Using Machine Learning to Understand Cosmic Gas Evolution

Rogelio Ochoa
Bryson Stemock
Alexander Stone-Martinez
rochoa9@nmsu.edu
bstemock@nmsu.edu
stonemaa@nmsu.edu
New Mexico State University Dept. of Astronomy
Las Cruces, New Mexico

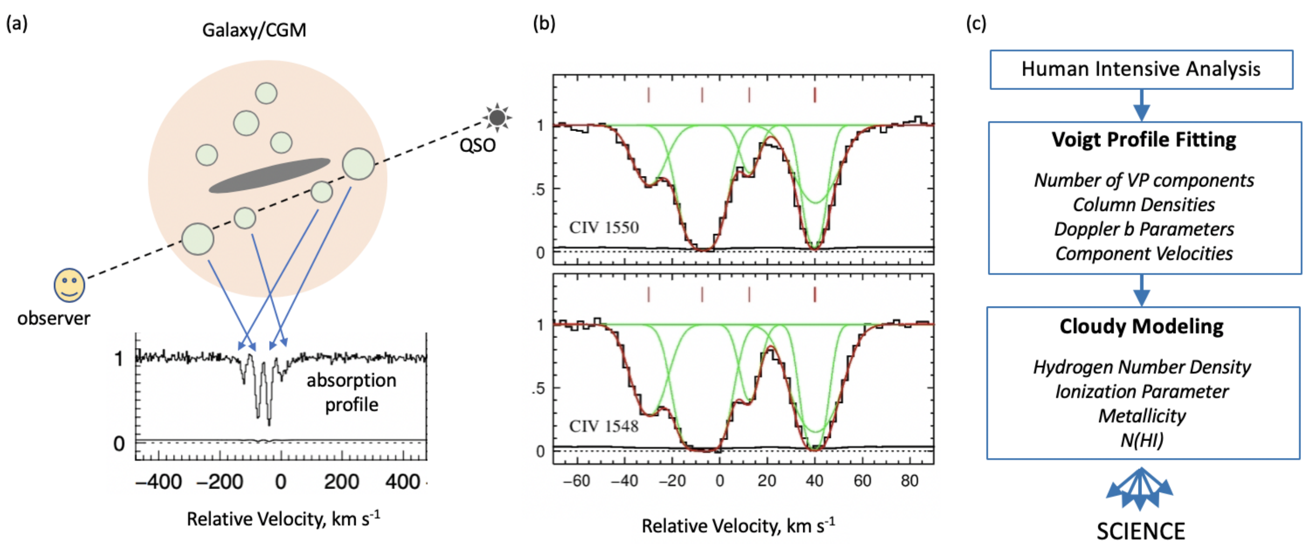

Figure 1: (a) Schematic of a quasar line of sight (dotted line) showing absorption from four distinct CGM clouds. The conventional interpretation: each absorption component corresponds to a cloud. (b) The Voigt profile (VP) model (red curve) of observed CIV $\lambda\lambda 1548, 1550$ doublet absorption profiles. Each VP component is shown as a green curves. (c) An analysis-to-science flowchart: VP fitting provides column densities that constrain ionization models, which yield physical quantities (in italics) for the science.

ABSTRACT

This project represents a small step in Bryson Stemock's Ph.D. thesis, to which Alexander Stone-Martinez and Rogelio Ochoa have contributed a significant amount already. The project focuses on an intermediate step in the thesis, the Voigt profile fitting of absorption line spectra. Synthetic spectra (i.e. the training and test sets) have been simulated already by Stemock. The goal of the project is to accurately predict physical parameters associated with absorption line spectra containing one or two clouds.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS CONCEPTS

• **Computing methodologies → Machine learning**.

KEYWORDS

datasets, neural networks

ACM Reference Format:
Rogelio Ochoa, Bryson Stemock, and Alexander Stone-Martinez. 2020. Using Machine Learning to Understand Cosmic Gas Evolution. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 MOTIVATION

Since 1962, quasars (highly luminous accreting supermassive black holes in distant galaxies) have been used to probe the cosmos [1].

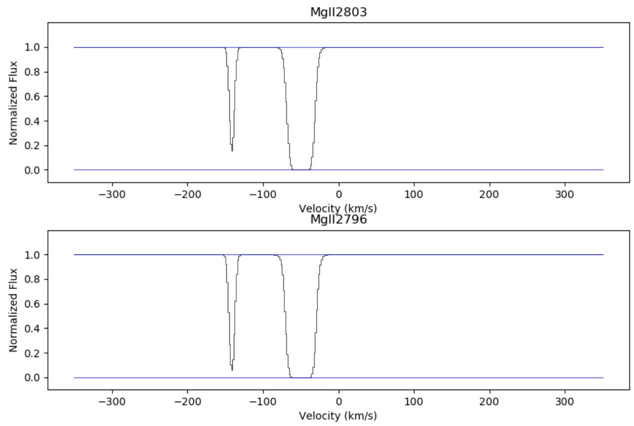

Figure 2: A sample two-cloud absorption line system. The only ions present here are from the MgII doublet at $\lambda\lambda 2796, 2803$**. Note that no noise is present in the data for this project.**

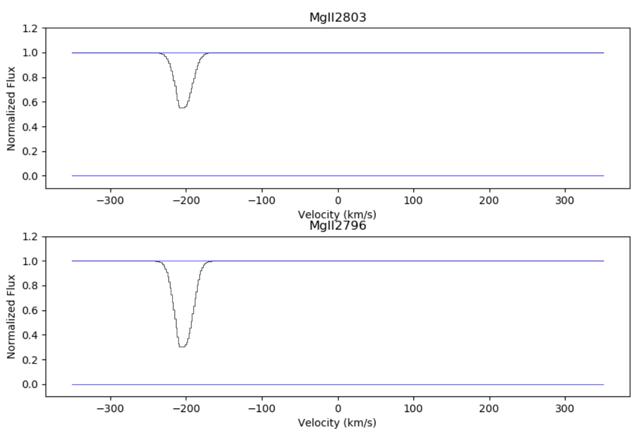

Figure 3: A sample blended two-cloud absorption line system. Note that this system contains TWO clouds on top of each other. This occurs when the clouds are very close together in velocity-space.

When light from a quasar passes through a gas structure, that gas leaves an imprint on the spectrum of the quasar. This imprint is called an absorption line system. Detailed analysis of these absorption line systems across cosmic time provide direct insight into the physics that governs the universe and how both its large-scale structure and the galaxies that inhabit it have evolved.

In order to extract gas properties from absorption line systems, one traditionally fits the data with Voigt profiles and then uses a Markov chain Monte Carlo (MCMC) method to model the gas. This process is illustrated in Figure 1 (borrowed from our proposal to the National Science Foundation). During the process, simplifying assumptions and human subjectivity creep in, introducing some level of bias into this line of work. In addition, the typical rate at which a given researcher is able to complete this process for a single absorption line system is around 1-2 systems per week. At this rate, it is estimated that our current sample of approximately 3,500 systems would require around **70 years** of human effort.

One group member, Bryson Stemock, has taken on as his Ph.D. thesis the task of automating this process through the design, training, and implementationg of a convolutional neural network (CNN). As part of a larger group within the NMSU Astronomy Department, the remaining group members, Alexander Stone-Martinez and Rogelio Ochoa, have contributed greatly to the progress of the overall project, primarily through the design of various CNNs. The overall project goal is to create a globally-available tool that will vastly accelerate the analysis of absorption line systems and, therefore, of the evolution of the universe.

2 THE PROBLEM

Clearly, the entirety of this project is too much to accomplish in a single semester, especially with a variety of data-related subtleties that were not mentioned above which arise from the complexity of the gas we observe (and simulate). Therefore, only a minute, refined portion of the overall thesis will be tackled here. The data, which will be explored more thoroughly in Section 4 of this report, consist of two-cloud absorption line systems with parameters (which were used to generate the spectra) drawn using Latin Hypercube Sampling. Each instance consists of two spectra, displayed in Figure 2, one for the MgII2796 transition and one for the MgII2803 transition. Our goal is to design a CNN that can accurately (better than 0.93 accuracy) predict the log of column density, the Doppler b parameter, and the velocity position of each cloud.

One obstacle that will need to be addressed is the occurrence of "blending". Blending is the overlapping of absorption lines and usually refers to an extraneous absorption line impeding on a separate system. However in our case, since the CNN will be searching for two lines and may see one large line (example in Figure 3), we may run into an issue.

REFERENCES

[1] Maarten Schmidt. [n.d.]. Spectrum of a Stellar Object Identified with the Radio Source 3c 286. *ApJ* ([n. d.]).