Draft version March 16, 2022
Typeset using LaTeX **twocolumn** style in AASTeX63

An End-to-End Simulation Analysis Pipeline for Interpreting Synthetic Observations of the Circumgalactic Medium

Brendan Boyd,[1] Devin Silvia,[2] and Brian O'Shea[1,2]

[1] *Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA*
[2] *The Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824, USA*

ABSTRACT

The hot, low density gas surrounding galaxies, called the circumgalactic medium (CGM), is vital to understanding the structure and evolution of galaxies. The diffuse nature of the CGM makes it difficult to observe by direction detection in emission, so much of our understanding comes from studying absorption features in the spectra of light from distant quasars that passes through intervening galaxies. As a result, careful study of simulated galaxies is a critical part of interpreting these limited observations. The complexity of physics-rich simulations provide far more data to analyze than absorption line studies alone. However, the analysis of these simulations still needs to be connected to observations to maximize our knowledge of the CGM. Using multiple open-source software packages, we developed an analysis pipeline that can be used to study these data-rich simulations and compare our results directly to the most current observational surveys. These tools provide us a whole new avenue to explore our simulations and should provide further insight into the physical mechanisms that produce the observed CGM properties. One direct application of this tool is to test whether observers can distinguish between metals (e.g. carbon, oxygen, silicon, magnesium) that have been ionized through collisional ionization versus photoionization, which will help us to understand the density and temperature structure of the CGM. The pipeline can also be used to explore how well observers can determine whether the absorption features were produced in the CGM or the interstellar medium of the galactic disk. These findings will advance CGM research and uncover the intricate nature of galaxies.

Keywords: Circumgalactic medium (1879), Quasar absorption line spectroscopy (1317), Computational methods (1965), Hydrodynamical simulations(767), Galaxy evolution(594)

1. INTRODUCTION

1.1. *What is the CGM?*

The circumgalacitc medium (CGM) is the hot, diffuse gas that surrounds galaxies. The temperature ranges from 10^4K to $\gtrsim 10^6K$ and the density is so low that direct detection of emission is extremely difficult or impossible in many cases (Tumlinson et al. 2017). The CGM acts as a buffer between the interstellar medium (ISM) and the intergalactic medium (IGM). A further understanding of the CGM may solve many of the most pressing issues with our current theories of galaxy evolution. This includes the apparent lack of baryons and metals in galaxies, as well as how galaxies sustain star formation and how that star formation is then quenched (Tumlinson et al. 2017). Outflows of gas bringing baryons and metals from the ISM into the CGM have been all but confirmed by observations (Tumlinson et al. 2017), and it is likely that the CGM mediates the inflows of cool gas from the IGM that help fuel star formation in the disk of the galaxy. This collection of interactions, as well as others, make the CGM a very complex region to study.

The diffuse nature of the region makes studying the CGM very challenging, especially through the direct detection of emitted radiation. One of the main ways researchers study the CGM is by looking at the spectra of quasar sightlines. As light travels from the quasar to our telescopes, it passes through intervening galaxies. Light is absorbed at specific wavelengths by the gas in the CGM of these galaxies, depending on what ionized atomic species are present. These absorption patterns are then studied to reveal what ions and how many ions are present in the CGM of that galaxy. By studying these absorption line features, we can probe different

Corresponding author: Brendan Boyd
boydbre1@msu.edu

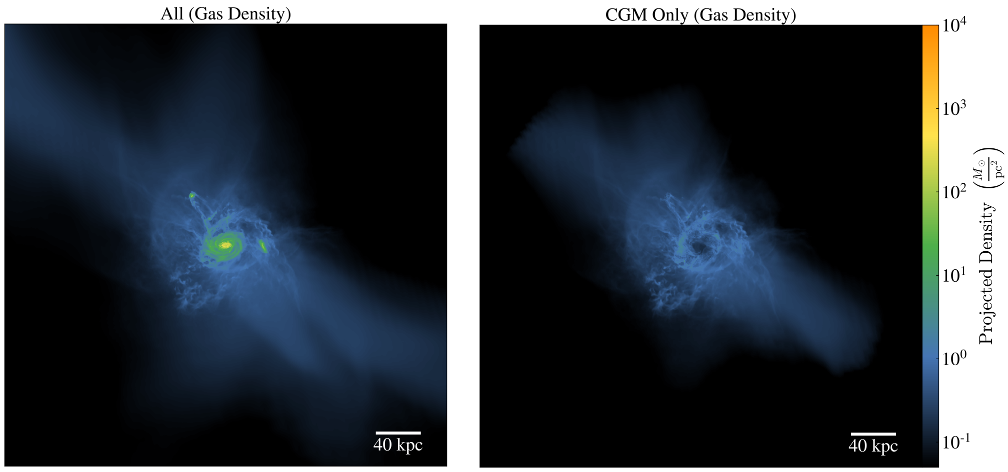

Figure 1. A density projection of one of the galaxy halos created by FOGGIE collaboration. This galaxy is at redshift 0.5. The left shows the density accounting for all of the cells in the simulation. The right image shows just the CGM by applying a filter to the data. In this case, the CGM is defined as all the cells between 10-200 kpc from the center of the galaxy with a temperature greater than $1.5 \times 10^4 K$ **or** a density less than $2 \times 10^{-26} g/cm^3$. The disk of the galaxy and the remnants of satellite galaxies are all filtered out. This leaves the hot, diffuse gas that characterizes the CGM.

temperature regions of the gas, estimate the metallicity of the CGM, and learn much more about the interactions that make the CGM so important (Lehner et al. 2019; Howk et al. 2017).

A galaxy, however, will typically have only one intervening sightline. So to infer properties about the CGM, a large survey of sightlines must be combined to create a statistical representation (Lehner et al. 2018). This gives researchers a better picture of the CGM, but the information available is still limited. To assist in studying the CGM, it is thus important to also look to hydrodynamic simulations of galaxies which hold far more data than these absorption line studies alone.

1.2. *Why use Simulated Observations?*

Galactic simulations give researchers access to a wealth of data that can be used to study the CGM. Although observational studies are inherently limited, they are still vitally important to understanding the CGM. It is best to use both observations and simulations, in conjunction, to further our understanding of the CGM. In this effort, we generate synthetic spectra from simulated gas to directly compare to observations of absorption lines.

Synthetic observations extract data from the simulation from an observers perspective. This allows us to look at both what observers can see and what is physically there in the simulation. Additionally, this approach reduces some of the barriers which make it difficult to compare simulation and observational studies. This enables us to verify absorption line studies or potentially find biases in these studies. Similarly, by comparing to observational studies, we can verify whether the simulation is accurately representing the CGM and which areas need further work (e.g. different stellar feedback models, other additional physics). Streamlining the comparison between simulations and observations, while still using all the extra information held in the simulations, will ensure faster, more accurate study of the CGM and it's role in galaxy evolution.

The layout of this paper is as follows. Section 2 goes into further detail about the simulations used for this work including their construction and the unique properties that make them exceptionally well-suited for this effort. Then in Section 3, we introduce the individual software tools that are utilized in the full analysis pipeline. Section 4, describes the full analysis pipeline and how it extracts absorber information. In Section 5, we present the analysis work done using this pipeline. Section 6 discusses the conclusions of this analysis and suggestions for future work.

2. SIMULATION

2.1. *Enzo*

Enzo is an adaptive mesh refinement (AMR) code that uses a block structure to evolve hydrodynamic simulations (Bryan et al. 2014). The simulation is divided into individual cells, each containing information about the gas inside (e.g., density, temperature, velocity). Enzo advances the system in time and tracks the movement and evolution of each cell and how it interacts with its surrounding cells using a variety of physics such as radiative heating/cooling, chemical evolution and many more. The key component to Enzo, however, is it's use of adaptive mesh refinement. AMR effectively changes the resolution of the simulation in order to properly resolve important areas while also conserving computational power. For example, in a simulation of a galaxy halo, interactions in and near the disk of the galaxy need much higher resolution than interactions in the IGM.

The resolution is changed by dividing existing cells into smaller, refined cells. This occurs when a given cell meets some specified criteria by the simulator (e.g., baryon mass, Jeans length, and/or cooling time). Enzo uses machine-vision to place the new "child" cells inside the previous cell, now called a "parent" cell. This allows Enzo to maintain a hierarchical structure of parent and child cells which help keep the simulation properly organized and aid in the evolution. Enzo can move or reconstruct cells to track interesting areas that spatially move over time. AMR allows Enzo to simulate large spatial and temporal ranges that would otherwise be computationally prohibitive for a code using a fixed cell structure.

2.2. *FOGGIE*

The FOGGIE simulations are cosmological hydrodynamic simulations generated using Enzo (Peeples et al. 2019). These simulations are used to study the CGM and its impact on the evolution of the galaxy. The unique property that makes these simulations ideal for this is a fixed refinement box around the galaxy. This cubic box has edges of 200 comoving $h^{-1}kpc$ in length and is centered at the center of the galaxy. This keeps the CGM spatially resolved at sub-kpc scale. This level of resolution allows for finer filaments and smaller clouds to form in the CGM and overall gives a more accurate representation of the CGM.

3. SOFTWARE TOOLS

In recent years, there have been many software tools developed to analyze astrophysical simulations and generate synthetic observations. These open-source software projects have collaborators from all around the world, constantly improving each of these tools to better aid researchers. Below, we will describe the main tools we incorporated in our own pipeline to analyze simulations.

3.1. *yt*

`yt` is an open-source python package used to visualize volumetric data (Turk et al. 2011). `yt` can quickly and easily load in simulated data from many different codes often used by astronomers, including Enzo. This functionality alone makes it extremely useful to astronomers. One of the uses of `yt` is creating slices and projections of data as seen in Figure 1 and 2. As well as making visualizations, `yt` makes it easy to access the many underlying fields in the simulation dataset (e.g., metallicity, density, velocity). It also has a fairly easy method for adding new fields to the dataset, such as defining a radial velocity relative to the center of a galaxy. In addition, `yt` can filter the dataset using logical cuts to the fields to return only the cells which satisfy a particular criteria (e.g., density < 10^{26} g/cm^3). This is incredibly useful for isolating the CGM from the rest of the galaxy (Figure 1). These tools and many more makes `yt` invaluable and many other tools which analyze simulations are built on top of `yt`.

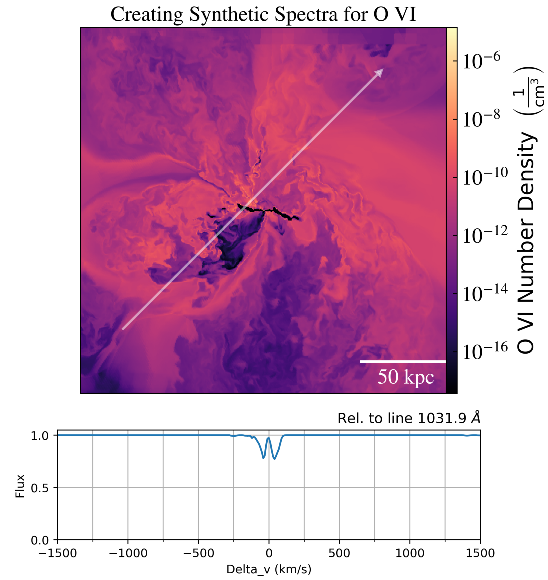

Figure 2. he top plot is a slice of O VI number density created using `yt` This is an edge on view of the disk, the black region in the middle. The lightray is created by Trident and the resulting spectra is plotted beneath. The spectra is plotted in velocity space where 0 corresponds to the 1032 Åline would be in a lab frame. Trident finds two separate absorbers with column density Log(N_{OVI})= 13.5 each (N_{OVI} in $cm^{-}2$).

3.2. *Trident*

One such tool is Trident, an open-source python package that creates synthetic spectra from astrophysical simulations (Hummels et al. 2017). Trident does this by first adding new ion species fields to the simulation dataset. Individual ion tracking is not typically done during the simulation as it can be computationally expensive. Instead, the amount of ions in a cell must be calculated afterward assuming ionization equilibrium. Trident accounts for both collisional ionization and photoionization based on a uniform ultraviolet radiation background model (e.g. Haardt (2012)). An ionization table estimates the number of ions by using density, temperature and metallicity fields.

Trident can then use these ion species fields to create synthetic spectra. Trident does this by first constructing a light ray that passes through some region of the simulation. This is simply a 1D object that connects continuous cells together in a straight line (see slice in Figure 2). Once this light ray is constructed, a spectra can be calculated by using the ion number density in each cell and the path length through that cell. Trident allows you to include or exclude any ions you want from the spectra as well as what wavelength range you wish to look at. This makes analysis of individual ions far easier since you can isolate their spectra from the other absorbers.

3.3. *Spectacle*

Spectacle[1] is an open-source python package that creates models of spectral data (Earl & Peeples, in preparation). It can use these models to fit spectra, like those created by Trident, and then extract the "observed" absorption features. Some of the features that can be extracted include: (1) full width at half maximum (FWHM), the width of the absorption line at the half way point between the continuum and the trough, (2) column density, the area density calculated by integrating over the volumetric density along the line of sight, (3) Doppler broadening, the increasing of the width of the line due to Doppler effects, and (4) equivalent width, the width of a rectangle with an area equal to the absorption line. This automated fitting of spectra allows us to process thousands of lightrays and extract meaningful data from the spectra.

In Figure 3, absorption spectra for 3 different ions (H I, C IV, and O VI) are shown as well as the fits done by Spectacle. Fitting spectra is not an easy process as absorption features can be embedded in one another. Deciding how many lines to fit on a given source can prove challenging and creates another layer of uncertainty when it comes to the fit.

column density, the area density calculated by integrating over the volumetric density along the line of sight.

4. ANALYSIS PIPELINE

In order to facilitate the statistical analyses necessary to compare directly to observational studies, we integrated the aforementioned software tools into a single analysis pipeline. The main purpose of the pipeline is to extract and categorize "absorbers" in the simulation. Absorbers are loosely defined as contiguous clouds of gas that create an absorption feature in the spectra. However, the column density necessary for a cloud to be consider a significant absorber varies as a function of ion species (see Table 1).

Absorber extraction is done via two methods which we will refer to as the iterative cloud extraction method, or ICE method, and the Spectacle method. ICE uses an algorithm we developed to extract absorbers by directly looking at simulated lightray data. The Spectacle method uses Spectacle to fit a synthetic spectra generated from the lightray using Trident. The Spec-

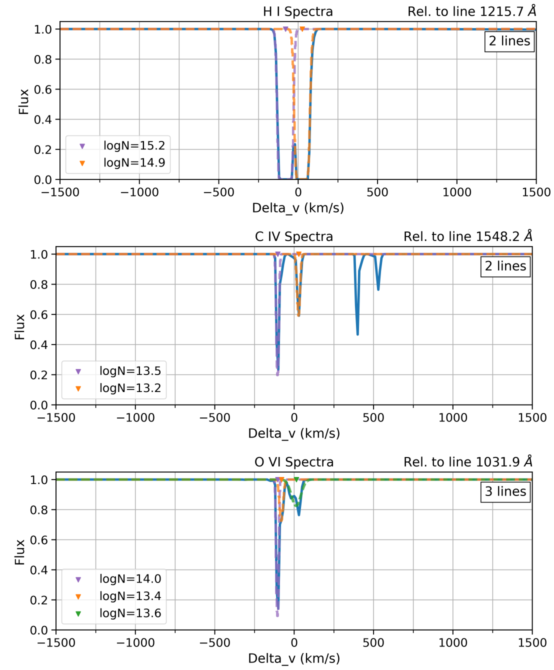

Figure 3. A single lightray passing through a galaxy at z=2.5 generates these spectra. The blue line shows the spectra, while the dashed lines are fits of individual lines by Spectacle. The bottom left shows the column density corresponding to each absorption line that was fit. C IV's doublet feature can be seen but Spectacle was only tasked with fitting the 1548 Åline.

Ion Species	Minimum Log Column Density
H I	12.5
C IV	13
O VI	13

Table 1. Absorbers were only extracted if they met the above minimum log column density. We empirically found that absorbers with lower column density did not generate significant absorption features in the spectra.

[1] https://spectacle-py.readthedocs.io/en/latest/

tale method is more analogous to the way information is extracted from observed spectra, however ICE leads to greater access to the underlying simulated data and thus more information about the properties of the absorber.

The full pipeline is partly based around creating a multi-panel plot (see Figure 4) that displays each piece of the analysis. The multi-panel plot allows you to visually connect the absorber's physical location and structure (top plot), how the absorber appears in the lightray's number density and velocity profiles (middle two plots) and finally the resulting absorption line in the spectra (see the bottom plot). This gives an inside look of what the machinery is doing and allows you to compare the two methods, ICE and Spectacle.

4.1. *Spectacle Absorber Extraction*

This pipeline streamlines the process of using Spectacle to extract absorbers. Using Trident we create a synthetic spectra and then immediately feed that to Spectacle to create a fit and extract information about the absorbers found. This allows us to process thousands of lightrays efficiently and in a manner similar to observers.

4.2. *ICE Absorber Extraction*

The ICE method extracts absorbers directly from the lightray object created by Trident. This retains all of the information in the fields of the simulation. This method allows us to analyze the absorbers at far greater detail than the Spectacle method because we have access to all of these different fields.

The algorithm that ICE uses is an iterative process designed to identify absorbers in the lightray. We first start by finding the number density, n_{ion}, such that the column density of the cells with number density greater than n_{ion} account for 80% of the column density along the lightray. We then combine all the contiguous cells that are above that threshold. This process is analogous to the cloud identification process used by the FOGGIE collaboration (Peeples et al. 2019). Then we mask out the cells already contained in "clouds" and repeat the process. We find a new n_{ion} for the cells left in the lightray and create a new set of clouds. Next we combine the new set of clouds with what we already have based on the clouds position and velocity information. Two clouds are combined only if they are next to each other, contiguous in space, and their difference in velocity is below a threshold of 10 km/s. This prevents clouds who are moving at very different speeds, and thus likely separate features in the CGM, from being combined together. This process repeats until the column density left in the lightray is below what we define as the observable limit for that ion (see Table 1). Finally, only the clouds that meet the observable limit of column density are returned.

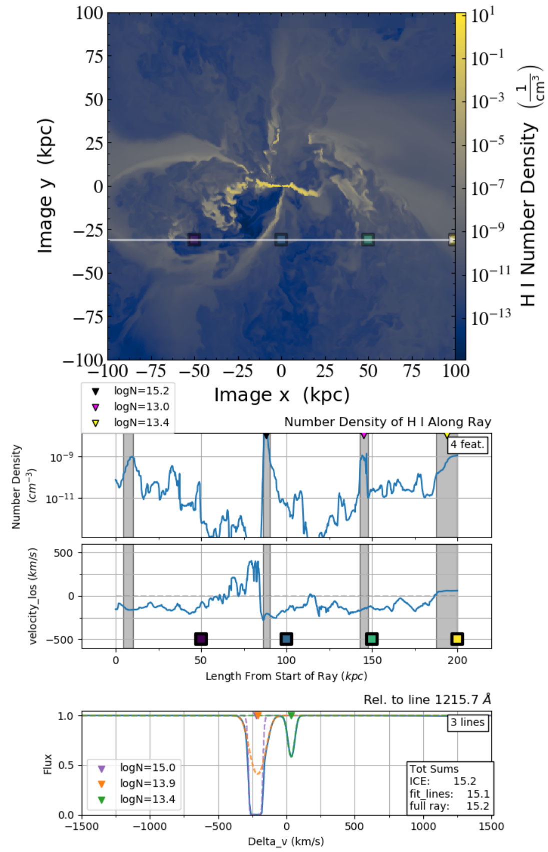

Figure 4. The multi-panel plot allows you to track where the absorber is in relation to the galaxy (see the top slice), where it is along the lightray and how fast it is moving (see the middle two plots), and then finally what the absorption line looks like. By looking at the number density and velocity, you can identify which absorption line corresponds to which absorber in the lightray. The column density of the absorbers are displayed on the left hand side in appropriate sections for ICE and Spectacle methods. The bottom left holds a box showing the total column density found by ICE, Spectacle's fits, and then the total column density in the lightray itself.

This process allows us to extract the largest absorbers as well as smaller absorbers that still produce significant features in the spectra. This also allows us to extract a great deal of information about the absorber that is lost when looking at just the spectra, (e.g., temperature, radial velocity, spatial location). In turn, we can conduct a large amount of unique analyses looking at the different distributions of absorber properties from many lightrays.

4.3. *Synthetic Surveys*

To do more comprehensive analyses, we need to look at a collection of synthetic observations and create a synthetic survey much like real, observational surveys. To do this, we generated a large sample of lightrays from a FOGGIE simulation at seven different snapshots, ranging from z=2.5 to z=0.3. For each snapshot, we generated 1500 sightlines by randomly sampling projected impact parameter from 0 to 200 kpc. This ensures there is no a sampling bias when comparing to observational studies. After these lightrays were generated, we used `yt` to apply a filter to the lightrays so that only CGM gas contributed to absorbers (see Figure 1 for details on cuts). Finally, we used our analysis pipeline to extract absorbers for further analysis. This gave us a large sample of absorbers that we then used to study the CGM and verify our methods.

The use of `yt`'s filtering allows us to conduct additional interesting analyses. One of these is to look at the differences between inflowing and outflowing absorbers. To do this, we applied a filter to the lightrays based on the radial velocity and then used the pipeline to extract absorbers. This gave us two distributions of absorbers, inflowing with radial velocity < 0 and outflowing with radial velocity > 0. This opened a new avenue to conduct analysis and test some of the current theories about galaxy evolution (see Section 5.2).

5. ANALYSIS

We chose to study the distribution of O VI absorbers due to its prevalence in observational studies. This is because of the unique doublet feature of this ion which has identical absorption patterns at 1032Å and 1038Å, with the 1038 line simply being smaller. This makes it far easier for observers to fit lines and then extract absorption features. Neither Spectacle or ICE take advantage of this feature, but by studying O VI we can more easily compare our findings to observations.

5.1. *Comparing Spectacle and ICE*

To look at how Spectacle and ICE compare, we can look at individual lightrays, as in Figure 4, but we can also look at a large sample, as in Figure 5. This brings to light one of the problems with the Spectacle method, and studying spectra in general. That is, some absorber information is lost in the spectra. This can be because a large absorber drowns out other, smaller absorbers. Or, if two different absorbers are moving at the same velocity along the lightray's line of sight, they will be redshifted/blueshifted the same amount and combine to create one large absorption feature. Since ICE looks at the simulation data and not the spectra, it can differentiate these different cases. Another difference is Spectacle tends to be less sensitive to low column density absorbers (see Figure 5). This is expected, though, because it is more difficult to fit smaller absorption lines. Overall however, Figure 5 shows the two methods have similar column density distributions despite not always agreeing.

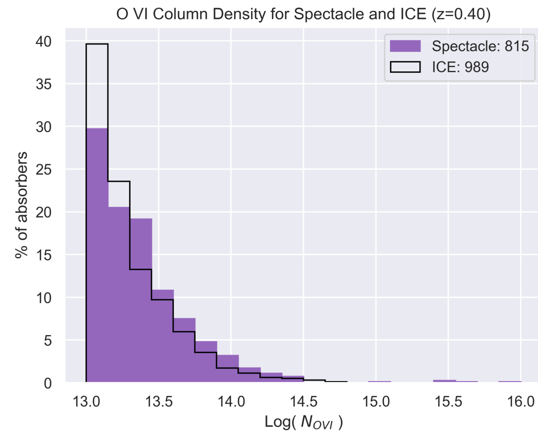

Figure 5. Combined O VI column density from seven snapshots at different redshifts (0.3, 0.35, 0.4, 0.5, 1.0, 2.0, 2.5). The histograms show the percentage of absorbers for a given bin compared to the total number of absorbers extracted by that method. The black steps indicate those extracted by ICE while the purple bars show those extracted by Spectacle. The total number of absorbers extracted by each method is indicated in the upper right corner.

5.2. *Inflows and Outflows*

As described in Section 4.3, we extract inflowing absorbers and outflowing absorber separately. This allows us to look for differences in their properties. Looking at their column density distributions, we can quickly see that there are far more inflowing absorbers than outflowing absorbers (see Figure 6). Despite the large discrepancy in number, the column density distribution of inflowing and outflowing absorbers tend to be very similar. Also note, the number of outflowing absorbers decreases substantially as a function redshift, this trend continues with the even lower redshift snapshots. Because of the low statistics at lower redshift, we restrict the rest of our analysis to a snapshot of the galaxy at redshift 2.

To explore the potential difference between inflowing and outflowing absorbers, we created a corner plot, sometimes called a pair plot, to look for correlations between different variables (see Figure 7). This style of plot creates a series of scatter plots that pair two different absorber properties. This can be used to show a correlation between these two properties. Additionally, along the diagonal, histograms are plotted to show the distribution of a single absorber property. This al-

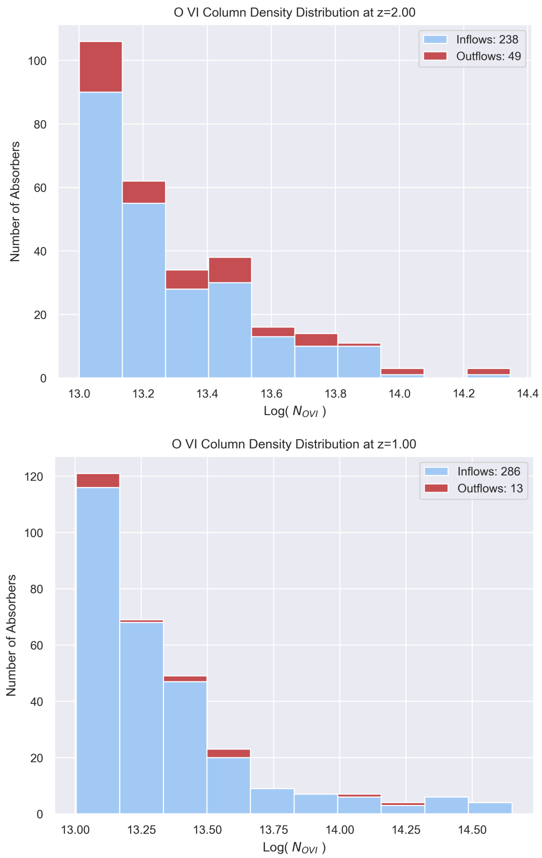

Figure 6. These two plots show the distribution of O VI column density for absorbers extracted by ICE at two different snapshots. Blue represents the inflowing absorber while red represents the outflowing absorbers. The total number for each are in the right hand corner of each plot. The absorbers were extracted from 1500 sightlines as described in Section 4.2.

lows us to look at the many different properties of these absorbers at once and find interesting correlations or differences between the populations of inflows and outflows.

Looking at the histograms for metallicity and for radial distance, we can see that the outflows tend to have higher metallicity on average and tend to be closer to the galaxy than inflows. The temperature and density distributions appear to have slightly different structure, but nothing quite as distinct as the metallicity and radial distance. Investigating the outflowing absorbers further, there appears to be a group of absorbers with different properties from the rest of the outflows.

These outliers are fairly distant from the galaxy, large radial distance, and have a lower metallicity than the other outflowing absorbers. By tracking this group of outflowing absorbers we can notice that they all have similar densities, temperatures and metallicities. This indicates these absorbers likely exist in the same region of space. But since we have access to the full simulation data, we can go further than just hypothesize. We can instead verify that assumption by marking the location of these observers in projections of the galaxy (see Figure 8). This shows that the absorbers appear to all be on the outer edge of a fairly hot bubble moving away from the galaxy. This is a significant benefit of synthetic observations. Having access to the simulation data at hand lets us know if this group is showing a larger trend of low metallicity outflows or if they are just from an isolated region.

6. DISCUSSION

In our analysis of inflows and outflows, we found that at redshift z=1 and lower, the number of outflowing absorber decreases substantially. This indicates that the galaxy has reached a quiescent state, with few to no mergers with satellite galaxies and other interactions that would perturb material to flow out of the galaxy. The more interesting analysis, however, was the tracking of outflows and inflows when the galaxy was at a redshift of z=2, with similar findings in the z=2.5 snapshot.

One of the prevalent theories of the CGM and galaxy evolution is the idea of gas flowing in and out of the galaxy through the CGM (Tumlinson et al. 2017). Outflowing gas is typically thought to be from supernovae and thus would be at a higher metallicity, while inflowing gas comes from the IGM and is that at a much lower metallicity. This lines up precisely with our findings where one of the biggest differences between outflows and inflows was the metallicity.

We also investigated some outflowing absorbers that appeared to be outliers in metallicity and radial distance. This analysis showed one of the key benefits of synthetic observations, that is, we have the simulation data to look back to in order to better understand our results. We were able to verify that these outlier all came from a similar region in the CGM and this region can now be further studied to see why it has outflowing absorbers where other regions at this distance do not.

7. CONCLUSION

The CGM has been revealed to be a complex and interesting region of space, though much of its dynamics and effects on galaxy evolution are still not well understood. Advances in observational studies will help answer more and more of these questions about the CGM, but it is increasingly apparent that hydrodynamic simulations can, and will, play a significant role in further research. It is important to study these simulations in a way that complements the observational studies and

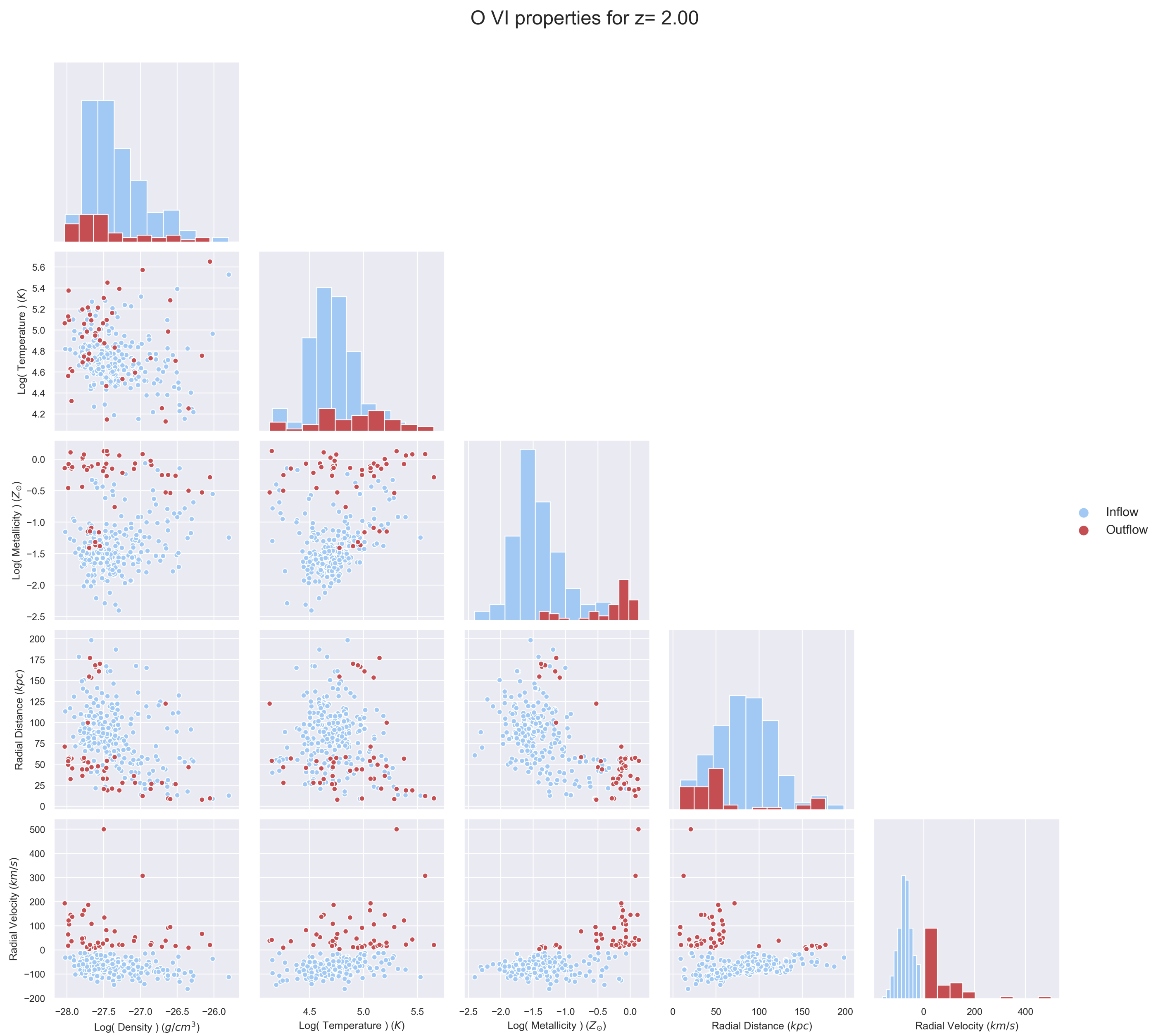

Figure 7. A corner plot, or pair plot, of different properties of the absorbers at a redshift, $z = 2$. The diagonal shows histograms for a given variable while scatter plots show the correlation of two variables, where each absorber is represented as a point. Density, temperature and metallicity are plotted in log space while radial distance and velocity are plotted linearly.

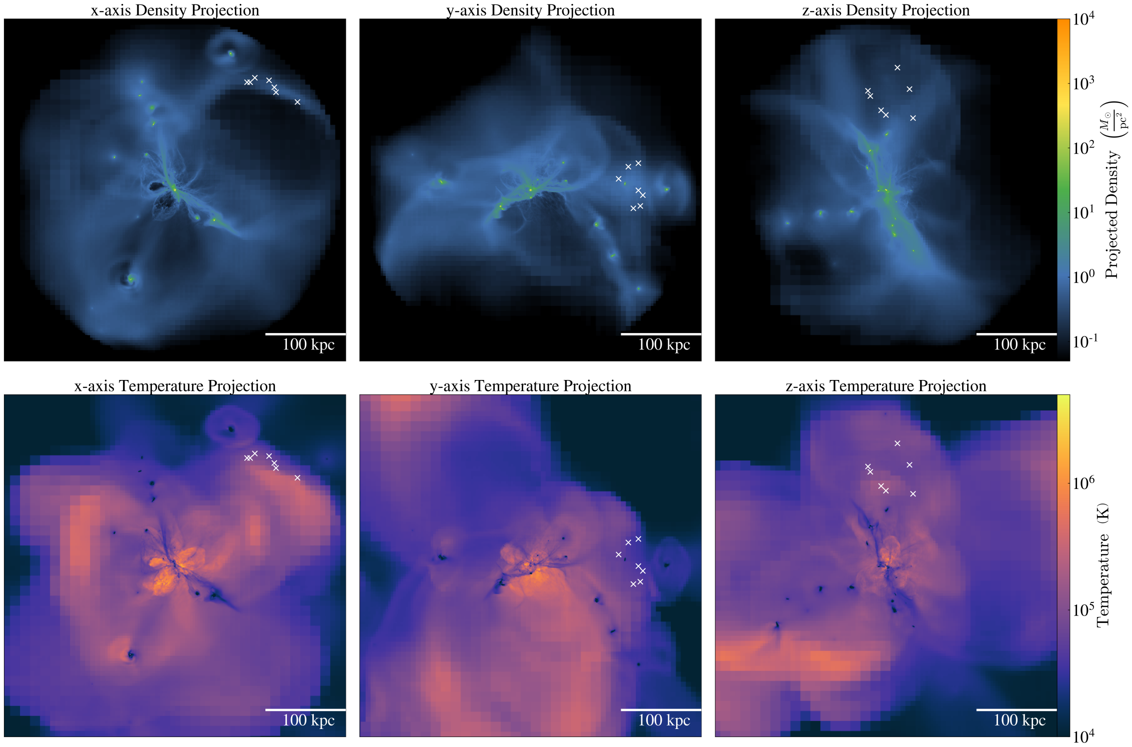

Figure 8. Density projections and a density weighted temperature projections of the z=2 snapshot. The white "X" markers indicate where the location of the "outlier" outflowing absorbers.

makes use of all the resources at hand. Synthetic observations and this pipeline can achieve both of those goals and bring us closer to fully understanding the CGM.

The pipeline will be used to further investigate some features of the CGM. Similar to our inflow/outflow absorber analysis, we plan to investigate the differences of hot and cold absorbers, ideally to find an observable difference in the spectra of absorbers in collisional ionization equilibrium versus those photoionization equilibrium. Additionally, the pipeline will be used to more directly compare simulated surveys to observational studies. The adaptability of this pipeline makes it perfect for many different future analyses in regard to absorbers in the CGM and future work will only enhance its performance and accuracy.

REFERENCES

Bryan, G. L., Norman, M. L., O'Shea, B. W., et al. 2014, ApJS, 211, doi: 10.1088/0067-0049/211/2/19

Haardt, Francesco; Madau, P. 2012, ApJ, 746, doi: 10.1088/0004-637X/746/2/125

Howk, J. C., Wotta, C. B., Berg, M. A., et al. 2017, ApJ, 846, doi: 10.3847/1538-4357/aa87b4

Hummels, C. B., Smith, B., & Silvia, D. W. 2017, ApJ, 847, doi: 10.3847/1538-4357/aa7e2d

Lehner, N., Wotta, C. B., Howk, J. C., et al. 2018, ApJ, 866, doi: 10.3847/1538-4357/aadd03

—. 2019, ApJ, 887, doi: 10.3847/1538-4357/ab41fd

Peeples, M. S., Corlies, L., Tumlinson, J., et al. 2019, ApJ, 873, doi: 10.3847/1538-4357/ab0654

Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, ARA&A, 55, doi: 10.1146/annurev-astro-091916-055240

Turk, M. J., Smith, B. D., Oishi, J. S., et al. 2011, ApJS, 192, doi: 10.1088/0067-0049/192/1/9