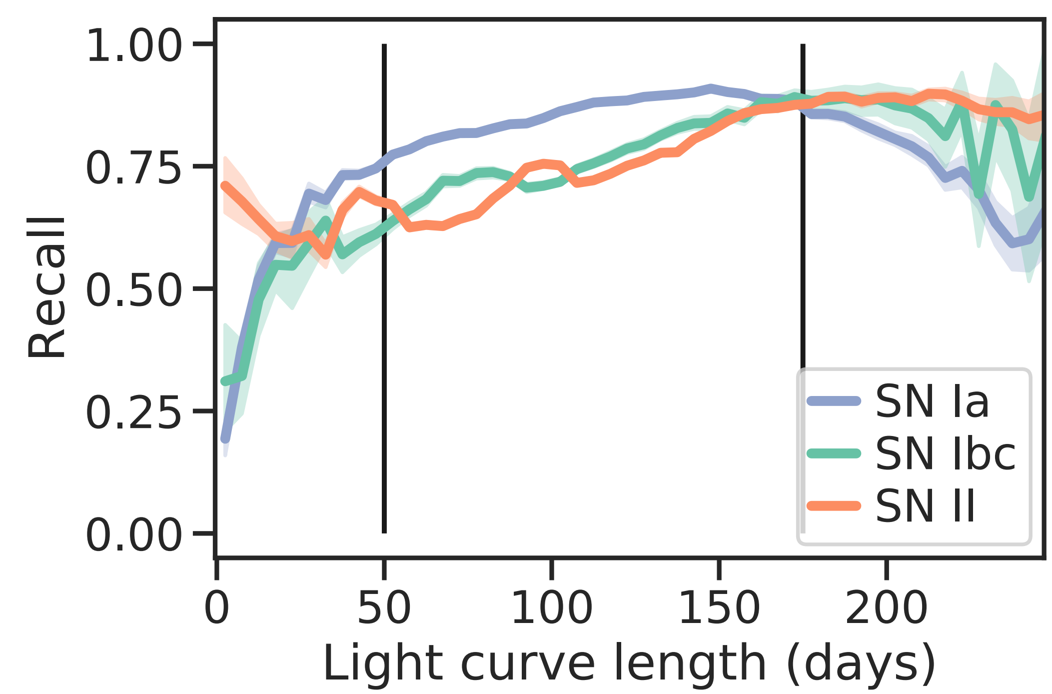
Light curve length

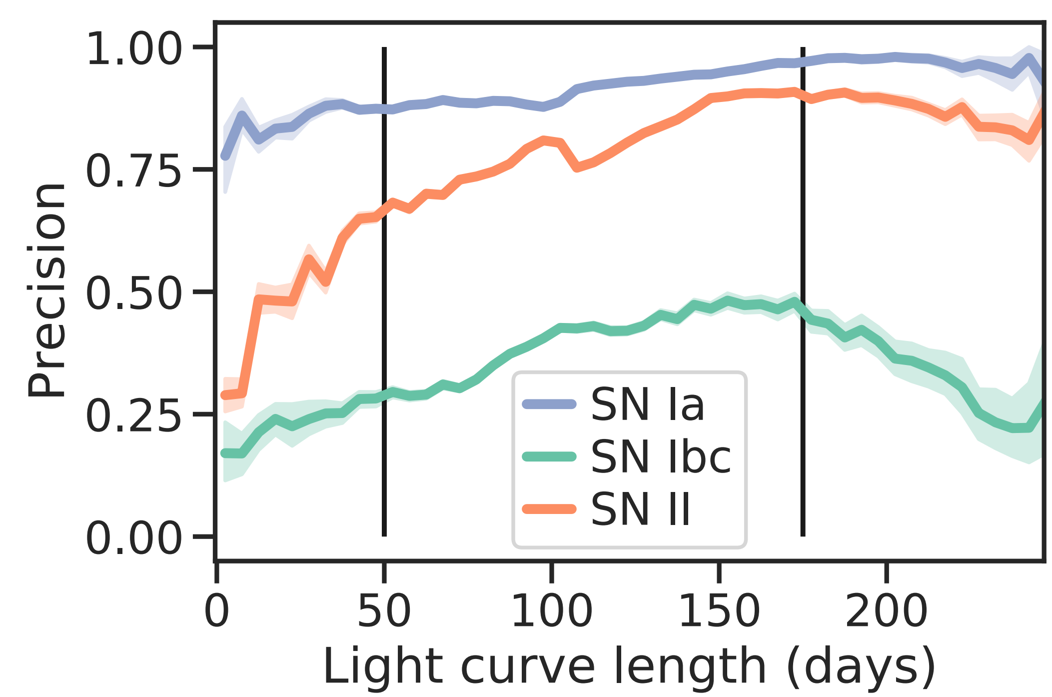
- Season length → tuned by taking additional observations in suboptimal conditions

- Light curve length → proxy for season length

- Focus on light curve lengths between 50–175 days; small-number effects outside that range

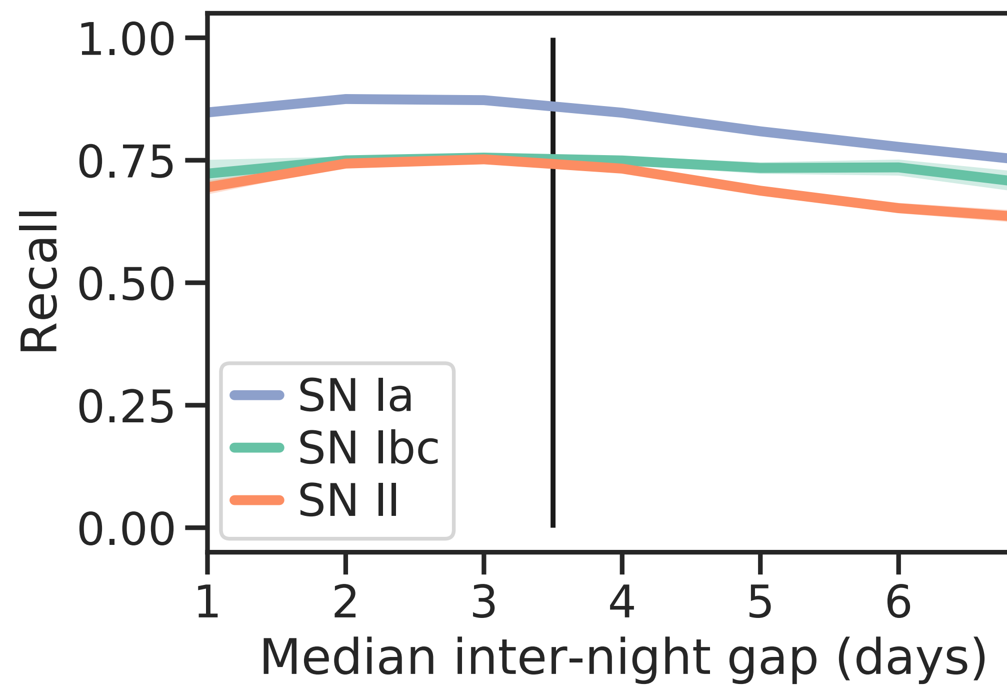
- Events observed for longer
  → better characterization by the feature extraction step
  → higher recall and precision

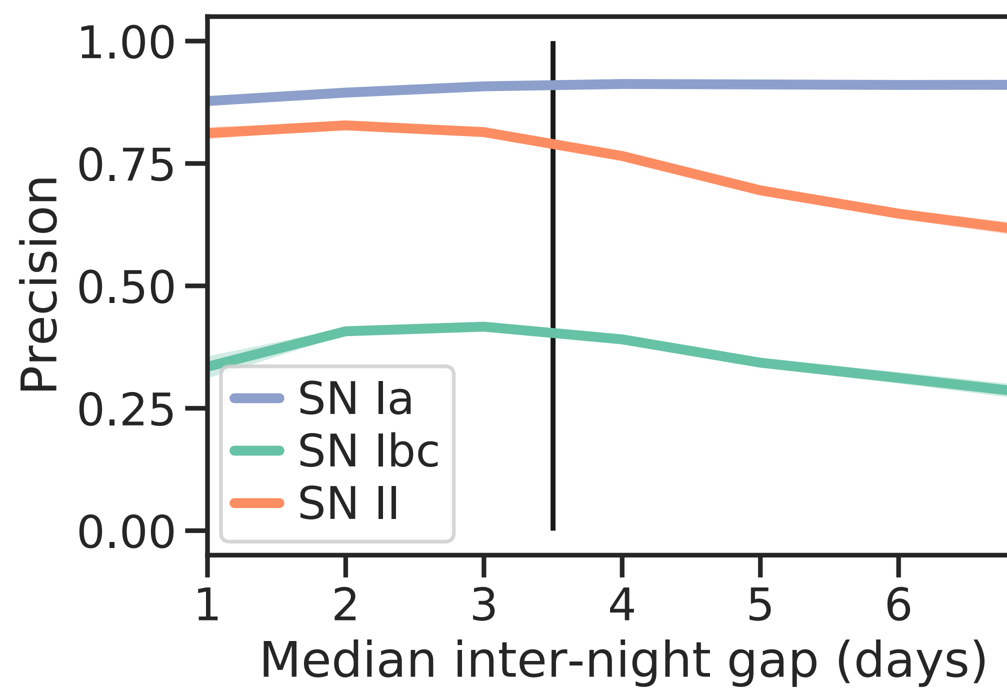




Median inter-night gap

- Cadence of observations → impacts all transient science goals
- Inter-night gap → quantifies the cadence
- Events whose median inter-night gap is < 3.5 days
  → better sampled events
  → higher light curve quality
  → higher recall and precision

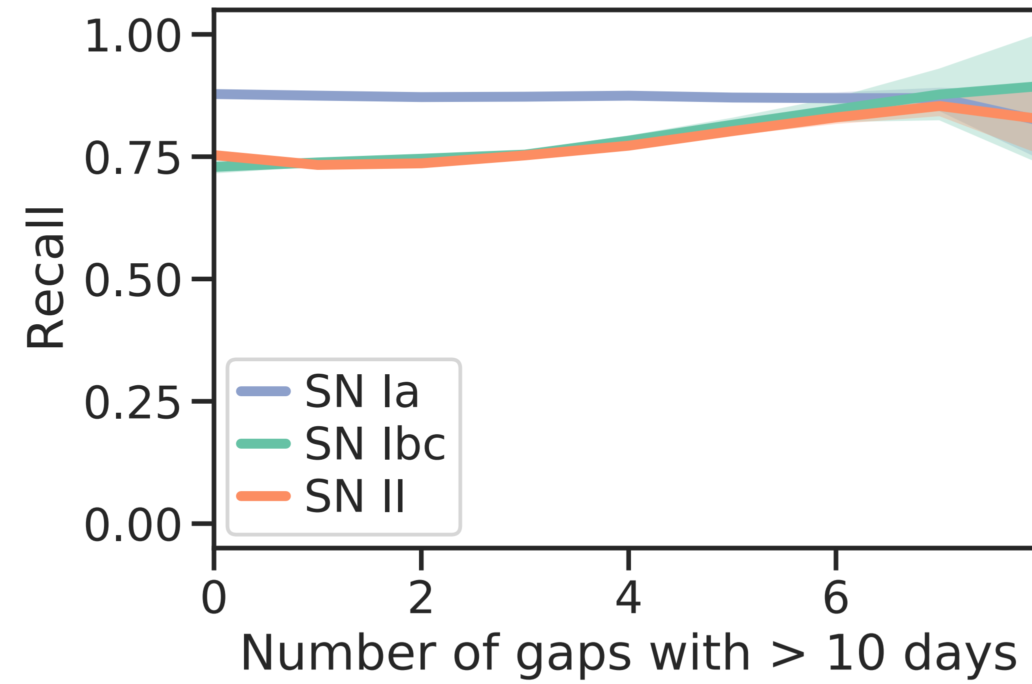


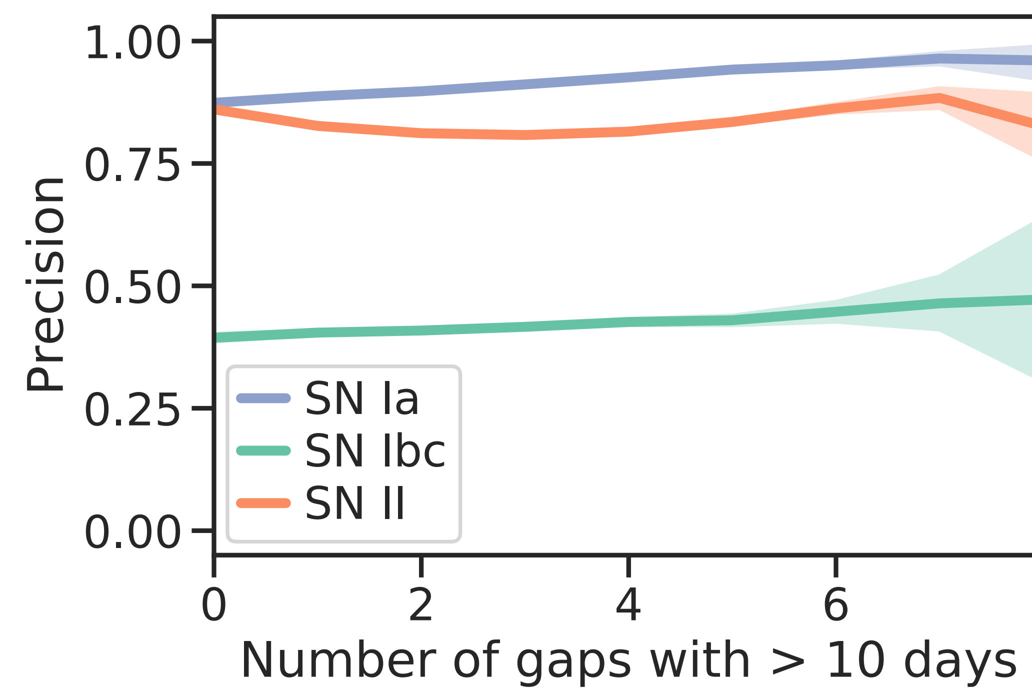


Large inter-night gaps

- Effect of the number of large gaps in events with a median inter-night gap < 3.5 days

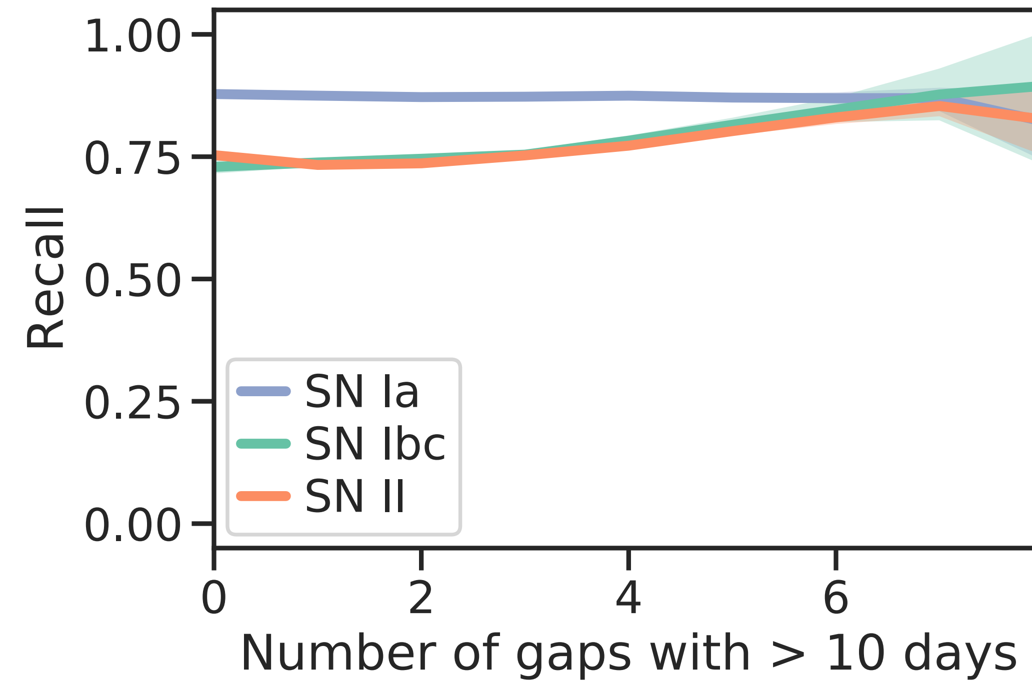
- GP fits can interpolate large gaps if median inter-night gap < 3.5 days
  → recall and precision independent of the number of large gaps

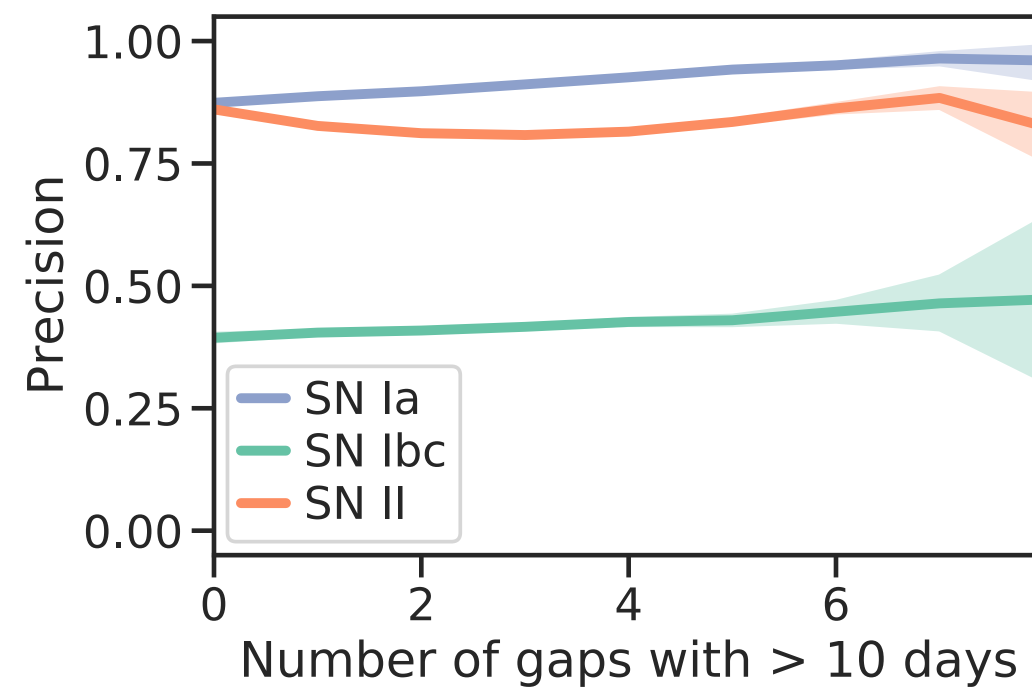




Large inter-night gaps

- Effect of the number of large gaps in events with a median inter-night gap < 3.5 days
- GP fits can interpolate large gaps if median inter-night gap < 3.5 days
  → recall and precision independent of the number of large gaps
- At which point does the performance degrades due to inability of GP fits to constrain a light curve fit?



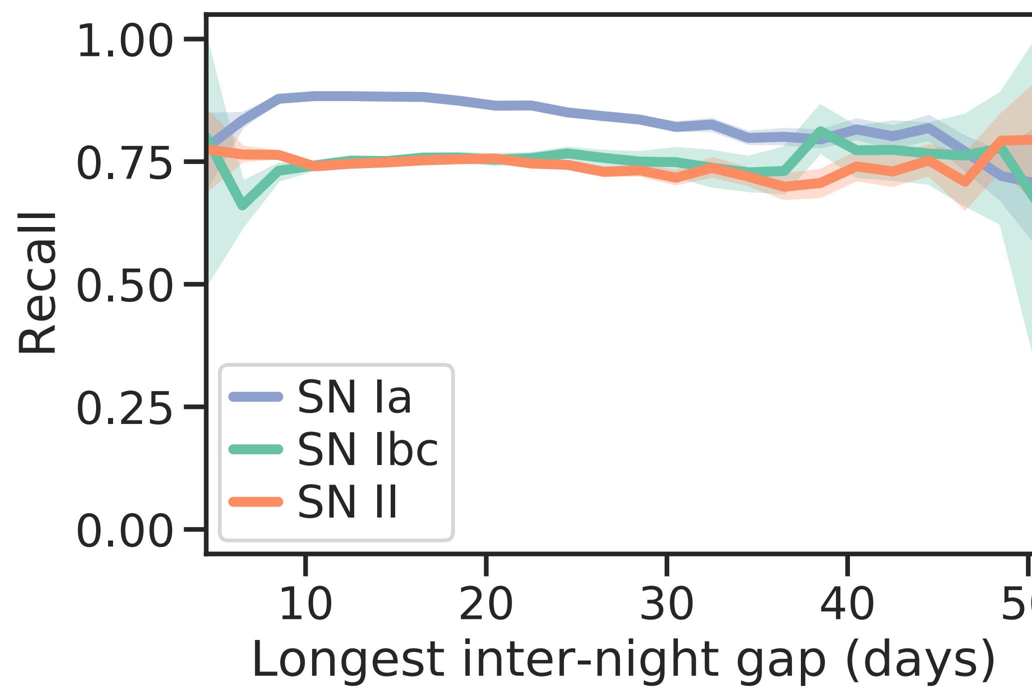


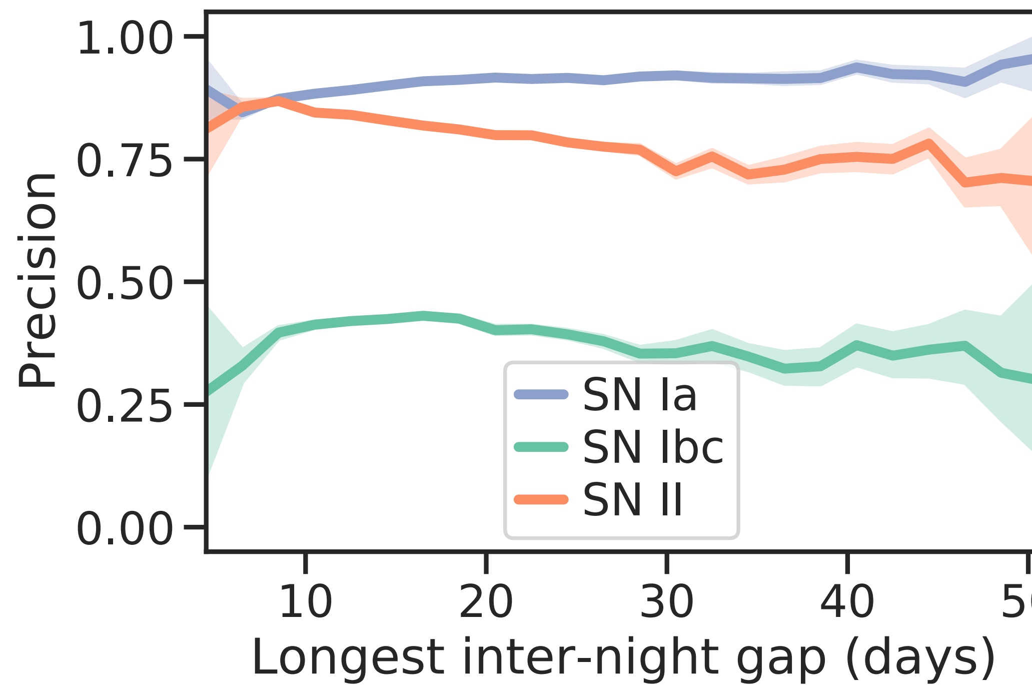
Longest inter-night gap

- Effect of the length of the longest gap in events with a median inter-night gap < 3.5 days

- Increase of the length of longest gap
  → recall and precision either slowly decrease or remain constant

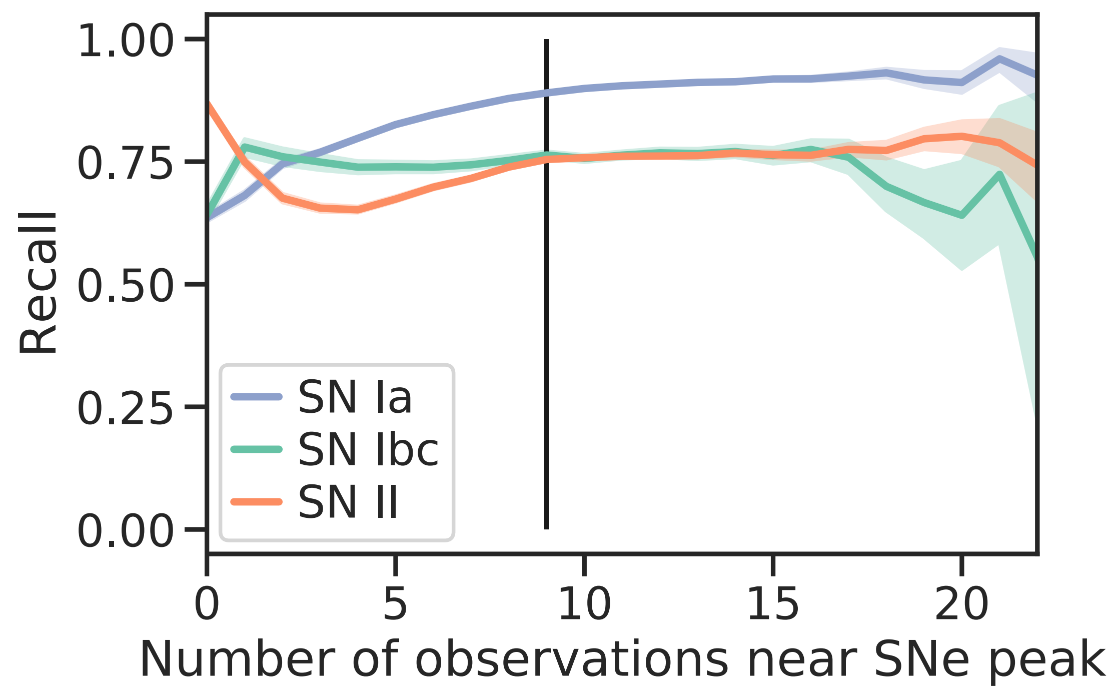
- Results show that a median inter-night gap of < 3.5 days is sufficient for photometric classification

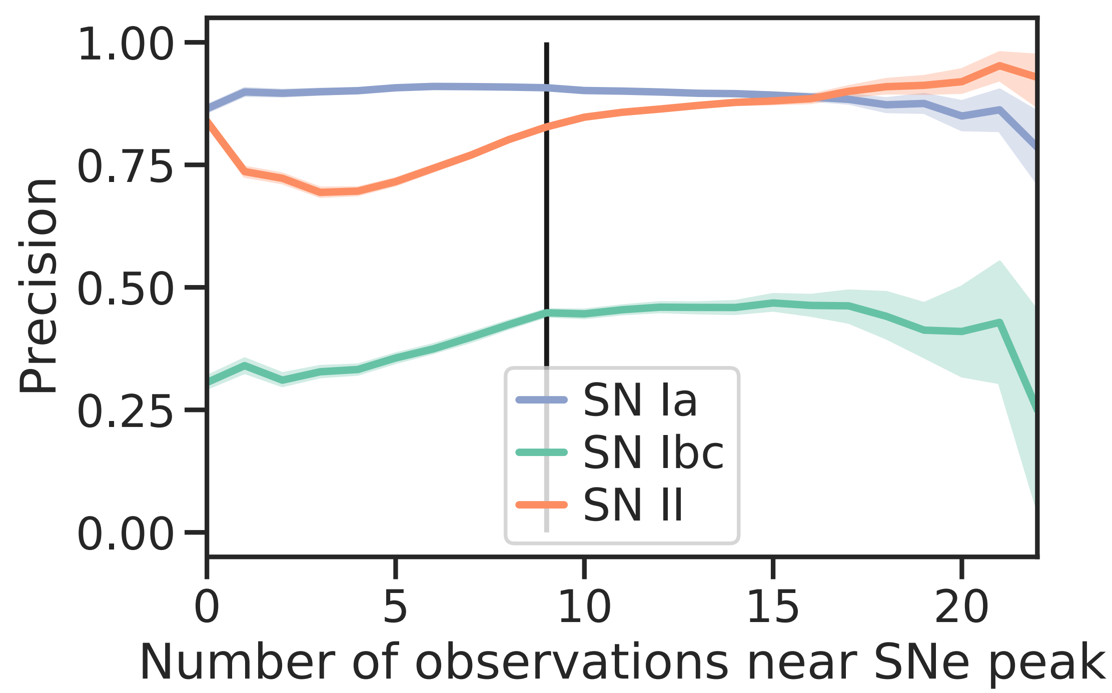




Observations near peak

- Observations near SN Ia peak → reliable cosmological distances
- Near peak: 10 days before and 30 days after the SNe peak
- Events with more observations near peak
  → better characterization of light curve shape
  → higher recall and precision
- Constant performance for > 9 observations





Conclusion

- Augmentation is crucial to obtain a representative training set
- First study of how observing strategy impacts photometric classification:
  - longer light curves → higher performance
  - median inter-night gap of < 3.5 days → higher performance
  - number of inter-night gaps > 10 days → no impact
  - observations near SNe peak → higher performance
- The results provide guidance for further refinement of the LSST observing strategy on the question of SNe photometric classification
- Public release of snmachine

Future work

- Investigate the dependence of classification performance on different observing strategy (OS) simulations
    - for each OS, simulate multi-band SNe light curves using SNANA software (Kessler et al. 2009)
    - apply methodology developed in this paper
    - quantify the difference in performance between the different OS
    - produce OS recommendations regarding SNe photometric classification

- Links
    - Paper: https://arxiv.org/abs/2107.07531
    - `snmachine` library: https://github.com/LSSTDESC/snmachine

Redshift distribution per class