USING COMPUTER PROGRAMMING TO SEARCH FOR TRENDS IN THE ATMOSPHERIC COMPOSITIONS OF EXTRASOLAR PLANETS

Shushmitha Radjaram, Hielen Enyew, and Karina (Aria) Xin-Yi Li

ABSTRACT: Since the discovery of extrasolar planets (planets orbiting another star) in the early 1990's, more than 4,350 exoplanets have been confirmed to exist as of February 2021 according to the NASA Exoplanet Archive. Given the large amount of new data generated on exoplanets, our research team has been searching for trends between exoplanets' atmospheric compositions and their physical and orbital properties. We have developed computer programs to automate the steps in our process of searching for these trends. These steps include gathering exoplanets' physical and orbital property measurements, identifying planets within our search scope, studying the atmospheric composition of exoplanets, and searching for potential trends. In addition to developing programs, we place a heavy emphasis on the documentation of our programs and processes. By providing extensive documentation, we strive to make our programs and processes easily understandable and usable for the astrophysics research community. Our team aims for our results, programs, and documented research processes to be published on forums such as GitHub, together with research journals.

1. Introduction

In 2001, the first detection of sodium absorption in an exoplanet's atmosphere was made (Charbonneau et al., 2002). Since then, a growing number of more than 4,100 exoplanets have been confirmed to have published atmospheric data (NASA Exoplanet Archive). Transmission spectroscopy is used to determine the existence of an element or molecule in a planet's atmosphere. If an exoplanet contains an element/molecule in its atmosphere, the light from its host star passing through its atmosphere at a wavelength corresponding to the element/molecule will be absorbed.

In addition, there has been a large amount of physical and orbital data collected and published on websites such as the NASA Exoplanet Data Explorer. However, few studies have been conducted with the purpose of determining trends between physical/orbital data and atmospheric data. Our research is focused on finding these trends.

With the increase of data on exoplanets needed to analyze in our search for trends, we needed to develop a system to automate the extraction, organization, and analysis of data. Our research process has time intensive and data heavy tasks that allow for human error. The ability for computer programs to process large amounts of data and automate processes will help us to reach accurate conclusions efficiently.

2. Methods

Our goal with our computer programming is to automate the search for trends between the physical/orbital properties and the atmospheric composition of exoplanets. In Section 2.1, we show the automatic method we have developed to collect data on planets within a specific scope. In Section 2.2, we explain our program that follows the requirements of a criteria to automatically conclude whether a planet has atmospheric absorption of an element or molecule. In Section 2.3, we describe how our plot-making programs help us to search for trends. Section 2.4 explains the database we are developing to store and manage the most recent data needed for and generated by our research. Lastly, in Section 2.5, we explain the documentation for our computer programs, which help make our programs accessible and easy to understand.

2. 1 Planet Selector Program

Our primary source of data on exoplanets' physical and orbital property measurements is the Habitable Zone Gallery (Kane et al., 2012). This website, developed by Dr. Stephen Kane and Dr. Dawn Gelino, extracts the names and measurements of physical and orbital properties of exoplanets from the NASA Exoplanet Data Explorer and calculates, among other things, the extent of the habitable zone around each exoplanet's star and the percentage of time the exoplanet spends there during its orbit. The website lists this information in a table, which can be downloaded in the form of a .csv (comma-separated value) file. To collect our data, our team downloads this .csv file and uses a computer program to find which planets fall into our search scope. The search scope is an area of focus consisting of a range of two physical or orbital properties.

Our team developed a Python computer program, named the Planet Selector Program, to read the .csv file and create a Microsoft Excel Sheet that lists the names of the planets within our search scope along with their physical and orbital property measurements given in the Habitable Zone Gallery (Maloney et al., 2017; Weber et al., 2019; Broussard et al., 2020). In addition, this program creates a scatter plot that plots all planets included in the HZG table based on any two physical properties and draws a red box indicating the search scope of our current search. Figure 1 is the scatter plot created by the Planet Selector Program to represent our

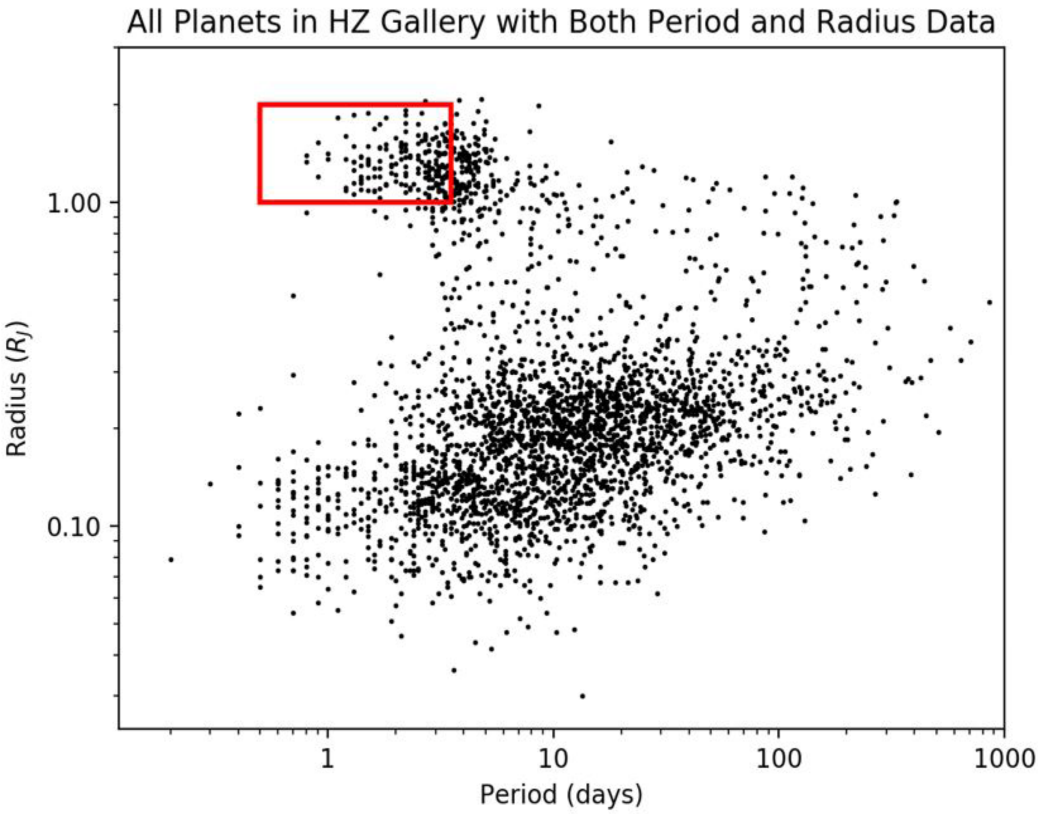

Figure 1 (above): Plot of all the planets in the Habitable Zone Gallery based on orbital period and radius measurements with a red box indicating the search scope.

current search scope of exoplanets with orbital periods of less than 3.5 Earth days and radii between 1 and 3 times the radius of Jupiter. We have developed a functionality within the Planet Selector Program that allows the user to input the two physical properties they would like planets to be graphed based on and the ranges of these physical properties they would like to represent as the search scope. After inputting these values in the indicated section of the program and running it, both the scatter plot and the Excel Sheet are saved to the files specified in the user inputs section of the program.

To keep up with the fast pace of exoplanet and exoplanet property findings that is a result of the improvement of detection technology, we are currently modifying the Planet Selector Program to directly read from the Habitable Zone Gallery website. This will ensure that we are working with the most recent data and eliminate the need to download the .csv table. We are in the process of creating a series of programs to make scheduled data extractions from the website. These data extractions will automatically occur every few days and update the data and scatter plot generated by the Planet Selector Program if there are any changes to the data in the Habitable Zone Gallery since the last extraction.

2. 2 Planet Absorption Criteria Program

After using the Planet Selector program to find which exoplanets are within our selected search scope, the next step in our search for trends is concluding the exoplanets' atmospheric absorption characteristics.

After reading through published literature on exoplanets in our search scope to find atmospheric absorption data, we note which wavelength ranges overlap. We choose a wavelength range that contains absorption of one or more elements to focus on. We then collect data on published literature on planets with spectral data in the chosen wavelength range, specifically looking for conclusions that the authors of the paper make on whether a

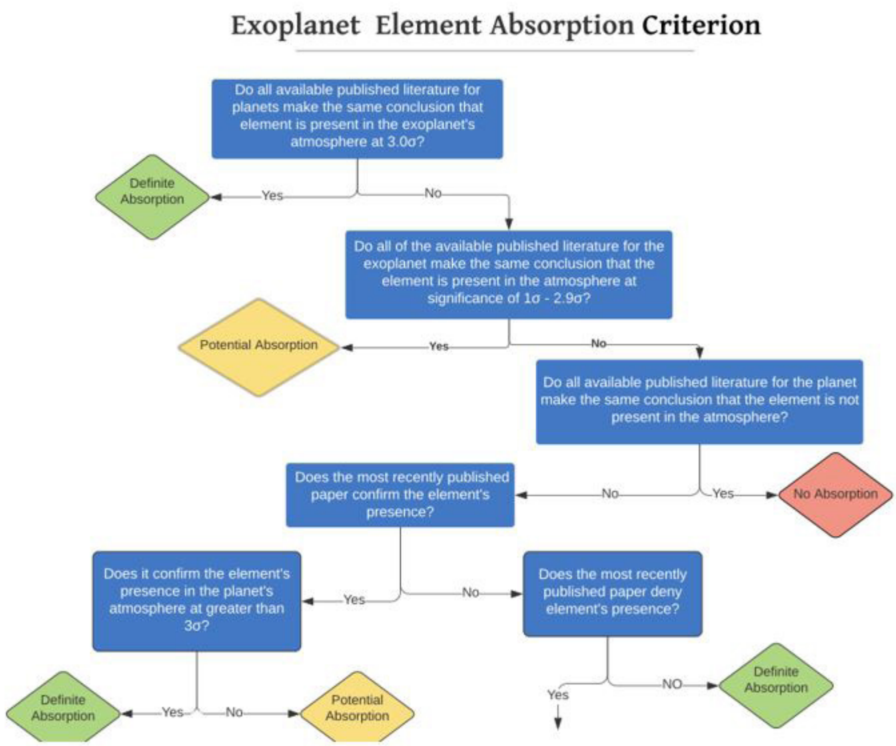

Figure 2 (above): Section of the Criteria flow for determining sodium and potassium absorption in an exoplanet's atmosphere.

planet has atmospheric absorption of an element or molecule.

Oftentimes, multiple papers written on a single planet make contradictory conclusions. This can be caused by the instruments used, the time the data was collected, or variability in the presence of haze/clouds in the atmosphere of the exoplanets. We have developed several criteria to classify whether a planet has atmospheric absorption of sodium, potassium, or water when contradicting conclusions are made between different papers. The criteria take into account the conclusion authors from each paper make on a planet's absorption of sodium, potassium, or water; the number of papers with authors that make a specific conclusion, the dates that the papers were written, and the instruments used to make these detections. The criteria we developed are designed to consider each piece of information listed above in a specific order of requirements. After following the instructions of the criteria, a planet can be concluded to have definite, potential, or no absorption of sodium, potassium, or water. Figure 2 shows a section from the criteria we developed to determine an exoplanet's atmospheric absorption of sodium and potassium.

Currently, members need to go through the criteria manually. This is a very time-intensive and detailed process, which can easily lead to human error. To solve this issue, we decided to develop a program that follows the requirements of this criteria and outputs the classification of absorption each planet has of sodium, potassium, and water. The Planet Absorption Criteria Program reads data needed for the criteria directly from a Google Sheet in the research team's shared Google Drive. To achieve this functionality, the Python API for Google Sheets, "gspread" is imported into the program. This allows the program to open, read, and write to the specified Google Sheet. To direct the program on which Google Sheet to write to, a JSON file that corresponds to the Google Sheet is saved to the same folder in which the program is saved to. The file is used in the program to verify that the person running the program has the credentials to open, read and write to the Google Sheet. Researchers will not be able to direct their program to access the Google Sheet if they do not have access to the JSON file.

By implementing a connection between a Google Sheet and the Planet Absorption Criteria Program, any updates made to the data happens in one place only and the program reads the most recent data.

2. 3 Property Plot Program

Our next step is to determine if there are correlations between the physical/orbital data and the atmospheric data. To do so, we create scatter plots that graph planets based on two physical or orbital properties and indicate each planet's absorption characteristics (Broussard et al., 2020). As it can be seen in Figure 3, one physical or orbital property is used to represent the x-axis, and another physical or orbital property is used to represent the y-axis. The colored points are used to represent planets. The green circles around the points represent potassium absorption in the atmosphere of the planet represented by that particular point. The blue circles around the points represent sodium absorption in the atmosphere of the planet represented by that particular point. Solid green and blue circles around points represent definite absorption of these elements. Dashed green and blue circles around points represent potential absorption of these elements. No green and/or blue circles around a point represent no absorption of these elements.

We are in the process of solidifying a set for criteria for water and have not yet classified planets based on their water absorption. Therefore, Figure 3 does not include water absorption.

We use these scatter plots created by the Property Plot Program to look for trends by

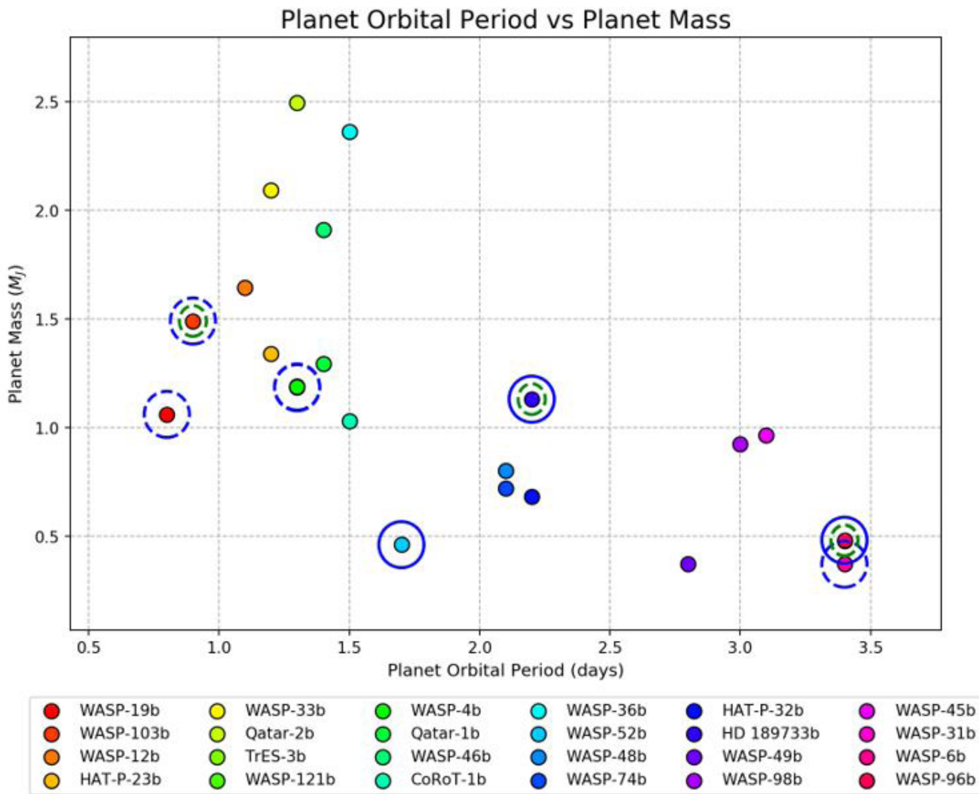

Figure 3 (above): Plot of planets based on orbital period and planet mass with sodium absorption indicated by blue circles around points and potassium absorption indicated by green circles around points. Dashed circles around points correspond to potential absorption and solid circles around points correspond to definite absorption.

noting whether all or most points with blue and/or green circles around them show to be in one specific section of the plot. For example, if most of the points with green circles around them fall in the section of a plot that indicates a relatively large mass and small radii, we can conclude that exoplanets with potassium absorption in their atmospheres have large masses and small radii. These scatter plots provide a way to bring our physical/orbital data and our atmospheric data together and see the correlation between the two. Statistical analysis, specifically 1D and 2D ks-testing, is conducted beyond this point to confirm the trends concluded.

2. 4 Database

We are currently in the process of creating a PostgreSQL database to hold all of the collected data. Interfaces that increase accessibility will be implemented so team members with various research backgrounds can easily access and use the data. Implementing the database will require us to switch from using Google Sheets as our shared data forum to a SQL database. The switch to the database would enable our team to easily access the same, most recent data in one place. However, this affects all of our software and we will need to modify each of the programs to use the database to input and extract our data. More specifically, this means the location(s) of the inputs and outputs of our programs will be changed to the SQL database.

2. 5 Documentation

All our published programs will be documented using Sphinx. Sphinx is a documentation tool that can be used for various software languages. The documentation would make our programs and processes more understandable to researchers that are not as acquainted with our research. As a future goal, we would like to publish our work and software to GitHub and other public platforms for others to use, and our documentation would allow them to do so with ease.

3. Conclusion

Computer programs have shown to improve our search for trends by automating our steps and returning accurate data and results. With them, we have been able to establish a search scope and collect data on exoplanets within that scope efficiently, automate the process of determining whether a planet has absorption of an element or molecule, and make conclusions on trends between the atmospheric composition and physical/orbital properties of exoplanets. The addition of our database ensures the most recent data exists in a localized platform that is easy to access and is reliable. Finally, our computer program documentation will help our team members and other researchers efficiently learn how to use our programs.

By integrating computer programming as a vital component of our research, we show how astrophysics and computer programming go hand-in-hand, better acquaint members of the research group with programming, and provide the astrophysics research community with highly versatile and applicable programs to be used for different types of research on exoplanets.

All members of our research team are required to learn basic programming in order to develop, run, and understand what our computer programs do. By guiding them through this process, we are educating research members on skills that are relevant in several STEM fields. Our team aims for our programs and documented research processes to be published on public forums such as GitHub, as well as research journals. The results of our search for trends in the atmospheric compositions of extrasolar planets will be included in the Habitable Zone Gallery. This way, our programs, processes, and results will be accessible to the community of researchers who wish to use this information as well.

References

Broussard, W., et al. (2020) "Searching for Trends in the Atmospheres of Exoplanets". *The University of Washington Bothell CROW*: Vol. 5.

Charbonneau, D., et al. "Detection of an extrasolar planet atmosphere". *The Astrophysical Journal* 568 (Mar. 2002): 377-384.

"Exoplanet and Candidate Statistics." *NASA Exoplanet Archive*, CALTECH, exoplanetarchive.ipac.caltech.edu/docs/counts_detail.html.

Kane, S. R. and Gelino, D. M. "The Habitable Zone Gallery". *Publications of the Astronomical Society of the Pacific* 124.914 (Apr. 2012): 323–328.

Maloney, Troy. "A Search For Trends In The Atmospheric Composition Of Exoplanets From Archived Spectral Data". National Conference of Undergraduate Research 2017, 7 Apr. 2017.

Weber, K., et al. (2019) "Searching for Trends in Atmospheric Compositions of Extrasolar Planets," *IdeaFest: Interdisciplinary Journal of Creative Works and Research from Humboldt State University:* Vol. 3.