BARYONS IN THE INTERGALACTIC MEDIUM: A HUBBLE SPECTROSCOPIC LEGACY ARCHIVE SEARCH

by

TREVOR BRUNNENMEYER

A DISSERTATION

Presented to the Department of Physics
and the Division of Graduate Studies of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2022

DISSERTATION APPROVAL PAGE

Student: Trevor Brunnenmeyer

Title: Baryons in the Intergalactic Medium: A Hubble Spectroscopic Legacy Archive Search

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics by:

Dietrich Belitz	Chairperson
Greg Bothun	Advisor
Spencer Chang	Core Member
Darren Johnson	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies.

Degree awarded June 2022

DISSERTATION ABSTRACT

Trevor Brunnenmeyer

Doctor of Philosophy

Department of Physics

June 2022

Title: Baryons in the Intergalactic Medium: A Hubble Spectroscopic Legacy Archive Search

Galaxies are not closed systems but are constantly interacting with their environment which generally contains a) other nearby galaxies and b) some form of warm gaseous medium existing between galaxies (known as the Intergalactic Medium or IGM). As a result, a number of physical mechanisms exist to liberate baryons (stars and gas) from within a galaxy and transfer them to the IGM. Currently there is a well-documented "missing baryon" problem in Cosmology as known galaxies do not contain enough baryons to be consistent with Big Bang Cosmology. This implies a substantial amount of Baryons must exist in the IGM, but most of this remains undetected. Using absorption in Quasar (QSO) spectra from the Hubble Spectroscopic Legacy Archive (HSLA) we can detect individual species present in the IGM.

In this work, I build up and refine a pipeline to analyze 688 QSO spectra from the HSLA. At the end of the pipeline, there are multiple possible species identifications for each absorption feature. Looking at models for the extremes of the possible absorptions puts bounds that between 1.6% and 48% of the total baryon density from Big Bang Cosmology can be found in extragalactic clouds. These results do not solve the Missing Baryon Problem, but do provide evidence that baryons liberated from galaxies could be part of the complete solution.

This dissertation includes previously unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Trevor Brunnenmeyer

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Utah, Salt Lake City

DEGREES AWARDED:

Doctor of Philosophy, Physics, 2022, University of Oregon
Master of Science, Physics, 2017, University of Oregon
Bachelor of Science, Applied Mathematics, 2015, University of Utah
Bachelor of Science, Physics, 2015, University of Utah

AREAS OF SPECIAL INTEREST:

Astrophysics Data Science
Physics Education Research

PROFESSIONAL EXPERIENCE:

Graduate Teaching Assistant, University of Oregon, 2015-2022

PHYS 201 Instructor, University of Oregon, Fall 2019

Science Literacy Program Fellow, University of Oregon, Fall 2018

GRANTS, AWARDS, AND HONORS:

Graduate Teaching Award, University of Oregon Department of Physics, AY 2018/19

Graduate Teaching Award, University of Oregon Department of Physics, AY 2017/18

ACKNOWLEDGMENTS

There are more people that have helped me get to the point of finishing this dissertation than there is room to list. This work would not have been possible if not for the support of the many teachers, coaches, and mentors I've had throughout my life.

I would like to thank my advisor, Greg Bothun, for his support and guidance throughout this project. I would like to thank my committee for their support in writing this dissertation.

I would like to thank my parents, Sandy and Kim Brunnenmeyer, and my sister, Tessa Brunnenmeyer, for their support throughout my whole life and for being a sounding board when I was stuck on something in this project.

I would also like to thank the friends I made in the physics department and across campus. In particular, my cohort members Amanda Steinhebel, Kris Schobert, Mark McKnight, Alice Greenberg, Tzula Propp, Conor Rowland, and Sofiane Merkouche who have provided both personal and academic support throughout my entire time in the program. I'd like to thank Ellen Olsen for sharing cat photos, cider, and commiserations of grad school that were needed throughout the last few years.

I would also like to thank my comrades in the Graduate Teaching Fellows Federation (GTFF 3544). Through our union I have forged amazing friendships and become a better person. In particular, I would like to thank the members of the 2018-2019 GTFF bargaining team: Mike Magee, Ellen Kress, Youssef Benasser, Curtis Dlouhy, Miche Dreiling, Michael Hurst, Alexis Kiessling, Andy Labuza, Pearl Lee, Alberto Lioy, Rita Ludwig, Adam Morse, Juan Ospina, Hannah Pell, Larissa Petrucci,

Kyle Raze, Amy Robbins, Morgan Sosa, and Sarah Stach; the long hours of debates we had have changed me for the better.

Much of my time as a graduate student has been spent learning and practicing teaching. The developments that I have made in becoming a better teacher would not have been possible without Elly Vandergrift, Julie Muller, and the everyone else from SLP/TEP. Additionally, I have gained valuable experience working with the PHYS 20x teaching team, in particular Billy Scannell and Stan Micklavzina.

This research has made use of the HSLA database, developed and maintained at STScI, Baltimore, USA.

This research has made use of the NASA/IPAC Extragalactic Database (NED), which is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology.

I acknowledge the use of the Legacy Archive for Microwave Background Data Analysis (LAMBDA), part of the High Energy Astrophysics Science Archive Center (HEASARC). HEASARC/LAMBDA is a service of the Astrophysics Science Division at the NASA Goddard Space Flight Center.

For Codex and Penny

This dissertation is the product of unionized labor as part of the Graduate Teaching Fellows Federation, AFT Local 3544.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Approximately one third of this chapter was written by Greg Bothun to round out the history and interest of the problem. I prepared the original draft and did the final editing.

1.1 Background

About 300,000 thousand years after the Big Bang, recombination occurred and the Universe became matter dominated. At this point, radiation pressure was removed from the matter field allowing for any overdense concentration of matter (could be mostly dark matter) that managed to survive being entirely smoothed out by radiation pressure, to begin to gravitationally clump. One potentially important aspect is that dark matter doesn't interact electromagnetically with photons and so likely coalescing clumps are mostly dark matter. How baryons "creep" into these dark matter clumps remains mysterious, and a variety of mechanisms have been proposed (Bryan et al. 2012; Kochanek and White 2001; Pedrosa, Tissera, and Scannapieco 2010). Of key concern to this dissertation is the idea of whether or not all baryons get incorporated into these dark matter potentials which will eventually form galaxies. Two observational clues, however, shed some light on the possible complexity of galaxy formation:

a) Galaxies are observed to have a very wide range of masses (approximately 6 orders of magnitude) and thus there is a galaxy mass spectrum. The conditions that determine this large range of masses are largely unknown, but this range suggests there were likely multiple ways in which galaxy masses could have been assembled.

b) It is now well known that galaxies are very highly clustered. Thus, in the proto-galaxy state of galaxy assembly, there are likely lots of tidal interactions (and some direct mergers) of the various sub-clumps. Qualitatively, such tidal interactions can remove baryonic matter (mostly gas) from forming galaxies. Depending on the depth of the gravitational potential at the time, this liberated debris could fall back into the galaxy at later times (and there is observational evidence for this (Mihos 2004)) or be liberated altogether to form some kind of intergalactic medium (IGM). This diffuse IGM could then be another repository for baryons in addition to those baryons now contained in galactic potential. A question probed by this work is whether or not more baryons exist in the IGM than exist in galaxies.

We can measure the number of baryons that existed in the early universe through looking at the cosmic microwave background (CMB)(Hinshaw et al. 2012). As the universe expanded, the density of baryons dropped, since the total number is constant but the volume they exist in increases. Tracking the density forward to the present, we can calculate the expected density of baryons in the modern universe. In addition, the rate of big bang nucleosynthesis (BBN) is dependent on the overall baryon dependence (particularly for the case of deuterium). Models coupled with observations of the primordial abundance of Helium and Lithium can yield approximate baryon densities. While there is good overall agreement, the abundance of Lithium (Fields 2011) remains a thorn in the side of any closed argument that this is now well understood, however, there is recent observational evidence that this problem has an astrophysical resolution (Fields and Olive 2022). When you add up all the baryons that are in galaxies,

in the form of stars, gas, and dust, as well as those inferred by various BBN observations, the accounting comes up short of the expected number by 30-50% (Gregory D Bothun 2003; Cen and Ostriker 1998; Fukugita 2003; J. Michael Shull, Smith, and Danforth 2012).

This is known as the Missing Baryon Problem. There are a few possible ways to resolve this problem:

a) Our accounting of galaxies may be incomplete and not all galaxies might be detectable with current instrumentation. One example of class of such objects were originally termed Low Surface Brightness (LSB) galaxies (O'Neil and Bothun 2000) and that terminology has now morphed to become Ultra Diffuse Galaxies (Piña et al. 2019).
b) The BBN and CMB models are simply incomplete or inaccurate.
c) There really is a population of IGM baryons. Such a population would be very difficult to detect. Indeed, the only real all sky method of detection would be to use background quasi-stellar-objects (QSOs) as point sources shining through intervening material to produce absorption lines where there is now visible galaxy. That is the method used in this dissertation.

If there is indeed a population of IGM baryons, the question remains if their origin is associated with gas liberated during the galaxy formation process or baryons liberated after galaxies formed as a result of continuing interaction in clusters and groups. There is plenty of evidence of liberated debris in interacting galaxies, but we now know that many galaxies, including the Milky Way have significant amounts of infalling material (Gregory D. Bothun et al. 1992; dell'Antonio, Geller, and Bothun 1996; Rocha, Peter,

and Bullock 2011) so it is not clear that baryons removed in this manner become freely floating in the IGM. Indeed, a rather dramatic example of baryon liberation in the galaxy group environment was discovered by White et al. (White et al. 2003), in which the system HCG 90 has more than 50% of its total baryonic light no longer in galaxies as shown in Figure 1.

Figure 1: HCG 90. An envelope of diffuse light surrounds the system of galaxies. Image taken from White and Bothun (2003)

1.2 Getting Baryons Out of Galaxies

There are a couple different ways for baryons to be liberated from a galaxy. Internally, supernova explosions can fountain material out of the galaxy's gravitational

well and into the IGM, but this mechanism is highly dependent on the depth of gravitational potential. Supernova ejection of baryonic material is likely only possible for systems with velocity dispersions of < 100 km/s. A typical galaxy, like the Milky Way, has a velocity dispersion scale (velocity dispersion maps directly onto the dynamical mass of a galaxy) of 250-300 km/s. Additionally, the observed rates of supernovae are not high enough to explain the 30% of missing material on the scale of the universe. Individual galaxies that have high supernova rates (like NGC 5253, NGC 6947) may indeed have a fair amount of liberated debris, but these objects have supernova rates 10-100 times the average rate.

For galaxies located in clusters of galaxies that have a hot intracluster medium (ICM), gas and dust can be stripped through ram pressure, where the galaxy collides with a cloud of gas in the IGM. The collision can liberate the material in the galaxy and directly inject that into the ICM, thus increasing the mass of the ICM while decreasing the mass of individual galaxies. This leads to an expectation that the abundance of external baryons may be higher in clusters of galaxies than it is in more average locations in the Universe.

Material can also be lost through gravitational interactions with other galaxies. When two galaxies interact, they form long tidal tails that can extend 10's of galaxy radii. In the field, these tidal tails eventually fall back into the host galaxies if that system is relatively isolated. In the case of galaxies in a cluster, further interactions can occur and thus there can be multiple gravitational interactions and scatter of material that serve to increase their velocities which then allows the tidal tails to completely escape the gravitational potential of the host galaxies.

1.3 Tidal Interactions to Strip Material

When galaxies interact, the tidal forces can strip large amounts of material far away from the center of the gravitational potential. Most of this material will be gas; stars are much harder to strip out of galaxies. Essentially a tidal interaction with differential gravitational forces acts like a dynamical fluid. Stars are hard bodies which are not well coupled to the gas, which acts like a fluid. The amount of ionized gas also matters in principle, but most interactions involve removing material from the outer edges of galaxies where ionization is low or non-existent. When galaxies merge in the field, most of the material is not stripped in the merger (<20%), whereas in a cluster 40% or more of the material can be stripped (Mihos 2004). From simulation and observation, it is likely that almost all galaxies have had several tidal interactions in their lifetimes, with galaxies in groups and clusters having more frequent interactions than galaxies in the field (Fried 1988). With such a large fraction of the material being removed in mergers, coupled with mergers being common, it is reasonable to expect that there exists a large number of baryons that don't reside in the gravitational potential of galaxies.

In particular, there have been observed low surface brightness features in Coma that are thought to be the result of tidal interactions between galaxies in the cluster. There is evidence that about 20% of the luminosity of one of the large galaxies at the center of Coma has been removed, assuming that the current removal rate has stayed constant, though it was likely higher in the past (Gregg and West 1998). As this material is removed, it becomes part of the diffuse, intergalactic medium. More recently, an analysis of the intracluster light in Coma has found a velocity dispersion consistent with the interpretation that "stars in these regions are tracing the gravitational potential of the

cluster, instead of any individual galaxy" (Gu et al. 2020). Another study found 22,426 globular clusters in Coma's intracluster medium (Madrid et al. 2018). These globular clusters are concentrated near larger galaxies, but there is still a low density distributed throughout the interior of the Coma cluster. In the Virgo cluster, there is evidence of intracluster stars that are tidally stripped from galaxies (Durrell et al. 2002).

These observations give good evidence that a significant fraction of a galaxy's matter can be stripped into the intergalactic medium and remain there for a long time. These studies however are only sensitive to the light emitted by stars, so they miss any gas or dust that is too diffuse to emit enough light to be measurable.

1.4 Techniques to Observe the IGM

Since the clouds of gas that we'd like to detect are too diffuse to direct with emitted light, we need techniques that absorb or scatter background light.

The Sunyaev-Zel'dovich effect is the name given to inverse Compton scattering of photons off hot, free electrons (Birkinshaw 1999). In particular, when a CMB photon passes through a hot plasma, the photon can scatter off an electron and gain energy, resulting in it having a higher frequency. If there is no peculiar motion of the cloud of plasma, the scattering is referred to as the thermal Sunyaev-Zel'dovich effect. If there is peculiar motion of the cloud, there is an additional Doppler shift referred to as the kinetic Sunyaev-Zel'dovich effect. There have been several studies that stack hundreds of thousands of pairs of galaxies and measure the cosmic microwave background radiation, looking for evidence of photons that are scattered due to both the thermal and kinetic Sunyaev-Zel'dovich effects (de Graaff et al., 2019; Lim et al., 2020; Tanimura et al., 2019). These studies have found evidence of large-scale filaments between galaxies.

However, since this technique relies on scattering of photons off of free electrons, it is ambivalent to the nuclear composition of the intervening material. In addition, it is only the hot X-ray cluster environment that can produce sufficiently large numbers of free electrons with significant energy to enable this effect.

The most direct way to observe baryonic content in the IGM would be through detectable absorption lines produced as a background light source shines through intervening material. Since most IGM content should be simply clouds of hydrogen then the simple expectation is the detection of hydrogen absorption lines redshifted to the cosmological epoch of the cloud's existence. Fortunately, nature has provided us with a natural background light source.

Quasars (Quasi Stellar Objects or QSOs) are the active nuclei of distant galaxies that effectively become point sources of light. As the light travels from the Quasar to our telescopes, light is absorbed by intervening material. A typical QSO sight line might experience absorption at several wavelengths due to varying redshifted IGM components along that line of sight. The higher the redshift of the QSO the more sightlines as a function of redshift become available. This situation is schematically shown in Figure 2 where the Hubble Space Telescope spectroscopically observes a distant quasar with hypothetical intervening IGM material along that sightline (the name IGM was later changed to the Cosmic Web for visibility and importance as a cosmological tool (J. M. Shull et al. 1999)).

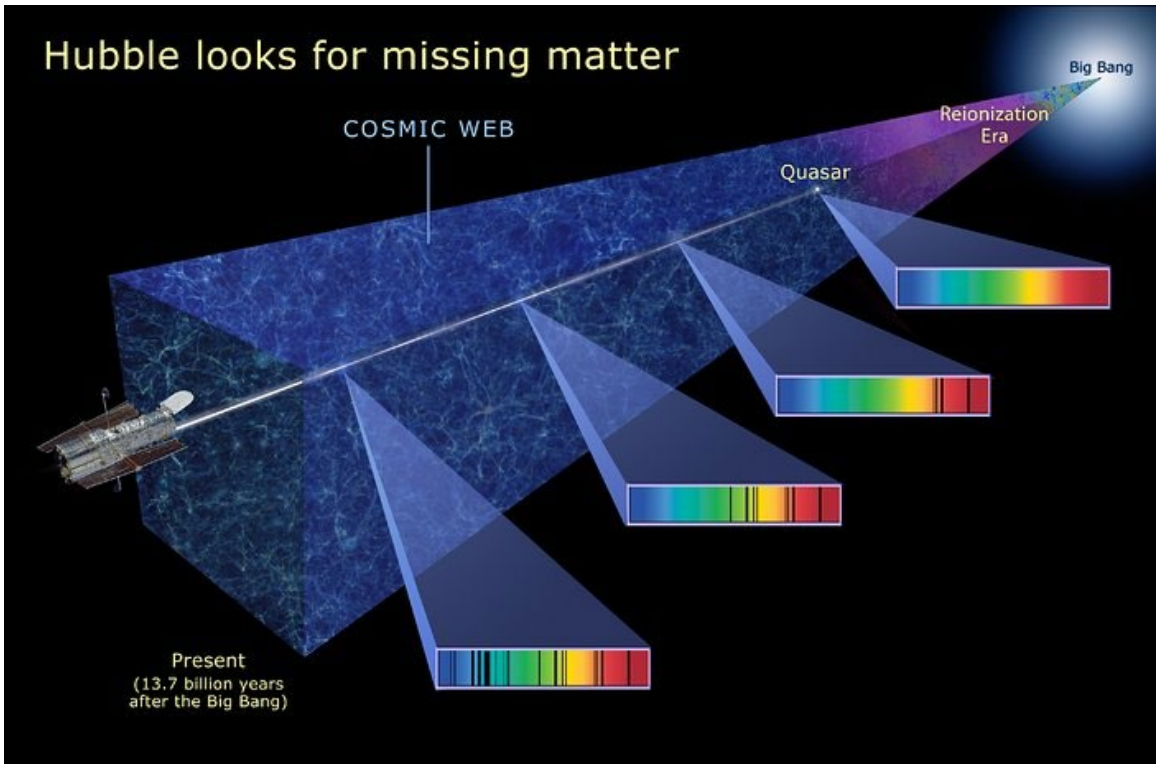

Figure 2: Cartoon showing absorption lines in a QSO spectrum caused by clouds at different points in the cosmic web. Image taken from NASA, ESA, and A. Feild (STScI)

While most of those absorption lines shown above may well be due to redshifted hydrogen clouds, there can be other chemical lines present if those clouds contain any processed material. The actual amount of elements in QSO absorption lines that are not hydrogen can also become an important cosmological probe, particularly if a redshift dependence on heavy element content can be shown. However, this is difficult and complicated as the ability to produce an absorption line depends both on the ionization state of the cloud (ionization can occur from the general UV background in the vicinity of the cloud - indeed much of that UV background comes from the combined continuum emission of distant QSOs) and on the total column density of the cloud. Thus, there can be heavy elements in some clouds which don't have the right physical property to produce the required chemical transition to yield a detectable absorption line.

Since the light along the sightline is redshifted by the expansion of the universe, by identifying the wavelength of the absorption and the element that produced it, we can track where in the universe that cloud of gas was. The sightlines to QSOs typically travel through regions of space that allow them to have 50-100 absorption lines. A typical QSO spectra that has been redshifted into the visible/infrared is shown in Figure 3.

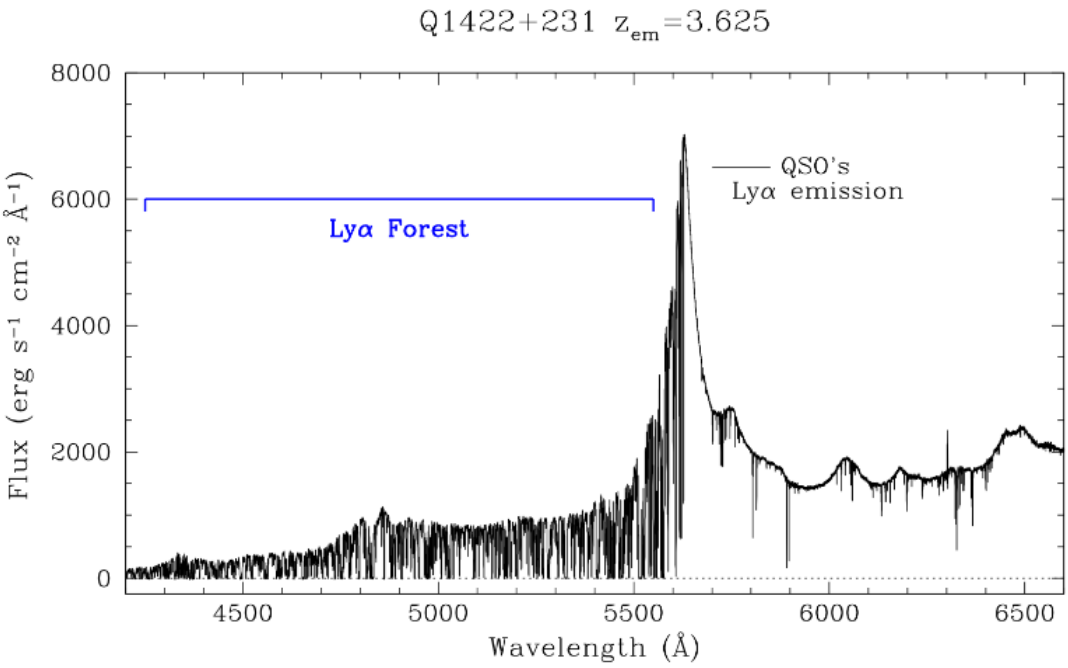

Figure 3 An example spectrum for a distant QSO. Figure from (Ellison 2000)

These absorptions happen in 3 distinct regions, near the QSO, in the Milky Way, and in intervening material. The crowded absorption lines near the redshift of the QSO is known as the Lya Forest. These lines are due to redshifted absorption from the Lyman alpha transitions at 1215 angstroms due to various intergalactic neutral hydrogen (H I) clouds in the general vicinity of the QSO itself. When searching for baryons in the IGM, we only care about the intervening material and not the material in the Milky Way or near the QSO. For the spectrum above, those IGM lines are the ones to the right (longer

wavelength) of the intrinsic Lyman alpha emission of the QSO itself. In this spectrum, there are some obvious strong absorption lines and quite a few weaker lines. The challenge is now line identification. For example, are all of the absorption lines due to only hydrogen clouds that appear at many redshifts along the line of sight or do some lines come from other elements present at other redshifts? To be more specific, suppose there is an easily detectable line at an observed wavelength of 6000 angstroms. Absorption caused by the H-beta transition is common in QSO spectra. H-beta is intrinsically at 4861 angstroms and so a redshift (1+z) hydrogen cloud could produce a line at 6000 angstroms for z = .234. But Silicon II is often detected in QSO spectra and SI II has an intrinsic transition at 2072 Å implying a redshift of z = 1.89. Hence in any QSO absorption line spectrum there is the potential for much redundancy which has to be sorted out and this process is the key piece of this work, because in order to trace out the history of where the material is from, we need to know if it has been processed or not. More than that, we need correct identification of as many non-hydrogen absorption features as possible.

When the light from a distant QSO reaches earth, it needs to be separated into a spectrum in order to see what has been absorbed. Many of the transitions that are useful to trace out ions are in the ultraviolet (UV). Since Earth's upper atmosphere absorbs almost all light in the UV, ground-based telescopes are not able to be used. In order to do this spectroscopy, we must use a space-based telescope that will not have local absorption in the UV.

The Cosmic Origins Spectrograph (COS) that was installed on the Hubble Space Telescope (HST) in 2009 is a high-resolution UV spectrograph in space. This satisfies all

that is needed to be able to study QSO absorption in the IGM. COS has 2 channels in the UV, one in the Near UV that covers 1700-3200 Å and one in the Far UV that covers 900-2050 Å. With a spectral resolution of 2,000-3,000 at the edges of the spectrum and 16,000- 24,000 at the center of the spectrum, COS is an incredibly powerful instrument (James 2022). It is able to get the sub 0.1 Å resolution needed to probe the forest of narrow lines left by intergalactic gas.

All spectra taken by COS is made public in the Hubble Spectroscopic Legacy Archive (HSLA) (Peeples et al. 2017). We use all of the QSO, Active Galactic Nuclei, and Seyfert galaxy spectra in this work. There are around 800 sources that have spectra in the HSLA covering most of the sky, though there is a concentration of sources in the Northern Hemisphere. (This is where people have historically done more observations, not that there are intrinsically more sources in the Northern Hemisphere.) The large number of sightlines and their wide distribution allows us to interrogate large volumes of the universe.

1.5 The Coma Cluster

The initial target for investigation was the Coma cluster. The relatively nearby Coma cluster is a collection of around 1,000 Milky way mass galaxies at a redshift of 0.0231 (about 336 million light-years away). This cluster is the most nearby rich cluster of galaxies and is surrounded by an extensive supercluster which contains a large population of infalling galaxies (Gregory D. Bothun et al. 1992). The central regions of the cluster host a hot X-ray medium which is capable of providing the kind of ram pressure stripping mentioned above. The ancient observations of Bothun & Dressler 1986 (G. D. Bothun and Dressler 1986) were among the first to identify real examples of ram

pressure in action on individual spiral galaxies. Over the combined cluster plus supercluster area there is approximately a 4 orders of magnitude range in local galaxy density. The baryonic removal mechanisms previously discussed mostly likely depend on local galaxy density and a thorough survey of all of this environment may reveal that the number of QSO absorption line features at the Coma redshift may reveal this dependency on local galaxy density. There is abundant evidence already (Adami et al. 2005; Goldsmith and Silk 1972; Kim et al. 1989; Matsushita et al. 2013; Peng et al. 2011) for features in this overall structure consistent with dynamical processes that produce an intergalactic population of baryons. Example evidence, from the deep CCD imaging of Coma (Gregg and West 1998), is shown in Figure 4 where the diffuse collection of light in the center of the image likely represents stars that have been removed from the surrounding galaxies. The overall distribution of local galaxy density in the overall Coma supercluster is shown in Figure 5 (Seth and Raychaudhury 2020).

Overall, this environment is unique among nearby ($z<0.2$) galaxy cluster environments and therefore represents an excellent location to start looking for QSO absorption by the intergalactic medium. The linear feature of galaxy density roughly running from the top left to the bottom right in Figure 5 is known as the Great Wall (dell'Antonio, Geller, and Bothun 1996) and represents one the largest filamentary structures in the known Universe. The gravitationally bound nature of the two main clusters likely points to the existence of dark matter as the binding mass as first pointed out by Zwicky in 1933 (see (Andernach and Zwicky 2017) for a recent translation of the article originally written in German).

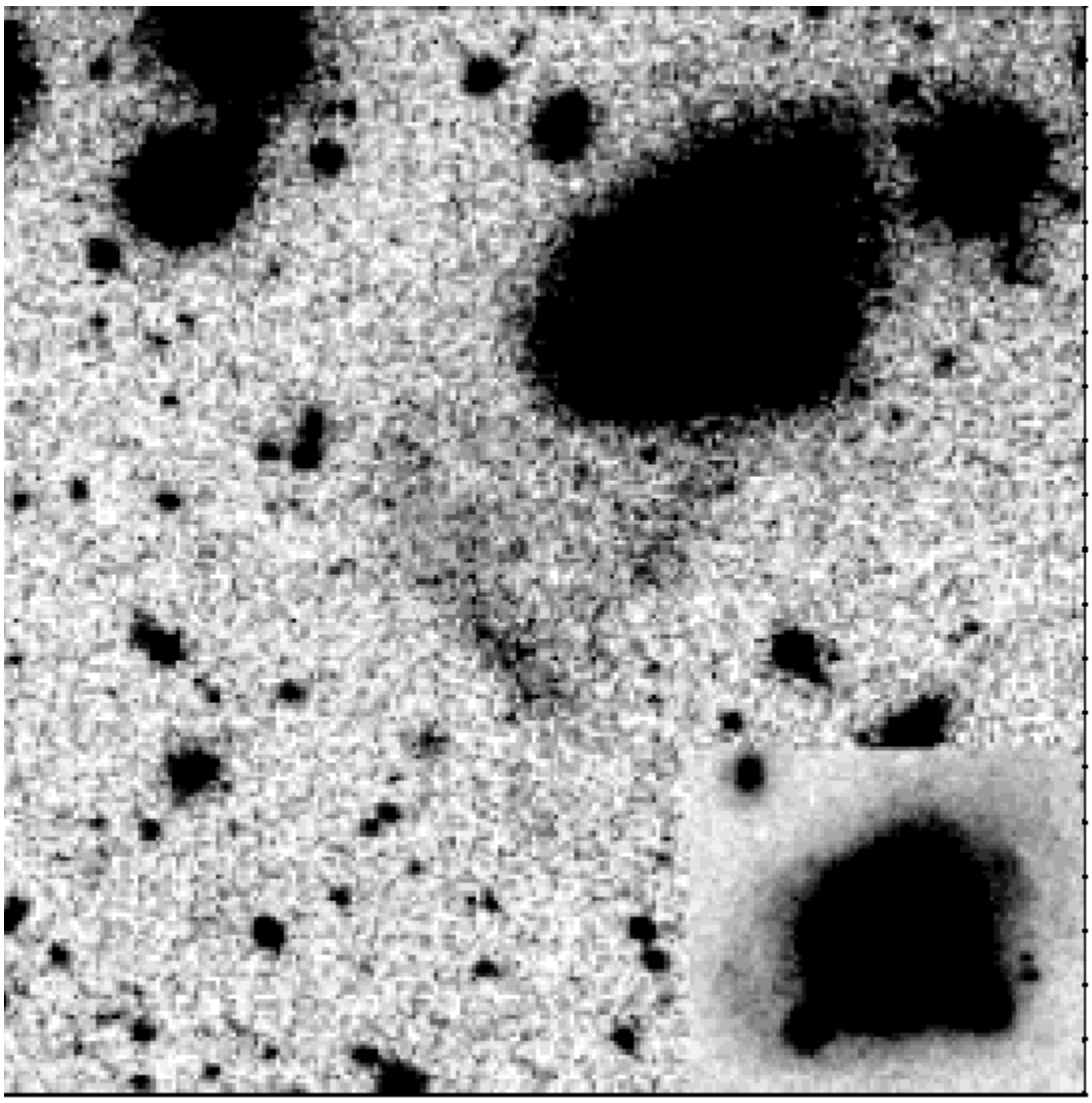

Figure 4: Image of the interior of the Coma cluster showing low surface brightness features. Figure taken from (Gregg and West 1998)

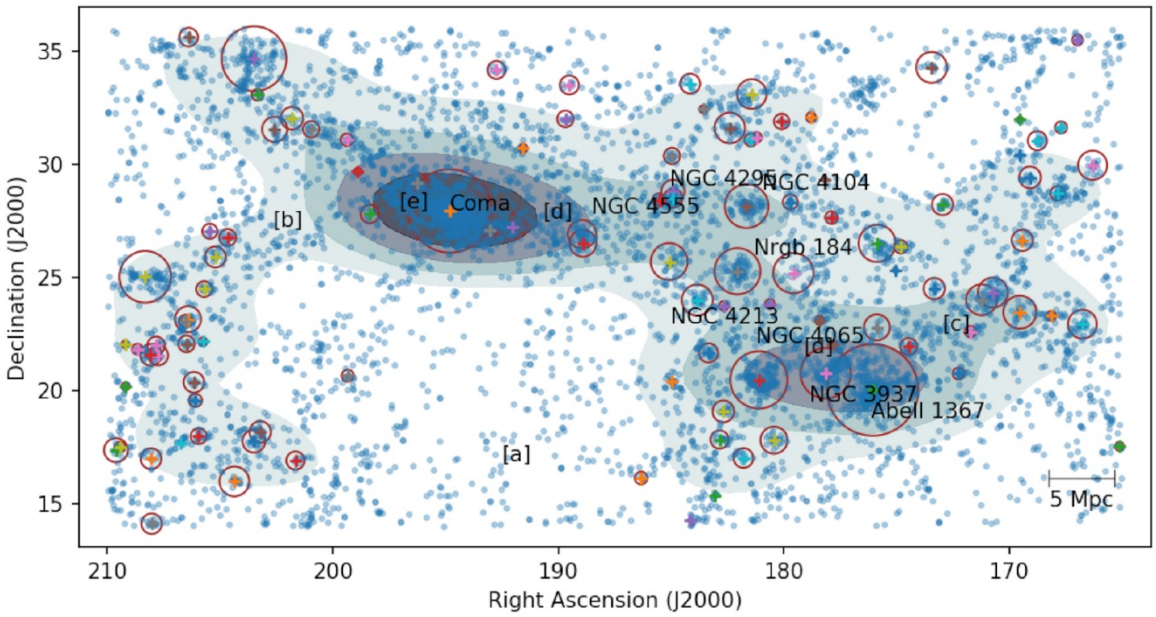

Figure 5 The Coma Supercluster, consisting of the Coma and Leo (Abell 1367) clusters. The galaxy density ranges from 31.7 N Mpc^{-2} at [e] to 1.6 N Mpc^{-2} at [a]. Figure reproduced from (Seth and Raychaudhury 2020).

In sum, the Coma supercluster appears to be the best nearby location to look for the signature of intergalactic baryons in the form of liberated baryonic gas from frequent galaxy-galaxy tidal encounters throughout the supercluster region. We would expect that most of this liberated gas has been processed by prior star formation within the host galaxies. Hence, we expect to find several absorption lines due to ionized metals (particularly Carbon, Silicon, and Iron) at the redshift of Coma.

Chapter II of this dissertation covers the pipeline used to take spectra from HST and return a list of possible absorption features. This includes the overall structure of the pipeline, along with optimizations to make the pipeline more accurate and faster than hand fitting lines.

Chapters III and IV of this dissertation cover the results of our analysis, limitations of the process, and a discussion of the implications of our results.

Chapter V gives ideas of how the search for extragalactic baryons should continue along with limitations of the current generation of telescopes for doing this work. This chapter contains coauthored work. The second paragraph in 5.1 and the final paragraph of 5.2 were written by Greg Bothun. I wrote the rest of the chapter and edited the entire chapter.

CHAPTER II

PIPELINE

2.1 The Hubble Spectroscopic Legacy Archive

The Hubble Spectroscopic Legacy Archive (HSLA) contains all the spectra that Hubble's Cosmic Origins Spectrograph (COS) has taken and is sorted into target type. The data has been processed to account for detector noise and systematics and then all observations are added together to make one clean spectrum. This processed spectrum, along with the raw files that went into making it, are available to be downloaded from the HSLA database. In order to probe the IGM with these spectra, the spectra need to be analyzed. This consists of finding where in the spectra are absorption lines of interest, calculating the properties of those lines (the line center and the equivalent width), and identifying what ionic species was responsible for the absorption. Overall, it's necessary to have a pipeline that takes in a spectrum and returns a list of each absorption line for each object, along with the column density[1] and species for each line.

2.2 Find Absorption Lines

Spectra taken by COS are processed with the calcos pipeline (Soderblom et al. 2021). This pipeline takes the raw spectrum detected and processes it into a usable spectrum through many steps. These steps correct for instrument systematics and return a time-weighted spectrum. This spectrum is high-quality, but not quite in a form where we can extract features from it. The beginning of my pipeline takes the spectra from HSLA and processes it further in order to extract features.

[1] The column density is the total number of particles in some area divided by the area. It is a result of the number density integrated along the line of sight. Because interactions happen along an extended line of sight, the column density ends up being the quantity that's measurable.

The first step is to smooth the spectra. Smoothing blends the noise from several adjacent pixels, which has the effect of smearing out noise and improving the signal. There are multiple ways to smooth, but they all involve taking a kernel, centering it over each pixel, and convolving the kernel with the spectrum. To convolve the kernel and the spectrum, you sum the product of the spectrum's pixel and the kernel's pixel over the full kernel. This sum is then divided by the sum of the pixels in the kernel (to find a weighted average) and set as the central pixel value in the smoothed spectrum. The simplest smoothing kernel is a boxcar, where the value of all pixels in the kernel is 1, but there are uses for more complex kernels. (For example, Gaussian kernels are often used.) In this analysis we've used boxcar smoothings, since the noise doesn't have significant local fluctuations and boxcar smoothing gives a clean enough spectrum to be useful.

QSOs have features that make their overall spectrum not flat. A prominent feature in QSO spectra is due to Lyman Alpha emission. The feature gets spread over hundreds of Angstroms due to velocity dispersion of the emitting material and then gets stretched further due to cosmological redshift. Impacts of other dynamics of the QSO can make other regions of the spectrum uneven as well. It's much easier to analyze spectra that are flat, since then any large variation from the flat background can be detected as a feature. To make the spectrum flat, the pipeline first fit a baseline to the spectrum (that is, what the spectrum would look like if there was no intervening absorption or emission) and then divides the measured spectrum by the baseline. Ideally, this gives a smoothed spectrum that is normalized to 1. In practice it gives us a spectrum that is normalized well enough to be useful.

To find the baseline the pipeline takes a large range of the spectra (on the order of hundreds to 1,000 pixels) centered at each pixel and calculates the mean and standard deviation of all pixel values in the range. It then excludes any pixels with a value more than 1 standard deviation away from the mean. The pipeline calculates the mean for the remaining pixels and sets that as the baseline for the central pixel. After calculating the baseline value of the flux, the baseline is boxcar smoothed.

The pipeline then takes the smoothed spectrum and divides each pixel by the baseline at that pixel. This normalizes the spectra to 1, which smooths out the features from the QSO spectra that are related to the region of space near the QSO and makes absorption and emission from intervening material the only data far below 1. This also makes calculating equivalent widths easier, since you divide by the continuum value while finding equivalent widths and the continuum value is 1.

Now that the spectrum is smooth and normalized, it is easy to find absorption features. The pipeline takes slices of the spectrum that are hundreds to 1,000 pixels wide. In each slice, it calculates the average excess flux, with excess flux being defined as any flux with a value between 1 and 10.5 in normalized units. The excess range is cut off at 10.5 to eliminate geocoronal Hydrogen Alpha emission that is more than an order of magnitude brighter than the light from the QSO, but not of any interest to our survey as it is emitted by Hydrogen in the Earth's exosphere. With the average excess in a region calculated, the pipeline selects any pixels that are more than 3 times the average excess below 1 as potential absorption lines.

The pipeline takes this list of potential absorption lines and compiles all pixels that are next to each other into a single line with line center at the average pixel value.

Any potential lines that only have a single pixel below the baseline flagged are rejected as noise. (Assuming that the noise follows a normal distribution, there is a 0.3% chance a single pixel will be more than 3 standard deviations from the mean. Assuming independence, the chance of 2 pixels next to each other both being 3 standard deviations below the mean is $9 \cdot 10^{-4}\%$, which would mean we have at most single digit false positives.)

At this point, we now have a list of potential absorption features. For each of these potential features, the pipeline finds the equivalent width of the line. The equivalent width is found by taking a box centered on the line center (on the tens to hundreds of pixels scale) and finding the area between the baseline and the measured flux. Since the flux is normalized, this can be calculated as

$$eqw = \sum 1 - F(\lambda)\Delta\lambda$$

with F(λ) the measured flux and Δλ the spacing between adjacent pixels. There are two ways to calculate this equivalent width, naively and by fitting the absorption feature. In the naive method, the function F(λ) is the normalized value of the flux at λ. You can instead fit a function to the potential feature and then use fit as F(λ). The difference between the naïve method and fitting the absorption as a gaussian is analyzed at in depth in Section 3.3.

2.3 Remove Geocoronal Lines

After finding all lines, I make a histogram of the number of lines as a function of wavelength with quarter wavelength resolution. I then calculate the median number of counts in the histogram and the standard deviation of those counts. Any bin that is more than 2.5 standard deviations above the median gives a wavelength to be removed from

our set of lines, given in Table 1. Any line that is within half an Angstrom of one of these lines is removed. This removes lines at a redshift of 0, which accounts for both absorption in the atmosphere of the Earth (called geocoronal lines) and absorption in the Milky Way. For this analysis, I do not distinguish between the two since this study is concerned with material far away from the Milky Way. Lines marked with a question mark are likely to be caused by the element listed. The wavelength bins at 1300.25 Å and 1300.50 Å do not correspond to an element listed in Morton's spectral list, but they have been removed since they meet the 2.5σ condition. A histogram of all detected lines is show in Figure 6. The lines flagged as geocoronal and Milky way absorption are easily seen in the plot.

2.4 Identify Lines

With the geocoronal and Milky Way lines removed, we now have a list of absorption lines from clouds of gas between the QSO and the Milky way. There are certain transitions that are possible and likely for intergalactic clouds. Using Morton, York, and Jenkins' search list for QSO Object Absorption lines (Morton, York, and Jenkins 1988) the pipeline tries to identify each absorption line. To do this, it takes the wavelength of each absorption line and divides it by the wavelength of each transition line in the spectra list. If this division gives a redshift[2] within 0.001 of the QSO's redshift, the line is removed from the analysis, since it is likely from material in the same galaxy as the QSO. Otherwise, if this division gives a value between 1 and 1.2, it is kept as a possible line with redshift between 0 and 0.2. This gives each absorption line many (a few to 10's) possible identifications. The remaining steps in the pipeline aim to reduce

[2] Redshift (z) is defined as 1 + z = wavelength detected/wavelength emitted.

Wavelength (Å)	Element	Wavelength (Å)	Element	Wavelength (Å)	Element
1143	Fe II	1253.75	S II	1369.75	Ni II
1144.75	Fe II	1259.25	S II	1370	Ni II
1152.5	Fe II/P II/Fe II	1259.5	S II	1393.25	Ni II/Si IV
1152.75	Fe II/P II/Fe II	1259.75	S II	1393.5	Si IV
1190	S III/Si II	1260	Si II/Fe II/ C I	1393.75	Si IV
1190.25	Si II	1260.25	Si II/Fe II/C I	1402.5	Si IV
1193	Si II	1260.5	Si II/Fe II/C I	1402.75	Si IV
1193.25	Si II	1277	C I	1526.25	Si II
1199.25	N I	1277.25	C I	1526.5	Si II
1199.5	N I	1280	C I	1526.75	Si II
1200	N I	1293.25	Ti III?	1547.75	C IV
1200.25	N I	1300.25		1548	C IV
1200.5	N I	1300.5		1548.25	C IV
1206	Si III	1302	O I	1550.5	C IV
1206.25	Si III	1304.25	Si II	1550.75	C IV
1206.5	Si III	1317	Ni II	1608	Fe II
1238.5	N V	1328.75	C I	1608.25	Fe II
1238.75	N V	1334.0	C II	1656.75	C I
1239.75	Mg II	1334.25	C II	1670.5	Al II
1250.25	S II	1334.5	C II/P III	1670.75	Al II
1250.5	S II	1335.5	C II	1741.25	Ni II
1253.5	S II	1347	Cl I?		

Table 1 Wavelength bins for identifying geocoronal and Milky Way lines with the ionic species that caused them at z=0.

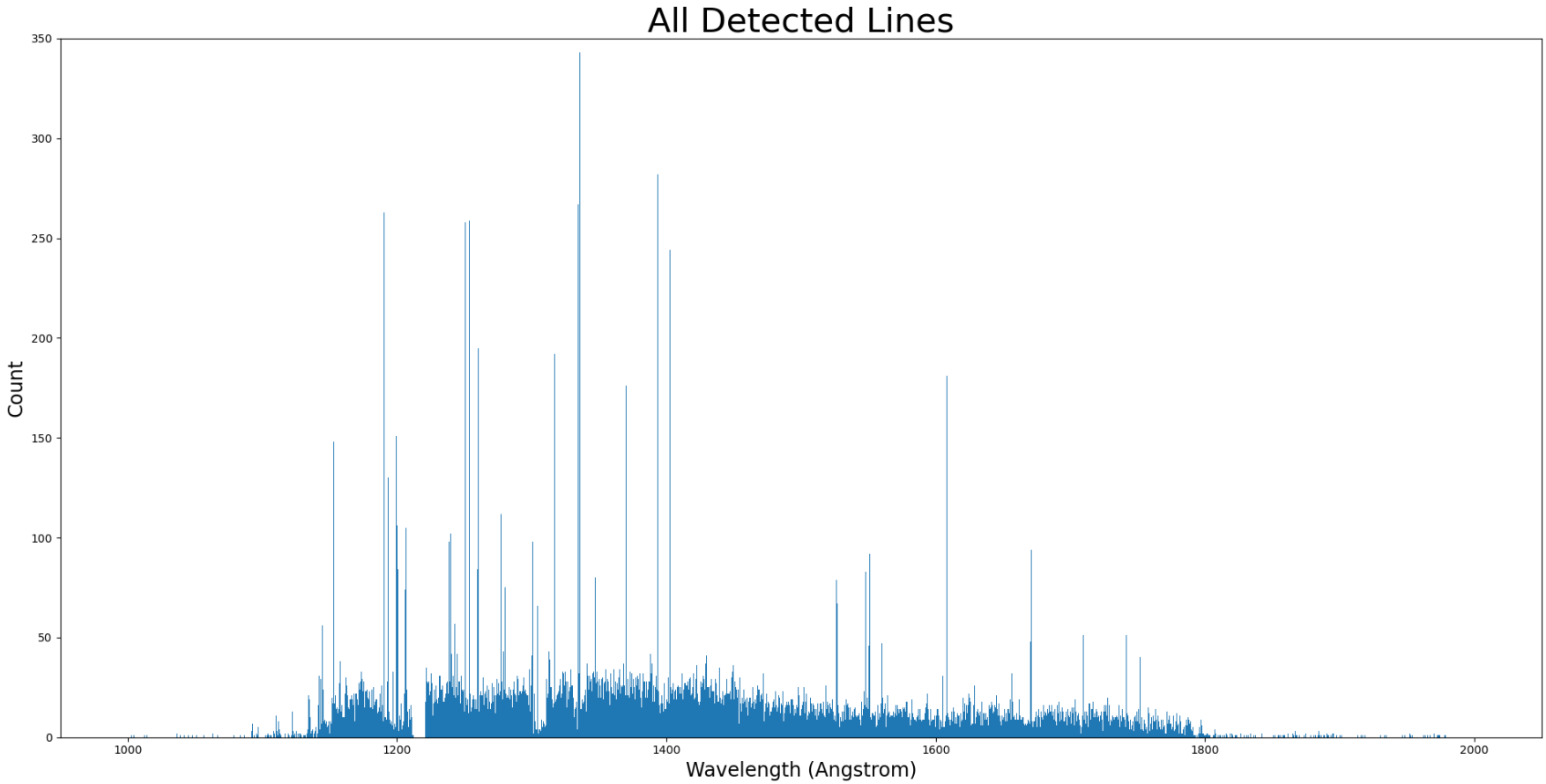

Figure 6 Histogram of all detected lines. The large spikes are the bins of absorption lines that are flagged as geocoronal and Milky Way.

the number of identifications for each line to 1, as each line is typically caused by a single species of ion. (There is a chance for more than one line to be around the same wavelength. This causes a blended line that is harder to deal with. This happens infrequently enough in this dataset that the pipeline does not check for blended lines.)

At this point there is now enough information to determine the column density that would be necessary to give a line of the given equivalent width by that element. The column density N (cm^{-2}) is given by:

$$\log N = 20.053 + \log\left(\frac{W_\lambda}{\lambda}\right) - \log(\lambda f)$$

with W_λ the equivalent width of the line, λ the vacuum wavelength of the transition, and f the oscillator strength(Morton, York, and Jenkins 1988).

2.5 Reduction Pipeline

Since species do not only have one transition line, a way to give a larger certainty in a detection is to find multiple lines at the same redshift for the same element and ionization state. The pipeline allows for an error in z of 0.001 since the line center for each line has some uncertainty in it. Any line that has another line with a similar redshift is kept.

We assume that any cloud of gas that has processed material must also have unprocessed material since a large fraction of matter does not get processed in stars. Namely, this means that any cloud of gas must have Hydrogen in addition to whatever metals are present. Since clouds of gas are subject to a meta-galactic UV flux, the presence of unionized Hydrogen (HI) can give insight to the prevalence of ionizing UV radiation. In particular, if HI is not in a system, then there should not be any other species

which are ionized at or below the ionization energy of HI (13.6 eV). The species that are removed because of this and their ionization energies are list in Table 2 (Kramida, Ralchenko, and Reader 2022).

Element	Ionization Energy (eV)	Element	Ionization Energy (eV)	Element	Ionization Energy (eV)
C I	11.2602880	Mg I	7.646236	Fe I	7.9024681
Cl I	12.967633	Mn I	7.4340380	Ti I	6.828120
S I	10.36001	Al I	5.985769	Li I	5.391714996
O I	13.618055	B I	8.298019	Na I	5.13907696
Be I	9.322699	Ti II	13.5755	Sc II	12.79977
P I	10.486686	Cr I	6.76651	V I	6.746187
Si I	8.15168	Sc I	6.56149	K I	4.34066373

Table 2 Ionic states with ionization energy below the ionization of neutral Hydrogen.

Previous studies have found that intergalactic clouds are primarily Carbon and Silicon so the final pass checks to see if a line has a possible ID for Carbon. If Carbon is among the likely IDs, all Carbon IDs are kept, since they are most likely the correct ID. If there are no possible IDs for Carbon, the pipeline then checks to see if it has a possible ID for Silicon. If it does, all of the Silicon lines are kept as possible IDs. If Carbon or Silicon are not a possible ID, the pipeline keeps all the rest of the possible IDs.

At the end of the spectra processing pipeline, we are left with a series of detected lines with 1 to many possible elements that could have caused that line. There are several parameters in the pipeline that can be tuned to get better detections of lines and better measurements of their parameters. These refinements are discussed in the following chapter.

CHAPTER III

REFINING THE PIPELINE

3.1 Chapter Overview

I've now built up a pipeline to identify lines from spectra. In order for this pipeline to be useful, it needs to reliably identify lines with the same or better accuracy as identifying them by inspection and using available, manual tools. Initially, I used PyRAF (a Python wrapper of IRAF[3]) (Tody 1986, 1993) to manually select lines, fit them with a gaussian profile, and find an equivalent width and center. This process was tedious, slow, and only found absorption lines that were dominant enough to easily see by eye. For these spectra, the automated pipeline should find the same lines and fit them in the same way as manual fitting in PyRAF. There are several parameters in the pipeline that can be varied. By finding the optimal parameters to recover the manually fit spectra, I can then use the pipeline on all spectra with confidence that it is finding all absorption lines.

3.2 Spectra Used For Testing

In our initial exploration of spectra that intersected Coma, I had fit around 60 spectra by hand. These spectra were the ones that were available in the HSLA in Summer 2017 that intersected near the center of the Coma supercluster. In particular, they had right ascension between 160° and 220° and declination between 20° and 30°. To test the pipeline, I randomly chose 10 spectra from the initial 60. The chose spectra are listed in Table 3. After an initial inspection of all 10 spectra by eye, I removed 1 of them from the test as it had low signal to noise. The following sections give the results of various tests

[3] IRAF is the Image Reduction and Analysis Facility. It is a standard software collection from the National Optical Astronomy Observatory.

on these spectra. The next section (3.3) looks at the difference between different methods for calculating the equivalent width of lines. Section 3.4 explores how the equivalent width and line center measurements compare between lines that are hand measured and lines that are measured by the pipeline.

Target	Used/Rejected
QSO-B1229+204	Used
J132652.44+292534.8	Rejected due to low SNR
J114903.47+220424.3	Used
Q1135+2414	Used
HB89.1259+281	Used
CSO-0873	Used
J130100.86+281944.7	Used
2MASS-J13250381+2717189	Used
SDSSJ105958.82+251708.8	Used
SDSS.J125846.67+242739.1	Used

Table 3 QSO targets randomly selected from the first sightlines of interest.

3.3 Methods For Finding the Equivalent Width

When calculating the equivalent width of an absorption line, there are two ways to algorithmically do it, the naive way and by fitting the line and finding the equivalent width of that line. For the naive way, I simply add up the total flux around an absorption feature and subtract it from 1 added along the same interval, since the spectrum has been normalized. This works as long as the fluctuations in the spectrum are the same both above and below the signal. If the noise is asymmetric or there are other features nearby, this method will over or under calculate the equivalent width.

The other way to calculate the equivalent width is to fit a line to the absorption feature and then find the equivalent width of the fit line. I have done this by fitting a gaussian absorption feature of the form:

$$f(\lambda) = 1 - \exp\left(\frac{-(\lambda - \lambda_0)^2}{2\sigma}\right)$$

Where λ_0 is the central wavelength of the feature and σ is the gaussian width.
The resulting equivalent widths for using both methods for finding equivalent widths is shown in the Figure 7. In general, there is no difference between the equivalent width given by either method. Figure 8 shows the measured line center for both methods of analyzing a line. The center wavelength for both lines match extremely well. From these two tests, both methods of analyzing a line are reasonable. Because they give similar accuracy and fitting the line prevents overcounting other lines near the wings of the line of interest, I have fit lines with a gaussian going forward.

3.4 Pipeline Parameter Tests

There are many parameters used in the pipeline that affect the analysis. The pipeline needs to detect lines with the highest accuracy it can and find the correct equivalent width and line center for those lines that it does detect. These parameters are in the steps of smoothing, fitting a baseline, removing pixels that are some deviation from their local mean as possible lines, and finding their equivalent widths. There are 4 parameters for the kernel size used at each stage in the pipeline. These kernels are the box sizes for smoothing, fitting a baseline, detecting features, and fitting a gaussian.

Figure 9 through Figure 17 show the relationship between the number of detected lines, the average difference in wavelength center, and the average difference in equivalent width as functions of the kernel size for smoothing, fitting a baseline, and

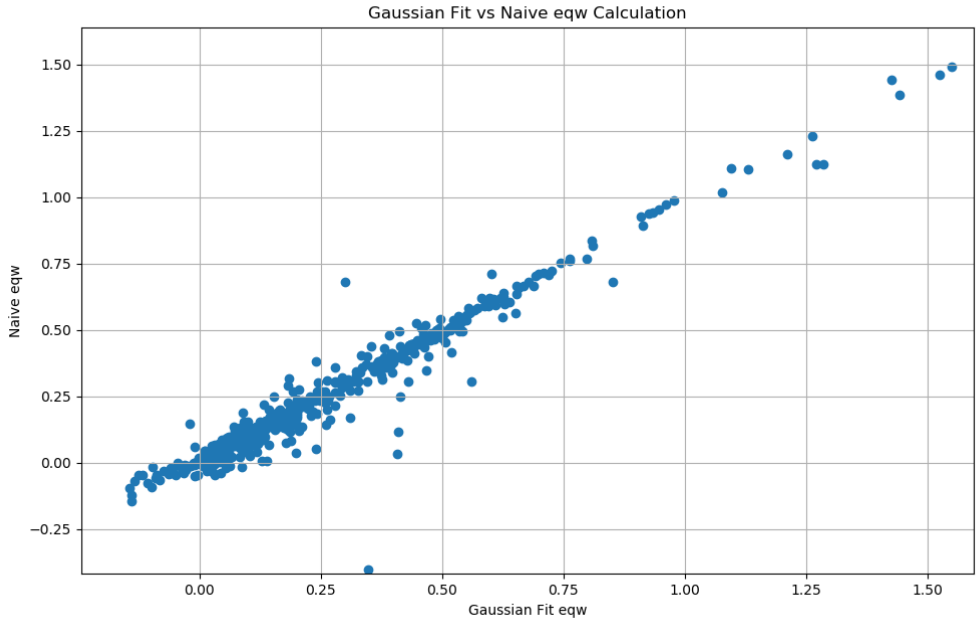

Figure 7 Comparison for naively measured equivalent widths and calculated equivalent widths from gaussian fits.

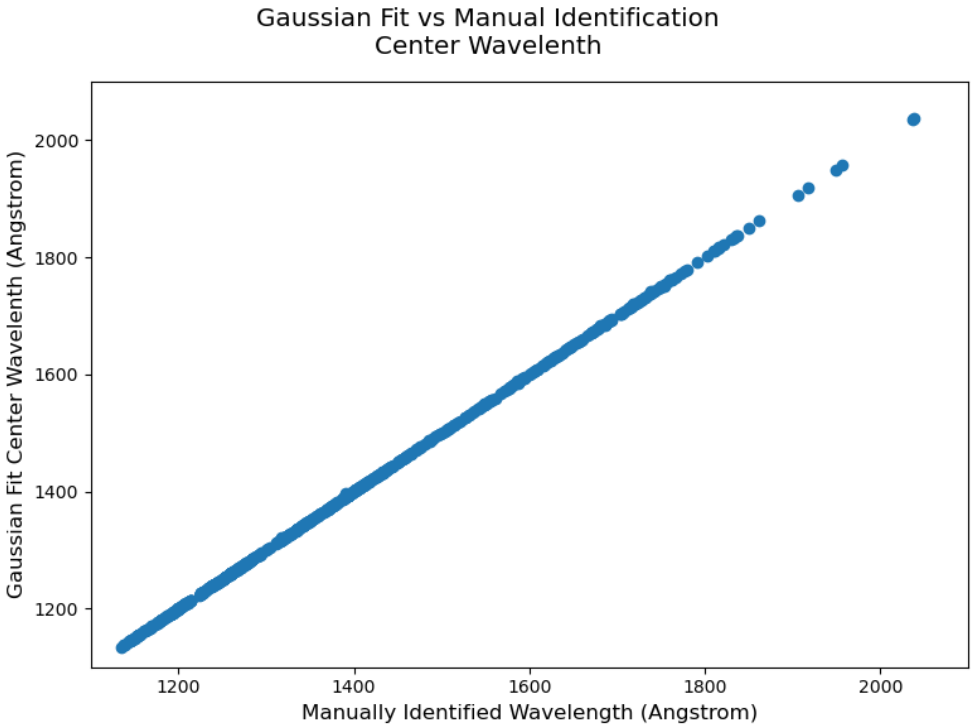

Figure 8 Wavelength identified using PyRAF versus wavelength identified using a gaussian fit.

detecting features. For most of the spectra, the number of lines found does not exactly equal the number of lines found by a manual fit. Because of this, the value of the errors for the wavelength center and equivalent width are not useful, however, the relative magnitude as the parameters vary is. For these tests each of the parameters was a randomly chosen integer in the range shown in the plots. This was done to have good coverage of the range of parameters of interest while still being able to run in a reasonable amount of time.

From these tests, there is no correlation between the size of kernel used for boxcar smoothing or the size of kernel used for fitting a baseline with any of our figures of merit. This means that any size of kernel should be reasonable for use in the pipeline. When initially writing the code I used a 7-pixel kernel for boxcar smoothing and a 500-pixel kernel for fitting the baseline. Since any size kernel is reasonable, these values are used in the final analysis.

The size of the kernel used to find features seems to not matter for the error in the equivalent width or the error in the identified wavelength, however, a large kernel does correlate with less identified lines. This makes sense because as the range you look over gets larger, there are more possible features included and a larger chance of wavelength specific noise variations that are included in the calculation. At small kernel sizes, there are patches of noise that happen to be far outside the normal deviations from the baseline in that range, which will get falsely flagged as a signal. With larger kernel sizes there is a risk of features getting missed because there are too many features nearby that make the standard deviation from the baseline in that region large. There is some turnover between

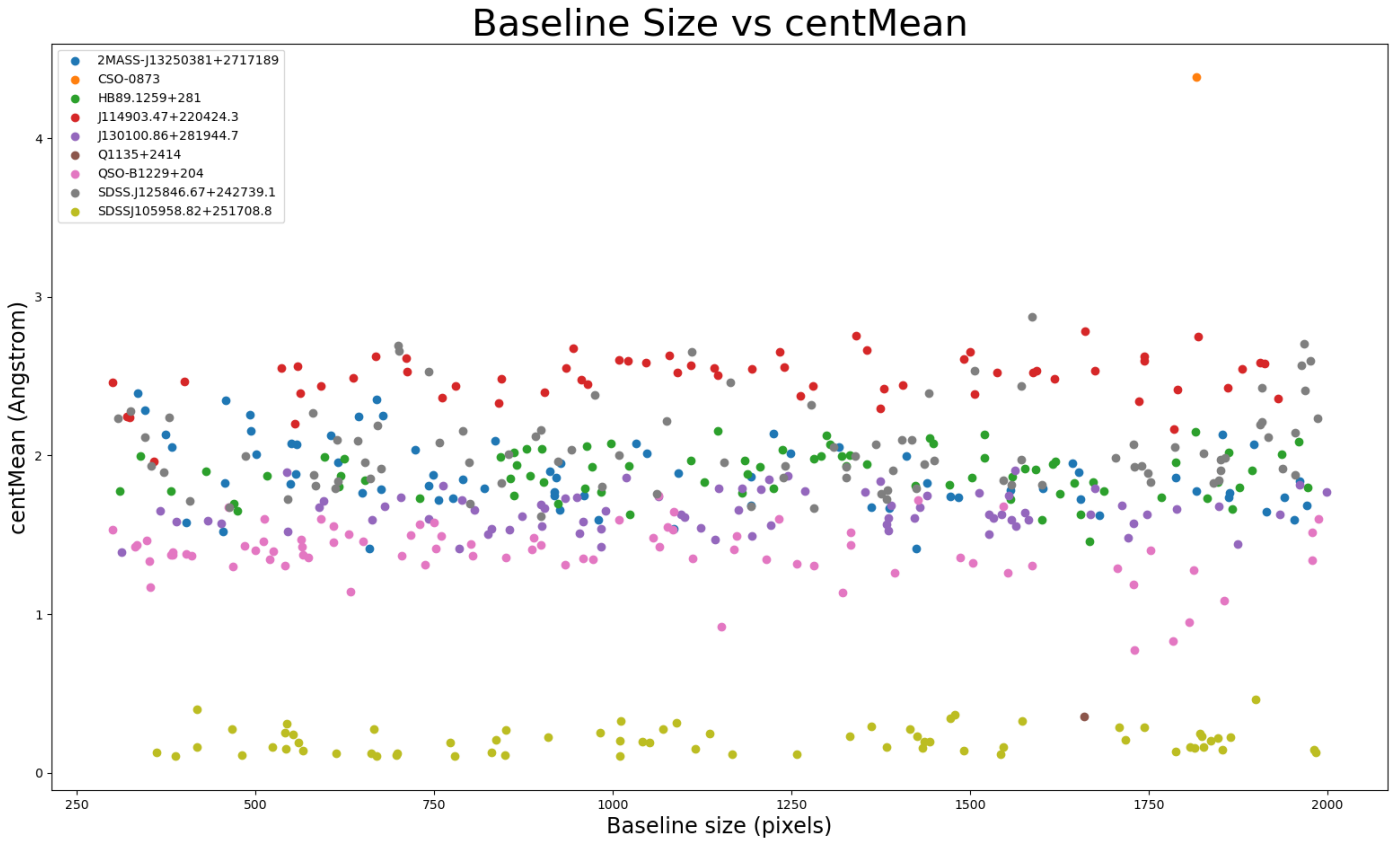

Figure 9 Plot of the average center wavelength deviation for each test spectrum versus the size of the baseline kernel in pixels.

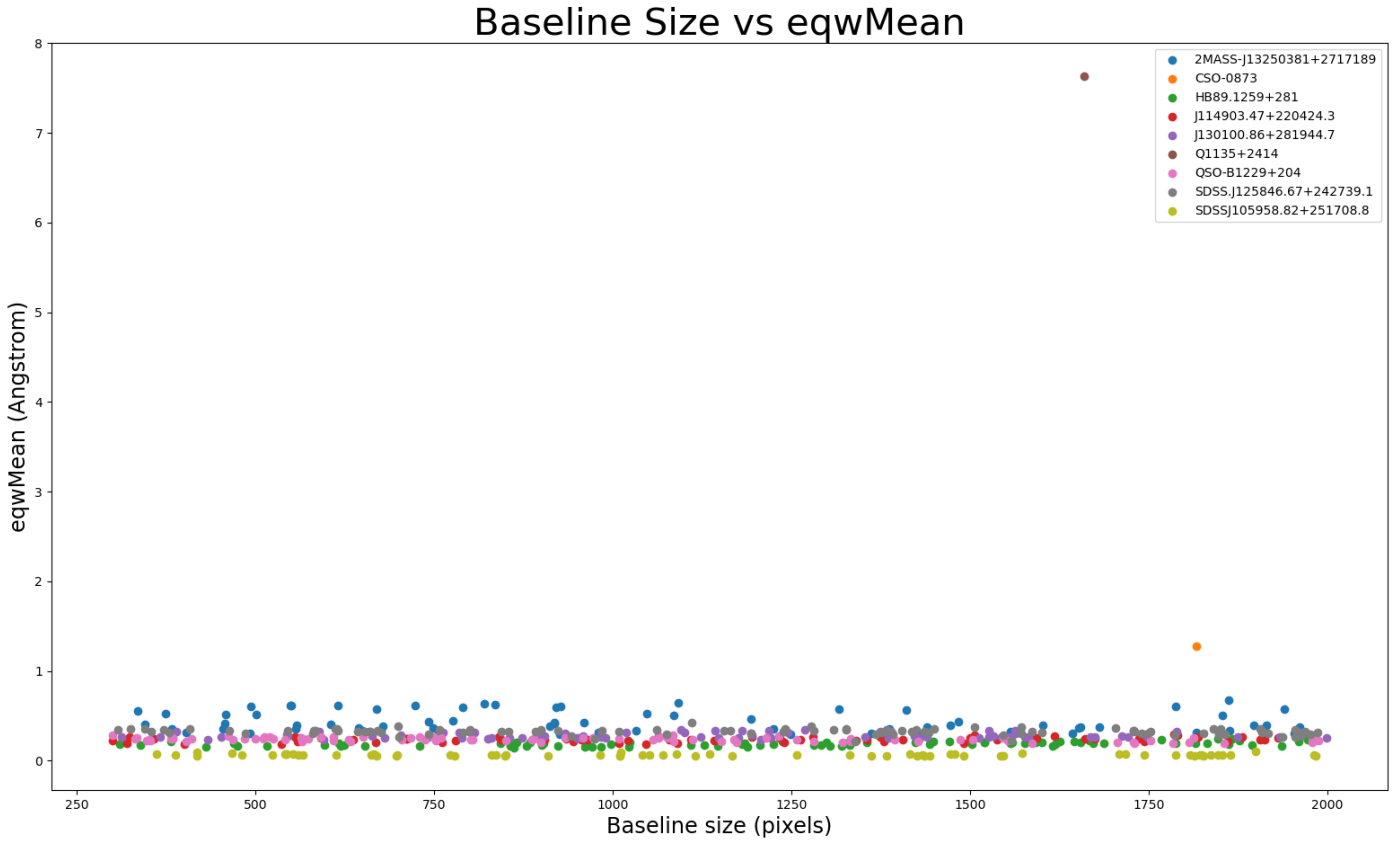

Figure 10 Plot of the average equivalent width deviation for each test spectrum versus the size of the baseline kernel in pixels.

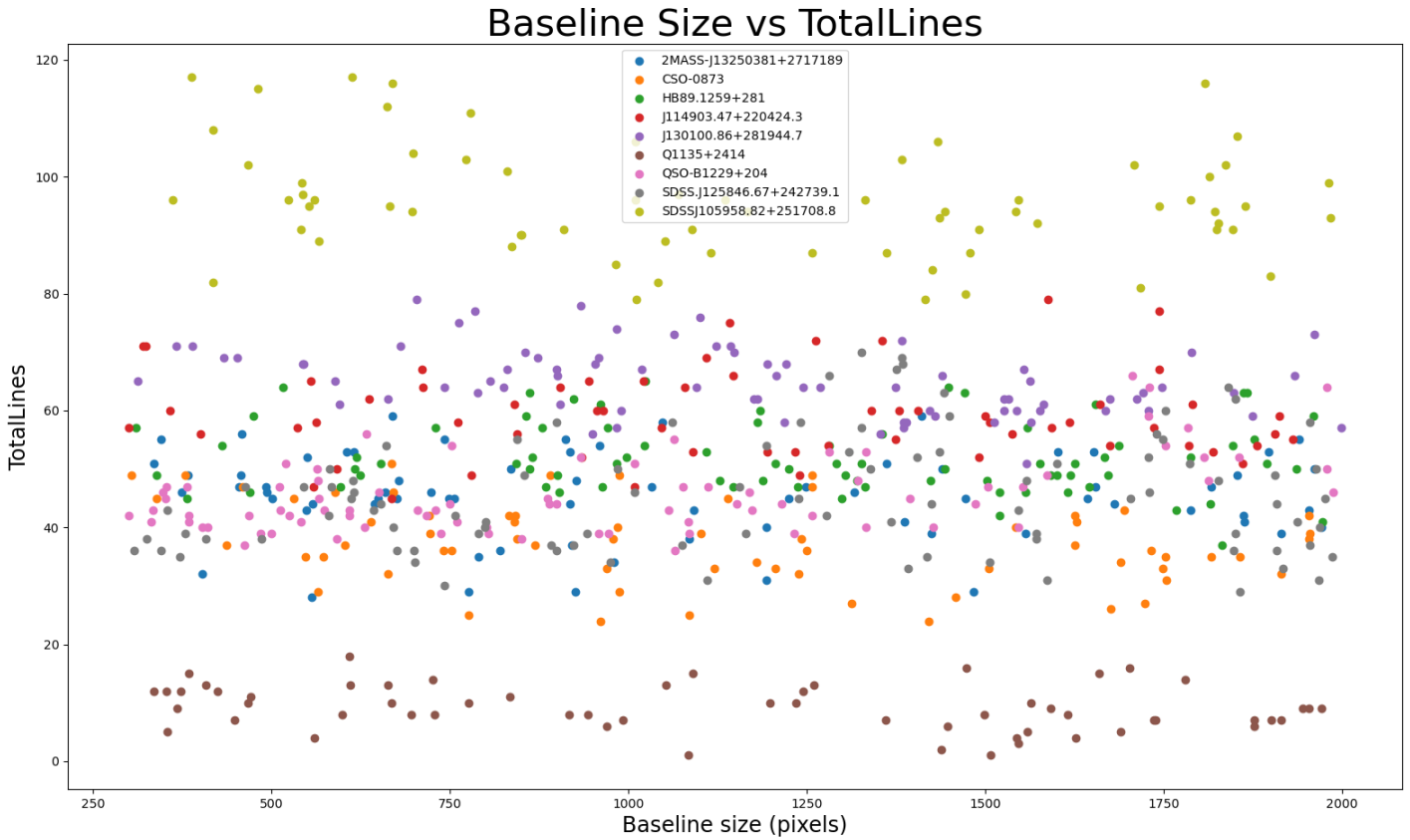

Figure 11 Plot of the total number of lines for each test spectrum versus the size of the baseline kernel in pixels.

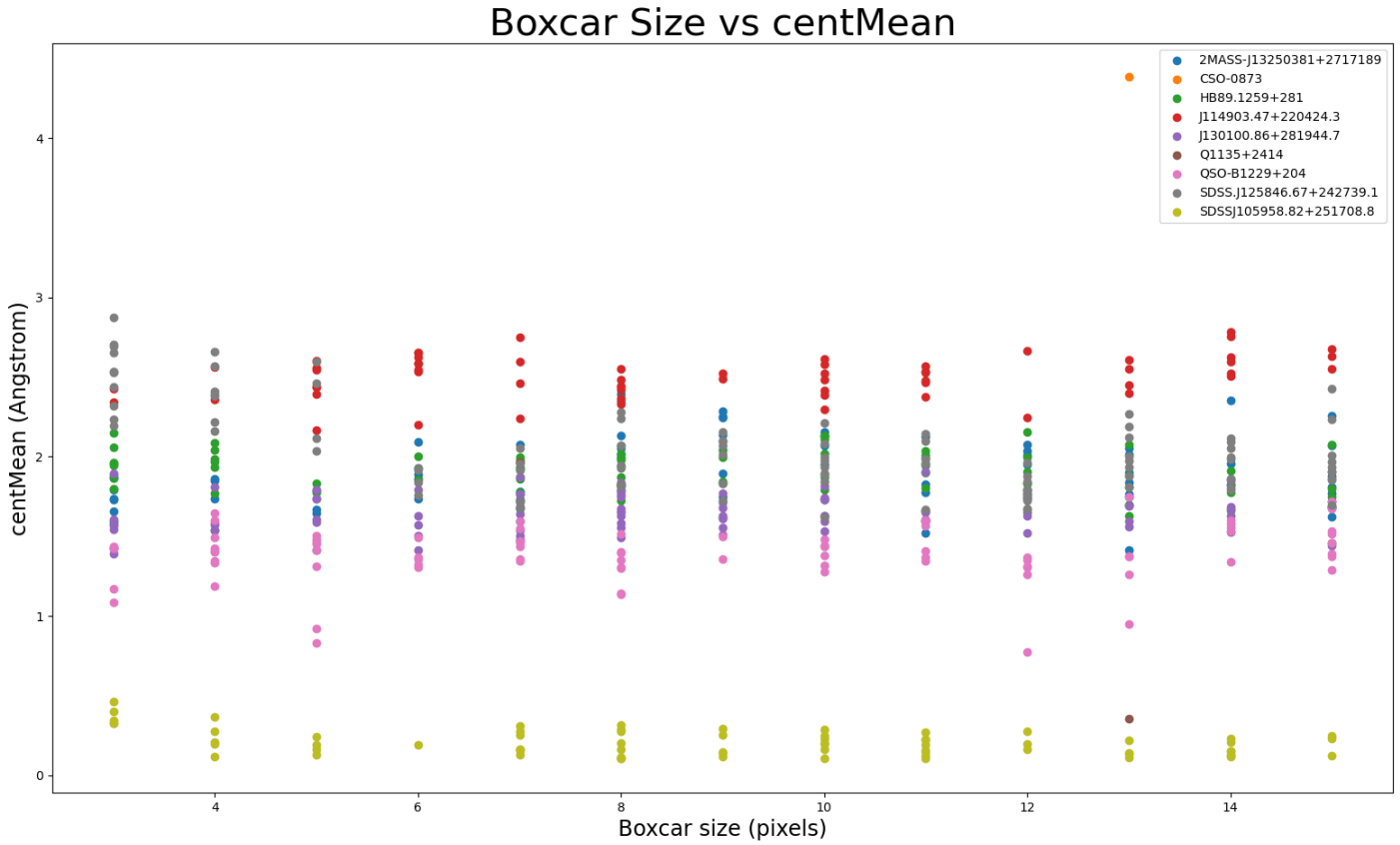

Figure 12 Plot of the average center wavelength deviation for each test spectrum versus the size of the boxcar smoothing kernel in pixels.

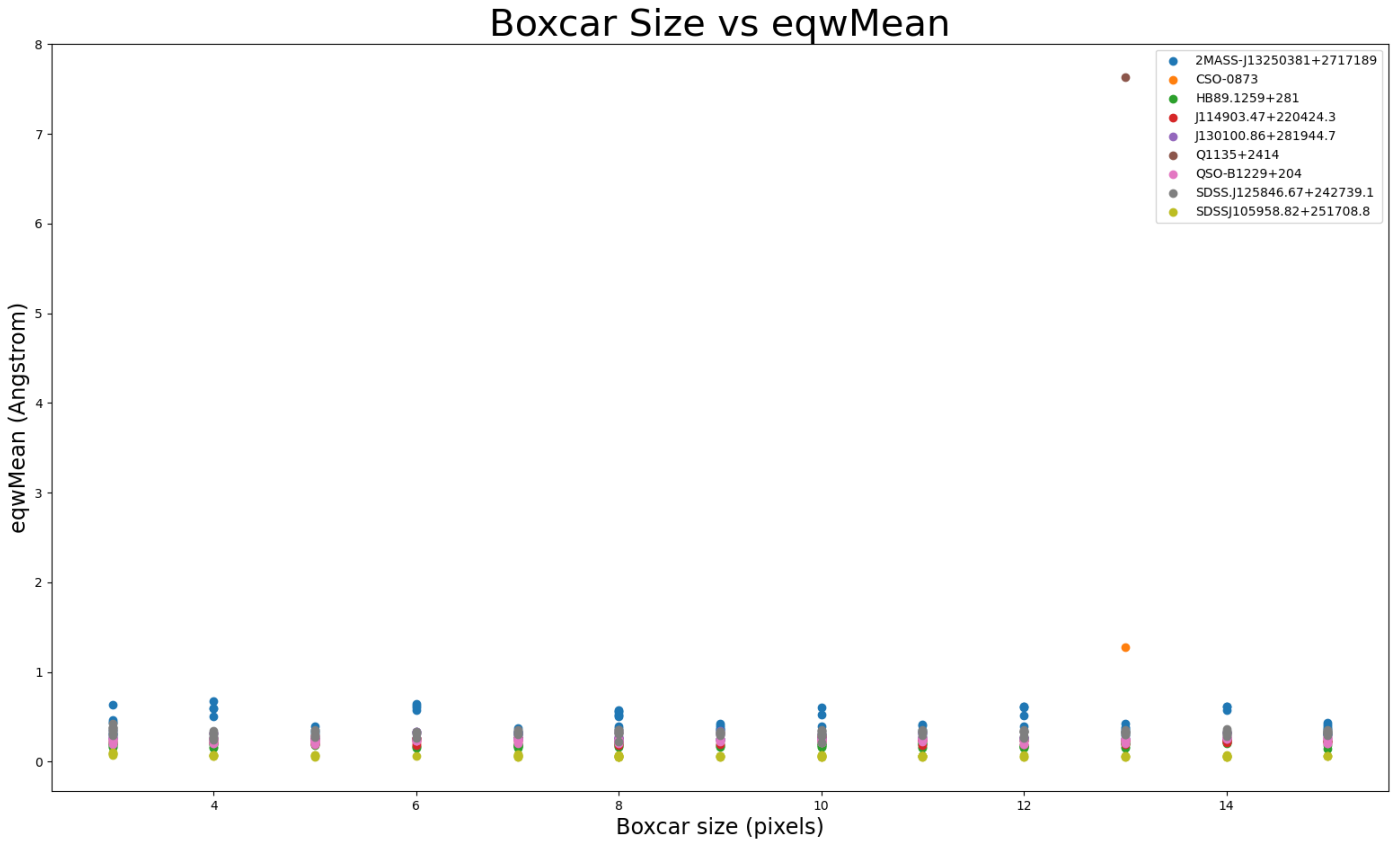

Figure 13 Plot of the average equivalent width deviation for each test spectrum versus the size of the boxcar smoothing kernel in pixels.

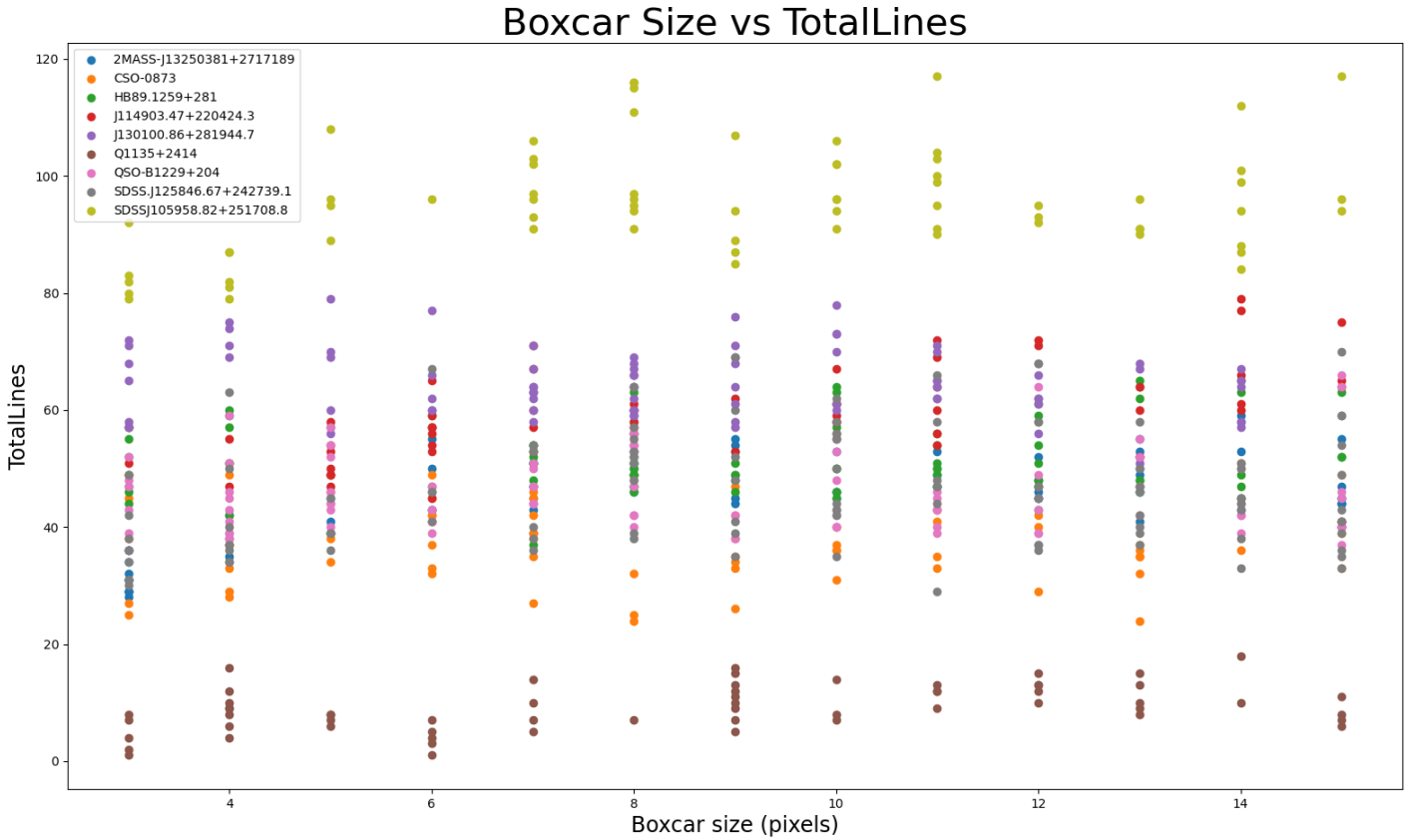

Figure 14 Plot of the total number of lines detected for each test spectrum versus the size of the boxcar smoothing kernel in pixels.

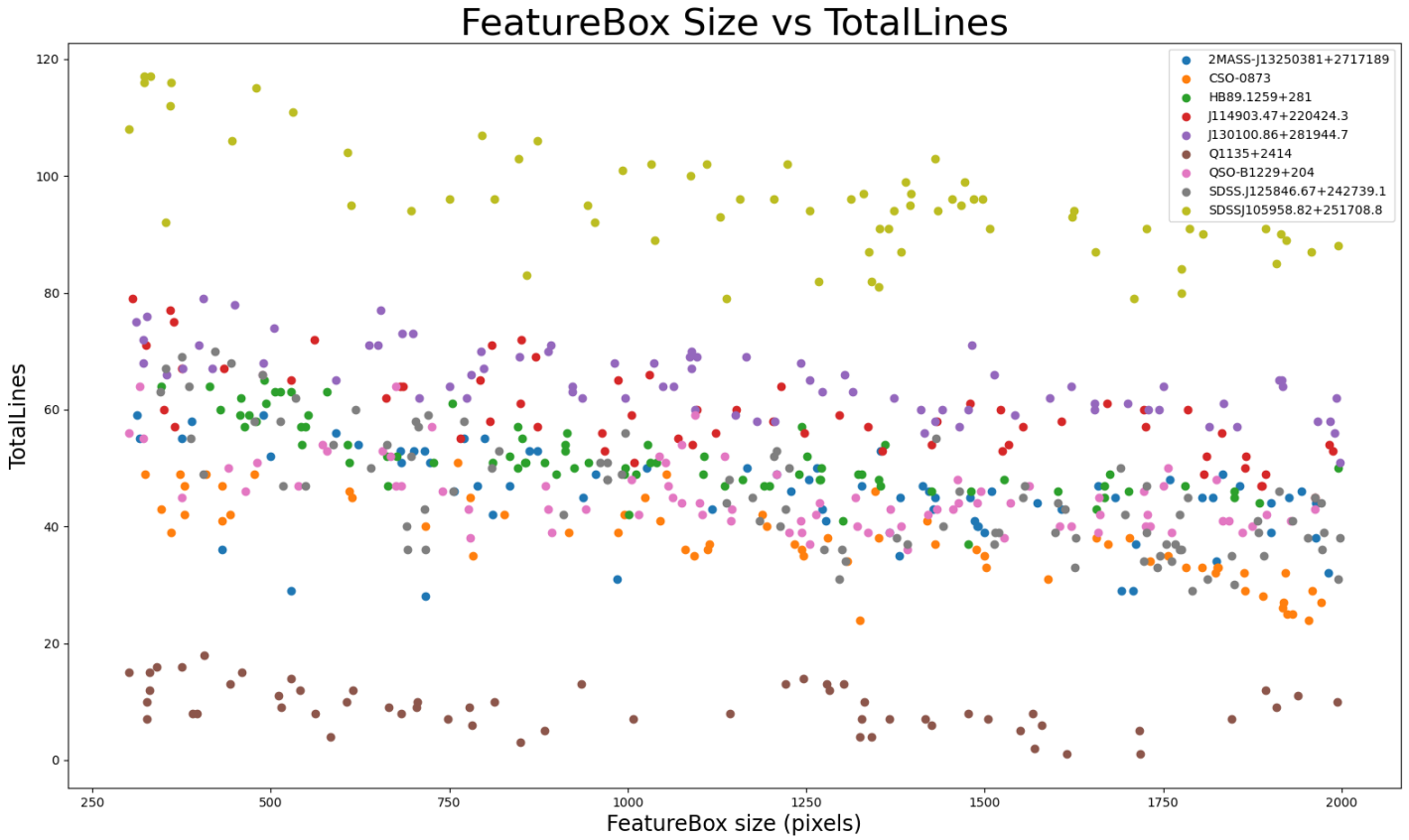

Figure 15 Plot of the total number of lines detected in each test spectrum versus the size of the detecting features kernel in pixels.

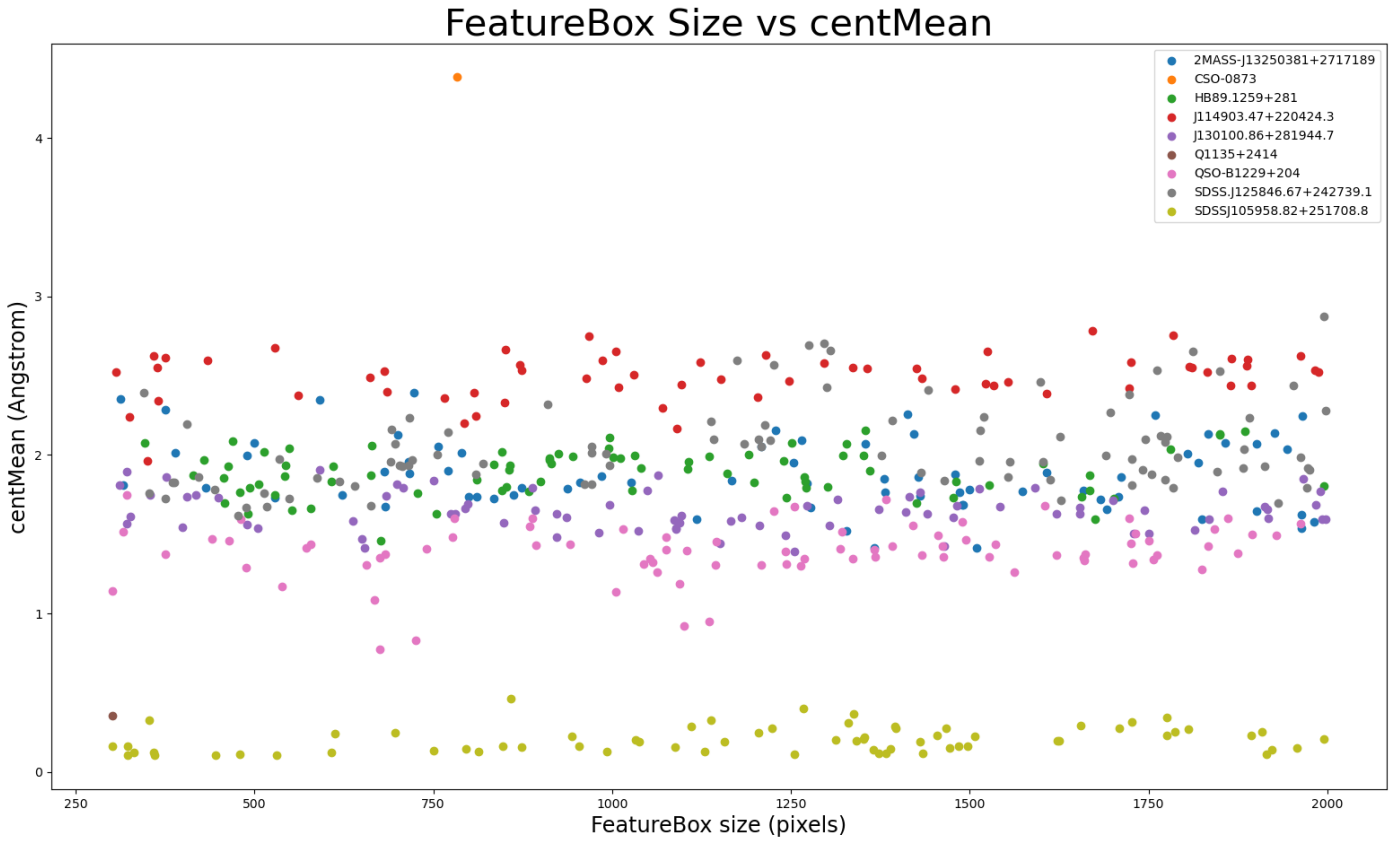

Figure 16 Plot of the average center wavelength deviation for each test spectrum versus the size of the detecting features kernel in pixels.

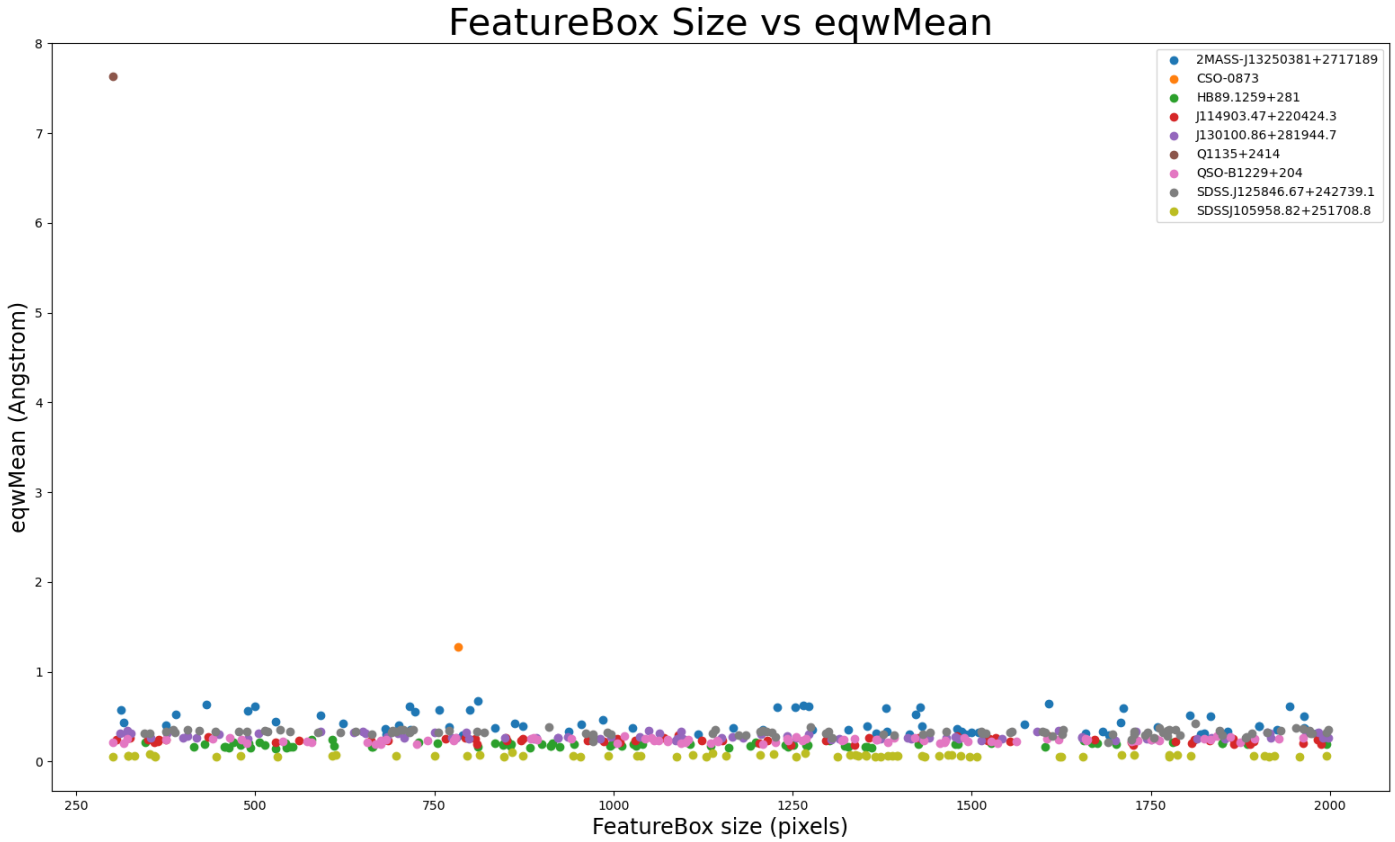

Figure 17 The average difference between the manual fit and pipeline fit equivalent widths versus the size of the kernel used to find features.

too small and too large of a kernel, but since we don't have an independent way to measure the intervening material, there is not good way to figure out specific noise variations that are included in the calculation. At small kernel sizes, there are patches of noise that happen to be far outside the normal deviations from the baseline in that range, which will get falsely flagged as a signal. With larger kernel sizes there is a risk of features getting missed because there are too many features nearby that make the standard deviation from the baseline in that region large. There is some turnover between too small and too large of a kernel, but since we don't have an independent way to measure the intervening material, there is not good way to figure out where that turnover is. I have historically used a 1000-pixel kernel, which corresponds to around 10 Angstrom in a typical spectrum. Since a typical feature has an equivalent width less than 1 Angstrom and there is not a particular reason to use a different kernel size, I will continue to use the 1000-pixel kernel.

The last parameter to test in the pipeline is the kernel size used to fit a gaussian and calculate the equivalent width of absorption features. This test is run with the parameters decided above. A 7-pixel kernel for smoothing spectrum, a 500-pixel kernel for fitting the baseline, and a 1000-pixel kernel for identifying features. The gaussian kernel size was a randomly selected integer between 50 and 500. Results from these tests are given in Figure 18, Figure 19, and Figure 20.

As I vary the size of the gaussian kernel varies, the total number of lines essentially doesn't change. For SDSSJ105958.82-251708.8 there is a very slight decrease in the number of lines for a large range. The only way the pipeline loses lines when

fitting them with the gaussian is if either the fit doesn't converge, or it converges to something about the baseline. Since it is only at large ranges used to fit that lines are lost, the total number of lines only puts a weak constraint that the range should be less than around 400 pixels. Most of the spectra tested don't seem to have any real difference between being fit with a 50-pixel kernel and a 500-pixel kernel. Two of the spectra do have significant differences. 2MASS-J13250381-2717189 has lower errors as the box size is increased from 50 to around 100. CSO-0873 has essentially constant errors until about 100 pixels and then increases steadily.

Since most of the tests are ambivalent to the size of box used and both 2MASS-J13250381-2717189 and CSO-0873 are towards their lowest at 100 pixels, I will use a box size of 100 pixels when fitting a gaussian. This sets an upper limit on the equivalent widths that I can measure of 100 times the pixel spacing. For the most detailed spectra this limit is a little under 1 Angstrom. Since a typical line that I've measured has an equivalent width of around 0.3 Angstrom, an upper limit of 1 Angstrom shouldn't be an issue.

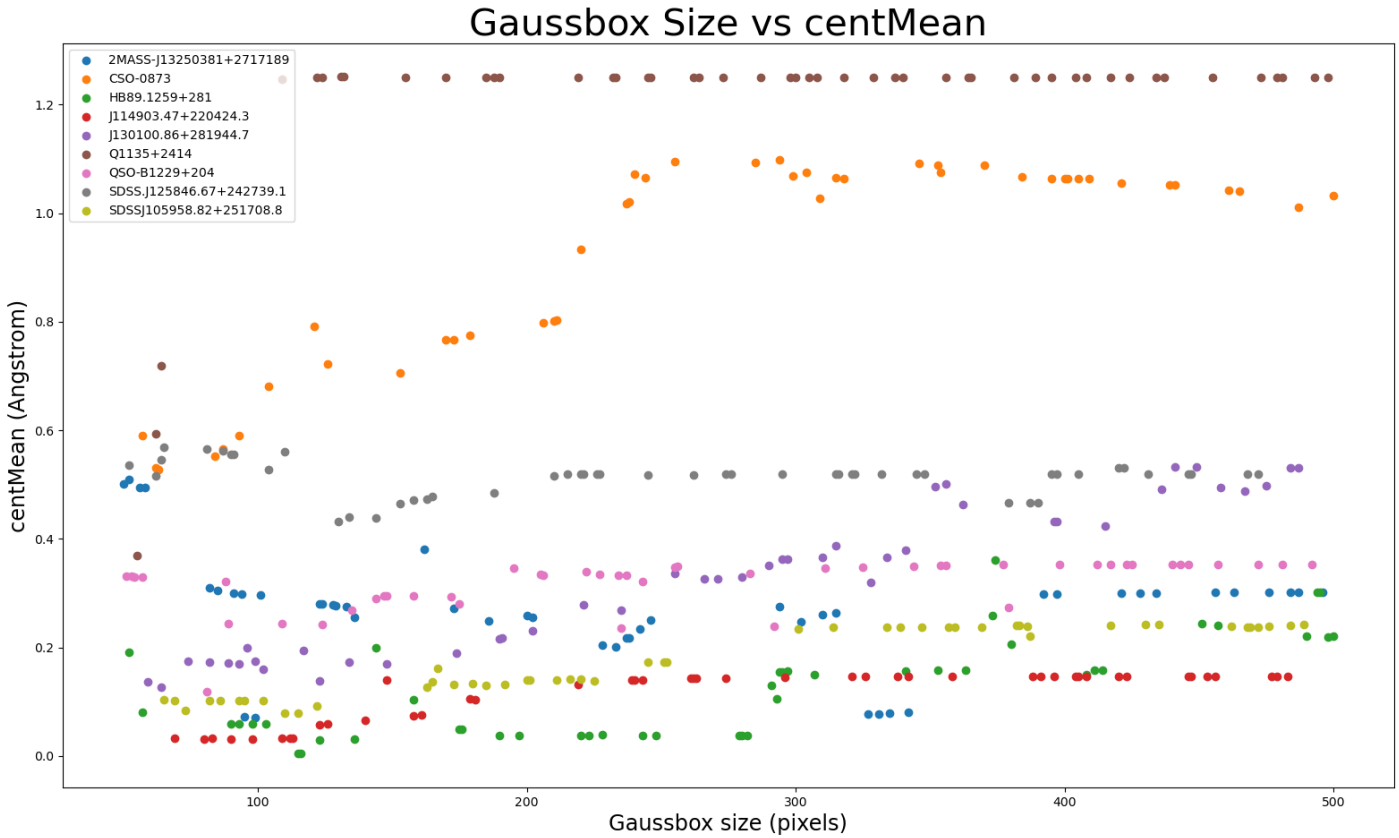

Figure 18 Plot of the average center wavelength deviation for each test spectrum versus the size of the gaussian kernel in pixels.

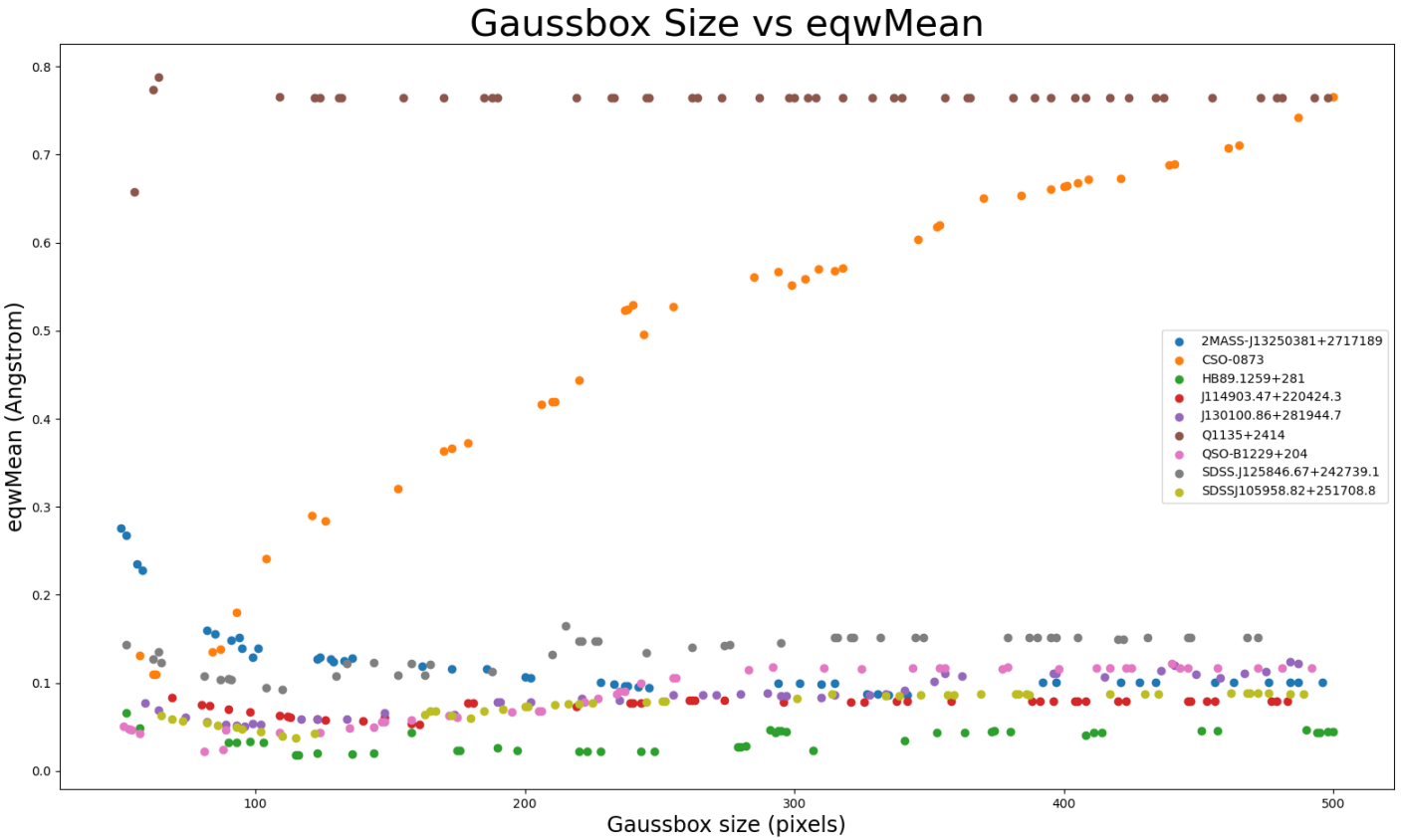

Figure 19 Plot of the average equivalent width deviation for each test spectrum versus the size of the gaussian kernel in pixels.

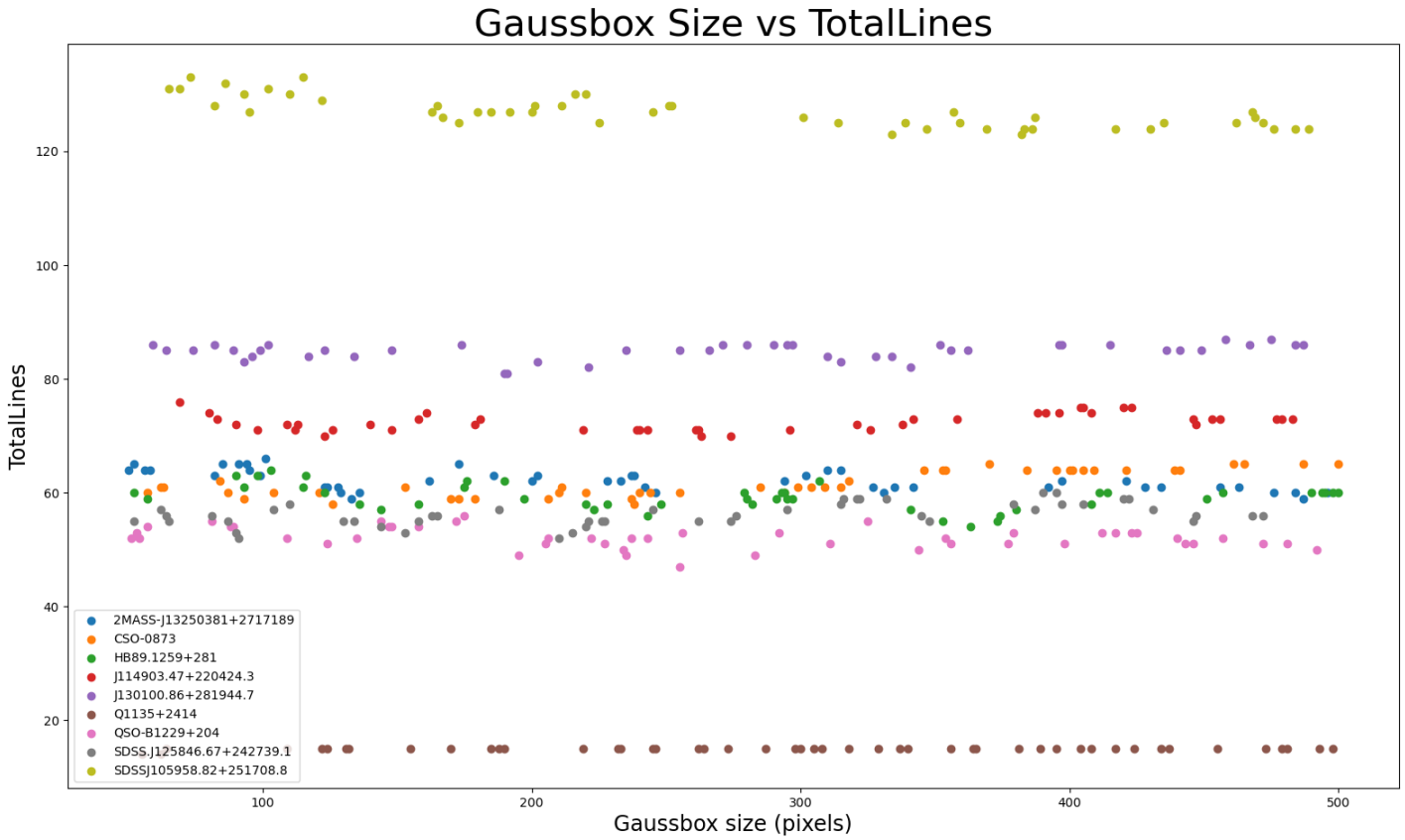

Figure 20 Plot of the total number of lines for each test spectrum versus the size of the gaussian kernel in pixels.

CHAPTER IV

RESULTS

4.1 Results From the Pipeline

With ideal parameters for the pipeline, I can now analyze all COS QSO spectra. Results are given for 3 regions of the sky: Sightlines going through the Coma and Leo clusters and their local neighborhood, sightlines going through the region in the extended Coma Supercluster neighborhood, and sightlines in the rest of the sky. The local neighborhood of Coma and Leo is defined as any sightline having a right ascension between 165° and 210° and declination between 15° and 35°. The Coma Supercluster neighborhood is defined as any sightline having a right ascension between 100° and 260° and declination between 0° and 70°. These sightlines are shown in Figure 21. There are 78 sightlines through Coma and Leo, 393 in the extended Coma Supercluster region (neglecting those also in the Coma and Leo neighborhood), and 217 in the rest of the data set for a total of 688 sightlines used in our analysis.

The initial attempt at fitting each line with any possible line in Morton's spectra list that corresponds to a redshift between 0.0 and 0.2 gives 909,798 possible IDs for 30,138 lines. The pipeline steps off checking for other elements at the same redshift, checking for neutral Hydrogen, and then preferring Silicon and Carbon if they are present reduces this to 66,062 possible IDs for 29,985 lines. This takes the average of 30 possibilities for each line to a little over 2 possibilities for each line. The number of lines and IDs at the beginning and end of these 3 steps are summarized in Table 4. There are 153 lines that are removed in this portion of the pipeline. These lines could be removed either because any possible IDs don't have another ID for the same element at a close enough redshift or because they don't have the right ionization energy for the state of

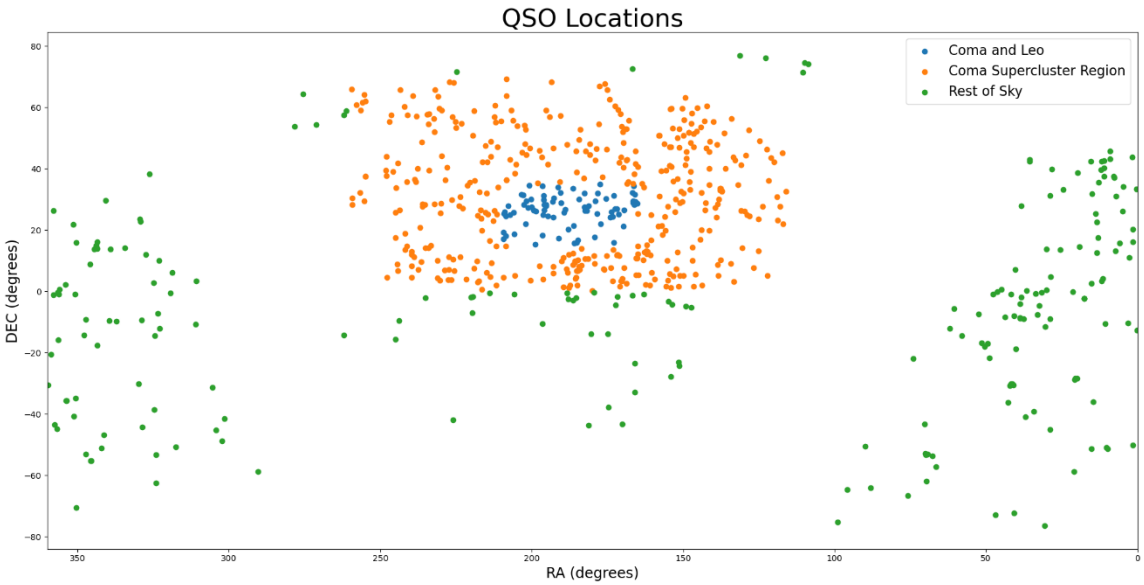

Figure 21 Sky locations for analyzed QSOs.

Hydrogen detected at that redshift. There is not a good way to determine if these lines should be removed from the dataset without additional data sources. This is out of the scope of this work. The removal of these may introduce errors up to the order of half a percent in the final calculation.

At the end of the pipeline slightly under half of the remaining lines have only one possible ID, a smaller fraction have 2 possible IDs, and an even smaller fraction have more than 2 possible IDs. This is good! For the lines that have multiple possible remaining IDs, there is not a good way to take that down to 1 using only results from this data set. (There may be other data sources that could augment this data set that would be useful, but they are beyond the scope of this study. More about this can be found in Chapter V.)

Since there is no perfect way to whittle the possible IDs down to 1, I will instead look at possible extremes and the distribution in between the extremes to set limits on the distribution of extragalactic baryons. The extremes will put the greatest possible bounds

	Coma and Leo	Coma Supercluster Region	Rest of Sky	Total
Sightlines	78	393	217	688
Possible IDs at beginning of pipeline	97,825	504,926	306,503	909,798
Unique Lines at start of pipeline	3,110	16,372	10,635	30,138
Unique Lines at end of pipeline	3,097	16,291	10597	29,985
Possible IDs at end of pipeline	7,028	36,461	22,573	66,062
Lines with unique ID	1,419	7,367	4,870	
Lines with 2 possible IDs	908	4,864	3,164	
Lines with more than 2 possible IDs	770	4,060	2,563	

Table 4 An overview of the statistics for analyzed spectra, broken down into each of the three regions of interest.

on this method, while the distribution will give an idea of what a more likely distribution will be.

If a line has only 1 possible ID, it isn't necessary to make any discrimination, so that ID is taken as the correct one. For each line with more than 1 possible ID I take either the highest column density of any possible ID, the lowest column density of any possible ID, or pick one randomly. Taking the highest density for each line gives the strict upper bound for the column density. Similarly, taking the lowest density gives the strict lower bound for the column density. To get a distribution of the possible distributions of material, I take a random ID for each line and then repeat the process many times to build up statistics.

4.2 Calculating the Mass of Detected Baryons

With having a process to give each spectral line a single ID, I can now calculate an average density of extragalactic baryons in the Coma and Leo neighborhood, in the extended Coma Supercluster region, and throughout the local universe. However, it's only possible to measure the column density of material, not the volume density. The column density is the product of the volume density and the distance along the line of sight; therefore, the volume density is the column density divided by the path length that the light traveled through the IGM. For the Coma and Leo region, this path length is given by defining anything with a redshift between 0.016 and 0.030 as part of the supercluster, for a path length of $1.84 \cdot 10^{26}$ cm for each sightline. For the Coma Supercluster region and the all sky search the path length was given by the distance from a redshift of 0 to 0.2, or $2.54 \cdot 10^{27}$ cm for each sightline[4].

The size of an individual cloud of intergalactic gas is not well constrained, however, since it is the total mass of baryons that's interesting the average density over the entire volume is useful. Adding up the column density along all lines of sight in each region and dividing by the total path length[5] of those sightlines gives the average density of material in the IGM in those regions. The estimated upper and lower bounds on the total mass of extragalactic baryons based on the three regions of interest are shown in Table 5. The first two columns show the lower and upper bounds on the total mass of matter in the universe under the assumption that the overall density of the universe is the same as in the sample volume. This analysis primarily detects metals, that is elements

[4] These distances were calculated with $H_0 = 69.6$ km/s/Mpc, $\Omega_M = 0.286$, and $\Omega_{vac} = 0.286$ (Bennett et al. 2014) using Ned Wright's Cosmology Calculator (Wright 2006).
[5] The total path length is the path length an individual sightline travels through the region of interest multiplied by the number of sightlines in that region.

that are heavier than Hydrogen. (In the low estimates, there are trace amounts of Hydrogen detected, but they make up around half a percent of the total mass of detected baryons.) Typical abundances of Hydrogen to Helium to metals in the Milky Way Galaxy are found to be 74/24/2. This means that metals only account for at most 2% of the total mass of extragalactic baryons. If material in the IGM has been processed and then removed from galaxies, then it should have a similar distribution. Under this assumption, I have multiplied the estimated mass of metals by 50 to get the total mass of extragalactic baryons. This is done in the final column of Table 5.

	Low Estimate Metals (g)	High Estimate Metals (g)	Estimated Total Baryonic Mass (g)	Estimated Total Baryon Fraction
Coma and Leo	$6.42 \cdot 10^{53}$	$1.44 \cdot 10^{54}$	$3.21 \cdot 10^{55}$ - $7.2 \cdot 10^{55}$	21% - 48%
Coma Supercluster	$5.20 \cdot 10^{52}$	$2.98 \cdot 10^{53}$	$2.60 \cdot 10^{54}$ - $1.49 \cdot 10^{55}$	1.7% - 10%
All Sky	$4.88 \cdot 10^{52}$	$2.74 \cdot 10^{53}$	$2.44 \cdot 10^{54}$ - $1.37 \cdot 10^{55}$	1.6% - 9.1%

Table 5 Summary of the estimated metal baryonic mass and total inferred baryonic mass from each of the 3 sample volumes.

The expected baryon mass can be found from looking at the Friedmann equations and measurements of the baryon density parameter. In the Friedmann equations, the critical density is

$$\rho_c = \frac{3\,H^2}{8\pi G} = \frac{3\left(100\frac{km}{s}Mpc^{-1}\right)^2}{8\pi G}h^2$$

with H the Hubble constant and G the universal gravitational constant. The baryon density is the critical density multiplied by the baryon density parameter.

$$\rho_B = \Omega_{\mathrm{B}}\rho_c = \frac{3\left(100\frac{km}{s}Mpc^{-1}\right)^2}{8\pi G}\Omega_{\mathrm{B}}h^2$$

There are multiple measurements of the baryon density parameter. For this analysis, I will use the weighted average of WMAP and Planck of $100\Omega_b h^2 = 2.239$ (Hinshaw et al. 2012; NASA/LAMBDA Archive Team 2022; Planck Collaboration et al. 2018). With this value of the density parameter, the baryon density is $4.21 \cdot 10^{-31}$g/cm^3. Multiplying by the volume of the observable universe gives a total mass of $1.5 \cdot 10^{56}$g. This means that the estimate of baryon mass in extragalactic clouds is between 1.6 and 48% of the total mass of the universe.

The total mass of extragalactic baryons is highest when only looking at absorption through Coma and Leo, is lower when looking through the Coma Supercluster region, and lower still when looking throughout the whole sky. The detected density being higher in the Coma and Leo region could be the result of the density actually being higher or could be the result of a systematic preference for IDing lines with higher column densities. If the true density of intergalactic clouds was not higher in the Coma and Leo clusters, when making the estimate by selecting the highest column density there would be no reason to expect that the highest density ID would be in Coma and Leo. This would have the effect of lowering the estimate of the mass of the region. Since the density is significantly higher, it would instead support that the density is truly higher in the cluster. This is not surprising.

Figure 22, Figure 23, and Figure 24 show histograms of 500 trials for randomly picking an ID for each line that has more than one possible ID and then calculating the mass of extragalactic baryons in the universe based on that sample. The mean and standard deviation for each distribution is given in the first two columns of Table 6. The

last two columns in the table give the estimated universal mass and contribution to the overall baryon density. For the Coma Supercluster region and the all sky search, the distribution of total masses from these trials are much narrower than the broad constraints placed by the minimum and maximum estimates. The distribution for Coma and Leo is bimodal, which each mode laying close to the extreme bounds from the high and low metal estimates. This makes the mean less meaningful. It is included in Table 6 for to provide the same comparison as the other regions, but the entire distribution should be kept in mind while drawing results.

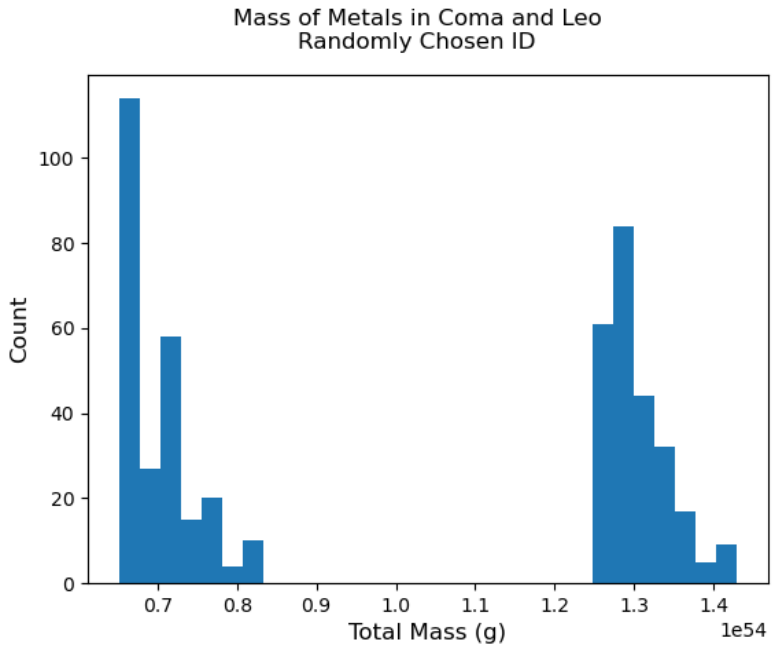

Figure 22 Histogram of the total mass estimated from selecting an ID at random for the Coma and Leo neighborhood.

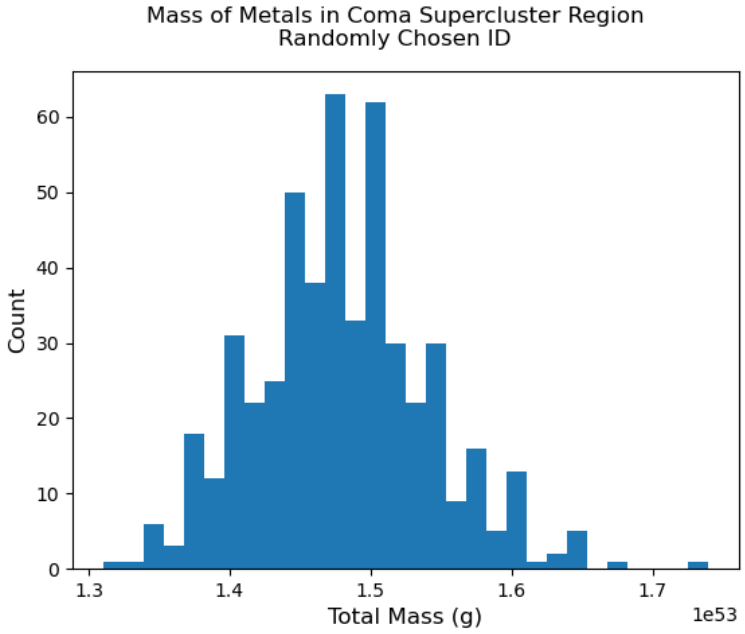

Figure 23 Histogram of the total mass estimated from selecting an ID at random for the Coma supercluster neighborhood.

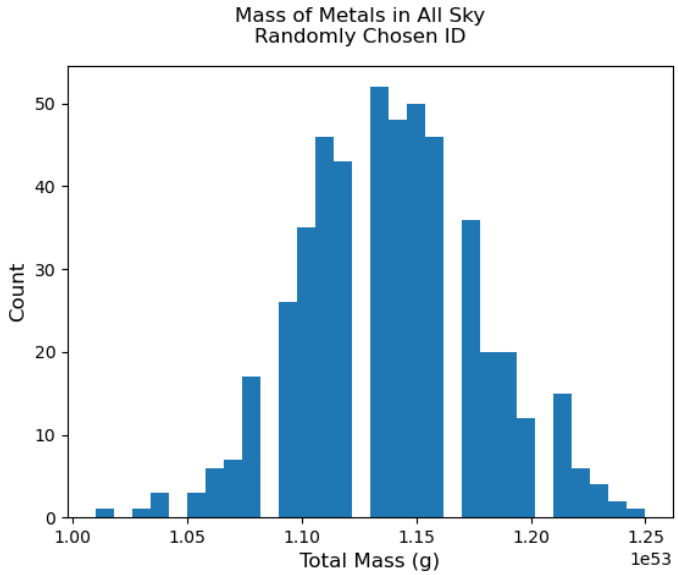

Figure 24 Histogram of the total mass estimated from selecting an ID at random for the all sky search.

	Mean (g)	Standard Deviation (g)	Total Estimated Mass (g)	Baryon Fraction
Coma and Leo	$1.00 \cdot 10^{54}$	$3.07 \cdot 10^{53}$	$5.00 \cdot 10^{55}$	33.3%
Coma Supercluster	$1.48 \cdot 10^{53}$	$6.22 \cdot 10^{51}$	$7.40 \cdot 10^{54}$	4.93%
All Sky	$1.14 \cdot 10^{53}$	$3.90 \cdot 10^{51}$	$5.70 \cdot 10^{54}$	3.65%

Table 6 Mean and standard deviation of the distributions for 500 random samples of randomly assigning IDs to lines with more than one possible ID.

The Coma and Leo region's randomly picked IDs have only a slightly smaller range than wide constraints given by the minimum and maximum estimates. These estimates are dominated by Fe II, which accounts for around 90% of the total mass in these estimates. Since Iron is only produced in stars that are heavily processed, it seems that the Coma and Leo environment is either extremely processed, or there is a selection effect in the pipeline that significantly over selects Iron over other elements.

The high estimate of the total mass of the universe is also problematic. Having enough baryons to account for 48% of the mass of the universe would over solve the Missing Baryon Problem and would create an Excess Baryon Problem. Both of these are reasons to think that Iron would not be as prevalent as detected. The solution to this problem needs either more data or data from other sources to know if there is a significant amount of Iron in the IGM, or if there is some other element that is being wrongly identified as Iron. This will have to be left to other studies.

In the supercluster region and the all sky search, the problem of too much Iron also exists, but is smaller. For both regions, Fe II contributes around 30% of the total mass. This seems to be too high but is less problematically high. Again, further work is

needed to determine if this is an incorrect fitting of the data or if there is something interesting going on.

4.3 Column Density vs Galaxy Density

Figure 25 and Figure 26 show the total column density for lines identified in the Coma and Leo clusters versus the local galaxy density. The galaxy density was calculated by taking a cube centered on the earth that extended to a redshift of 0.2 along any axis and separating it into 200 bins along each axis. (For a total of 8 million bins) The total number of galaxies in each bin was then divided by 557.64 Mpc^3 to find the local density in that bin. The total column density for each element was binned in the same way.

There are bins closest to $z = 0$ have been excluded from the plot area. These bins contain galaxies in the close vicinity of the Milky Way, where we are able to observe more dwarf galaxies. There are also several hundred galaxies that are slightly blueshifted in NED, putting them into the smallest bin. Since these galaxies are moving towards the Milky Way, their redshift doesn't give the correct information about their radial distance. Most of the elements that are detected have very slight positive correlations. The overall column density also seems to increase very slightly with galaxy density, though it is not statistically significantly different from not correlating.

Figure 27 and Figure 28 expand the same analysis to the Coma supercluster region. Figure 29 and Figure 30 show the same plots with the horizontal range narrowed to focus on low density areas where most of the detections are. Figure 31 and Figure 32 extend the analysis to the entire sky. Figure 33 and Figure 34 show the all sky plots with the horizontal range narrowed to focus on the low density areas.

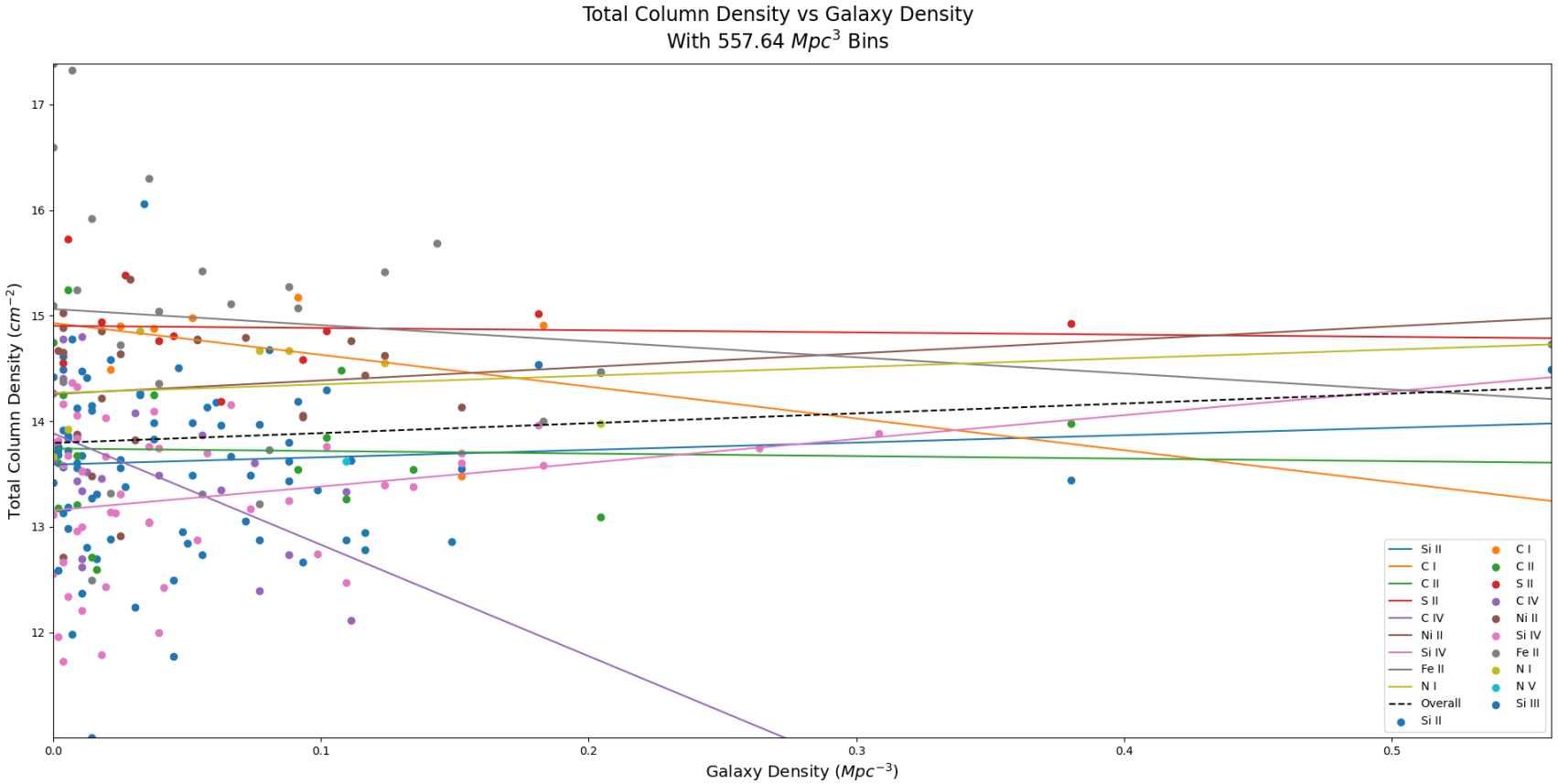

Figure 25 Total binned column density versus local galaxy density in the Coma and Leo neighborhood. Line ID decided by picking the highest column density.

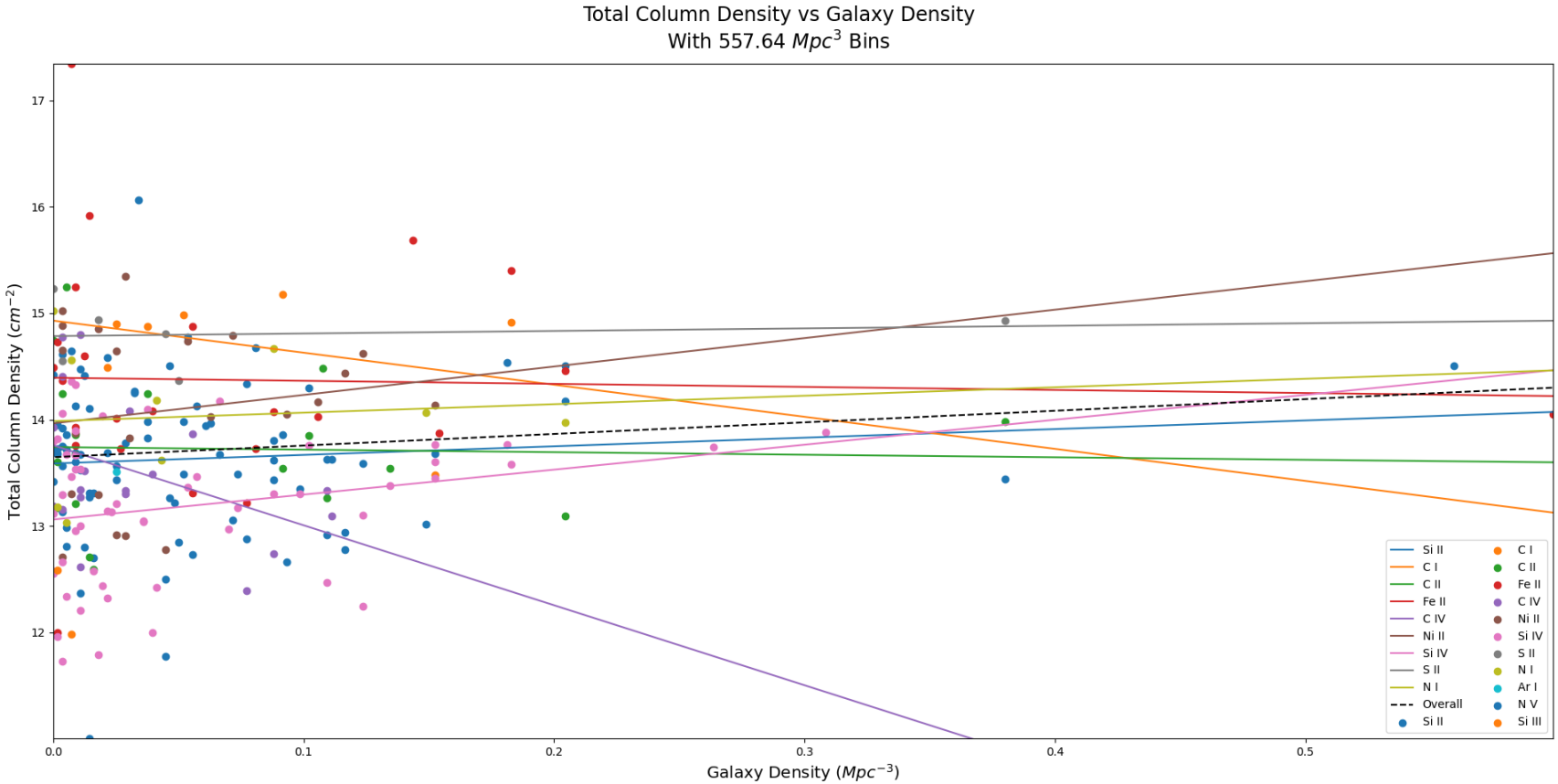

Figure 26 Total binned column density versus local galaxy density in the Coma and Leo neighborhood. Line ID decided by picking the lowest column density.

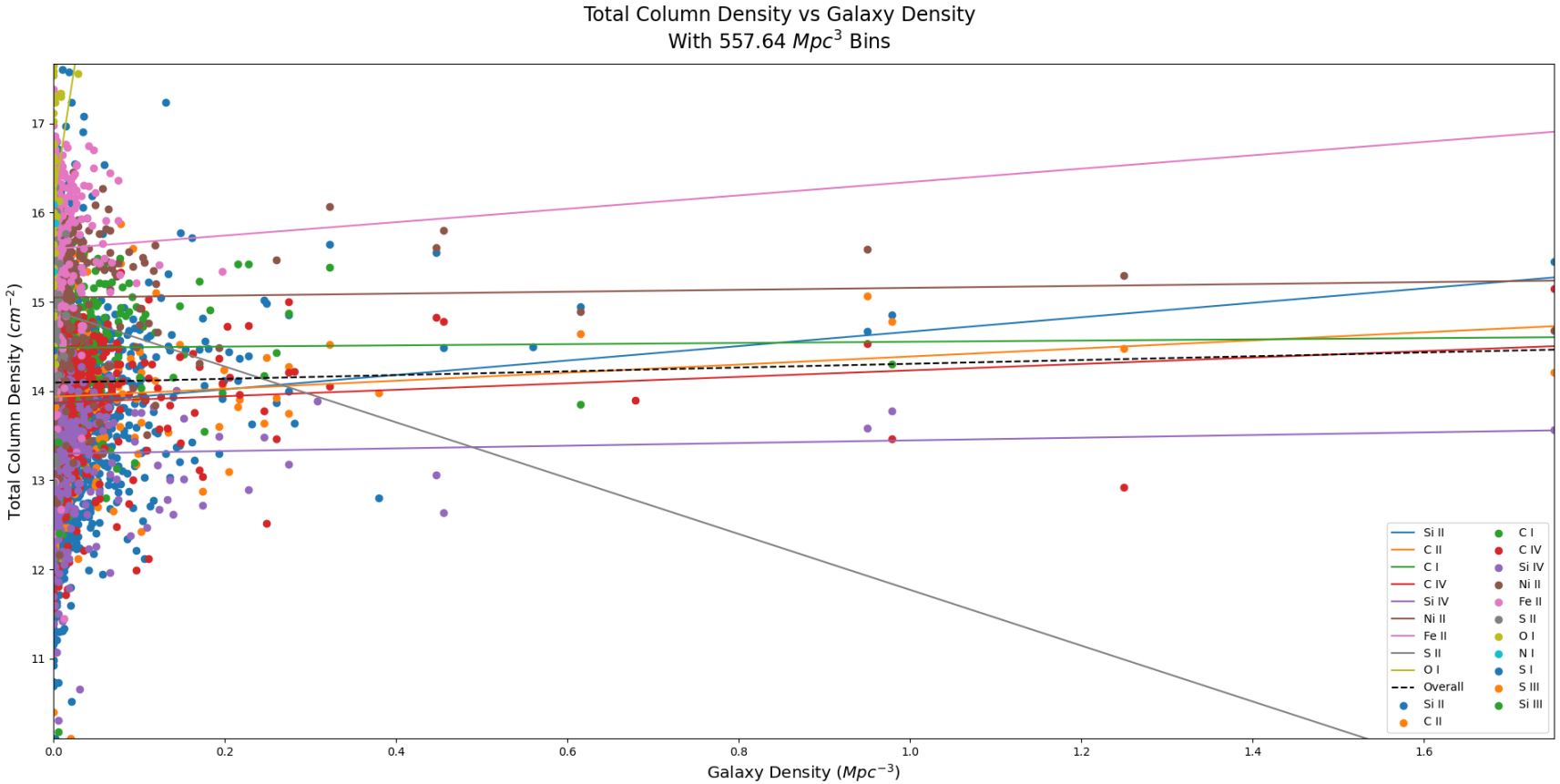

Figure 27 Total binned column density versus local galaxy density in the Coma supercluster neighborhood. Line ID decided by picking the highest column density.

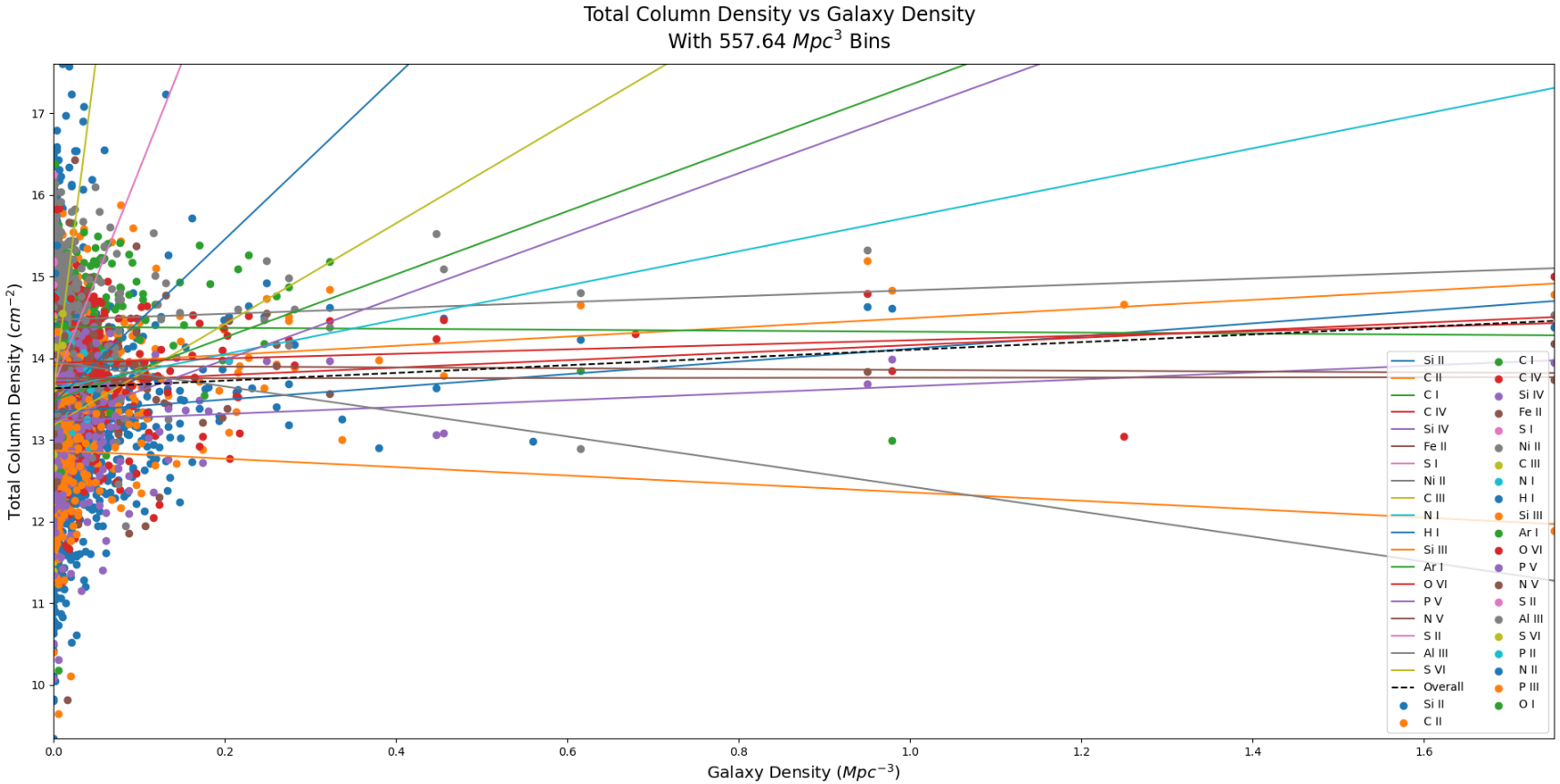

Figure 28 Total binned column density versus local galaxy density in the Coma supercluster neighborhood. Line ID decided by picking the lowest column density.

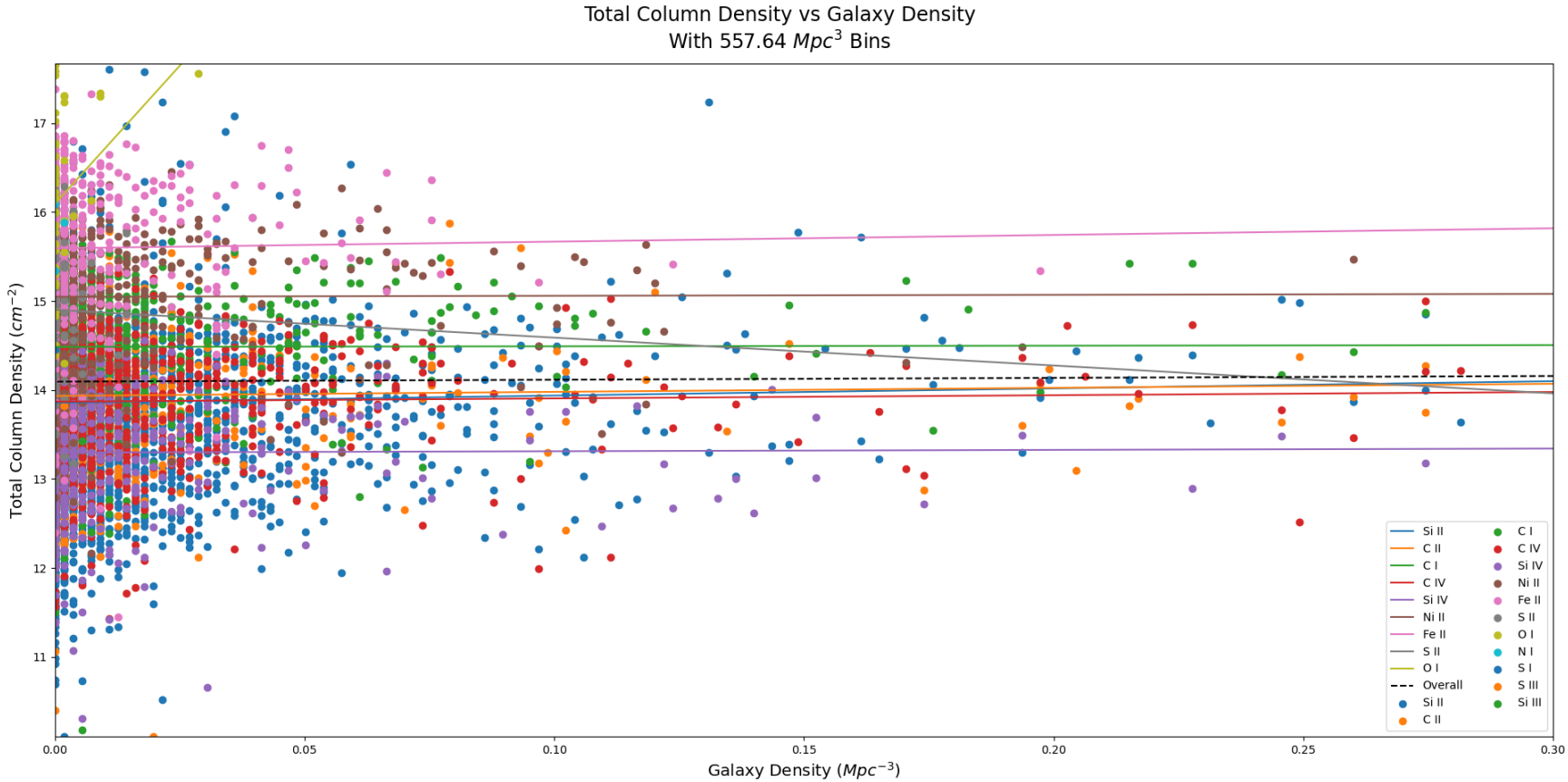

Figure 29 Total binned column density versus local galaxy density in the Coma supercluster neighborhood. Line ID decided by picking the highest column density. Zoomed in on low density to show detail.

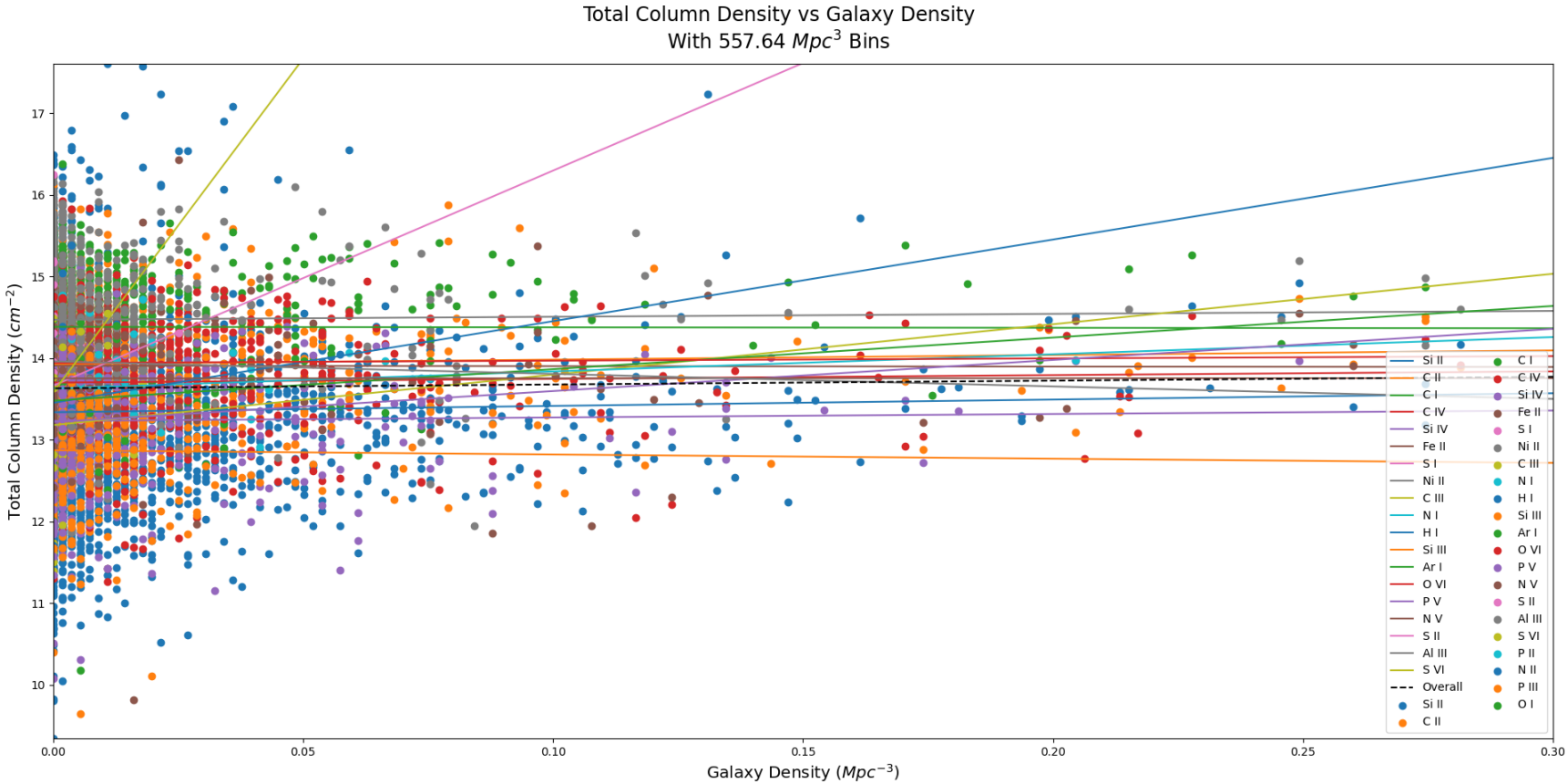

Figure 30 Total binned column density versus local galaxy density in the Coma supercluster neighborhood. Line ID decided by picking the lowest column density. Zoomed in on low density to show detail.

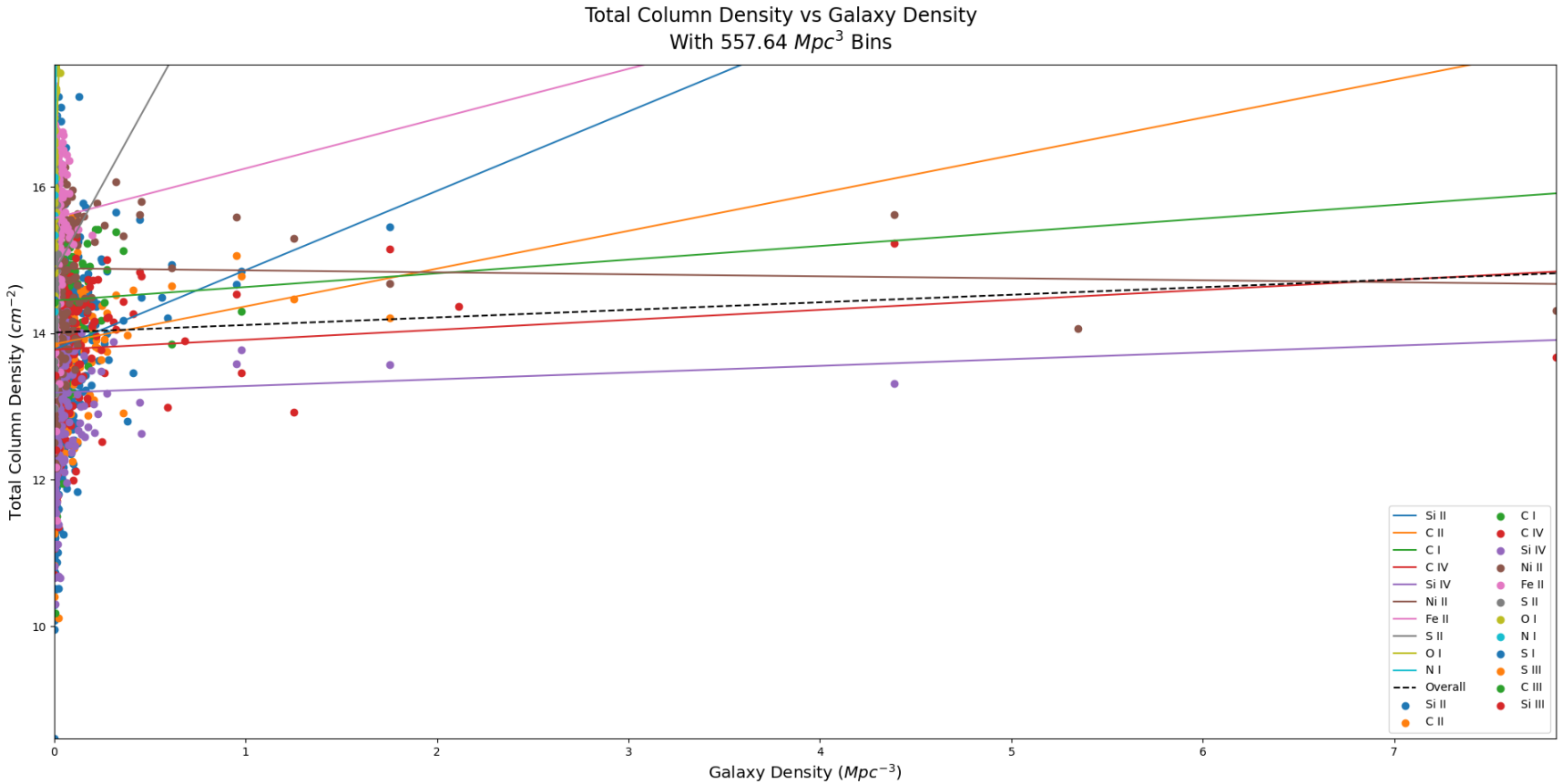

Figure 31 Total binned column density versus local galaxy density for the all sky search. Line ID decided by picking the highest column density.

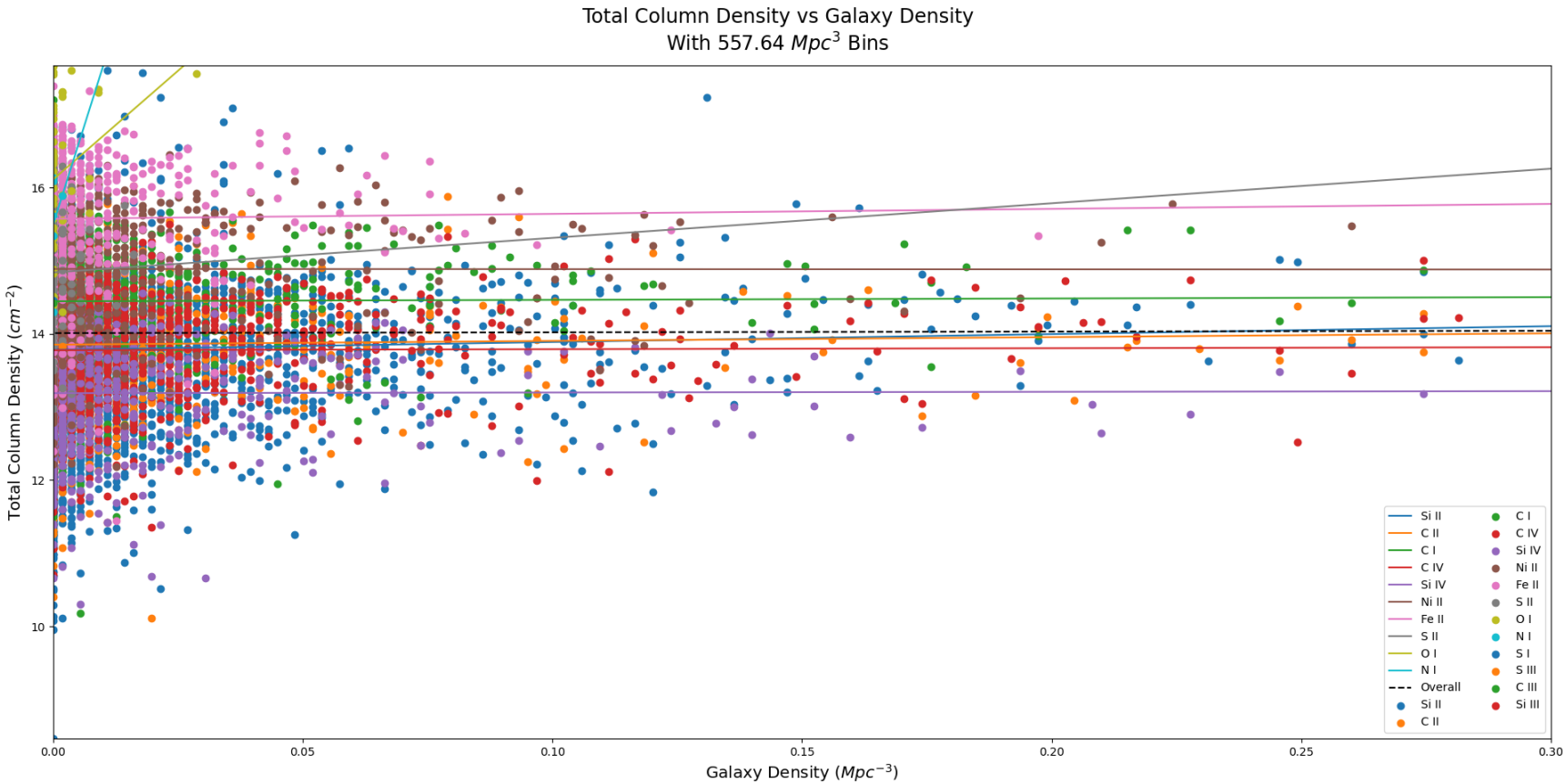

Figure 32 Total binned column density versus local galaxy density for the all sky search. Line ID decided by picking the highest column density. Zoomed in on low density to show detail.

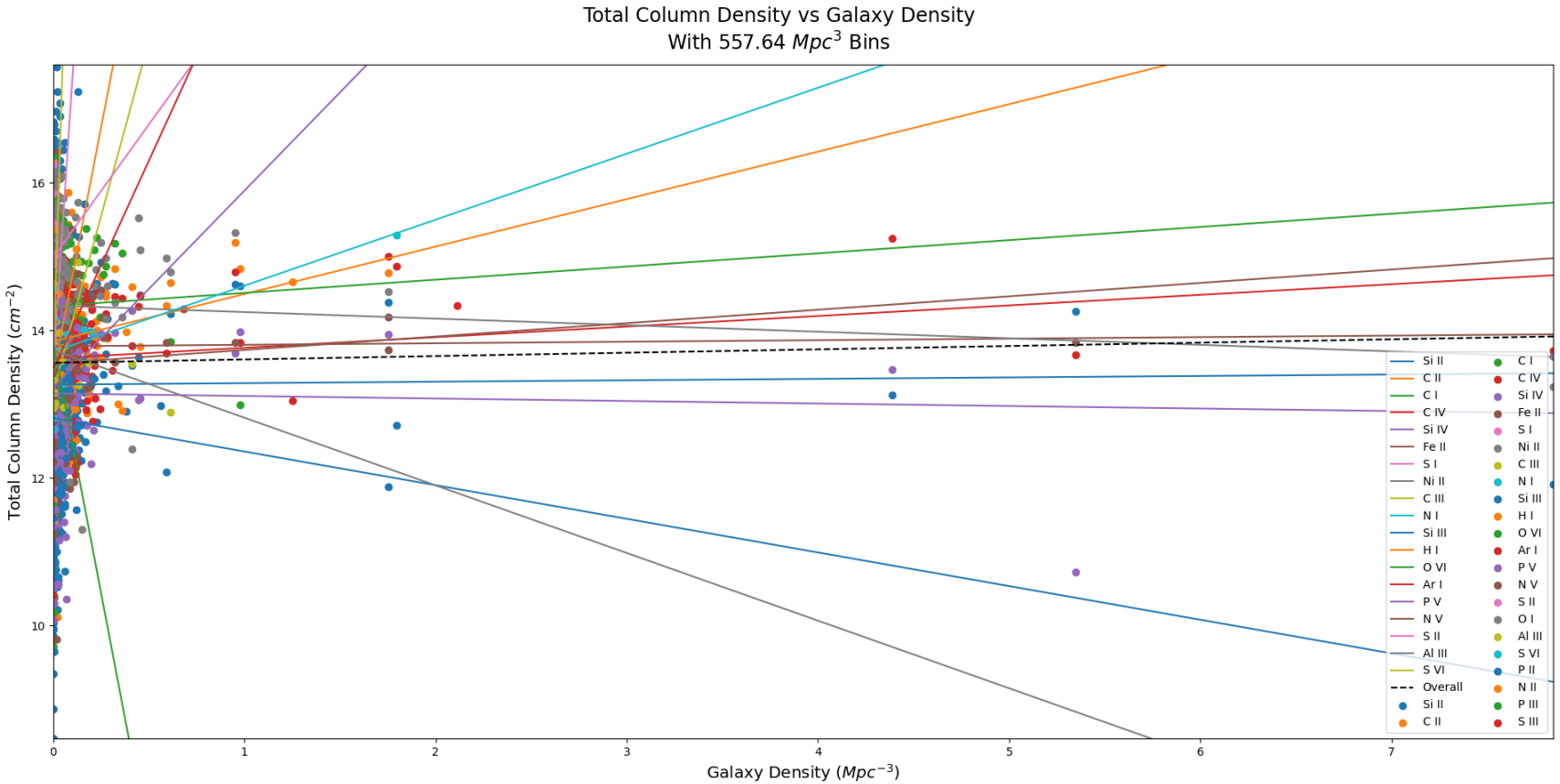

Figure 33 Total binned column density versus local galaxy density for the all sky search. Line ID decided by picking the lowest column density.

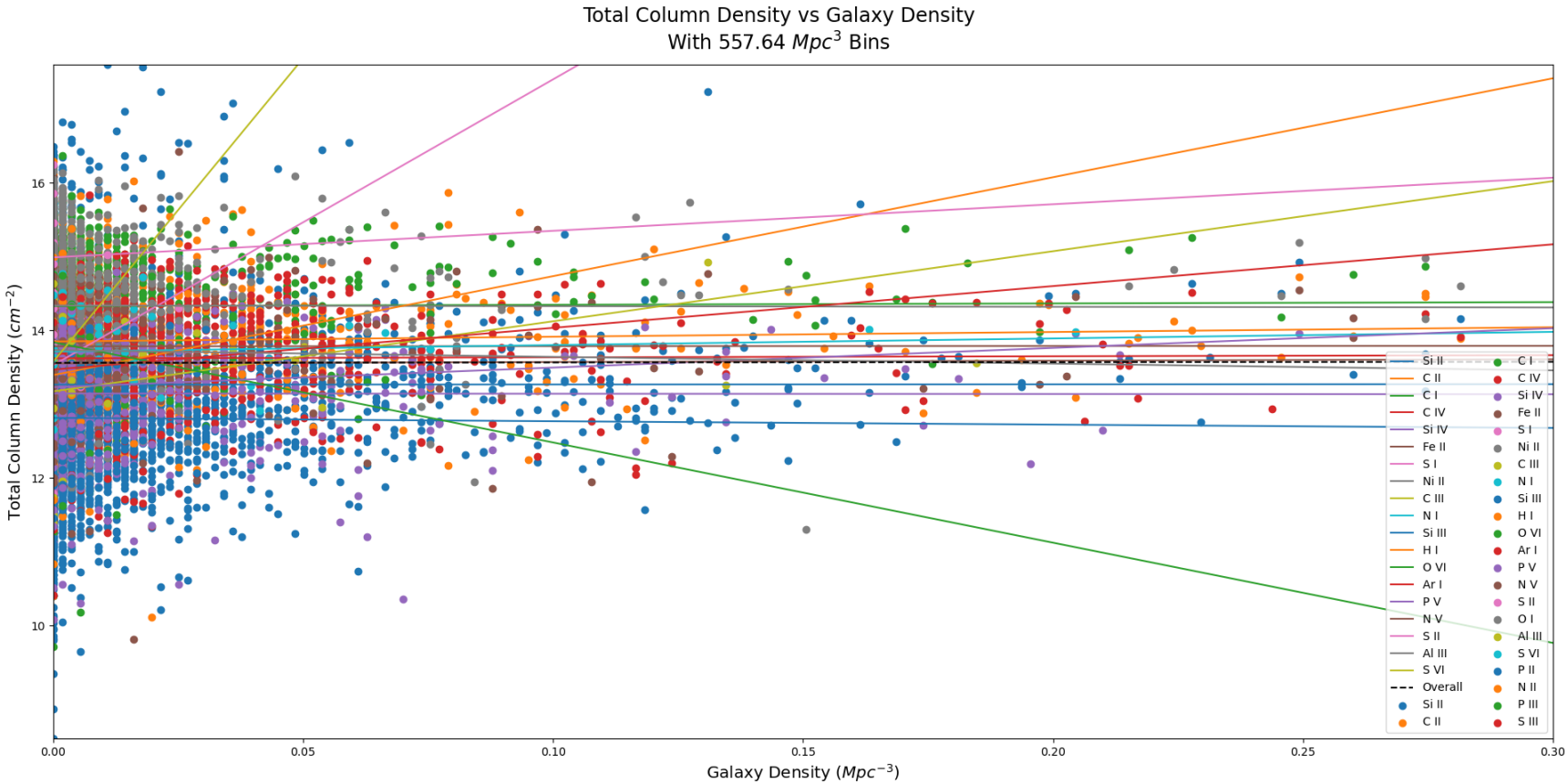

Figure 34 Total binned column density versus local galaxy density for the all sky search. Line ID decided by picking the lowest column density. Zoomed in on low density to show detail.

There is at best a small increase in detected column density with increasing galaxy density, but it is not statistically significant in most of the fits. All of the fits have low correlation coefficients. This suggests either an error in analysis, an observation bias, or that the population of extra-galactic clouds does not correlate well with the local galaxy density. For individual species, there are few enough data points that small fluctuations dominate the fit. This can be seen with the large number of fit lines that are at very steep slopes (both positive and negative). Overall, I am not convinced that there is sufficient data in this data set to make good judgements about the density of material as a function of galaxy density. With more data from UV and other sources, as well as better surveys of galaxies, this type of analysis would be more useful.

4.4 Distance to Nearest Galaxy

One interesting question and a possible way to validate these detections is to look at the distribution of distances identified clouds are from the nearest galaxies. We can then compare this to how close random points in space are to that same set of galaxies. If the clouds of gas are recently liberated from galaxies and are only transient in the intergalactic medium, then the distribution of distances should be shifted down towards lower distances compared to a random sample.

Figure 35 shows histograms for the distance from all detected clouds in our dataset to the nearest galaxy. The list of galaxies is taken from NED[6]. The mean and standard deviation for each of these distributions is listed in Table 7. The distribution for all 3 regions does not vary significantly for both the high and low estimates of column densities in that region. However, the mean distance does differ significantly between the

[6] The NASA/IPAC Extragalactic Database (NED) is funded by the National Aeronautics and Space Administration and operated by the California Institute of Technology.

regions. In the Coma and Leo sample, the mean distance is more than a factor of 2 lower than in the all sky sample.

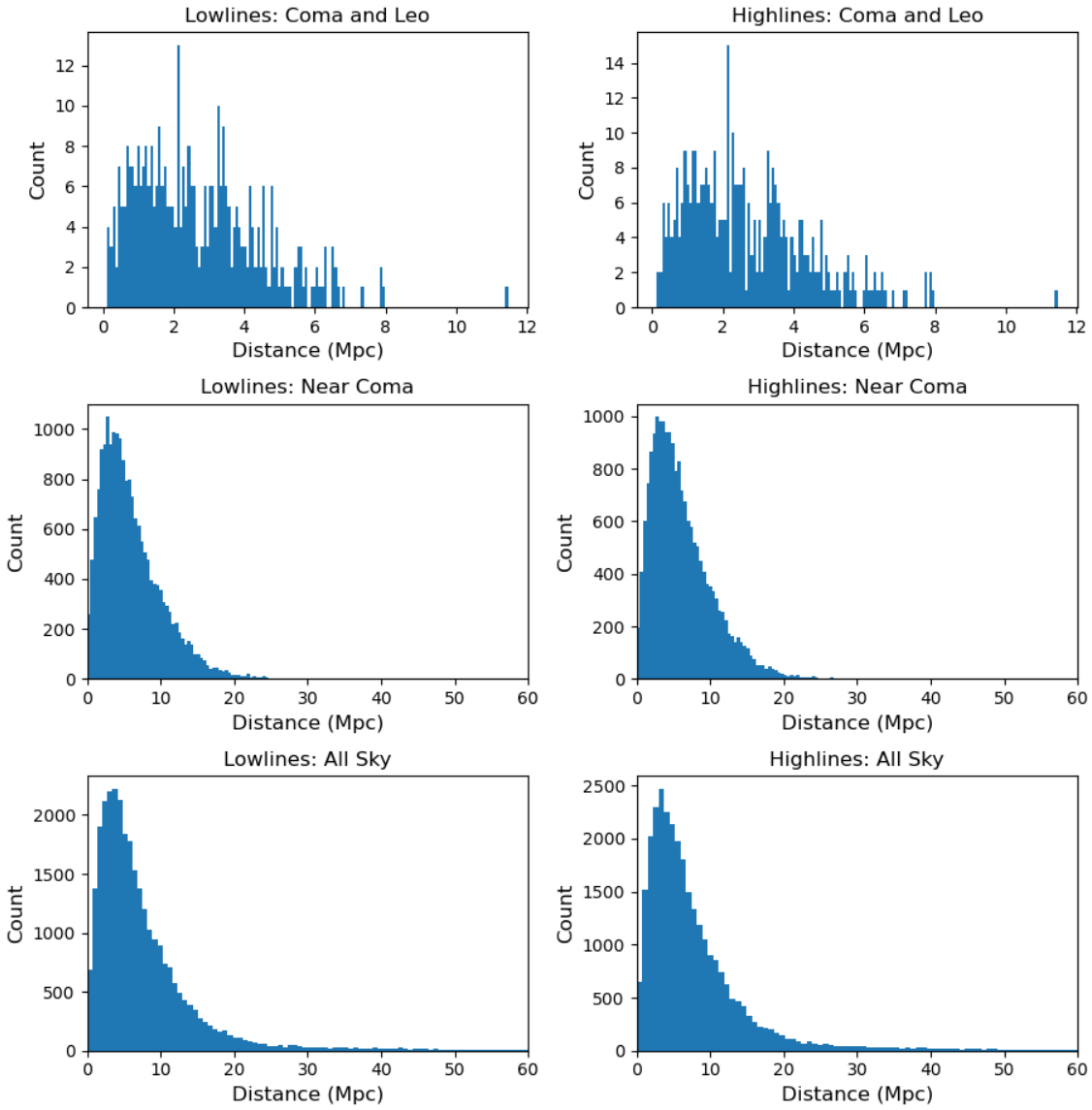

Figure 35 Histograms of the distances from detected clouds of gas to the nearest galaxy in NED. The left column is for the low bound estimate and the right column is the upper bound estimate.

	Observed		Randomly Generated	
	Mean (Mpc)	Standard Deviation (Mpc)	Mean (Mpc)	Standard Deviation (Mpc)
Coma and Leo (Highlines)	2.70	1.74	2.82	1.63
(Lowlines)	2.71	1.74		
Supercluster (Highlines)	6.21	4.42	14.0	17.7
(Lowlines)	6.08	4.45		
All Sky (Highlines)	8.06	8.33	13.9	17.4
(Lowlines)	7.854	8.22		

Table 7 Descriptive statistics for the distance clouds of gas are from the nearest galaxy. For observed lines shows both the lower and upper bound estimates. For randomly generated clouds shows one overall statistic.

To put the observed distance from detected clouds of gas to the nearest galaxy in perspective, I have sampled random points in space and found the distance from those points to the nearest galaxy in each of the regions. To ensure that the points were uniformly random, I first selected points in a cube that corresponded to all RA and DEC and overshoots on redshift. (The maximal extent along any cardinal direction was set to a redshift of 0.2, but this allows some points at a further distance if they don't lie along one of the cartesian axes) I then converted the cartesian coordinates into spherical coordinates and kept all of the points that satisfied the RA, DEC, and redshift of the 3 regions of interest. For the Coma and Leo region, I selected points until there were more than 10,000 points in that region to get a smoother distribution of distances. For the supercluster region and the all sky search I used 100,000 points. The Coma and Leo region had a smaller number of points but was still large enough to have a reasonably

smooth distribution. The larger number of points for the other 2 regions was probably unnecessarily large but run time for both was quick enough that the large size of the sample was not a worry. Results for the Coma and Leo neighborhood, the Coma supercluster neighborhood, and the entire sky are show in Figure 36, Figure 37, and Figure 38 respectively. The mean and standard deviation of these distributions are listed in Table 7.

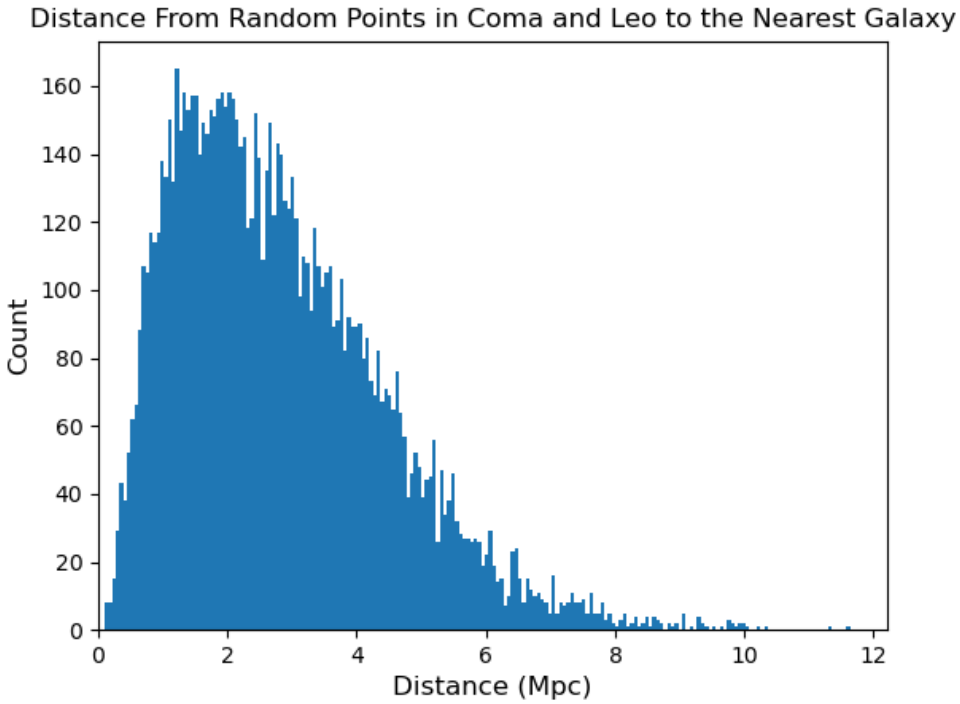

Figure 36 This shows a histogram of distances from randomly sampled points in the Coma and Leo region to the nearest galaxy. It has an average of 2.82 Mpc and a standard deviation of 1.63 Mpc.

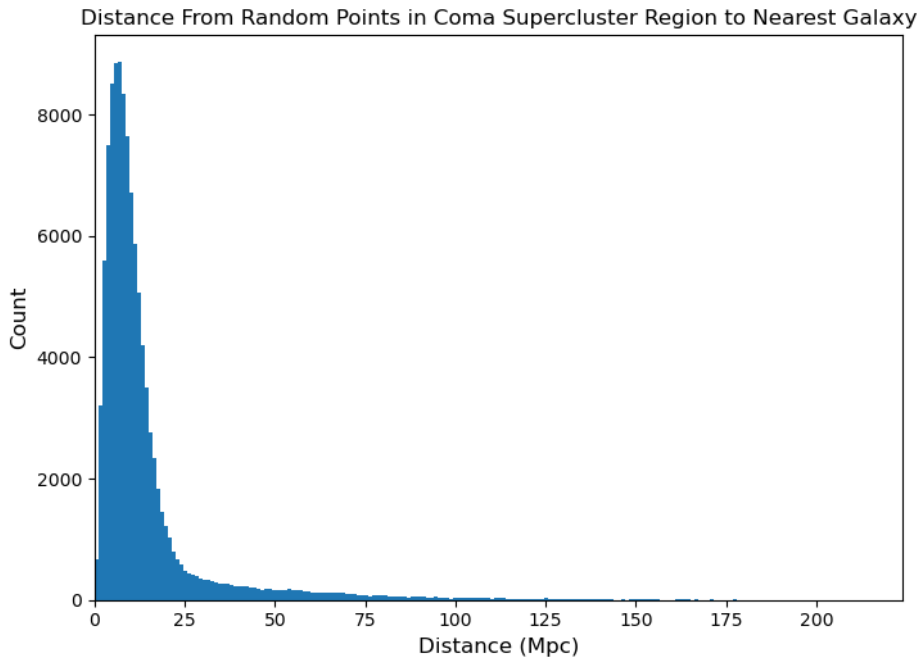

Figure 37 This plot shows a histogram of the distance from points in the extended supercluster region to their nearest galaxies. It has a mean of 14.0 Mpc and a standard deviation of 17.7 Mpc.

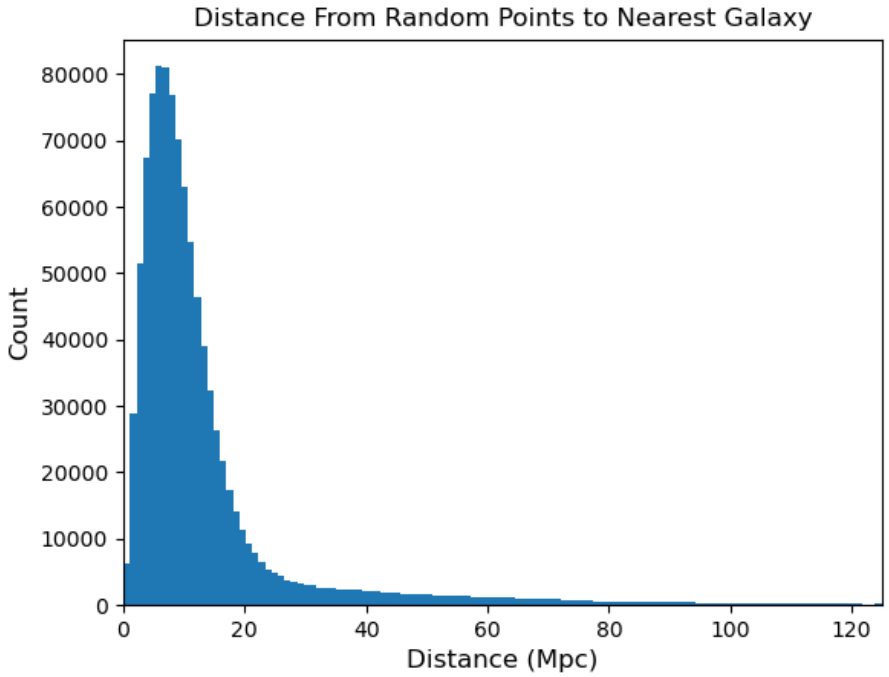

Figure 38 This plot shows a histogram of random points within a redshift of 0.2 to the nearest galaxy. The mean 13.9 Mpc and the standard deviation is 17.4 Mpc.

The distribution for the whole sky has some intrinsic error. When selecting points, I used points from anywhere in the sky closer than a redshift of 0.2. Because it is hard to observe through the disk of the Milky Way, the number of known galaxies has a heavy bias. In directions that aren't through the disk of the Milky Way, there are more than an order of magnitude more known galaxies than there are in directions through the Milky Way. This would naively mean that a random point in the regions that we can't see galaxies well should have a larger distance than for those in regions where we do see galaxies well. However, in this sampling the distance for the supercluster region (which lies away from the plane of the galaxy) and the whole sky have similar averages and standard deviations. (And in fact, the all sky is marginally smaller than the supercluster average.) Because there doesn't seem to be a strong issue with using all of the sky instead of the whole sky minus areas through the Milky Way, I will use the full sky distribution in the analysis.

If clouds of gas were held in the gravitational potentials of galaxies, it would be expected that the detected distance would be less than the distance for a randomly distributed system. If the distribution of distances is nominally the same as for a random distribution, that would mean that the clouds of gas are not gravitationally bound by individual galaxies but are instead bound by a larger cluster or not bound at all. If the average of the detected distribution was larger than the random distribution that would mean that the clouds are preferentially found away from galaxies, which would either mean that there was some strong observational bias or some methodological error in the analysis.

The distributions for these three sample areas all have smaller average distances than the randomly sampled points in the same area. In the Coma and Leo clusters this average is very slightly less, but in the supercluster and all sky regions is around half of the random sample. This means that the Coma and Leo clusters have clouds of gas that are very close to randomly distributed, giving good evidence that they trace out the intracluster potential and are not bound to individual galaxies. As we go away from the dense center of the Coma and Leo clusters, the average distance rises as would be expected, but it rises less quickly than it would if the clouds of gas continued to be free from the gravitational influence of individual galaxies.

A typical galaxy has a radius on the scale of kiloparsecs, with Milky Way sized galaxies having a radius between 50 and 100 kpc. In the Coma and Leo clusters, the average cloud of gas detected is on the order of 20-30 galaxy radii away from the nearest galaxy. In the supercluster region and all sky the average detected cloud is even further away, on the order of around 100 galaxy radii. These large distances from galaxies provide good evidence that these clouds are not tightly bound to a galaxy.

CHAPTER V

DISCUSSION AND FUTURE WORK

The second paragraph in 5.1 and the final paragraph of 5.2 were written by Greg Bothun. I wrote the rest of the chapter and edited the entire chapter.

5.1 Does This Solve the Missing Baryon Problem?

The results from this analysis provide evidence of a large number of baryons in intergalactic gas, but there is likely not enough material to account for all of the missing baryons. The highest estimate, that calculates that the extragalactic material has 50% of the mass in the universe is higher than the estimated missing baryon mass. However, this is very likely a gross overestimate as much of the estimated mass is dominated by relatively few sightlines. The lowest estimate that has the extragalactic mass around 0.5% are also likely underestimates. These low estimate identifications are dominated by the assumption that weaker lines are much more prevalent and have identifications of species that are not commonly seen in other studies.

But perhaps we are being overly dismissive here. We began this investigation with the explicit assumption that many of the missing baryons may be the result of various baryon removal processes that affect galaxies. The more highly clustered the galaxies are, the more repeated interactions will occur which should increase the number of baryons (mostly gas) that are removed. The results shown in Table 6 are therefore intriguing. Here we have straightforward evidence that, indeed, the missing baryon content in the richest structures (Coma and Leo) have significantly higher detection of a missing baryon component. Indeed, the galaxy density is much higher in Coma and Leo

than it is in the extended Coma Supercluster or in the general field. Thus, a prime result of this thesis is the detection of missing baryons in regions of high galaxy density.

Overall, this work gives evidence that there are large collections of extragalactic baryons that have been processed. The error bars on how much material is contained in these reservoirs are large, with upper limits with the intergalactic medium consisting entirely of processed material and the lower limits being that the intergalactic medium contains processed material, but only a small fraction of this material has been processed within galaxies. Further work is needed to put better constrains on this.

One additional study that should be done on this dataset is to carefully look carefully to see how well these absorption lines would be fit by Lyman alpha absorption. Since this survey was only looking for absorption between a redshift of 0 and 0.2, there is the possibility that the absorption was caused by a cloud outside of that redshift range. In particular, this could be a cloud of Hydrogen. This complication leads to an interesting question that is beyond the scope of this study.

5.2 The Distribution and Density of Material

This analysis has detected evidence of large numbers of extragalactic baryons in QSO spectra throughout the whole sky. The sightlines that intersect the Coma and Leo clusters give an estimate of the total mass of extragalactic baryons that is between a factor of around 2.5 to 30 higher than the estimate from sightlines from the entire sky. Additionally, the material in the all sky search is a factor of ~3 times farther away from the nearest galaxy compared to the material when just looking in the Coma and Leo neighborhood. This is consistent with expectations. It is expected that the cluster environment would have more interactions between galaxies that can strip material into

the intergalactic medium. Further, because the galaxy density is higher in a cluster, an extragalactic cloud is more likely to be close to some galaxy, even if it isn't in the gravitational potential of that galaxy. In the field where the galaxy density is much lower, clouds of gas are much closer to galaxies than would be expected if they were randomly distributed.

This evidence is somewhat consistent with predictions from simulations and prior observations of the Coma cluster. Simulations estimate ~40% of material could be stripped away in mergers in a cluster, whereas only ~20% could be stripped away in a merger in the field (Mihos 2004). The tidal tails of a typical Milky Way size galaxy in these simulations can have a velocity of 0.2 Mpc/Gyr for the part of the tail 200 kpc from the center of the galaxy. Given the universe is order 10 Gyr old, this would mean that material should be at most around 2 Mpc from its original galaxy. Our observed distribution in the all sky search is, on average, a factor of 4 further than this. This could be from material that was further than 200 kpc after 0.5 Gyr, and thus had a larger velocity. In the Coma and Leo cluster, the average distance is little over 2 Gpc. This is at least the right order for the distance that material is from galaxies. For clouds of material in the cluster environment, the simulations seem to agree with the observations. For material in lower density regions of space the simulations give an underestimate of the observed distances. More work is needed to determine if there is a something missing from the simulations or if the true distribution of intergalactic clouds is different than measured by this work.

Now, the issue becomes how representative are Coma and Leo of the general universe. This is a difficult question to answer. Certainly, within the z =0.1 volume of the

Universe, Coma is a unique rich structure. However, it has been interpreted that >90% of all galaxies with redshift between z = 0.2 and z = 0.5 are part of larger structures with overdensities similar to Coma (Small et al. 1999). This basic qualitative feature of the galaxy distribution seems to always be the case as data from more and more redshifts is acquired. In this case, most of the galaxies are contained in environments conducive to baryon removal. We further expect that during the cluster formation process (anywhere from z = 1 to z = 4) these removal processes would be more robust. Hence, a reasonable conclusion to draw from this study is that intergalactic baryons, resulting from multiple galaxy interactions, contained in rich galaxy environments can make up a substantial fraction of the missing baryon problem if this is the most common kind of environment in which galaxies reside.

5.3 Concluding Remarks

This project gives evidence of large numbers of processed baryons in the intergalactic medium, but there are still wide error bars that need to be constrained. There are several things that could be looked at in future studies to help constrain the error bars. These include further observations by Hubble along different sightlines and looking in different wavebands along the same sightlines.

This study looked at spectra from around 700 QSOs. There have been more than a million quasars discovered (Secrest et al. 2020). I would expect that as we look at other QSO sightlines, the same overall trends would be found in the distribution of clouds of intergalactic gas. With QSO sightlines that are close enough together it could be possible to identify a cloud of gas in multiple spectra which would confirm the existence and location of a cloud, along with put constraints on the size of these clouds, since a single

sightline only gives information about the density integrated to become the column density.

Currently, COS is the best instrument we have for doing UV spectroscopy and it will be for the near future. COS can take high resolution spectra of QSOs with a magnitude of ~17.5. This gives an average angular separation of QSOs on the order of 10's of arcminutes (J. M. Shull et al. 1999). At the redshift of the Coma Cluster, this separation is on the order of a megaparsec. Since a typical galaxy is on the order of 100 kiloparsecs wide, the average separation of sources with high resolution spectra needs to come down by about an order of magnitude. To get average separation on the order of arcminutes, we need a space-based telescope with a larger aperture than Hubble and high-resolution detectors. This would be a significant technological challenge and is the same problem that has been discussed since the turn of the 21st century (J. M. Shull et al. 1999).

It would also be useful to study these same sightlines at different wave bands. These additional wavebands could be used to confirm clouds of gas at a particular redshift. Since the intergalactic medium typically has a temperature of 10^5 and 10^7, these clouds would have black body spectra that extend into X-rays. This emission could be visible to the Chandra telescope. These observations wouldn't be useful for determining species, as blackbody radiation is emitted from cloud with charges, but it would be useful for confirming the redshift of clouds of gas.

A possibly interesting place to look is at longer wavelengths than the UV. The James Webb Space Telescope (JWST) has high resolution spectrometers in the infrared. Infrared spectroscopy finds transitions between states in molecules, rather than ions. As

the material in the detected clouds is typically ionized, the average energy would be too high for atoms to form molecular bonds. This puts harsh limits on the usefulness of infrared spectroscopy.

Overall, this study has found good evidence that reservoirs of processed baryons exist far removed from galaxies. The full solution to the Missing Baryon Problem still has not been found. With further analysis of existing archive spectra and new spectra from current and future telescopes, we will hopefully be able to solve the Missing Baryon Problem in the near future.

REFERENCES CITED

Adami, C. et al. 2005. “Large Scale Diffuse Light in the Coma Cluster: A Multi-Scale Approach.” *Astronomy & Astrophysics* 429(1): 39–48.

Andernach, Heinz, and Fritz Zwicky. 2017. “English and Spanish Translation of Zwicky’s (1933) The Redshift of Extragalactic Nebulae.”

Bennett, C. L., D. Larson, J. L. Weiland, and G. Hinshaw. 2014. “The 1% Concordance Hubble Constant.”

Birkinshaw, M. 1999. “The Sunyaev–Zel’dovich Effect.” *Physics Reports* 310(2–3): 97–195.

Bothun, G. D., and A. Dressler. 1986. “Blue Disk Galaxies in the Coma Cluster - Analogs to Z = 0.5 Cluster Members?” *The Astrophysical Journal* 301: 57.

Bothun, Gregory D. et al. 1992. “The Velocity-Distance Relation for Galaxies on a Bubble.” *The Astrophysical Journal* 395: 347.

Bothun, Gregory D. 2003. “Diffuse Baryons in the Universe: Galaxies, Warm Gas, or Cold Lonely Lumps.” In *The IGM/Galaxy Connection: The Distribution of Baryons at Z=0*, ed. Mary E Rosenberg Jessica L. and Putman. Dordrecht: Springer Netherlands, 11–20. https://doi.org/10.1007/978-94-010-0115-1_2.

Bryan, S. E. et al. 2012. “Influence of Baryons on the Orbital Structure of Dark Matter Haloes.” *Monthly Notices of the Royal Astronomical Society* 422(3): 1863–79.

Cen, Renyue, and Jeremiah P. Ostriker. 1998. “Where Are the Baryons?”

dell’Antonio, Ian P., Margaret J. Geller, and Gregory D. Bothun. 1996. “Peculiar Velocities for Galaxies in the Great Wall.II.Analysis.” *The Astronomical Journal* 112: 1780.

Durrell, Patrick R. et al. 2002. “Intracluster Red Giant Stars in the Virgo Cluster.” *The Astrophysical Journal* 570(1): 119–31.

Ellison, Sara L. 2000. “The Chemical Evolution of QSO Absorbers.” University of Cambridge.

Fields, Brian D. 2011. “The Primordial Lithium Problem.” *Annual Review of Nuclear and Particle Science* 61(1): 47–68.

Fields, Brian D., and Keith A. Olive. 2022. “Implications of the Non-Observation of 6Li in Halo Stars for the Primordial 7Li Problem.”

Fried, J. W. 1988. “How Frequent Are Tidal Interactions between Galaxies ?” *Astronomy and Astrophysics* 189: 42–44.

Fukugita, Masataka. 2003. “Cosmic Matter Distribution: Cosmic Baryon Budget Revisited.”

Goldsmith, Donald, and Joseph Silk. 1972. “Intergalactic Clouds in the Coma Cluster?” *The Astrophysical Journal* 172: 563.

de Graaff, Anna, Yan-Chuan Cai, Catherine Heymans, and John A. Peacock. 2019. “Probing the Missing Baryons with the Sunyaev-Zel’dovich Effect from Filaments.” *Astronomy & Astrophysics* 624: A48.

Gregg, Michael D., and Michael J. West. 1998. “Galaxy Disruption as the Origin of Intracluster Light in the Coma Cluster of Galaxies.” *Nature* 396(6711): 549–52.

Gu, Meng et al. 2020. “Spectroscopic Constraints on the Buildup of Intracluster Light in the Coma Cluster.” *The Astrophysical Journal* 894(1): 32.

Hinshaw, G. et al. 2012. “Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results.”

James, B. L., et al. 2022. *Cosmic Origins Spectrograph Instrument Handbook*. 14th ed. (Baltimore: STScI).

Kim, K.-T., P. P. Kronberg, G. Giovannini, and T. Venturi. 1989. “Discovery of Intergalactic Radio Emission in the Coma–A1367 Supercluster.” *Nature* 341(6244): 720–23.

Kochanek, C. S., and Martin White. 2001. “Global Probes of the Impact of Baryons on Dark Matter Halos.” *The Astrophysical Journal* 559(2): 531–43.

Kramida, A., Yu. Ralchenko, and J. Reader. “NIST Atomic Spectra Database (Ver. 5.9).” *Available: https://physics.nist.gov/asd [2022, April 18]. National Institute of Standards and Technology, Gaithersburg, MD. DOI: https://doi.org/10.18434/T4W30F*. https://physics.nist.gov/asd (April 17, 2022).

Lim, S. H., H. J. Mo, Huiyuan Wang, and Xiaohu Yang. 2020. “Detection of Missing Baryons in Galaxy Groups with Kinetic Sunyaev–Zel’dovich Effect.” *The Astrophysical Journal* 889(1).

Madrid, Juan P., Conor R. O’Neill, Alexander T. Gagliano, and Joshua R. Marvil. 2018. “A Wide Field Map of Intracluster Globular Clusters in Coma.”

Matsushita, Kyoko, Takuya Sato, Eri Sakuma, and Kosuke Sato. 2013. “Distribution of Si, Fe, and Ni in the Intracluster Medium of the Coma Cluster.” *Publications of the Astronomical Society of Japan* 65(1): 10.

Mihos, J. Christopher. 2004. “The Evolution of Tidal Debris.” *Symposium - International Astronomical Union* 217: 390–99.

Morton, Donald C., Donald G. York, and Edward B. Jenkins. 1988. “A Search List of Lines for Quasi-Stellar Object Absorption Systems.” *The Astrophysical Journal Supplement Series* 68: 449.

NASA/LAMBDA Archive Team. 2022. “LAMBDA - Baryonic Density.”

O’Neil, K., and G. Bothun. 2000. “The Space Density of Galaxies through MB(0) = 25.0 Magnitudes per Inverse Arcsecond Squared.” *The Astrophysical Journal* 529(2): 811–15.

Pedrosa, Susana, Patricia B. Tissera, and Cecilia Scannapieco. 2010. “The Joint Evolution of Baryons and Dark Matter Haloes.” *Monthly Notices of the Royal Astronomical Society* 402(2): 776–88.

Peeples, M. et al. 2017. *The Hubble Spectroscopic Legacy Archive*.

Peng, Eric W. et al. 2011. “THE HST/ACS COMA CLUSTER SURVEY. IV. INTERGALACTIC GLOBULAR CLUSTERS AND THE MASSIVE GLOBULAR CLUSTER SYSTEM AT THE CORE OF THE COMA GALAXY CLUSTER.” *The Astrophysical Journal* 730(1): 23.

Piña, Pavel E. Mancera et al. 2019. “Off the Baryonic Tully-Fisher Relation: A Population of Baryon-Dominated Ultra-Diffuse Galaxies.”

Planck Collaboration et al. 2018. “Planck 2018 Results. VI. Cosmological Parameters.”

Rocha, Miguel, Annika H. G. Peter, and James S. Bullock. 2011. “Infall Times for Milky Way Satellites From Their Present-Day Kinematics.”

Secrest, Nathan et al. 2020. “A Test of the Cosmological Principle with Quasars.”

Seth, Ruchika, and Somak Raychaudhury. 2020. “Evolution of Galaxies in Groups in the Coma Supercluster.”

Shull, J. M. et al. 1999. “The Emergence of the Modern Universe: Tracing the Cosmic Web.”

Shull, J. Michael, Britton D. Smith, and Charles W. Danforth. 2012. “THE BARYON CENSUS IN A MULTIPHASE INTERGALACTIC MEDIUM: 30% OF THE BARYONS MAY STILL BE MISSING.” *The Astrophysical Journal* 759(1): 23.

Small, Todd A., Chung-Pei Ma, Wallace L. W. Sargent, and Donald Hamilton. 1999. “Galaxy Clustering and Large-Scale Structure from Z=0.2 to Z=0.5 in Two Norris Redshift Surveys.” *The Astrophysical Journal* 524(1): 31–48.

Soderblom, D., and et al. 2021. *COS Data Handbook, Version 5.0*. Baltimore: STSci.

Tanimura, Hideki et al. 2019. “A Search for Warm/Hot Gas Filaments between Pairs of SDSS Luminous Red Galaxies.” *Monthly Notices of the Royal Astronomical Society* 483(1).

Tody, Doug. 1986. “The Iraf Data Reduction And Analysis System.” In ed. David L. Crawford. , 733.

———. 1993. “IRAF in the Nineties.” In *Astronomical Data Analysis Software and Systems II*, Astronomical Society of the Pacific Conference Series, eds. R.~J. Hanisch, R.~J.~V. Brissenden, and J Barnes. , 173.

White, P. M. et al. 2003. “Extraordinary Diffuse Light in Hickson Compact Group 90.” *The Astrophysical Journal* 585(2): 739–49.

Wright, Edward L. 2006. “A Cosmology Calculator for the World Wide Web.”