Draft version May 20, 2025
Typeset using LaTeX **default** style in AASTeX7

CTA200H Mini-Project: MW Analogues in cosmological simulations

Nicole Jiang[1]

[1] *University of Toronto*

ABSTRACT

We investigate the galaxy star formation main sequence by analyzing simulated stellar masses and star formation rates of galaxies within a subhalo from the EAGLE and IllustrisTNG databases. Single-snapshot and multi-snapshot samples of data are compared to investigate galaxies over varying redshifts. For each sample, we generate an SFR vs. stellar mass scatterplot and a two-dimensional histogram, then examine the distribution, scatter, and behaviour of the main sequence. The denser multi-snapshot sample exhibits a steeper slope and greater scatter. We discuss how observational limitations, particularly luminosity sensitivity and interstellar medium absorption, determine the maximum redshift of real galaxy surveys. Our results show that incorporating multiple redshifts yields a more realistic analogue to observational data, but the shortcuts that enable simulations to operate on such a large scale with such efficiency are also the reason the simulated plots differ from what is expected from real life.

Keywords: Cosmology (343) — Galaxies (573) — Redshifts (1378) —Astronomical simulations (1857) — Stellar masses (1614) — Star formation (1569)

1. METHODS

1.1. *EAGLE Database Querying*

An account was created to access the EAGLE databases and simulations. In a Jupyter notebook, an SQL query was run on the EAGLE database.

A query retrieves columns of data from a table within a database. In this case, the query retrieved the simulated stellar mass and star formation rates of galaxies within a subhalo (table RefL0100N1504_SubHalo) recorded in the EAGLE database, but with a condition. The condition placed on the query was 'SnapNum = 28', meaning that the data retrieved was restricted to the 28th snapshot such that the output data was from a unique moment of time. This is because snapshots in the EAGLE database refer to a certain value of cosmological redshift.

Because the universe is always expanding, distant objects become more distant with time. Light's wavelength stretches as it travels through the expanding universe, and the stretching is a phenomenon called cosmological redshift. The cosmological redshift of a snapshot indicates how much the universe has expanded since that snapshot (known as the expansion factor) and how far back in time the snapshot is (known as the lookback time). In this case, the 28th snapshot refers to a redshift of $z = 0.00$, so the simulation is of present day, where we look back in time 0.00 gigayears to when the expansion factor is 1.000.

If the same condition of 'SnapNum = 28' is applied to different columns of the table, this would mean that the data retrieved would be different properties (columns) but of the same galaxies (rows) in the same subhalo (table) at the same moment in time (SnapNum).

Email: nicolejitong.jiang@mail.utoronto.ca

The star formation rate and stellar mass data were used to plot an SFR vs. Stellar Mass scatterplot and two-dimensional histogram to observe the distribution and relationships. See **Code 1** for reference. Another query was run, but it covered all snapshots after redshift 0.5.

> To access all galaxies found in snapshots after a redshift of $z = 0.5$, the condition can be changed to 'SnapNum<= 28 and SnapNum> 23 '. This is because 'after' a redshift of $z = 0.5$ is a range of $0 <= z < 0.5$. According to Appendix C of the EAGLE public release of halo and galaxy catalogues, a redshift of 0 is the 28th snapshot while a redshift of $z = 0.5$ is the 23rd snapshot. It follows that '$0 <= z < 0.5$' in terms of snapshot number is '28 <= SnapNum< 23'.

An SFR vs. Stellar Mass scatterplot and two-dimensional histogram were also generated for this set of data, and a comparison to the single-snapshot plot was observed.

1.2. *IllustrisTNG Database Querying*

After applying for and receiving an API key, a helper function was programmed to make an http get request. A request to the API root was issued in order to look at all galaxies at redshift = 0. The star formation rates and stellar masses of the galaxies in this sample were retrieved and the plots they generated were compared to the previously created plots. The number of galaxies in this sample were noted and compared to the previous samples.

2. RESULTS

2.1. *EAGLE Database Querying*

See **Figure 1** for the single-snapshot SFR vs. Stellar Mass scatterplot and its corresponding two-dimensional histogram.

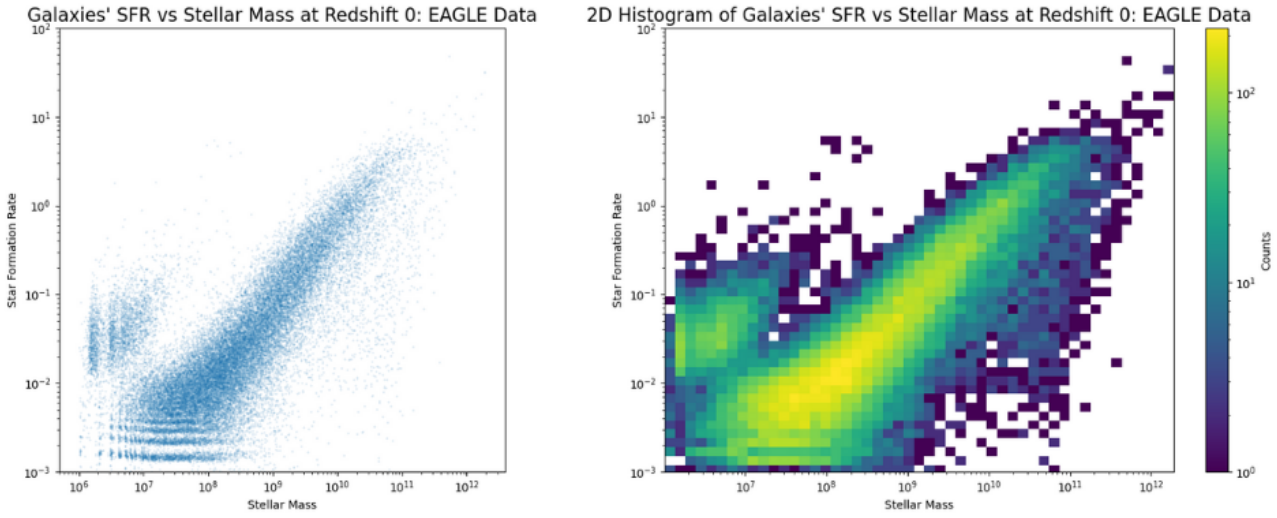

Figure 1.

It is evident that the SFR vs Stellar Mass relationship of this plot resembles an increasing, roughly linear trend: as the stellar mass increases, so does the star formation rate. This correlation is known as the galaxy star formation main sequence. The data points scatter around a diagonal line, known as the main sequence. Galaxies that deviate from the main sequence can be grouped as follows:

1. Starbursts

   > These galaxies lie above the main sequence and have a higher SFR compared to galaxies on the main sequence with the same stellar mass. This is from forming many stars in short periods of time, called "starburst events".

2. Green valley

   Located just below the main sequence, these galaxies are transitioning to red and dead.

3. Red and dead

   This group of galaxies has the lowest SFR and is believed to have been once on the main sequence, but through debatable processes, fell off into the red and dead state. It gets its name because galaxies with new, young stars are hotter and bluer while galaxies with stellar mass but no new stars appear redder.

It is believed that over time, galaxies evolve along the main sequence, but can be pushed upwards, deviating to starbursts, or below, falling through the green valley, quenching (ceasing star formation activity), and becoming red and dead.

In the simulated SFR vs Stellar Mass plot, the starburst galaxies seem to group in vertical stripes while the galaxies deviating below the main sequence form horizontal stripes. This differs from what we would expect from a plot using observed data from nature, which looks more continuous, smooth, and groups are cloud-like instead of stripes. These differing features of the simulated plot are likely due to the simplifications that make the simulations even possible on such a large scale, and efficient. The vertical stripes are likely because in order to keep track of less entities, the simulation groups stars that form in the same region into single particle masses. So the stellar masses of galaxies are grouped in levels instead growing continuously like with real data. Furthermore, another simplification that makes simulations efficient is large timesteps. Larger time steps means less integration steps so faster runtime, but also means the simulation is not continuous. This is another reason for the breaks, and is likely the reason for the horizontal stripes. These simplifications also exist along the main sequence but because the points are denser and more scattered, it is not visible. Additionally, observational studies tend to span more than a single redshift, so data is broader in that sense.

See **Figure 2** for the SFR vs. Stellar Mass scatterplot and its corresponding two-dimensional histogram using data after redshift 0.5.

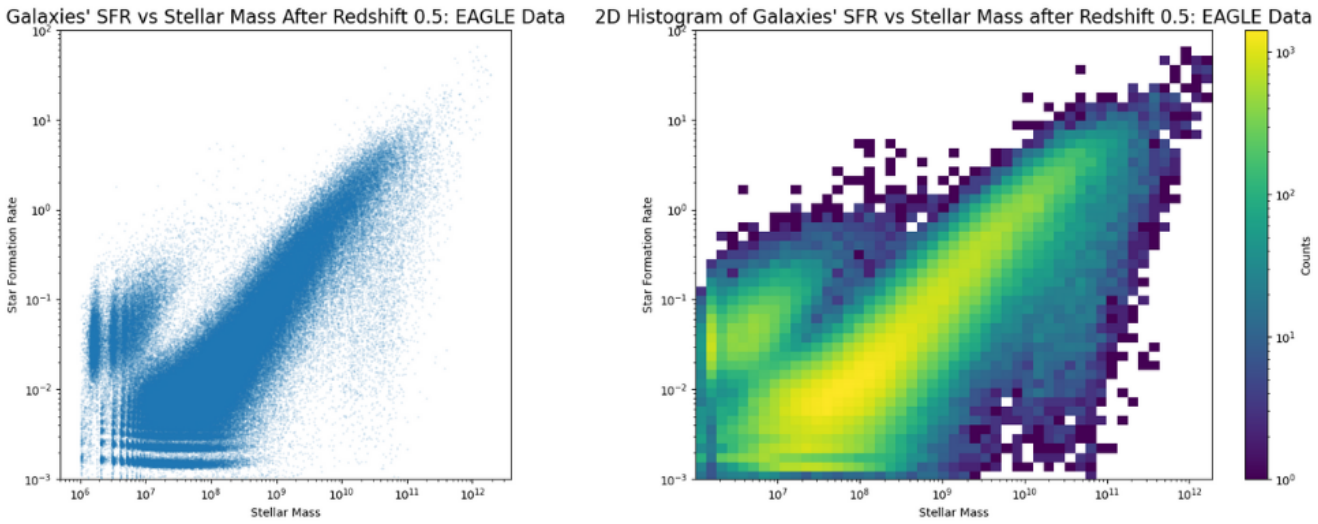

Figure 2.

It is clear from the figures that the plot of data after redshift 0.5 is denser everywhere compared to the plot of just redshift 0 (snapshot 28). The redshift 0 plot has 2,275,510 galaxies while the plot showing galaxies after redshift 0.5 has 11,819,412 galaxies. The plot after redshift 0.5 has more galaxies in the output and is therefore more dense and crowded. The distribution is comparatively more scattered, broadening the main sequence. It also reaches a higher SFR and thus the main sequence is steeper.

2.2. *IllustrisTNG Database Querying*

See **Figure 3** for the scatterplot and two-dimensional histogram of the IllustringTNG sample of data. It is evident that...

The number of galaxies in this sample is 53,939, which is significantly less than the EAGLE database 2,275,510 and 11,819,412 galaxy samples.

3. DISCUSSION

A simulation spanning several redshifts is more representitive of the data collected from an observational galaxy survey in real life because the line of sight of an advanced observational instrument can go so far that it can detect distant, redshifted galaxies where the light coming in is actually gigayears old. We do not observe the whole universe at a single instant, since the farther we look the older the information is.

The maximum redshift that a galaxy survey can see galaxies at depends on the capabilities of the observational instrument at use and the detectability of a distant galaxy. An instrument's sensitivity to luminosity and ability to filter out interfering light decides whether or not it detects a galaxy. A galaxy can also be harder to detect if the interstellar medium preceeding it absorbs a lot of its light or makes it harder to see through.

REFERENCES

Alton, P. 2016, Astrobites, Nov. 10, "The end of the line", https://astrobites.org/2016/11/10/the-end-of-the-line/

Alton, P. 2016, Astrobites, Apr. 18, "Reinventing star-formation", https://astrobites.org/2016/04/18/reinventing-star-formation/

Béthermin, M., Daddi, E., Magdis, G., et al. 2015, A&A, Jan. 7, "Evolution of the dust emission of massive galaxies up to z = 4 and constraints on their dominant mode of star formation" https://www.aanda.org/articles/aa/full_html/2015/01/aa25031-14/aa25031-14.html

Eales, S. A., Baes, M., Bourne, N., et al. 2018, Oxford Academic, Aug. 23, "The causes of the red sequence, the blue cloud, the green valley, and the green mountain", https://academic.oup.com/mnras/article/481/1/1183/5078380

McAlpine, S., Helly, J. C., Schaller, M., et al. 2015, arXiv:1510.01320