Classifying Stripped Envelope Supernovae with Properly Synthesized Low-Resolution Spectra

Willow Fox Fortino (she/her)

University of Delaware Department of Physics and Astronomy

Abstract

After first light for the Rubin Observatory, millions of transient events and hundreds of supernovae will be discovered each night. As a result, observatories around the world will have the chance to photometrically and spectroscopically follow up on these events. While certain supernovae can be classified through photometry alone, stripped envelope core collapse supernova must be classified through spectroscopy. Therefore, when deciding the resolving power, R, of new spectrographs, it is important to know the minimum R necessary to classify these supernovae to arbitrary accuracy. This work attempts to identify a critical spectral resolution $R = \Delta\lambda/\lambda$ at which spectral classification of subtypes of supernovae becomes impossible. In the process, we provide a rigorous method to degrade space to simulate future low resolution observations from existing high resolution data. To classify the spectra we follow previous work of Williamson (2019) and: first perform Principal Component Analysis (PCA) on spectra taken at similar phases of the supernova's evolution. Subsequently, a Support Vector Machine (SVM) classifier is used on some of the principal components (PCs). The SVM score for each group of supernovae is recorded as we artificially degrade the spectra. We confirm that even at R = 5, the SVM score remains at ~0.50, significantly above what would be expected for a random guess, ~0.25. Further work includes measuring the performance of different data-driven classifiers as a function of spectral resolution.

Introduction

Stripped envelope core collapse supernovae are a subset of supernovae whereby the outer layer of Hydrogen (and sometimes Helium) is removed before the actual supernova event. The four stripped envelope supernovae that are considered for classification in this work are Type Ib, Type Ic, Type Ic-BL, and Type IIb. While we expect spectra to be necessary for classification, in this work we ask what is the minimum spectral resolution that is needed for an accurate classification. This work builds on work by Williamson (2019) who provided an entirely data-driven classification scheme for stripped envelope supernovae and on prior work of our group (Umer Zubair's UD Masters Thesis, *Stripped Envelope Supernovae Classification at Low Spectral Resolution*). Notably, in our group's first exploration of the question, Zubair observed no critical spectral resolution below which the classifier was only as good as a random guess.

We attempt to reproduce the results of this thesis by examining the same dataset with the same classifier while using an adjusted method for degrading the spectra. The original method for degrading the spectra did not take the spectral PSF into account and instead simply re-binned the fluxes in order to maintain the original signal-to-noise ratio and also to avoid excessive information loss. The method we employ convolves the spectrum with a Gaussian with a FWHM that is proportional to the wavelength bin size and is inversely proportional to the degraded spectral resolution, R. The Gaussian serves as a generic spectral PSF, which enable the results of this work to be informative for those seeking to build new spectrographs.

SNePCA (Williamson 2019) is a machine learning based classifier that first performs Principal Component Analysis (PCA) on the dataset. The result of the PCA is the principal components (PCs) and the PC coefficients. SNePCA feeds these PC coefficients into a linear support vector machine (SVM) classifier which constructs hyperplanes in the feature space which optimally classify the spectra.

The first five PCs explain ~80% of the variance of the data, and as a result we do not consider any other PCs. However, while Williamson selected two optimal components from the top five by visual inspection to build the SVM classifier, we improve on this method by performing the classification in the 5-dimensional feature space of the top five PCs.

Methodology

The entire dataset of well observed stripped envelope supernovae classified at high confidence consists of about 200 spectra across the four phases of the light curve: 0±5, 5±5, 10±5, and 15±5 days from peak brightness. Figure 1 shows an example of a spectrum taken 1.8 days after peak brightness. Each spectrum begins with a spectral resolution of R = 738, and the wavelength bins are constantly spaced in log-space. After extracting the data, the following is performed in Python:

1. **Degradation:** In order to degrade the spectrum to some new spectral resolution, R, the spectrum is convolved with a Gaussian kernel with a FWHM that changes at each step of the convolution. The FWHM is always proportional to the wavelength bin size at each step. Figure 2 shows each step of the convolution process.

2. **PCA:** Using the SNePCA Python package, the 'eigenspectra' of all of the spectra corresponding to the same phase and the corresponding PCA coefficients are calculated. Only the first five PCs are considered in this work as they explain ~80% of the variance. It is possible that the other PCs are simply noise and including more in the classification would not be helpful.

3. **SVM:** Using the LinearSVC function from the Python package scikit-learn, we train an SVM on 70% of the data, reserving 30% for calculating accuracy. Because the dataset is so small, 50 rounds of cross validation are performed where the SVM is trained on another random training set. The SVM score we report, and its uncertainty, come from the average and standard deviation of these cross validation steps.

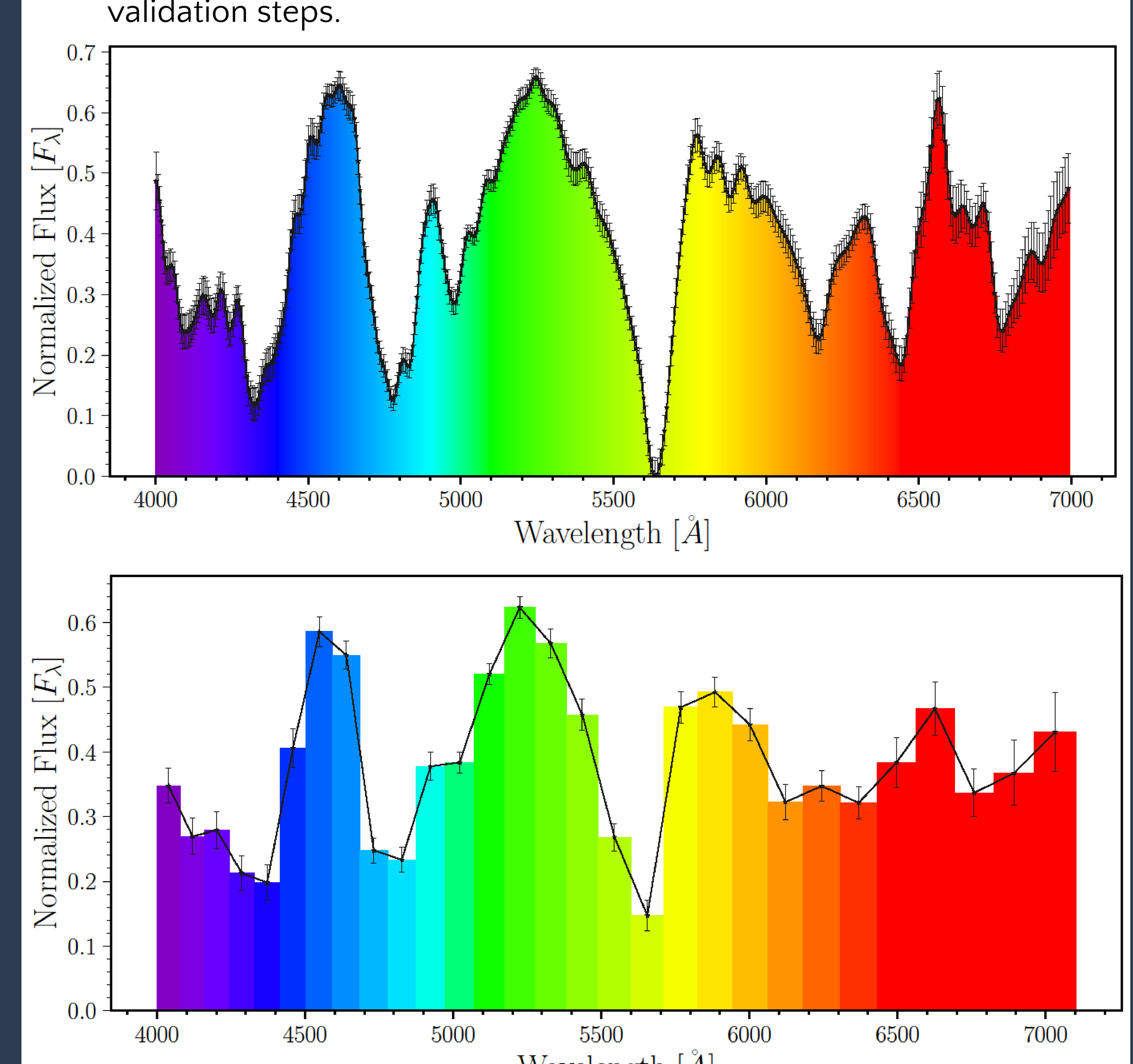

Figure 1: (Top) A spectrum of Type Ib-normal stripped envelope supernova sn1998dt taken 1.8 days after peak brightness, plotted with flux uncertainties. The center of each wavelength bin in the histogram corresponds to the color of the bar. The spectral resolving power, R, of this spectra is 738. (Bottom) The spectrum in the top panel has been artificially degraded such that it would appear the spectra was taken by an instrument with a spectral resolving power of 50.

Results

In an effort to assess the minimum resolution required for stripped envelope supernovae classification, we build on previous work by our group with the following improvements: (1) we performed the classification using the first *five* PCs while Williamson (2019) used two PCs for each phase which we shown to produce the best SVM score.

As expected, utilizing more PCs leads to a higher SVM score at every epoch and at every spectral resolution (see Figure 3), although an analysis of the feature space like what was done in Williamson & Modjaz & Bianco (2019) becomes impossible.

The SVM scores for the classification of spectra at the native spectral resolution when using all five PCs are:

0.80 ± 0.10 for Phase 0 ± 5 days
0.79 ± 0.09 for Phase 5 ± 5 days
0.78 ± 0.09 for Phase 10 ± 5 days
0.77 ± 0.09 for Phase 15 ± 5 days

The SVM scores for the classification of spectra at a spectral resolution of R = 5 (where there are only four wavelength bins total) when using all five PCs are:

0.46 ± 0.11 for Phase 0 ± 5 days
0.49 ± 0.10 for Phase 5 ± 5 days
0.57 ± 0.13 for Phase 10 ± 5 days
0.51 ± 0.14 for Phase 15 ± 5 days

As the amount of information in the spectra is reduced as the resolution tends to 0, we expect the classifier to have no classifying power. That is, the random guess classification power which takes dataset imbalance into account is calculated by inputting only random noise into the SVM which yields 0.21 ± 0.09, 0.21 ± 0.10, 0.23 ± 0.10, and 0.22 ± 0.09 for each phase, respectively.

In future work we will look at alternative classification techniques such as DASH (Muthukrishna et al. 2019), a deep learning based classifier, in order to further investigate the critical spectral resolution, as well as understand how accurate a classifier can be at this task.

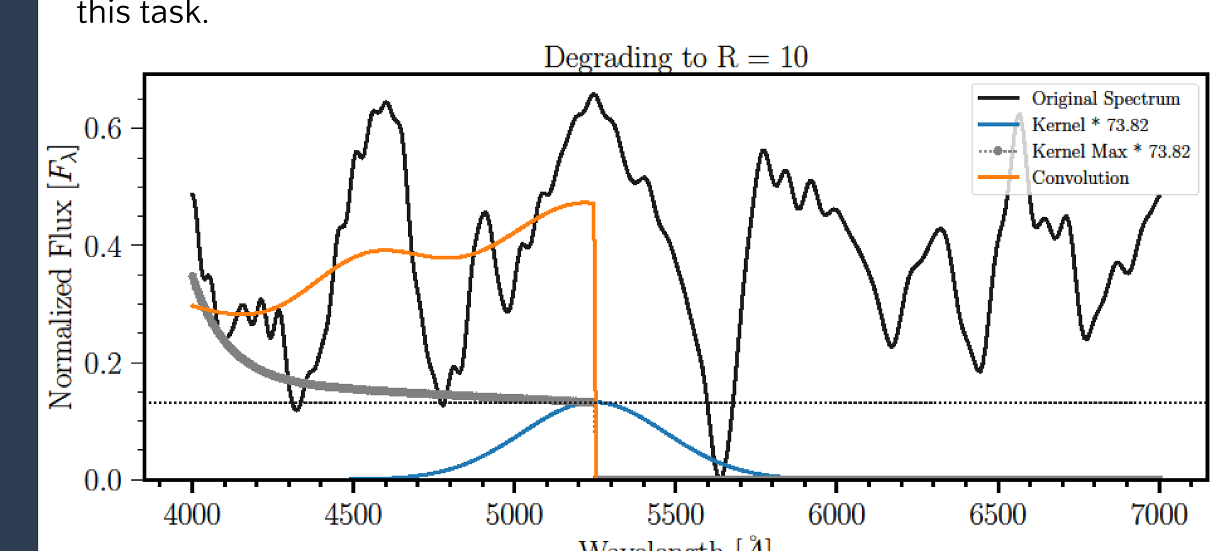

Figure 2: A still from a GIF (animated GIF at the QR code on the right) shows each step of the convolution process. The original spectrum is shown in black, and the resulting convolution is shown in yellow. The kernel at each step is shown in blue, the peak of the kernel is denoted by the horizontal dashed black line, and the peak of the kernel is procedurally plotted in gray. Note how the kernel gets wider as it moves to the right.

Conclusion

The results of this work surprisingly indicates that even at low resolution we retain some classification power. While there is an obvious degradation of the classification accuracy at R < ~30 we continue to retain more accuracy than a random guess (~0.25). Even with four spectra bins our classification power is >0.4. At R = 5 there are only four wavelength bins remaining in the spectra, which is less information than is contained in a photometric survey with the typical five passbands.

More work is required to test the limits of this observation. Adding more supernova to the dataset and trying different classifiers could help assess how robust this finding is.

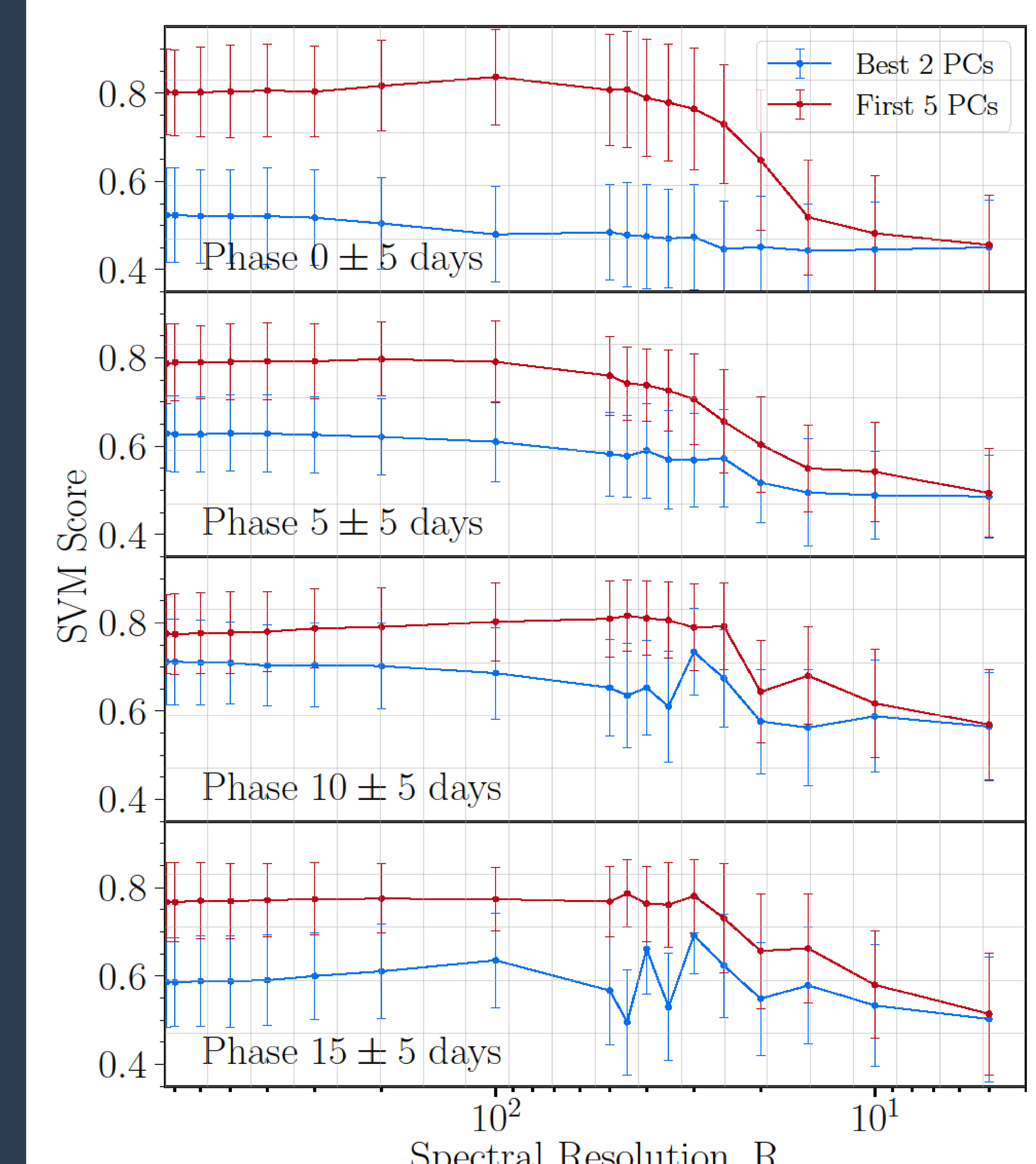

Figure 3: The SVM score for the classifier is plotted as a function of the spectral resolution, R, of the degraded spectra. Each panel represents supernova of epochs 0±5, 5±5, 10±5, and 15±5 days from top to bottom, respectively. The blue curves represent the SVM score when only two principal components were used. According Williamson (2019), the best two of the first five principal components were used as features in the classifier. The red curves represent the SVM score when all of the first five principal components are used in the classifier. Curiously, we are unable to find a minimum spectral resolution below which the SVM score is worse than chance, ~0.25.

Acknowledgements

With special thanks to my advisor Dr. Federica Bianco for her support, and to Dr. Veronique Petit for her expertise on spectroscopy.