# Predicting Fundamental Stellar Parameters from Photometric Light Curves

*Adam Miller*
*Jet Propulsion Laboratory / California Institute of Technology*

## Abstract

We present a new machine learning based framework for the prediction of the fundamental stellar parameters, $T_{\rm eff}$, $\log g$, and [Fe/H], based on the photometric light curves of variable stellar sources. The method was developed following a systematic spectroscopic survey of stellar variability. Variable sources were selected from repeated Sloan Digital Sky Survey (SDSS) observations of Stripe 82, and spectroscopic observations were obtained with Hectospec on the 6.5-m Multi-Mirror Telescope. In sum, spectra were obtained for ∼9,000 stellar variables (including ∼3,000 from the SDSS archive), for which we measured $T_{\rm eff}$, $\log g$, and [Fe/H] using the Segue Stellar Parameters Pipeline (SSPP). Examining the full sample of ∼67,000 variables in Stripe 82, we show that the vast majority of photometric variables are consistent with main-sequence stars, even after restricting the search to high galactic latitudes. From the spectroscopic sample we confirm that most of these stellar variables are G and K dwarfs, though there is a bias in the output of the SSPP that prevents the identification of M type variables. We are unable to identify the dominant source of variability for these stars, but eclipsing systems and/or star spots are the most likely explanation. We develop a machine learning model that can determine $T_{\rm eff}$, $\log g$, and [Fe/H] without obtaining a spectrum. Instead, the random forest regression model uses SDSS color information and light curve features to infer stellar properties. We detail how the feature set is pruned and the model is optimized to produce final predictions of $T_{\rm eff}$, $\log g$, and [Fe/H] with a typical scatter of 165 K, 0.42 dex, and 0.33 dex, respectively. We further show that for the subset of variables with at least 50 observations in the $g$ band the typical scatter reduces to 75 K, 0.19 dex, and 0.16 dex, respectively. We consider these results an important step on the path to the efficient and optimal extraction of information from future time-domain experiments, such as the Large Survey Synoptic Telescope. We argue that this machine learning framework, for which we outline future possible improvements, will enable the construction of the most detailed maps of the Milky Way ever created.