

**Finding Needles In The Haystack: Outlier Detection In Astronomical Datasets**

**Rafael Martinez-Galarza**  
Harvard/Smithsonian

**July 10, 2019**  
4:00 p.m.

Upcoming large observational time-domain surveys such as the Large Synoptic Survey Telescope (LSST) and the Transiting Exoplanet Survey Satellite (TESS) will produce millions of regularly- and irregularly-sampled astronomical light curves. The large volume of the resulting datasets, however, implies that their processing, classification, and interpretation will require sophisticated algorithms involving statistical learning. One important question is: how do we discover the unexpected when we are presented with a large dataset? How do we find scientifically interesting light curves (or any kind of astronomical data) that are not explained by current models? In this talk I will discuss state-of-the-art anomaly detection methods that use machine learning to find needles in this upcoming haystack of data, and will show the results of applying them to a dataset of Kepler, TESS, and Chandra objects. After a brief introduction to machine learning and its application in time-domain astronomy, I will delve into different methods for outlier detection. I will then show how these methods can be adapted for time-domain and for high energy astronomy, and present the results of applying them to a large dataset of TESS light curves and the Chandra Source Catalog 2.0. I will describe the astrophysical implications of our findings in terms of where the most extreme outliers live in the Hertzprung-Russell diagram, and discuss the potential of the algorithms for discovery in the era of large astronomical datasets.

**Wilson Hall, One West**