Two-Step X-Ray Transit Identification: Bayesian Block Simplification and Sequential Machine Learning Techniques

Shen, Lauren (School: Morgantown High School)

The transit method is versatile; it has been critical in not only the discovery of over four thousand Milky Way exoplanets but also the impressive discovery in galaxy Messier 51 of M51-ULS-1b, the first possible extragalactic planet. However, current methods of detecting transits involve visual identification, take significant time, and can be prone to human error. Combined with the large amount of data available, these observations naturally point to the use of computational techniques to aid the transit method. In this work, a two-step development of a machine-learning model was proposed to automate transit identification. In the first step, a simplified light curve was generated using the Bayesian blocks algorithm. Then, time-series datasets containing sections of event lists (sorted depending on the presence of a transit) were created. A training dataset was created from a source in 47 Tucanae containing many example transits; a validation dataset was created from the transit of M51-ULS-1b as a prime example of an extragalactic planet. In the second step, a random forest model was trained, optimized, and evaluated: it performed with high accuracy and was able to find the exact point in time of the transit for M51-ULS-1b. This method is unique because of its efficiency and applicability: it significantly focuses the approach to transit identification by reducing the time (from days to minutes) and possible errors involved in finding statistically significant transits and also allows astrophysicists to perform meaningful work without the need for an “intuition.”