

Sloan Digital Sky Survey Analysis and Data Engineering

Joeri Hermans
joeri.hermans@doct.ulg.ac.be

November 5, 2017

1 Introduction

The Sloan Digital Sky Survey¹ or SDSS, is a major multi-spectral imaging and spectroscopic redshift survey using a dedicated 2.5 meter wide-angle optical telescope at Apache Point Observatory in New Mexico, United States. The Sloan Digital Sky Survey has created the most detailed three-dimensional maps of the Universe ever made, with deep multi-color images of one third of the sky, and spectra for more than three million astronomical objects. The SDSS began regular survey operations in 2000, after a decade of design and construction. It has progressed through several phases, SDSS-I (2000-2005), SDSS-II (2005-2008), SDSS-III (2008-2014), and SDSS-IV (2014-). Each of these phases has involved multiple surveys with interlocking science goals. The three surveys that comprise SDSS-IV are eBOSS, APOGEE-2, and MaNGA.

In this project, you will be working with data from the eBOSS (Extended Baryon Oscillation Spectroscopic Survey) experiment. eBOSS precisely measures the expansion history of the Universe throughout eighty percent of cosmic history, back to when the Universe was less than three billion years old, and improve constraints on the nature of dark energy. “Dark energy” refers to the observed phenomenon that the expansion of the Universe is currently accelerating, which is one of the most mysterious experimental results in modern physics.

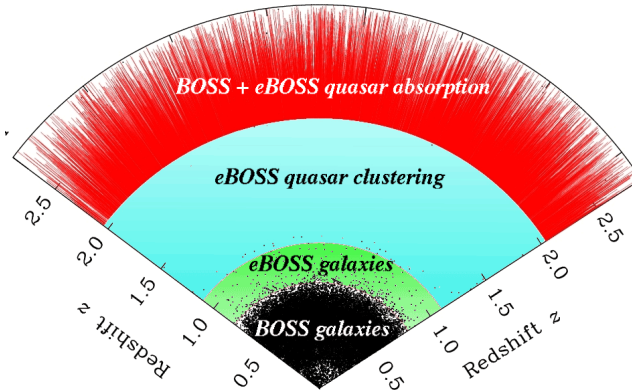


Figure 1: Convergence of the eBOSS experiment up to redshift (z) = 3.

In physical terms, redshift, described in Equation 1, happens when light or other electromagnetic radiation from an object is increased in wavelength, or shifted to the red end of the spectrum. In general, whether or not the radiation is within the visible spectrum, “redder” means an increase in wavelength – equivalent to a lower frequency and a lower photon energy, in accordance with, respectively, the wave and quantum theories of light. Some redshifts are an example of the Doppler effect, familiar in the change of apparent pitches of sirens and frequency of the sound waves emitted by speeding vehicles. A redshift occurs whenever a light source moves away from an observer.

$$z = \left(\frac{\lambda_{\text{observed}}}{\lambda_{\text{rest}}} \right) - 1 \quad (1)$$

In astronomy, redshift can be utilized to measure the *accelerating* expansion of the universe. This is exactly one of the key questions posed by the eBOSS survey. In principle, eBOSS measures this by identifying the wavelengths of emission and absorption lines, and then comparing with the known spectra in a vacuum for those elements and thereby obtaining an average redshift for all spectra using Equation 1.

¹<https://www.sdss.org>

Furthermore, these emission lines and their accompanying fluxes (number of photons that caused a certain amount of electrons to move) can be utilized whether the observed instance is a *star* (subclasses of the spectral types can also be identified²), *galaxy* or a *quasar*, as shown in Figure 2 and Figure 3.

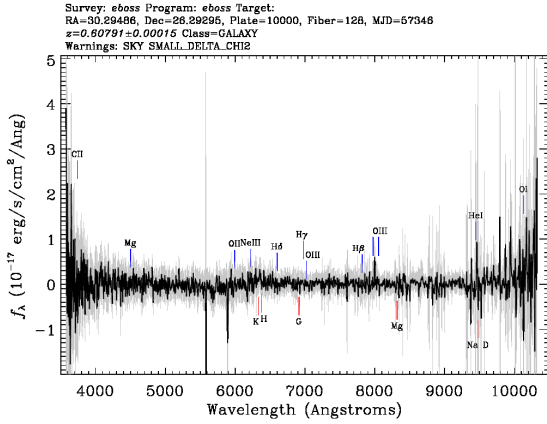


Figure 2: Spectra with accompanying fluxes of a galaxy with identified absorption lines.

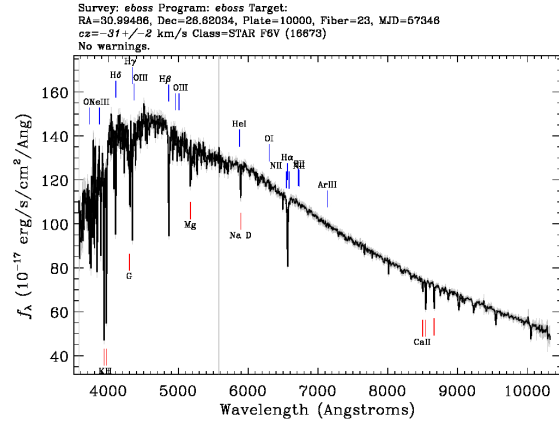


Figure 3: Spectra with accompanying fluxes of a star with identified absorption lines.

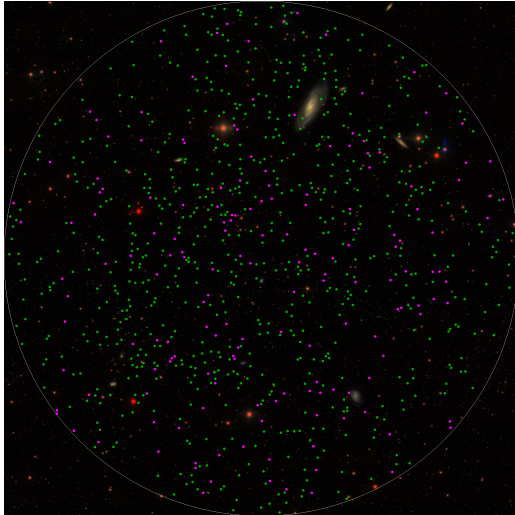


Figure 4: SDSS plate and target selection (green for galaxies and purple for quasars).

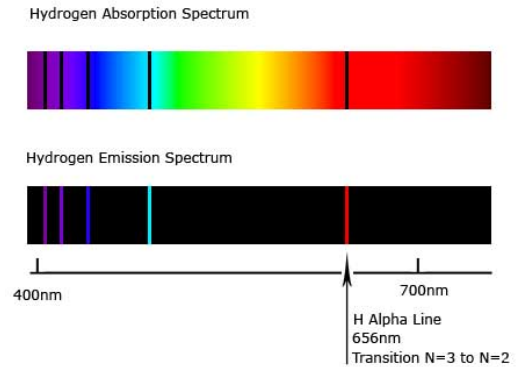


Figure 5: Balmer Series of the Hydrogen. Shows absorption and emission lines from orbital 2 to 3 and higher.

The data collection procedure is done by drilling several holes in an aluminum plate which are positioned in a particular way to obtain individual fluxes of target objects, as shown in Figure 4. This approach the eBOSS survey to observe multiple targets at once by attaching an optical *fiber* to every drilled hole, causing the individual observations to be independent of each other. Furthermore, this approach allows SDSS to filter out nearby galaxies which are known to be *blueshifted*, and obtain the individual background flux for every observed object.

2 Deliverables

To complete this project successfully.

3 Bonus

TO BE CONFIRMED BY GILLES

²en.wikipedia.org/wiki/Stellar_classification#Spectral_types

A bonus point (1/20) is awarded to the group which is able to provide the fastest average random query time over the complete eBOSS dataset (171 GB). The query that will be ran to evaluate the performance of your data architecture, will be concerned with selecting objects in a particular region in the sky for a particular redshift. Furthermore, we also might be interested in only selecting particular objects, i.e., *stars*, *galaxies*, and *quasars*. Your result might have practical implications as well, since the current SDSS query tool is *very* slow.