Using Machine Learning and Linear Regression Methods to Model Stellar Activity in RV Searches

Zoe de Beurs
Massachusetts Institute of Technology

Exoplanet detection with precise radial velocity (RV) observations is currently limited by spurious RV signals introduced by stellar activity (i.e. faculae, starspots). Here we show that machine learning techniques such as linear regression and neural networks can significantly improve RV measurements by separating the activity signals (primarily starspots/faculae) from real center-of-mass RV shifts. Many EPRV efforts have focused on carefully filtering out activity signals in time using Gaussian process regression. Instead, we separate activity signals from true center-of-mass RV shifts using only changes to the average shape of spectral lines, and no information about when the observations were collected.
We have tested our machine learning methods on both solar observations from the HARPS-N Solar Telescope and extrasolar observations from EXPRES and HARPS-N. For the solar observations, we find that these techniques can successfully predict and remove stellar activity and reduce the RMS by a factor of ~1.7 or about 40% (de Beurs et. al 2022). For extrasolar observations from EXPRES, we found a similar reduction in RMS for the most active stars in our sample. Lastly, we were able to successfully apply our methods to measure the mass of K2-167, a planet which was first found using the transit method in 2015. This promising result inspires us to apply these or similar techniques to solar-type (FGK) stars, help measure masses of planets, and eventually help us detect habitable-zone Earth-mass exoplanets.