Resampling the Spectral Time Series of a SN in the Time Dimension

Yuqian Liu, Department of Physics

1) The high-level problem

My project is to resample the spectral time series of a supernova in the time dimension using various methods of interpolation in machine learning. Core-Collapse Supernovae (CCSNe) are the explosions of massive stars at the end of their lives. CCSNe are intrinsically interesting not only because they are the most powerful explosions in the Universe, but also because they enriched the Universe with the oxygen we breathe, the calcium in our bones and the iron in our blood. Moreover, they are crucial tools for other areas of astronomy and astrophysics.

Stripped-envelope CCSNe come from stars whose outer layers of Hydrogen and Helium have been partly stripped or completely stripped. Following the empirical classification based on the presence or absence of certain lines in SN spectra, stripped CCSNe can be divided into several subtypes: Type IIb SNe (SNe IIb) which initially show strong H lines, but over time, the H lines become weaker whereas the He I lines grow stronger; Type Ib SNe (SNe Ib) which show conspicuous He I lines; Type Ic SNe (SNe Ic) which do not show prominent H lines nor He I lines; and broad-lined SNe Ic (SNe Ic-bl) which are similar to SNe Ic, but exhibit much broader lines.

One outstanding question in the field is to constrain the stellar systems that give rise to the different explosion types (which include but are not limited to those subtypes of stripped CCSNe mentioned above). The first step is to correctly classify different explosion types. For SNe, the current generation of surveys classify them using their spectra. However, the next generation of surveys cannot afford to take spectra of SNe but will take photometry of SNe instead and use the photometry to classify SNe. In photometric classification, a time series of average spectra of different SN subtypes are needed.

I have constructed such average spectra using a bin size of 5 days. Denser sampling is necessary but prohibited either because the spectra of SNe are sparsely sampled or because the lack of CCSNe. CCSNe are rare since there is roughly 1 CCSN per century per galaxy. Hence, interpolation in the time domain is often necessary.

I plan to apply different interpolation methods to the temporal spectra of a well-studied SN and find the method that works best based on cross-validation. If I have time, I'll apply the algorithm to other CCSNe and construct average spectra using a bin size of 1 day.

2) The data set

The data set consists of the spectral time series of SN 2008D, as shown in figure 1[1]. SN 2008D is a SN that was discovered in 2008 and is extremely well-studied.

The data set has three dimensions: time, spectral flux intensity, and wavelength, which are shown as the y axis, the color map, and the x axis in figure 1. For SN 2008D, we have 28 spectra (i.e. the spectral flux intensity as a function of wavelength over a certain range) sampled over 80 days. The observations (which are indicated via gray horizontal lines in figure 1) are irregular both in the time dimension and in the wavelength dimension. Especially, the observations are sparse at late times and at large wavelength ranges. Since a SN changes significantly during the first 1-2 months of the explosion and the sampling is denser during early times, I'm more interested in interpolating the first 24 spectra which are sampled over 50 days. Along the wavelength dimension, the observations are sampled roughly every 0.2 nm. Since observations at different epochs cover different wavelength ranges, each observation includes a different number of data points: from ~ 800 to ~ 2500 data points. Thus, the wavelength range over which to perform interpolation should be decided.

The noise in spectral flux density mainly consists of a constant gaussian noise (which is the readout noise of CCD) and a photon-count dependent poisson noise (which is the square root of the number of photons hitting the CCD). The poisson noise is approximated as a gaussian with a sigma given by the number of photons. The noise changes with the wavelength. For each spectrum, the noise is large at wavelength positions where non-SN emissions (e.g. HII region emission lines, etc.) happen. Moreover, the noise in edges of a spectrum is larger than in the center. Due to the low quantum efficiency of CCD around 400 nm (which is the small wavelength edge of our spectra), the small wavelength edge of a spectrum has a relatively high noise level. Compared to other wavelength ranges, the large wavelength edge of our spectra is more

[1] This figure was taken from Prof. Modjaz's proposal for Moore-Sloan Funding of Postdoctoral Researchers in Data Science

contaminated by sky lines, which results in a relatively high noise level at the large wavelength edge of a spectrum. However, the noise is not well known. Thus it is difficult to truly model the noise.

The data set has other characteristics: the spectral flux density should change continuously in both time dimension and in wavelength dimension; the low spectral flux density region (e.g. Helium absorption line feature) will evolve from small wavelength ranges to large wavelength ranges as time goes by, though the shift in wavelength is small so that it is not obvious in figure 1; there should not be periodic features along the time dimension; there are periodic features (e.g. Helium absorption line feature in figure 1) along the wavelength dimension.

3) How to evaluate the performance

Cross-validation will be used to evaluate the performance. Multiply rounds of cross-validation will be preformed using different partitions (or subsets of observed spectra in this case) and the validation results will be averaged over the rounds.

4) A "baseline algorithm"

A "baseline algorithm" is the simple linear interpolation between two observed spectra, without considering the noise in spectra.

5) Method and timeline

I'll mainly use Gaussian Process (GP) regression to fill the gap in the time dimension of the spectral time series of SN 2008D. I'll perform the interpolation over different wavelength ranges. I'll also take the noise in spectra into consideration.

I'll try different kernel functions and correlation lengths in GP regression from week 11 to week 13. I'll try other interpolation methods such as Poisson regression during week 14. I'll prepare the poster for my project and present the poster during week 15.

References

[1] Modjaz, M., et al. 2009, Astrophysical Journal, 702, 226

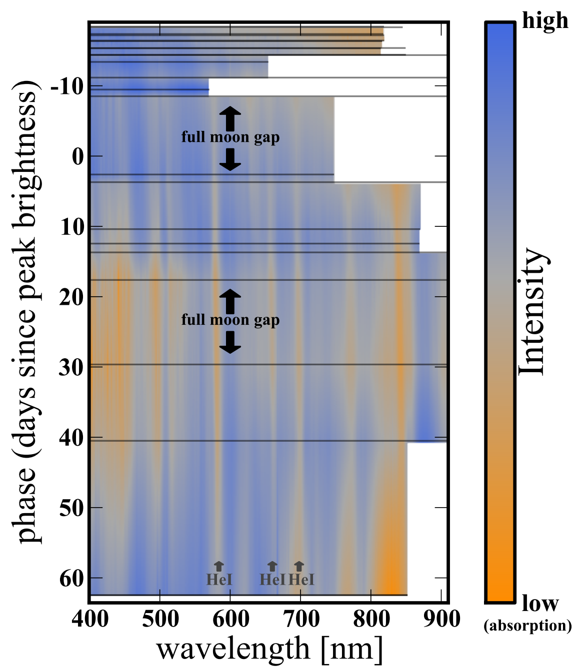

Figure 1: Spectral time series for SN 2008D, one of the best-studied Type Ib CC SNe (Modjaz et al. 2009). Gray horizontal lines indicate the epochs at which spectra were collected. White regions indicate missing coverage. Time and wavelength are naively interpolated linearly between observations, for illustrative purposes; however, a robust treatment is needed. Helium absorption-line features are indicated by arrows. Note that this figure was taken from Prof. Modjaz's proposal for Moore-Sloan Funding of Postdoctoral Researchers in Data Science.