Measuring the Influence of Stellar Neighbors on a Planetary System's Potential for Habitability

David Vaida

Received January 10, 2025
Accepted April 04, 2025
Electronic access April 30, 2025

This study explores the impact of stellar environments on the presence of rocky planets within the habitable zone (HZ) of planetary systems. Dense stellar clusters are known to disrupt protoplanetary disks and influence planetary system architectures, often favoring the formation of massive, close-orbit planets. However, the extent to which these environments affect the likelihood of forming HZ planets remains insufficiently studied. Using a dataset of 516 planetary systems, including 53 with HZ planets, this research employs logistic regression, correlation analysis, and the Kruskal-Wallis test to identify key environmental factors. The analysis highlights that host star properties, such as mass and temperature, and cluster characteristics, including astrometric weight and proximity to the neighboring stars, significantly influence the formation of HZ planets. Strong correlations were found between closer neighboring stars and average flux of the cluster with increased HZ habitability, while clusters with larger astrometric weight appeared to hinder HZ planet formation. The logistic regression model achieved a 94.6% accuracy rate in predicting the presence of HZ planets, emphasizing the critical role of stellar dynamics. The findings provide valuable insights into the environmental conditions conducive to supporting rocky planets in the HZ, advancing our understanding of habitability across the universe. This research also offers guidance for future telescope observations and mission planning, contributing to the search for potentially habitable worlds.

Keywords: Stellar environments, habitability, planetary systems, exoplanets, star clusters.

Introduction

Over the last few decades, astronomical observations have yielded a large number of exoplanetary discoveries, revealing a universe full of diverse planetary systems. The formation of these planetary systems is highly dependent on the interactions with the surrounding star cluster environment. Our own Solar System is believed to have been born in a relatively dense stellar cluster with at least one high-mass star that provided short-lived radioactive isotopes, which are still present in Solar System meteorites[1]. Planetary formation simulations show that protoplanetary disks in dense stellar environments, including those in the habitable zone, are highly susceptible to disruption from neighboring stars[2–4].

Recent studies have shown that the density of the surrounding stellar environment plays an important role in shaping the architecture of the planetary systems. Planetary systems in high density regions are more likely to be single-planet systems, in contrast to their counterparts in low density environments[5]. Furthermore, these dense clusters seem to favor the formation of massive planets that orbit closely to their host star[6]. Another notable difference has been observed in the distribution of planet radii. Planetary systems in areas with lower stellar density have a much larger population of planets with a radius less than twice that of Earth. In contrast, planetary systems in high-density stellar regions have far fewer of these smaller planets[7]. Orbital properties also show significant differences in systems across various cluster densities. Planets in dense clusters tend to have shorter, less eccentric orbits and smaller semi-major axes. These planets are typically larger in terms of mass and radii compared to planetary systems in less crowded clusters[6,8]. These differences highlight the significant impact of the stellar neighbors on the architecture and characteristics of planetary systems.

However, a crucial question of how stellar clusters influence the ability of a system to host rocky planets that orbit their stars in the habitable zone—where liquid water could be present—remains largely unanswered. This study proposes that certain characteristics of star clusters, such as lower average astrometric weight and larger distances between stars, may enhance the likelihood of planetary systems hosting such planets. While addressing this significant gap, it is important to note that the dataset used here is relatively small (516 planetary systems, only 53 with habitable-zone planets), which can affect the reliability of statistical models. Nevertheless, by investigating key stellar environmental factors, this research aims to deepen our understanding of the conditions necessary for the emergence of life in the universe and significantly advance the ongoing search for habitable planets beyond our solar system. Ultimately, these

© The National High School Journal of Science 2025

insights can guide future space exploration missions and telescope observations, bringing us closer to identifying potentially life-supporting worlds.

Methodology

To gain a deeper understanding of the conditions necessary for planetary habitability, it is important to analyze the diverse environmental forces that influence planets that orbit a host star within the habitable zone. To achieve this objective, a combination of four distinct methodologies were employed: logistic regression, correlation analysis, statistical testing, and exploratory data analysis. Logistic regression was used to develop a predictive model that estimates the likelihood of a planetary system having habitable-zone planets, based on 32 characteristics of its host star and its surrounding stellar environment. This model can be utilized to identify potential targets for telescope observations or exploratory missions. Correlation analysis was used to pinpoint cluster and host star attributes that are significantly associated with the existence of habitable-zone planets. Lastly, exploratory data analysis and statistical testing identified specific characteristics of a planetary system environment that are significantly different between planetary systems with habitable-zone planets and those lacking such planets.

Stellar environment and planetary systems data was collected from three separate sources. The Nasa Exoplanet Catalog hosted by NASA's Ames Research Center provided detailed information on more than 5,000 confirmed exoplanets, and over 3,000 distinct planetary systems including their stellar host characteristics[9]. This catalog has been used by other researchers to examine patterns in the creation and distribution of exoplanets based on the host star attributes[10,11]. The categorization of habitability, which represents the dependent variable, was obtained from the University of Puerto Rico Planetary Habitability[12]. This dataset identifies planetary systems with rocky planets that orbit within the habitable zone—defined as the region around a star where conditions may allow for the presence of liquid water on a planet's surface. Approximately 53 planetary systems have rocky planets with their habitable zone. Finally, the stellar neighborhood data was gathered from the European Space Agency Gaia Archive[13]. Gaia contains astrometric measurements for over 3 billion stars. The stellar environment of each planetary system was defined by considering all stars within a 20-parsec radius. This threshold has been selected to account for the fact that typical planetary clusters can span about 20 parsecs across, according to Australian National Telescope Facility[14]. The number of stars in the cluster along with the mean, minimum and maximum stellar properties, including luminosity, flux, mass, astrometric weight, radius, velocity, temperature, and distance from the planetary system were calculated to obtain the stellar environment variables. Astrometric weight measures observational precision in arcseconds^{-2}, with higher values for brighter, often more massive stars[15]. Gaia data has been used to study the relationship between stellar characteristics and the architecture of planetary systems in numerous studies[6,16]. The Gaia data had to be accessed through python, Jupyter notebook, and Gaia-specific query language. A separate query had to be submitted for each planetary system, calculating the number of neighbors, the minimum, maximum and average stellar properties, and distances from the host star. A sample of 516 planetary systems, 53 of which represent systems with rocky planets in the habitable zone was selected for this study. Data analysis was performed in python using statistical libraries including Sklearn, Scipy, Numpy, Pandas and Matplotlib.

Logistic Regression, the first analysis method used, is a statistical model well-suited for identifying binary outcomes such as whether a system can support planets within its habitable zone or not. Regression models have been previously employed to study the relations between planetary systems and their host star[17,18]. This model can identify the key environmental factors influencing the presence of habitable-zone planets and estimate the likelihood of a system hosting such planets.

Analyzing the correlations between stellar cluster properties and the occurrence of rocky planets within the habitable zone of a planetary system can pinpoint the environmental factors that are most strongly linked to the presence of these types of planets. Previous research has leveraged correlational analysis to explore the relationships within planetary systems[6,8,19].

Exploratory data analysis and statistical tests were used to further investigate the trends and differences between star cluster characteristics and their potential impact on HZ planets. Scatter plots and boxplots visually represented trends and patterns that may not be captured by logistic regression and correlation alone. The Kruskal-Wallis test compared the median values of independent variables, revealing whether planetary systems with habitable-zone planets exhibit statistically significant environmental differences from those without. Taken together, logistic regression, correlational analysis, statistical testing, exploratory data analysis provided a more complete understanding of the factors that impact habitability.

Data preprocessing involved discarding any columns with more than 40% missing data and applying a KNN imputation strategy with 2 neighbors and uniform weights to fill remaining missing values by averaging the two closest observations. All continuous features were scaled using the Robust Scaler, which is less sensitive to outliers by centering each feature on its median and scaling it according to its interquartile range. A logistic regression model was optimized using a grid search method to identify the optimal hyperparameters. The resulting best parameters included a regularization strength (C) of 0.01, class weights set to 1:2 to address class imbalance, a maximum iteration limit of 1000, and L2 regularization penalty with the liblinear solver. The dataset was divided into training (75%) and testing (25%) subsets using stratified sampling to preserve the original class

© The National High School Journal of Science 2025

distribution. The parameter tuning was performed through a stratified 20-fold cross-validation to ensure robust selection of hyperparameters and mitigate overfitting. Class imbalance was addressed through stratified sampling, ensuring the minority class remained proportionally represented in both training and test sets. The grid search optimization tested different class weight adjustments (None, 'balanced', 0:1, 1:3, 0:1, 1:5, 0:1, 1:2) to improve sensitivity to underrepresented cases. A 20-fold Stratified K Fold preserved class distribution across validation splits, preventing overfitting to the majority class. Finally, to address the risk of inflated type I errors when conducting multiple Kruskal-Wallis tests on numerous variables, Bonferroni or false discovery rate (FDR) corrections were applied. This step ensured that the rate of false positives remained controlled despite the large number of statistical comparisons.

Results

The logistic regression model achieved a high overall accuracy of 94.6%, indicating strong predictive ability. The model exhibited a high precision of 87.9%, meaning that predictions identifying positive cases were largely reliable, with only two false positives. Recall, however, was moderate at 79.9%, indicating that the model correctly captured a majority—but not all—of the true positive cases, resulting in five false negatives. The F1-score of 83.3% reflects a good balance between precision and recall, which is particularly relevant given the dataset's class imbalance. The confusion matrix further detailed the performance, showing 114 true negatives, 8 true positives, 2 false positives, and 5 false negatives, illustrating the model's overall effectiveness alongside its moderate limitation in fully identifying all positive instances. The model coefficients revealed that the top five most significant factors in distinguishing the two categories, based on their absolute coefficient values, were host star mass and temperature, the minimum star mass of the cluster, the star with the largest flux within the cluster, and the number of planets. Among the variables with the lowest assigned coefficients were the minimum, maximum and average distance between the stars in the cluster and the host star, host star proper motion and the star with the minimum temperature in the cluster. The regression coefficients represent how strongly each variable influences the model's prediction about a planetary system's habitability. Specifically, a larger absolute coefficient value indicates that a given feature has greater predictive importance, meaning even small changes in this feature significantly impact the predicted likelihood of hosting habitable-zone planets. The values of all 32 regression coefficients are illustrated in Figure 1 and Table 1 Panel A.

While the host star temperature and mass are identified among the most significant coefficients in the logistic regression analysis, it does not negate the influence of the stellar cluster on the planetary systems. The scatter plot in Figure 2 illustrates the relationship between the host star mass and a star withing the cluster with the smallest astrometric weight. As the star's minimum astrometric weight decreases, the host star mass drops more significantly for systems with habitable-zone planets compared to systems without habitable-zone planets. This difference in trends is clearly visualized by the linear regression lines, generated using NumPy's polyfit function to fit a least-squares line for each group. Since astrometric weight measures a star brightness, a lower minimum value could indicate a smaller or less disruptive stellar neighbor, which may help preserve stable protoplanetary disks and thus enhance the potential for habitability. Therefore, host star mass could also be influenced by the composition of the cluster.

To quantify the relationship between stellar environments and the likelihood of a planetary system hosting HZ planets, a point biserial correlation was calculated. This method was chosen due to the binary nature of the dependent variable. Correlation coefficients were computed by first testing if a relationship exists between each independent variable and the dichotomous dependent variable (habitable versus non-habitable planetary systems), and then measuring the degree of the relationships. Out of the 32 independent variables, 11 were found to be correlated with the planetary system classification. The significant correlation coefficients are presented in Table 1 Panel B. Figure 3 also shows the correlation coefficients alongside the distribution of the variable for habitable and non-habitable categories. Host star mass, magnitude, temperature, radius, and proper motion were found to have significant correlations with the planetary systems classification. The correlation coefficients for host star mass, radius, and temperature were negative, indicating that as these values decrease, the probability of a planetary system hosting habitable-zone planets increases. The finding that lower host star mass corresponds with an increased likelihood of hosting habitable-zone planets relies on the assumption that smaller stars provide less disruptive conditions, facilitating stable planet formation in their habitable zones. However, this relationship could be influenced by additional variables not directly considered in the current analysis, such as stellar age, metallicity, or planetary composition. These unaccounted-for factors may further shape the observed connection between host star mass and planetary habitability. The correlations for host star magnitude and proper motion were positive, suggesting that brighter and faster-moving stars are more likely to host habitable-zone planets. This finding aligns with the results of the logistic regression model, which also identified the same trend in the coefficient signs for most of these variables. To gauge the practical importance of these findings, we also examined the effect sizes of the correlation coefficients. Most significant correlations fell within a low-to-moderate range (absolute values approximately 0.2–0.4), indicating that while these factors are statistically meaningful, they are not the sole determinants of habitability.

The number of planets in a system positively correlates with

the likelihood of hosting habitable-zone planets, as systems with more planets are generally more stable. It is therefore less likely to find HZ planets in denser clusters, since denser clusters favor single-planet systems dominated by gas giants [8].

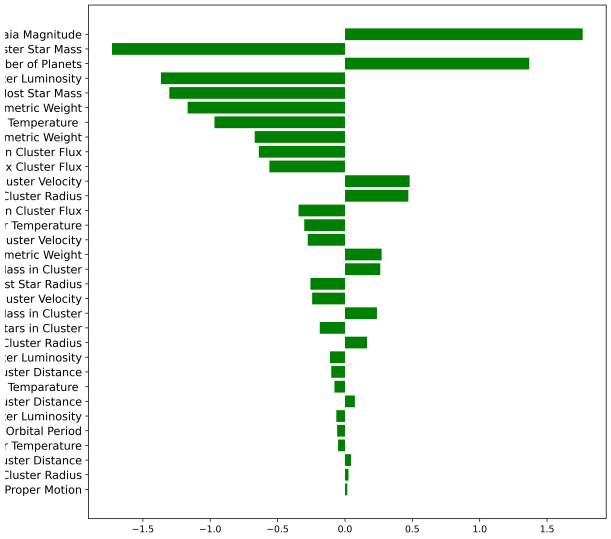

Fig. 1 Logistic regression coefficients. The importance of the variable is measured by the absolute value of the coefficients. The coefficients can be either positive or negative.

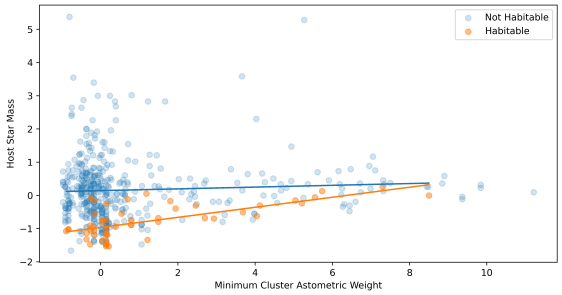

Fig. 2 Scatter plot of the host star mass relative to the minimum astrometric weight of the cluster. As the minimum astrometric weight of the cluster decreases, so does the mass of the host star for habitable -type systems compared with systems without habitable-zone planets.

Among the stellar cluster variables, the average and maximum distance between the stars in the cluster and the host star, and the average flux of the cluster were positively correlated with planetary systems classification. The mean astrometric weight had negative correlation coefficients. As the average and distance between the planetary system and neighboring stars increases, the likelihood of the system being habitable type also increases. Same trend has been observed for the distance between the host star and the farthest star within the cluster. Brighter clusters as measured by the average flux, also tend to have a higher likelihood of having planetary systems with HZ planets. In clusters with higher average astrometric weight, this probability decreases. Astrometric weight is defined as the relative size of the star compared with its surrounding stars. Therefore, having larger stars within the cluster can negatively impact a planetary system's habitability potential. This discovery aligns with existing literature which has shown that irradiation caused by massive nearby stars may disrupt a planetary system's ability to form and maintain [3,4,8]. One intriguing correlation result is the fact that distance between the host star and its closest neighbor is negatively correlated with the existence of habitable-zone planets. Having a nearby stellar companion may enhance a system's habitability by offering stabilizing gravitational effects and shielding it from disruptions caused by other, more distant stars. This aligns with evidence from the Solar System, where a nearby high-mass star is believed to have contributed short-lived radioactive isotopes that influenced planetary formation [1]. While dense stellar environments are known to disrupt planetary system formation, this finding suggests that in planetary systems that have already formed, the presence of a single close stellar neighbor may play a protective role, counteracting some of the destabilizing effects of the stellar cluster when considering habitability specifically.

Lastly, the Kruskal-Wallis test was used to identify significant differences in the median values of independent variables between planetary systems with and without habitable zone planets. All Kruskal-Wallis p-values were adjusted using a Bonferroni correction to account for multiple comparisons. Table 1 Panel C presents the 8 variables exhibiting statistically significant median values. Beyond the correlations previously identified, the Kruskal-Wallis analysis uncovered additional cluster attributes that distinguish planetary systems with HZ from those without. These key attributes include luminosity (corrected p-value = 0.032). and astrometric weight (corrected p-value = 0.023). As illustrated in Figure 2, decreasing minimum astrometric weight corresponds with a notable decrease in host star mass specifically for habitable-zone systems. Although the logistic regression assigned a moderate positive coefficient to minimum astrometric weight, this variable itself did not directly correlate with habitability. However, the Kruskal-Wallis test showed significant differences between habitable and non-habitable systems, likely because astrometric weight influences habitability indirectly by affecting host star properties, such as mass. The high ranking of cluster luminosity within the logistic regression coefficients confirms its importance in distinguishing between different planetary system types. Boxplots in Figure 4 show differences in the distribution of the independent variables between the two groups. Consistent with the logistic regression coefficients and

© The National High School Journal of Science 2025

correlational analysis, host star attributes seem to have better separation between planetary system types. Even though the cluster attributes are less visually separated, it is clear from the boxplots that non-habitable systems have many more outliers, which indicates that non-habitable clusters generally have a greater diversity of stars.

Correlation analysis and Kruskal-Wallis test identified 6 out of the top 10 variables with the highest regression coefficients. Interestingly, the correlation analysis also found the bottom 4 coefficients to be strongly correlated with planetary system type, while Kruskal-Wallis found the lowest regression coefficient as significant (Table 1). The logistic regression model has demonstrated that by utilizing star cluster attributes and host star characteristics, it can accurately identify, with a 94.6% success rate, a system that hosts habitable-zone rocky planets. The model coefficients quantified the importance of each variable in distinguishing between habitable and non-habitable planetary systems. These findings are corroborated by the correlation analysis and statistical testing, demonstrating that cluster and host star characteristics exhibit statistically significant differences between systems with and without habitable-zone planets.

While the existing research has demonstrated that the density of a star cluster influences the structure of planetary systems, surprisingly, the number of stars within a cluster does not seem to be a strong indicator for the presence of habitable-zone planets.

Conclusion

The architecture of planetary systems and the presence of rocky planets in habitable zones are heavily influenced by interactions between host stars, planet-forming disks, and environmental factors such as stellar flybys, radiation bursts, and gravitational disturbances from neighboring stars. Previous research studied the impact of the stellar environment density on the architecture of planetary systems and protoplanetary disks. However, there is a lack of comprehensive analysis in quantifying the precise stellar environmental parameters that have the greatest influence on the existence of habitable-zone rocky planets. Combining statistical modeling, testing, and exploratory data analysis can pinpoint specific stellar environment factors linked to the presence of rocky planets within habitable zones. This study quantifies how host star and cluster characteristics impact the likelihood of HZ planets, shedding light on critical environmental factors that shape habitable conditions.

The logistic regression model achieved 94.6% accuracy in classifying planetary systems with HZ planets, with host star mass and temperature, the star with the smallest mass and the star with the largest flux in the cluster identified as the most significant predictors. A positive correlation was observed between systems with more planets, which are often more stable, and the presence of HZ planets. These findings are consistent with existing literature, which has shown that cluster characteristics, such as high density, tend to favor single-planet systems dominated by gas giants which are less favorable for habitability. Additionally, environmental factors such as cluster luminosity, radius, and proximity to neighboring stars significantly influence habitability potential, implying that having a nearby stellar companion could boost a system's habitability. Clusters with greater average distances and lower astrometric weights are more conducive to HZ systems, while having a close stellar neighbor can sometimes stabilize protoplanetary disks, enhancing habitability.

This study underscores the dual role of host star and cluster attributes in determining the likelihood of HZ systems. While dense clusters generally reduce habitability potential by being more conductive of single-planet systems, specific characteristics, such as lower average astrometric weight and luminosity of neighboring stars, can exert stabilizing effects. These findings enhance our understanding of the complex interplay between stellar environments and planetary system evolution, offering insights into the conditions necessary for supporting life.

The results show that the properties of the stellar cluster and host star can be utilized to differentiate between planetary systems that have rocky planets within the habitable zone and those that do not. The logistic regression model, correlation analysis, and statistical testing identified significant connections between different attributes of a stellar cluster, host star, and the planetary system type. Host star characteristics may also be dependent on cluster dynamics, therefore the significant host star differences between the two planetary system types can be also attributed to the cluster dynamics.

A major limitation of the study is the limited dataset of 516 planetary systems, out of which only 53 have habitable-zone planets. This relatively small sample size and class imbalance could lead to optimistic performance estimates. While stratified cross-validation helps mitigate overfitting, the model's behavior in more diverse or significantly different datasets remains uncertain. Additionally, the sample may not be large enough to capture all the significant factors that influence their environments, given the small number of habitable-type systems. Additionally, while correlations exist between stellar environment characteristics and the presence of rocky planets in the habitable zone, these relationships do not imply a direct causal effect. The observed trends may be influenced by underlying factors not accounted for in this analysis. To better understand these associations, future research should incorporate additional data, explore potential confounding variables, and conduct simulation studies to assess the mechanisms driving planetary system formation and habitability. Nevertheless, this research offers valuable insights into the relationship between the conditions surrounding host stars and the existence of rocky planets within the habitable zone of a planetary system. This enhances our comprehension of the prerequisites needed for the presence of habitable exoplanets beyond our own solar system. These findings can provide important information for future space

A Regression Coefficients

32 independent variables were used to derive a regression equation that can separate between habitable and non-habitable planetary systems. The regression coefficients indicate the weights that the model assigns to each variable.

#	Independent Variables	Regression Coefficient
1	Host Star Mass	1.8710
2	Min Cluster Star Mass	1.5445
3	Host Star Temperature	1.2549
4	Number of Planets	1.0142
5	Max Cluster Flux	0.9183
6	Mean Cluster Luminosity	0.6969
7	Mean Cluster Astometric Weight	0.6691
8	Mean Cluster Flux	0.5323
9	Host Star Gaia Magnitude	0.5129
10	Max Cluster Astometric Weight	0.4500
11	Min Cluster Luminosity	0.4392
12	Min Cluster Astometric Weight	0.4141
13	Max Cluster Velocity	0.3855
14	Mean Cluster Radius	0.3460
15	Mean Cluster Temparature	0.2961
16	Min Cluster Flux	0.2534
17	Host Star Radius	0.2326
18	Max Cluster Temperature	0.2284
19	Max Star Mass in Cluster	0.1882
20	Planet Mean Orbital Period	0.1573
21	Max Cluster Radius	0.1347
22	Mean Cluster Velocity	0.1299
23	Min Cluster Velocity	0.1021
24	Min Cluster Radius	0.0633
25	Number of Stars in Cluster	0.0576
26	Mean Star Mass in Cluster	0.0399
27	Max Cluster Luminosity	0.0375
28	Min Cluster Temperature	0.0258
29	Host Star Proper Motion	0.0216
30	Max Cluster Distance	0.0133
31	Mean Cluster Distance	0.0064
32	Min Cluster Distance	0.0028

B Correlation Coefficients

Point biserial correlation between independent variables and dependent variable A negative correlation indicates that the changes of the system being habitable increases as the variable value decreases. A p-value smaller than .05 indicates that the correlation is significant.

#	Independent Variable	Correlation Coefficient	P Value
1	Host Star Gaia Magnitude	0.31	0.000
2	Host Star Mass	-0.30	0.000
3	Number of Planets	0.29	0.000
4	Host Star Temperature	-0.24	0.000
5	Mean Cluster Distance	0.19	0.000
6	Host Star Proper Motion	0.15	0.001
7	Min Cluster Distance	-0.13	0.003
8	Host Star Radius	-0.12	0.006
9	Max Cluster Distance	0.12	0.007
10	Mean Cluster Flux	0.11	0.017
11	Mean Cluster Astometric Weight	-0.09	0.041

C Kruskal-Wallis H-Test

Kruskal-Wallis H-Test comparing the median values of independent variables between planetary system types. A Bonferroni-adjusted p-value smaller than .05 indicates that the median values of habitable-type systems are different than non-habitable type systems.

#	Independent Variable	P Value
1	Number of Planets	0.000
2	Host Star Temperature	0.000
3	Host Star Radius	0.000
4	Host Star Mass	0.000
5	Host Star Gaia Magnitude	0.000
6	Mean Cluster Luminosity	0.032
7	Min Cluster Astometric Weight	0.023
8	Min Cluster Distance	0.024

Table 1: Panel A: Coefficient values for logistic regression. Coefficients highlighted in green are significantly correlated with the planetary system type. Significant differences in the median values between habitable and non-habitable systems are shown in red text. Panel B: Significant correlation coefficients ordered by their absolute values. Panel C: Variables that have statistically significant differences in median values between the two planetary system types, as measured by the Kruskal-Wallis test.

exploration missions and assist in the identification of targets for telescope observations, ultimately advancing the search for habitable planets.

Acknowledgments

I extend my gratitude to Dr. Andrew Winter, astrophysics researcher at the Observatoire de la Côte d'Azur, for his valuable guidance.

References

1 M. Reiter and R. Parker, *Dynamics of Young Stellar Clusters as Planet-Forming Environments*.

2 F. Adams, *The Birth Environment of the Solar System*.

3 M. Juan Ovelar, J. Kruijssen, E. Bressert, L. Testi and H. N.Bastian, *Can Habitable Planets Form in Clustered Environments?*

4 J. Jiménez-Torres, B. Pichardo, G. Lake and A. Segura, *Habitability in different Milky Way Stellar Environments: A Stellar Interaction Dynamical Approach*.

© The National High School Journal of Science 2025

5 S. Longmore, M. Chevance and D. Kruijssen, *The impact of stellar clustering on the observed multiplicity and orbital periods of planetary systems*.

6 A. Winter, D. Kruijssen, S. Longmore and M. Chevance, *Stellar clustering shapes the architecture of planetary systems*.

7 D. Kruijssen, S. Longmore and M. Chevance, *Bridging the planet radius valley: Stellar clustering as a key driver for turning sub-Neptunes into super-Earths*.

8 M. Chevance, D. Kruijssen and S. Longmore, *When the peas jump around the pod: How stellar clustering affects the observed correlations between planet properties in multiplanet systems*.

9 N. Archive, *NASA Exoplanet Archive*, `https://exoplanetarchive.ipac.caltech.edu`.

10 C. Dressing and D. Charbonneau, *The occurrence of potentially habitable planets orbiting M dwarfs estimated from the full Kepler dataset and an empirical measurement of the detection sensitivity*.

11 E. Petigura, G. Marcy, J. Winn, L. Weiss, B. Fulton, A. Howard and J. Johnson, *The California-Kepler survey. IV. Metal-rich stars host a greater diversity of planets*.

12 P. . U. Arecibo, *The Habitable Worlds Catalog (HWC*, `http://phl.upr.edu/hwc`.

13 A. Gaia Collaboration, *Gaia Early Data Release 3 (EDR3) - Data release note*, `https://gea.esac.esa.int/archive/`.

14 A. T. N. Facility, *Stellar evolution of star clusters*, `https://www.atnf.csiro.au/outreach/education/senior/astrophysics/stellarevolution_clusters.html`.

15 L. Lindegren, U. Lammers, D. Hobbs, W. O'Mullane, U. Bastian and J. Hernandez, *The astrometric core solution for the Gaia mission—Overview of models, algorithms, and software implementation*.

16 R. Sanders and G. Collaboration, *The Gaia satellite mission*.

17 S. Millholland, S. Wang and G. Laughlin, *Kepler multi-planet systems exhibit unexpected intra-system uniformity in mass and radius*.

18 G. Mulders, I. Pascucci and D. Apai, *An increase in the mass of planetary systems around lower-mass stars*.

19 L. Weiss, H. Isaacson, G. Marcy, A. Howard, E. Petigura, B. Fulton and J. Rowe, *The California-Kepler Survey. VI. Kepler multis and singles have similar planet and stellar properties indicating a common origin*.

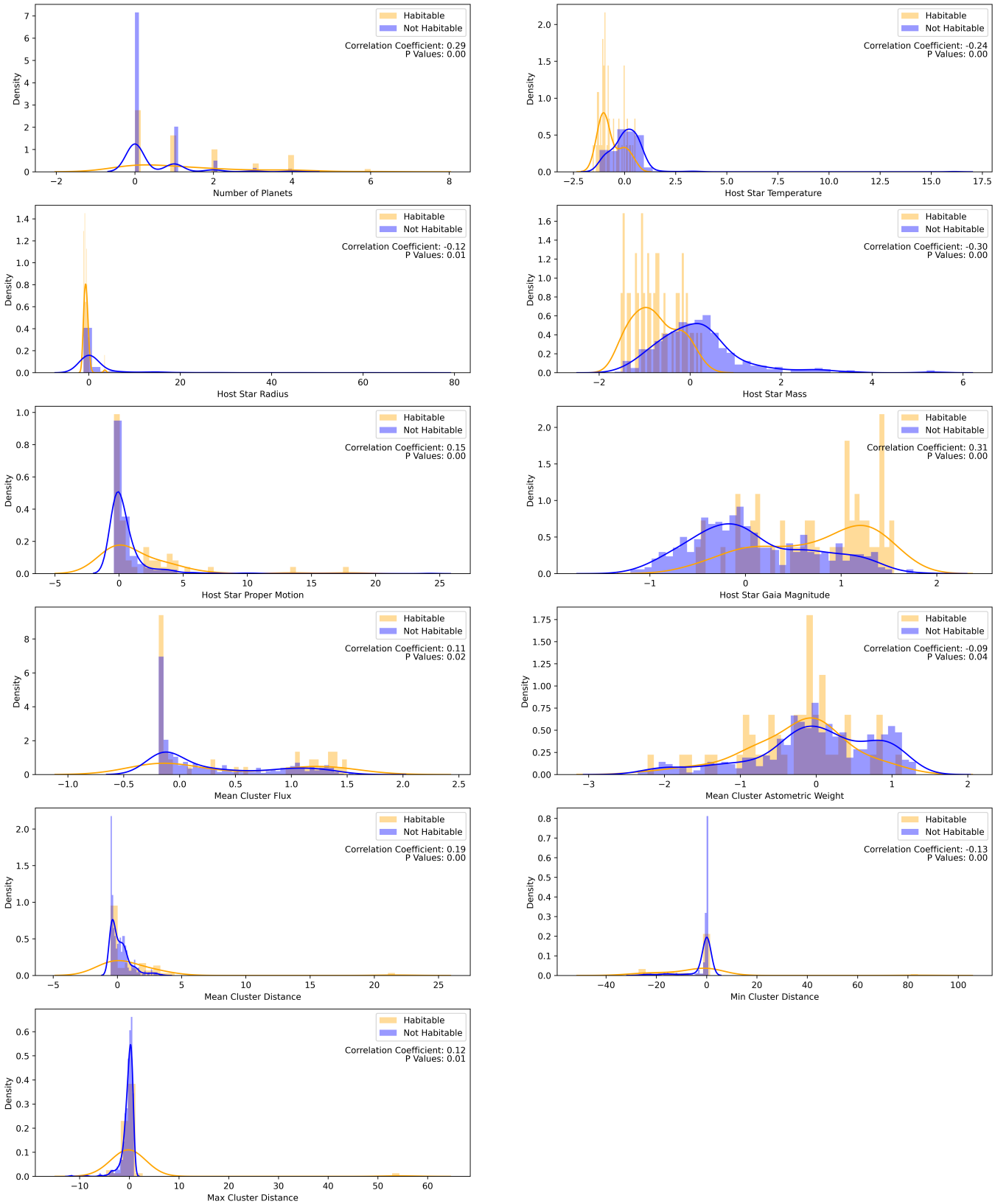

Fig. 3 Distribution plots of independent variables that have a statistically significant correlation with the planetary system type.

© The National High School Journal of Science 2025

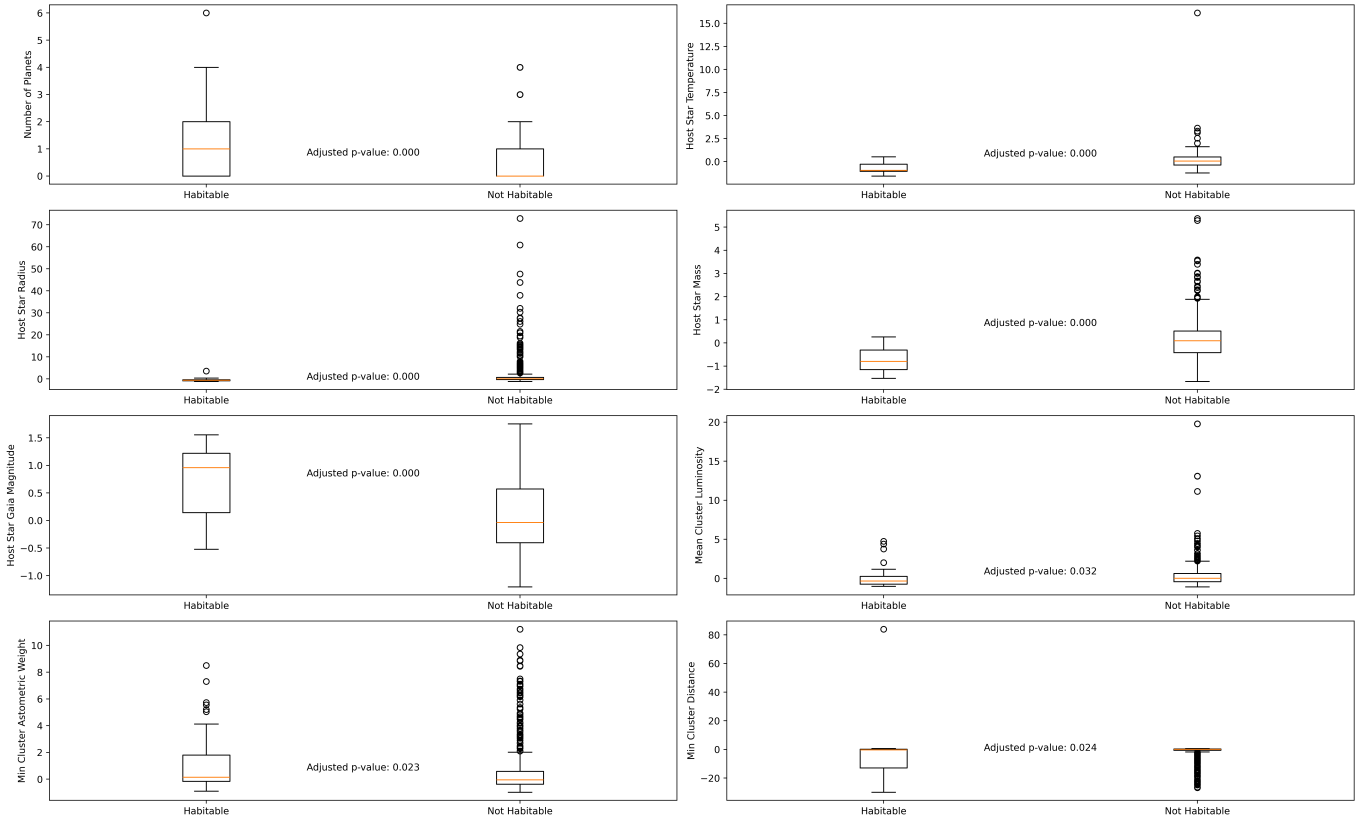

Fig. 4 Box plots with variables that have statistically significant differences in the median values for habitable and non-habitable planetary systems, according to Bonferroni-corrected Kruskal-Wallis test.