JOURNAL OF
EMERGING INVESTIGATORS

Article

# Exploring the Wonders of the Early Universe: Green Pea Galaxies and Light Flux

**Mina Mandic[1], Claudia Scarlata[2] , Karissa Baker[1]**
[1]St. Paul Academy and Summit School, St. Paul, Minnesota
[2]University of Minnesota, Minneapolis, Minnesota

## SUMMARY
**"Green Peas" are a unique set of galaxies characterized by low mass, low density, and high star formation rate. These properties are shared with Lyman alpha emitters, one of the first types of galaxies that existed in the early universe and played a role in reionization, a phase in the early evolution of the universe that is not well understood. Investigating the properties of Green Peas would improve the Lyman alpha emitters and their contribution to reionization because they are easier to see and measure. This project examines the light flux distribution emitted by Green Peas, separating the light from the stars and the surrounding gas. An algorithm implemented in Python was used to extract images of 80 galaxies from the Sloan Digital Sky Survey database and create images of the components produced by stars and oxygen gas. These images were used to make plots of Flux vs Distance from the center of each galaxy to show the regions from which the two components were emitted. The results show that the stars and the oxygen gas emit light from the same locations within the galaxy, hence providing new insight into the role of Lyman alpha emitters in reionization.**

## INTRODUCTION

"Green Peas" are galaxies discovered by volunteer citizen scientists in the Galaxy Zoo Project, a study to classify different types of galaxies (1). They are referred to as green because they emit strongly in the green part of the visible wavelength spectrum between 6,000 to 6,800 angstroms and are pea-shaped when seen from Earth due to unresolved atmospheric distortion (1). Green Peas have key characteristics that make them unique compared to other galaxies, such as their low mass, low density, and high star formation rate (2). These key characteristics are also shared with Lyman alpha (Ly$\alpha$) emitters, a type of galaxy that is thought to have an important role in the early evolution of the universe (3).

As the universe expanded and cooled, about 400,000 years after the Big Bang, protons and electrons combined to form neutral hydrogen, a process known as recombination. Neutral hydrogen absorbed most of the far-ultraviolet (UV) light present at the time, making the universe unobservable. Over time, the hydrogen formed gas clusters that eventually became the first stars and galaxies, including Ly$\alpha$ emitters. These stars and galaxies emitted high energy photons that were capable of reionizing the hydrogen atoms, splitting them back into protons and electrons (4), known as reionization (5).

How and when did these first galaxies form? What were their properties? What were the characteristics of the light they emitted? These are all unanswered questions to date. By studying reionization, it is possible to learn about how the structure of the universe began to form, transitioning from the uniform hydrogen gas to the complex galaxy structures observed today. However, directly observing Ly$\alpha$ emitters is difficult because they are far from Earth (about 12.8 billion light years away) and partly concealed by the neutral hydrogen gas. Green Peas, on the other hand, are relatively closer (about three billion light-years from Earth), and therefore easier to observe (6). We proposed to investigate the light emission from Green Peas, leveraging the similarity between Green Peas and Ly$\alpha$ emitters as an innovative approach to advance the understanding of the role of Ly$\alpha$ emitters in the early stages of the universe.

The spectrum of light emitted from a Green Pea includes two components. First, there is a flat continuum part of the spectrum with wavelengths in the range of 4,000-8,000 angstroms at a relatively constant spectral intensity (7). This is the natural constant flux emission from stars. Second, there is a sharp emission line near 6,000 angstroms (within the red part of the visible spectrum) that comes from the oxygen gas [OIII], whose distribution in and around the galaxy is not known (3). In a galaxy, the stars often occupy the central region while the interstellar gas extends much further out, sometimes to distances about 10 times farther than the stars. The interstellar gas flows from the outer regions of the galaxy towards its center and back out, and is recycled during the star formation. During the star formation process, the oxygen gas can flow far from the galaxy or can be recycled back into the galaxy (7). This project is the first to determine the location of the gas in or around a Green Pea galaxy by comparing the location of the source of the emission line with that of the continuum. Knowing the location of the gas would help explain how much light escapes from galaxies like Ly$\alpha$ emitters (and is not absorbed by the contents of the galaxy), resulting in the reionization of the universe. By studying the distribution of light from gas and from stars in a Green Pea galaxy, we hypothesized that the emission line would be emitted from a larger region than the continuum. This hypothesis is based on the expectation that the gas in a Green Pea galaxy takes up a greater volume than the stars.

We found that, in all observed galaxies, the continuum has a trend of having higher flux than the emission line, yet the emission regions of the two components were statistically similar.

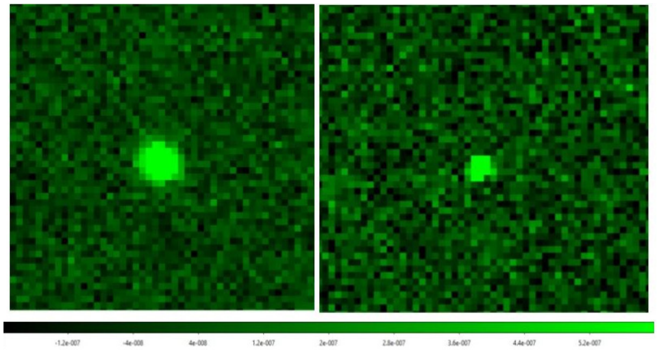

**Figure 1. Images of galaxy number 587735663159738526.** Shown are a continuum image (left), and an emission line image (right). The pixel size is the unit used to determine the distance of any pixel to the center of mass. The scale represents pixel intensity (or flux) in microJanksys with green representing high intensity and black representing low intensity. Created using Python.

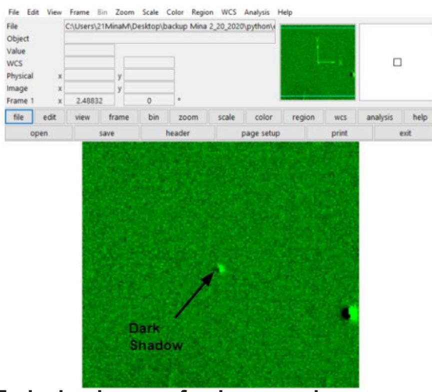


**Figure 2. Emission image of galaxy number 587724241767825591.** Rendered in the software used to view the images (DS9). The dark regions indicate imperfect alignment of images in different filters.

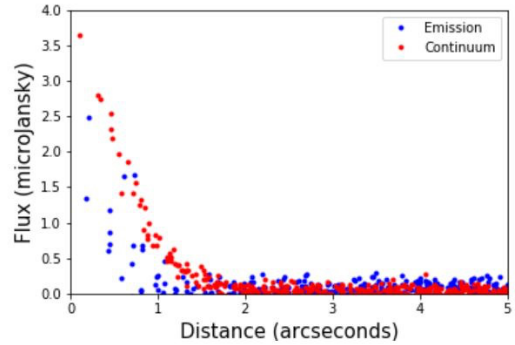


**Figure 3. The Flux vs Distance graph for galaxy number 587735663159738526 (same galaxy as in Figure 1).** The blue dots represent the emission line flux vs distance from the center of mass, and the red dots represent the continuum flux vs distance from the center of mass.

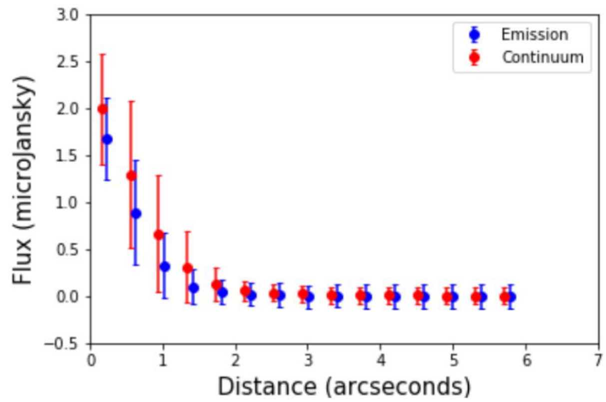


**Figure 4. Flux vs Distance graph averaged over all 27 galaxies.** The error bars correspond to the standard deviations in each of the distance bins. The points were shifted in the x direction by one tenth of a pixel for clarity.

## RESULTS

To measure the flux of Green Pea galaxies, we used telescope observations in different wavelength bands to create images of the emission line and continuum part of the spectrum for each galaxy. These images were then used to extract information about the location of the emission from stars and oxygen gas. The continuum exhibited a higher flux than the emission line, which was a common pattern in the analyzed 27 galaxies (**Figure 1**). Additionally, some galaxies from the original sample of 80 galaxies were rejected due to multiple factors that would alter their center of mass estimate. This includes images with imperfect alignment, galaxies with other nearby objects, galaxies outside of the redshift range, and galaxies that were either too dim or had too much background noise (**Figure 2**). In this case, the center of mass was altered by imperfect alignment. The center of mass must be accurate because it is used to calculate distance from each pixel to the center of mass, which allows for an understanding of the distribution of light flux.

The pixel intensity data from these images was used to generate Flux vs Distance graphs for each individual galaxy (**Figure 3**). These graphs also show that the continuum flux was greater than the emission line flux, as previously noted. As the distance increased beyond the extent of the galaxy, both the emission line and the continuum approached a flux near zero microJanskys, a unit of flux equivalent to 10-26 Watts per square meter per Hertz.

Next, we averaged the Flux (continuum and emission) vs Distance graphs from the 27 selected galaxies with ± 1 standard deviation error bars for both the emission line and the continuum (**Figure 4**). The mean continuum flux appeared to be greater than the emission line, but the two curves were similar to each other. The overlapping error bars indicated that the curves were not statistically different from one another. This means that the emission line and the continuum were emitted from similar distances from the center of mass of the galaxy. In other words, the stars (that emit the continuum) and the oxygen gas (that emits the emission line) occupied similar regions within the galaxy.

## DISCUSSION

Our project was the first study that examined the distribution of light flux from Green Peas, using the set of galaxies discovered by citizen scientists in the Sloan Digital Sky Survey (SDSS). Specifically, we measured the flux vs distance relationship for the continuum light component emitted by stars in the galaxy and the emission line component (around 6,000 angstroms) emitted by the oxygen gas in the galaxy. The results of this project indicated that the gas (emission line) and the stars (continuum) have similar distributions around the Green Pea galaxies, which provides new insight about the role of the first galaxies in the reionization phase of the universe.

Our study determined the spatial organization of Green Peas, which in turn constrains the amount and wavelength distribution of the light escaping the galaxy. This finding may serve as a key input for scientists constructing models and simulations of Green Peas because models in which the stars and gas occupy similar regions are more consistent with our observations than the models in which the gas extends much farther than the stars. Since Green Pea galaxies are similar to the early universe galaxies, such as Lyman alpha emitters, these models will likely help to indicate whether Lyman alpha emitter galaxies emitted enough high-energy photons to play an important role in the reionization phase of the universe. Furthermore, these models could establish the properties of the first galaxies in the universe, such as their mass, density, and star formation rate, and explain how the large-scale structure in the universe formed from the initial neutral hydrogen gas.

The results of this study were limited by the quality of the available data. The availability of the photometric SDSS data in only five wavelength bands provided limited information regarding the frequency spectrum of Green Peas. Having spectroscopic measurements that provide detailed frequency spectrum emitted by Green Peas could yield a much better estimate of the OIII emission lines that would not depend on manual alignment of images nor the subtraction of the continuum contribution. Additionally, the study was limited by the small number of observed Green Peas, which was further reduced by the imposed data quality cuts. Future observations by the upcoming Large Synoptic Survey Telescope (LSST) and the EUCLID space-borne telescope will provide many more images of Green Pea galaxies. Repeating our analysis with this increased sample size would likely reduce the uncertainties in the Flux vs Distance graphs and help reveal possible differences between the continuum and emission line components.

## METHODS

Python was used to extract and analyze data on Green Pea galaxies from the Sloan Digital Sky Survey (SDSS) database (1,8). The four primary coding libraries used were numpy, scipy, matplotlib, and astropy. The SDSS provides five pixelated images for each object in FITS format, one image for each of the five wavelength bands: u, g, r, i, and z (9). The OIII emission line of Green Peas falls in the range of 6,000-6,800 angstroms, which is known as the r-band. The bands right next to the r-band are the g- and i-bands, which represent the green and infrared wavelengths, respectively. Since the continuum section of the spectrum is flat, the average of these two bands provided an accurate estimate of the continuum.

Cardamone et al. found 80 Green Peas from the SDSS (1). Starting from this set of galaxies, 48 galaxies with an ideal distance from the Earth were selected. Specifically, the galaxy redshift (ratio of observed to emitted wavelength minus one) was required to be in the range [0.25, 0.35], to ensure that the OIII emission line falls within the r-band.

Pixelated images of the continuum and emission line for each galaxy were computed as follows. For each Green Pea galaxy, the continuum image was created by averaging the g- and i-band images. Since images in different bands were not aligned, they were manually shifted to result in perfect alignment. The emission line image was created by subtracting the continuum image from the r-band image for each galaxy.

To study the distribution of the flux for both the emission line and the continuum, the center of mass was first computed for both the continuum and OIII emission line for each galaxy using the astropy library in Python, which assumes pixel brightness as a proxy for mass. Then, for each pixel in the continuum image, its distance from the corresponding center of mass was computed using the distance formula (Eq 1):

$$d = \sqrt{[(x_c - x_p)^2 + (y_c - y_p)^2]}, \qquad \text{(Eq 1)}$$

where ($x_c$, $x_p$) are the coordinates for the center of mass of the galaxy, and ($y_c$, $y_p$) are the coordinates of a pixel. The same calculation was repeated for the emission line image. Finally, for both images, the pixel intensity (flux) was graphed as a function of its distance from the center of mass (**Figure 3**). The x-axis (distance) unit was converted from pixels to arcseconds, which are commonly used to describe distances in the sky. This was done by multiplying each point by 0.396 in Python before plotting.

We developed exclusion criteria based on the emission line and continuum images and the

Flux vs Distance graphs and applied them to identify the most accurate images. Galaxies with too faint emission line images had noisy Flux vs Distance graphs and were removed. Similarly, galaxies with imperfect alignment resulting in dark shadows around the galaxy in the emission line image were also removed from the analysis. Finally, galaxies whose images contained another nearby object within 20 pixels were removed from analysis because they had incorrect estimates of the center of mass. After applying the exclusion criteria, 27 galaxies remained.

The Flux vs Distance graphs of these galaxies were averaged on both the x-axis (distance) and the y-axis (flux) to reduce the uncertainty on each of the two curves (emission line and continuum), making the difference between them clearer. First, for each galaxy, the distances were split into 15 "bins" of one pixel width on the x-axis. For example, bin one contained distances between zero and one pixel from the center of mass, bin two contained distances between one and two pixels, and so on. As there are typically multiple pixels in each bin, the average flux and the standard deviation were computed for all the pixels within each bin. These flux values were further averaged over the 27 galaxies. Finally, the resulting Flux vs Distance curves for both the emission line and the continuum were plotted in a single graph including the error bars (**Figure 4**).

**Received:** September 16, 2020
**Accepted:** June 30th, 2021
**Published:** August 22, 2021

## REFERENCES

1. Cardamone, Carolin *et al.* "Galaxy Zoo Green Peas: Discovery of a Class of Compact Extremely Star-forming Galaxies." *Monthly Notices of the Royal Astronomical Society*, vol. 399, Issue 3, 2009, pp. 1191–1205. DOI: doi.org/10.1111/j.1365-2966.2009.15383.x.
2. Amorín, Ricardo *et al*. "Analogues of primeval galaxies two billion years after the Big Bang." *Nature Astronomy*, vol. 1, 2017, p. 0052. DOI: doi.org/10.1038/s41550-017-0052.
3. Tumlinson, Jason *et al.* "The Circumgalactic Medium". *Annual Review of Astronomy and Astrophysics*, vol. 55, 2017, pp. 389-432. DOI: doi.org/10.1146/annurev-astro-091916-055240.
4. Freedman, Roger *et al*. Universe. 10th ed., *WH Freeman and Company*, New York, 2015.
5. Byrd, Deborah. "Peering Toward the Cosmic Dark Ages." *EarthSky*, July 16, 2017. earthsky.org/space/cosmic-dark-ages-lyman-alpha-galaxies-lager. Accessed 1 June 2021.
6. Izotov, Yuri *et al*. "Eight per cent leakage of Lyman continuum photons from a compact, star-forming dwarf galaxy." *Nature*, vol. 529, 2016, pp. 178–180. DOI: doi.org/10.1038/nature16456.
7. Straub, Miranda. "Giving Citizen Scientists a Chance: A Study of Volunteer-led Scientific Discovery". *Citizen Science: Theory and Practice*, vol. 1, 2016, p. 5. DOI: doi.org/10.5334/cstp.40.
8. York, Donald. *et al*. "The Sloan Digital Sky Survey: Technical Summary". *The Astronomical Journal*, vol. 120, 2000, pp. 1579-1587. DOI: doi:10.1086/301513.
9. Doi, Mamoru *et al*. "Photometric Response Functions of the Sloan Digital Sky Survey Imager" *The Astronomical Journal*, vol. 139, no. 4, 2010, pp. 1628-1648. DOI: doi.org/10.1088/0004-6256/139/4/1628.