Can Rotation Data Predict Stellar Parameters in *Kepler* Dwarves?

S. Vissapragada[1] and D. Kipping[1]

sv2421@columbia.edu

Received _____; accepted _____

[1]Department of Astronomy, Columbia University, New York, NY 10027, U. S. A.

ABSTRACT

Newly-obtained rotation data from *Kepler* make it possible to investigate the relationship between rotation and stellar parameters using data science techniques. In this work, I explore one such technique: I develop a parametric polynomial model selection scheme using k-folding to determine the existence of trends in the envelope functions of various stellar parameters and rotation period. I find a negative quadratic relation for the upper envelope of mass, luminosity, and effective temperature, a linear relation for the upper envelope of metallicity, and a quadratic relationship for the lower envelope of mass. I discuss the implications of these results; specifically, I note that such trends for mass, luminosity, and effective temperature may fall in line with the theory that dwarf stars spin down as they eject ionized circumstellar material.

Subject headings: stars: rotation, methods: statistical

1. Introduction

The *Kepler* satellite is well-known for offering a way to measure exoplanet transit data. However, *Kepler* also offers insight into the rotation periods of dwarf stars. These rotation periods are not random; dwarf stars tend to spin down over time due to magnetized stellar wind (Kawaler 1988; Bouvier et al. 1997). Thus, we expect there to be a relation between the period of a rotating star and its age. This gyrochronological relationship has, in turn, generated interest in relationships between rotation parameters and other stellar parameters such as effective temperature $T_{\rm eff}$, metallicity [Fe/H], and surface gravity $\log g$.

McQuillan et al. (2014) have compiled rotation periods for 34,000 dwarf stars observed by *Kepler* using an automated autocorrelation function, which measures self-similarity in light-curves across many time lags to detect rotation periods. My goal was to take these rotation data and investigate trends with other stellar parameters. However, the data for those other stellar parameters given by McQuillan et al. (2014) were slightly outdated. Thankfully, Mathur et al. (2016) have provided the most recently revised catalogue of these stellar parameters that, among other things, provide uncertainties from posterior samples.

In this work, I describe the beginnings of a potential data science technique to determine whether stellar parameters are correlated with rotation data. In Section 2, I describe how I determined which trends to investigate further. In Section 3, I describe the parametric models tested and model selection techniques used to determine the existence of a trend in one sample relationship. In Section 4, I provide trend results for all tested stellar parameters. In Section 5, I comment on the existence and lack of various trends in light of the aforementioned relationship between rotation period and age, the shortcomings of my technique, and various selection biases.

2. Diagnosing Trends

First, I took the rotational data from McQuillan et al. (2014) and loaded them into a `numpy` array. Then, I created a dictionary of the updated stellar property data from Mathur et al. (2016) hashed by KIC number, and created a composite `numpy` array of McQuillan et al. (2014) data with Mathur et al. (2016) parameters. In doing this, I also included luminosity L as a calculated parameter. Luminosity is calculated as follows:

$$L = 4\pi\sigma R^2 T_{\rm eff}^4, \tag{1}$$

where R is the stellar radius, $T_{\rm eff}$ is the effective temperature, and σ is the Stefan-Boltzmann constant. Additionally, I instituted an upper limit on mass of $M = M_\odot$ on all Kepler objects to be investigated further. This mass cutoff ensured that there were no giants contaminating the sample.

I then made a diagnostic plot matrix using the `pandas` package, which is shown in Figure (1). This plot allowed me to investigate the correlation between every combination of two parameters by eye to determine which sets of parameters warranted follow-up investigation. The diagonal of this plot matrix contains normalized density distributions of each parameter to ensure that the majority of the data were being plotted.

There were many sets of parameters with clear trends such as mass vs. radius and surface gravity vs. density; however, these trends are more related to stellar structure rather than rotation period and are thus outside the scope of the work. As an aside, these parameters are largely dependent on each other - as such it would not make much sense to determine the relationship between parameters where it already exists (for instance, the mass-radius relationship).

To determine which parameters to investigate, I looked at the two columns with rotational parameters as independent variables, and checked whether there were variables

whose allowed range seemed to depend on a rotational parameter. Indeed, there were: [Fe/H], L, $T_{\rm eff}$, $R_{\rm per}$, M, and ρ all seemed to have some envelope dependence on $P_{\rm rot}$. While there seemed to be a pretty tight range on $R_{\rm per}$, it seemed to not be a very good predictor of the allowed range on any of the variables. Thus, I went ahead with fitting the envelope for the above parameters with $P_{\rm rot}$, performing model selection to prove optimality for each fit.

3. $P_{\rm rot}$ vs. $T_{\rm eff}$ Envelope Fitting and Model Selection

The upper envelope of the function was established using a moving 95th percentile and the lower envelope using a moving 5th percentile. The window size chosen for this moving percentile operation was 2000 data points, meaning that 1000 data points on either end of the $P_{\rm rot}$ axis were cut off from the envelope model. The initial upper and lower envelope bounds are shown for this dataset in Figure (2).

Then, four polynomial models for each part of the envelope were established. Each polynomial model was coded in `numpy` with the following prescription:

$$M_n = \sum_{i=0}^{n} C_i x^i \tag{2}$$

The null model was M_0, a 0th-order polynomial representing the 95th or 5th percentile of the smoothed data. If a model could outperform this null model, then I could say confidently that a trend existed. In addition, I had 3 other polynomial models: M_1, M_2, and M_3. By performing cross-validation between all four models, I was able to simultaneously tell whether a trend existed and, if so, approximately what order polynomial could explain it.

I performed cross-validation using a modified k-folds technique. I cut the data set along the $P_{\rm rot}$ axis into k equally-sized validation sets. On each fold, I treated the rest of

the data as a training set, and trained all four of my models using `numpy.polyfit`. Then, for each model M_n, I calculated the χ^2statistic as follows:

$$\chi^2 = \sum_{(x,y)\in V} \Big(\frac{y - M_n(x)}{\sigma_y}\Big)^2, \tag{3}$$

where V represents the points in the validation dataset. However, since the data was put through a moving percentile filter, it was impossible to use the measured errors from the Mathur et al. (2016) posteriors. To fix this, I set $\sigma_y = y$ for each data point. This process was repeated and averaged over all k-folds to determine the $\overline{\chi^2}$ for each model. The model with the lowest $\overline{\chi^2}$ had therefore been selected via k-folds cross-validation.

I chose $k = 100$ and performed the above cross-validation technique to select a model. The table of $\overline{\chi^2}$ for each model and envelope is included in Table (1). The model selection for the upper envelope follows a very familiar trend - the $\chi^2$is relatively high for the null model, gets lower with the first-order model, gets even lower with the second-order model, then begins to fit out noise in the third-order model. Thus, M_2 was selected for the upper envelope. The lower envelope had a much more spurious relation, as the none of the higher-order models were much better than each other and they were all only a bit better than the null model. Thus, we could only constrain the upper 95th percentile using a quadratic model. After re-fitting M_2 to the upper 95th percentile, I found that the relation:

$$T_{\rm eff} = (6172\ {\rm K}) + (7.707\ {\rm K/day})P_{\rm rot} - (0.7655\ {\rm K/day}^2)P_{\rm rot}^2 \tag{4}$$

constrains the upper 95th percentile of $T_{\rm eff}$ given $P_{\rm rot}$ for 3.019 days $\leq P_{\rm rot} \leq$ 38.235 days.

4. Envelope Fits for the Remaining Parameters

The procedure in Section 3 was then repeated for [Fe/H], L, $R_{\rm per}$, M, and ρ vs. $P_{\rm rot}$. The result for each procedure is given in the subsections below.

4.1. [Fe/H] vs. $P_{\rm rot}$

The envelope and median from the moving percentile filters are given in Figure (3). The upper envelope was found to be linear but it was difficult to distinguish this model from the cubic simply based on the mean χ^2. Keeping true to Occam's Razor and distinguishing the models by eye, I chose the linear model, which gave the relation:

$$[Fe/H] = 0.1026 + (.006461 \text{ day}^{-1})P_{\rm rot} \tag{5}$$

to constrain the upper 95th percentile of [Fe/H] given $P_{\rm rot}$ for 3.019 days $\leq P_{\rm rot} \leq$ 38.235 days. The lower envelope did not strongly prefer one model over another and was thus discarded.

4.2. L vs. $P_{\rm rot}$

The envelope and median from the moving percentile filters are given in Figure (4). The upper envelope was found clearly to be quadratic, with the relation:

$$T_{\rm eff} = (0.2573\ L_\odot) + (0.01591\ \text{L}_\odot/\text{day})P_{\rm rot} - (0.001084\ \text{L}_\odot/\text{day}^2)P_{\rm rot}^2 \tag{6}$$

to constrain the upper 95th percentile of L given $P_{\rm rot}$ for 3.019 days $\leq P_{\rm rot} \leq$ 38.235 days. The lower envelope did not strongly prefer one model over another and was thus discarded.

4.3. $R_{\rm per}$ vs. $P_{\rm rot}$

The envelope and median from the moving percentile filters are given in Figure (5). For this fit, I did something slightly different - I fit the two parameters in log-space rather than lin-space. I did this because both parameters were visualized in log-space on my diagnostic plot, so I figured I should stay true to that plot in my fit evaluation. While the χ^2values

for both the upper and lower envelopes clearly rejected the null model, they did not allow me to select between M_1, M_2 or M_3 for both the upper and lower envelopes. Thus, neither envelope could be constrained by this polynomial method.

4.4. M vs. $P_{\rm rot}$

The envelope and median from the moving percentile filters are given in Figure (6). Both the quadratic and cubic fits seemed to work about the same for both envelopes in this case, and I discarded the cubics by eye and chose the quadratics. The upper envelope was given by the relation:

$$M = (0.9761\ M_\odot) + (0.001773\ \mathrm{M}_\odot/\mathrm{day})P_{\rm rot} - (0.0001176\ \mathrm{M}_\odot/\mathrm{day}^2)P_{\rm rot}^2 \qquad (7)$$

to constrain the upper 95th percentile of M given $P_{\rm rot}$ for 3.019 days $\leq P_{\rm rot} \leq$ 38.235 days.

The lower envelope was given by the relation:

$$M = (0.5471\ M_\odot) + (0.002415\ \mathrm{M}_\odot/\mathrm{day})P_{\rm rot} - (0.0001837\ \mathrm{M}_\odot/\mathrm{day}^2)P_{\rm rot}^2 \qquad (8)$$

to constrain the upper 95th percentile of M given $P_{\rm rot}$ for 3.019 days $\leq P_{\rm rot} \leq$ 38.235 days.

4.5. ρ vs. $P_{\rm rot}$

The envelope and median from the moving percentile filters are given in Figure (7). While the χ^2values for both the upper and lower envelopes clearly rejected the null model, they did not allow me to select between M_1, M_2 or M_3 for both the upper and lower envelopes. Thus, neither envelope could be constrained by this polynomial method.

5. Summary and Discussion

Of all the upper limits I did end up selecting, the one that seemed most interesting to me was luminosity. As stated in the Section 1, dwarf stars spin down when circumstellar material is ejected along magnetic field lines. It seems plausible that the ejection of this circumstellar material (which is almost certainly ionized, and as such may be radiating) could then decrease both the mass and luminosity of the star. Though the two situations are hardly comparable, one could make a rough comparison to stellar winds driving mass loss in high-mass stars, as described in Smith (2014). While this proposal is a bit of a shot in the dark, it would explain a few things found in this project. The weakly negative quadratic trend in the upper envelopes of luminosity (and thus effective temperature, by the Stefan-Boltzmann law) and mass would make sense, because stars that are rotating the slowest may have undergone mass loss during spin down.

I would also like to make a few remarks on how the method in this project could be improved in the future. First, the mass cutoff seemed to do a good job of decontaminating data in all but the ρ vs. P_{rot}. In the future, it may be wise to incorporate some kind of density cutoff, since that impacted my ability to fit the trend. Second, it was clear throughout this whole project that the window size for the moving median strongly influenced the outcome of the fits. While I made the conscious decision to stick to the window size that worked for the first fit so as to keep my method repeatable, I realize that someone using different window sizes may get different results. Third, I also made the conscious decision to perform my fits across the entire range of P_{rot} that was included in the moving median calculation to keep my method consistent; however, I also realize that relaxing this requirement and fitting for certain ranges may have helped select models in cases where my k-folds technique otherwise failed. Fifth, my assumption that $\sigma_y = y$ in calculation of $\chi^2$was incorrect and may have influenced my results. Sixth, my polynomial

method was not very comprehensive; I certainly could have used polynomials up to a higher degree, or included other functions, but my reasoning was that polynomials could at least discriminate between the null model and a trend relatively well (and there didn't seem to be any odd correlations, like periodicity or something).

Additionally, it's clear that the lower envelopes I got for most of my fits weren't great, and sometimes didn't even perform better than the null model. This could be attributed to the failure of the moving 5th percentile filter to produce a clean curve., which is related to my second point above. However, in some cases, it could also be due to selection bias at the lower ends of the axis. For instance, for luminosity, we see that there are more luminous objects with small rotation periods than dim objects with small rotation periods, and indeed the dim region is sampled quite poorly for rotation periods under 10 days - potentially because these objects would be more difficult to detect.

Finally, I would like to express my gratitude to Professor Kipping and Emily for an excellent class, in which I learned **a lot** of things that I'm sure will help me no matter what career I go into. Thank you!

REFERENCES

Bouvier, J., Forestini, M., & Allain, S. 1997, A&A, 326, 1023

Kawaler, S. D. 1988, ApJ, 333, 236

Mathur, S., Huber, D., Batalha, N. M., et al. 2016, ApJS, submitted

McQuillan, A., Mazeh, T., & Aigrain, S. 2014, ApJS, 211, 24

Smith, N. 2014, ARAA, 52, 487

This manuscript was prepared with the AAS LaTeX macros v5.2.

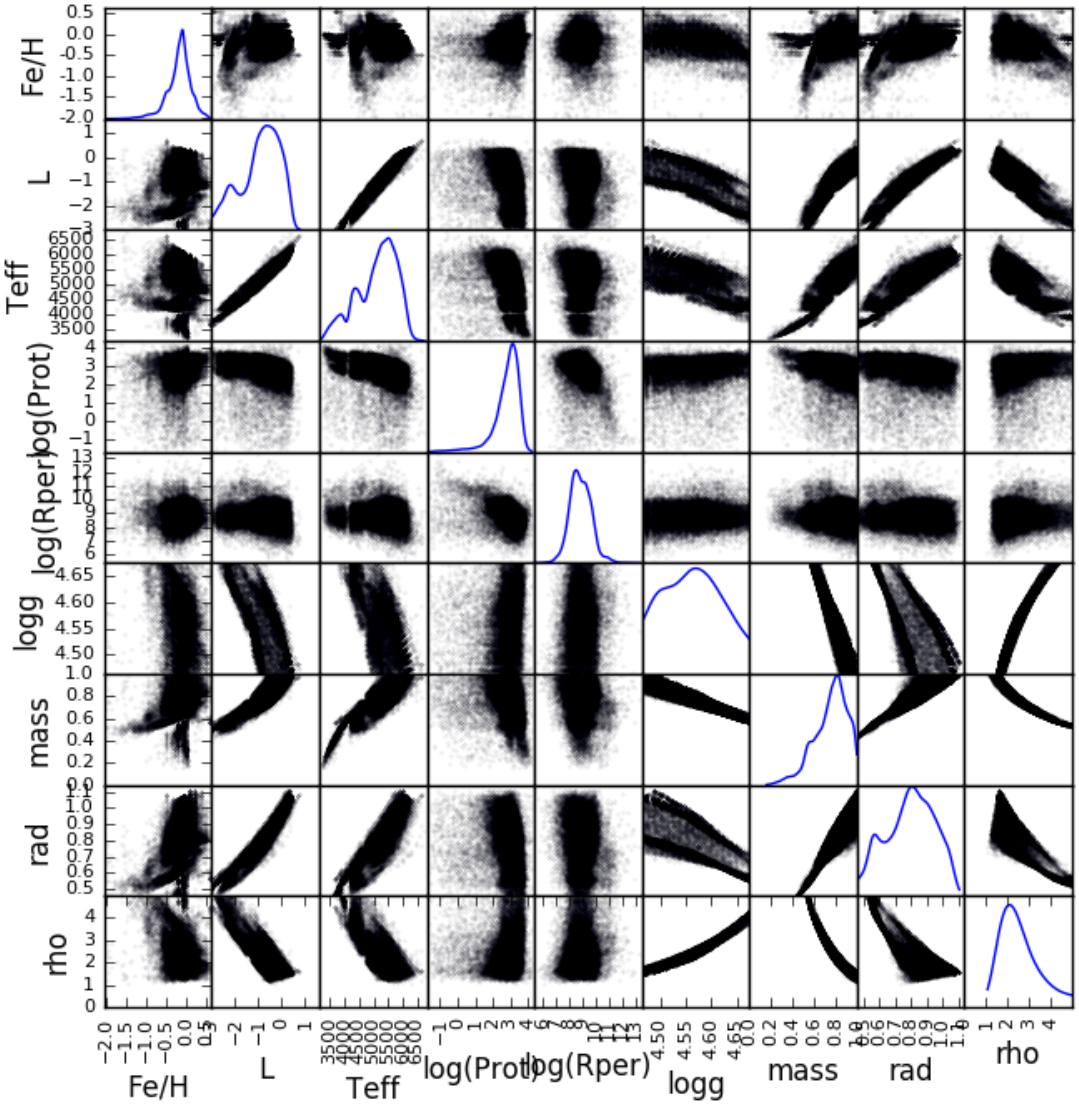

Fig. 1.— Diagnostic plot matrix for visualizing relationship between multiple parameters in this dataset.

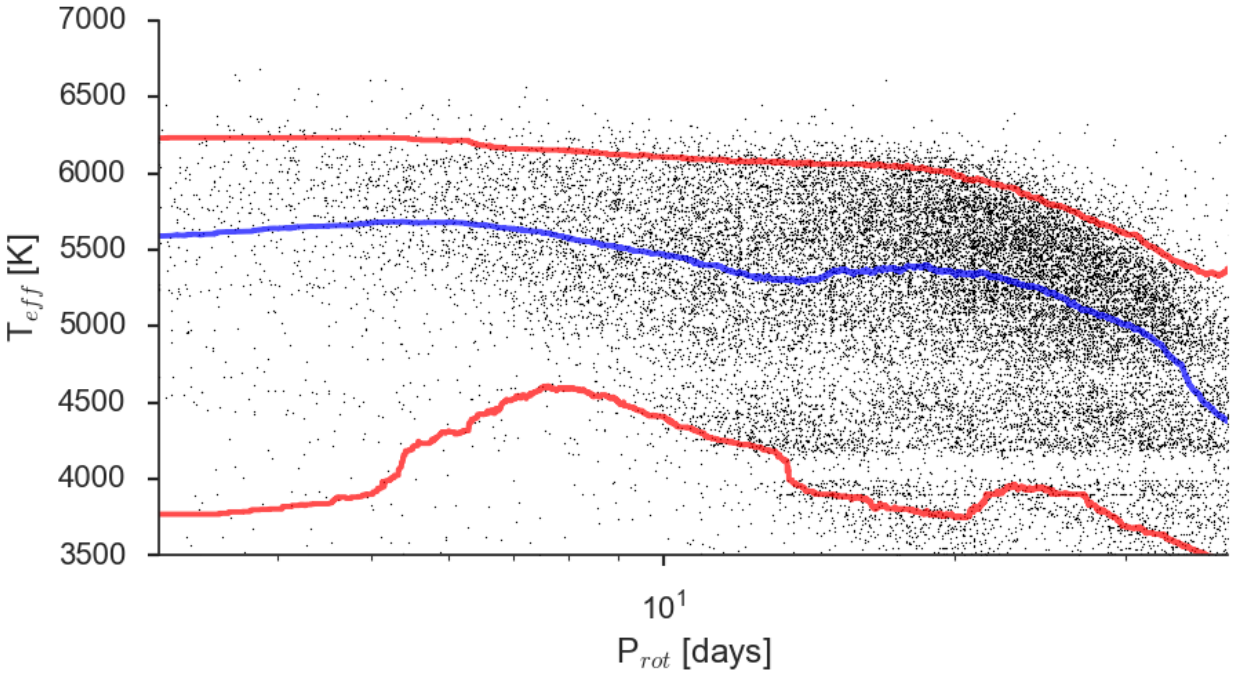

Fig. 2.— Effective temperature as a function of rotation period. The upper (95th) and lower (5th) percentiles are given in red, while the moving median is given in yellow.

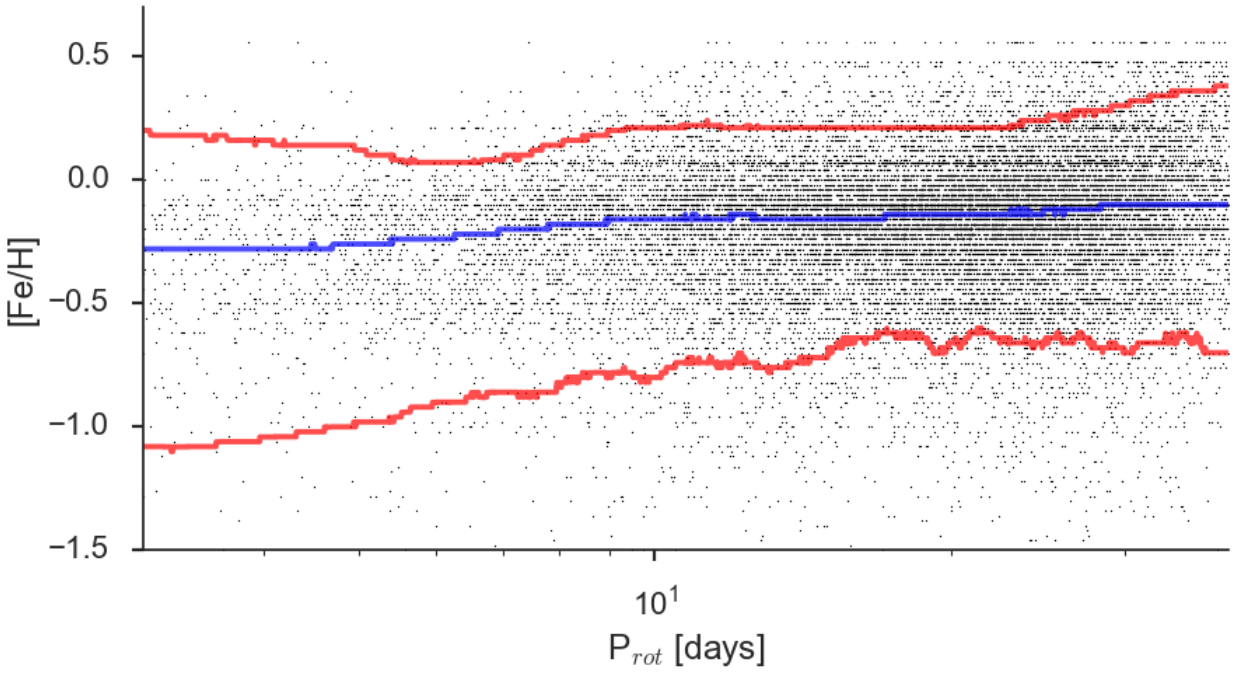

Fig. 3.— Metallicity as a function of rotation period. The upper (95th) and lower (5th) percentiles are given in red, while the moving median is given in yellow.

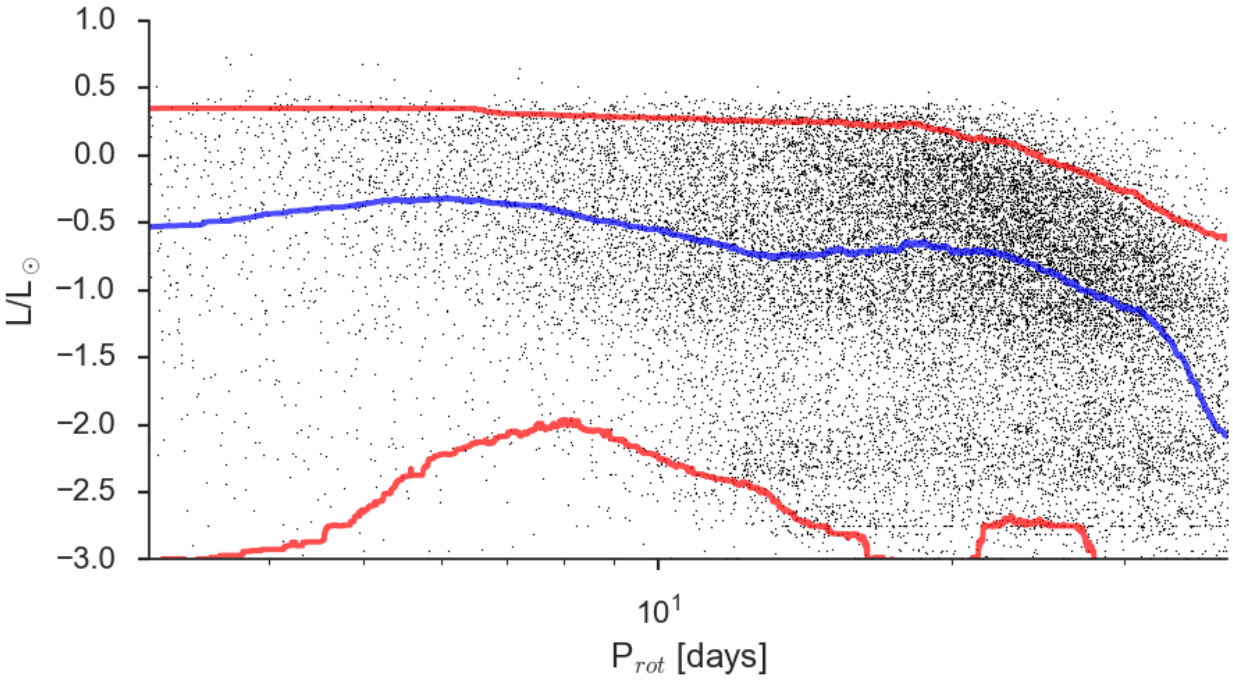

Fig. 4.— Luminosity as a function of rotation period. The upper (95th) and lower (5th) percentiles are given in red, while the moving median is given in yellow.

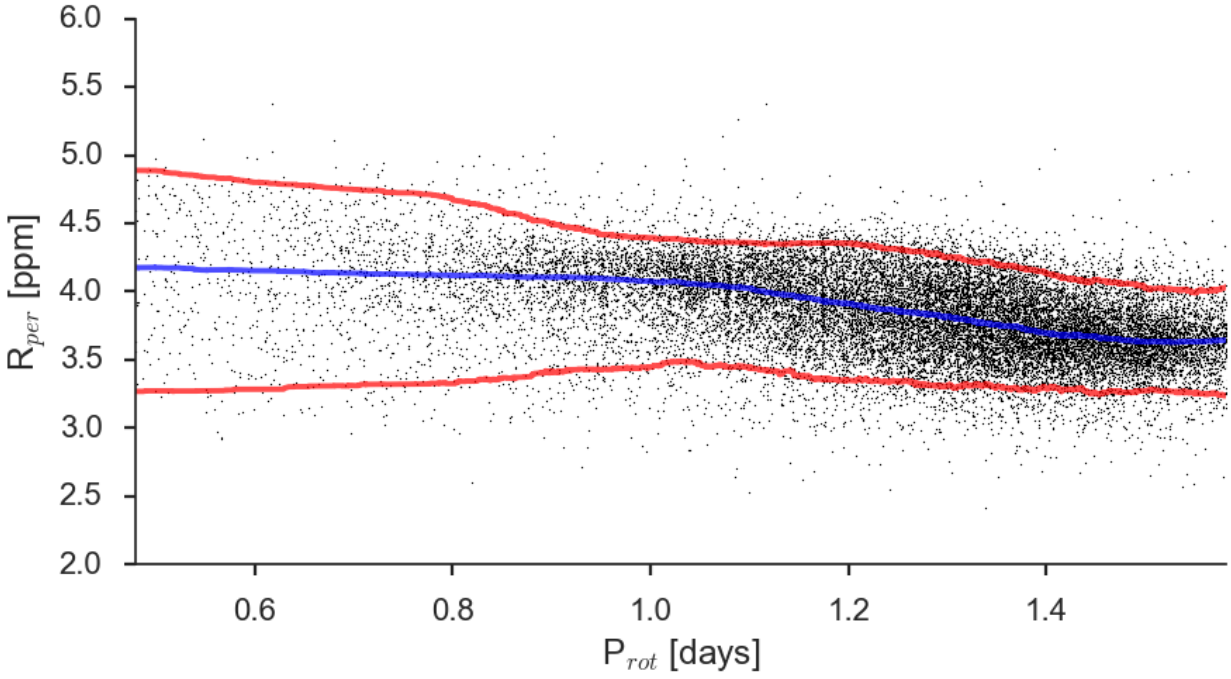

Fig. 5.— Log rotation amplitude as a function of log rotation period. The upper (95th) and lower (5th) percentiles are given in red, while the moving median is given in yellow.

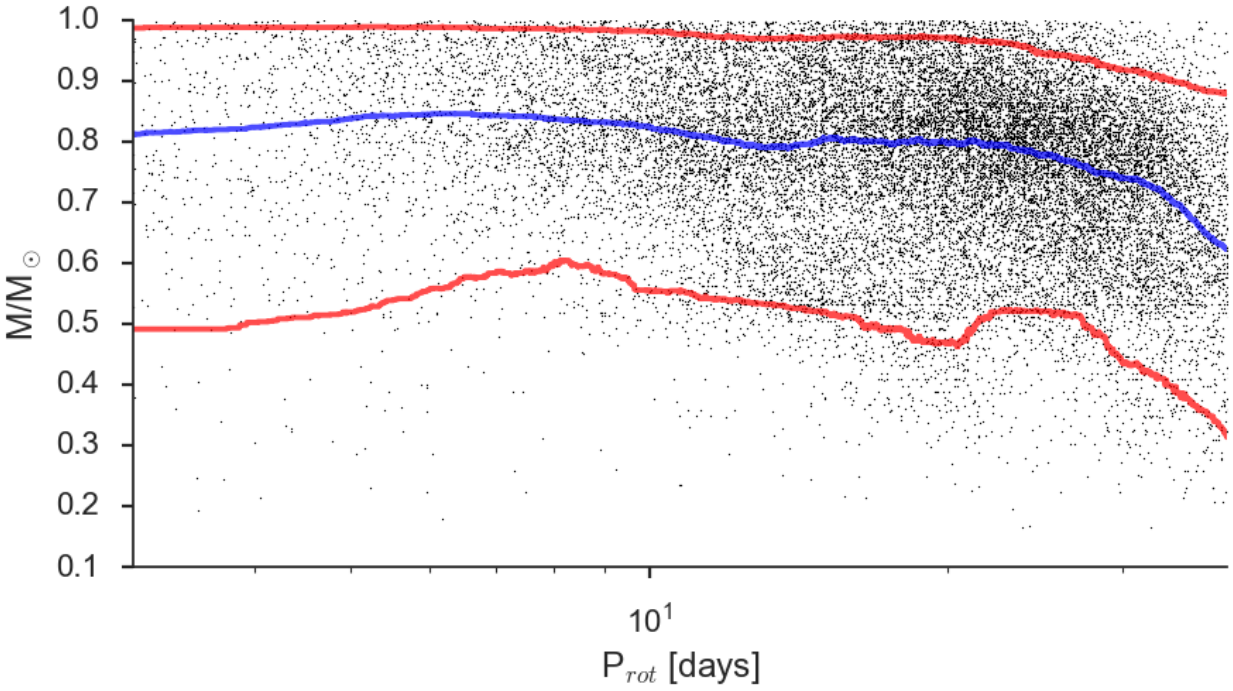

Fig. 6.— Mass as a function of rotation period. The upper (95th) and lower (5th) percentiles are given in red, while the moving median is given in yellow.

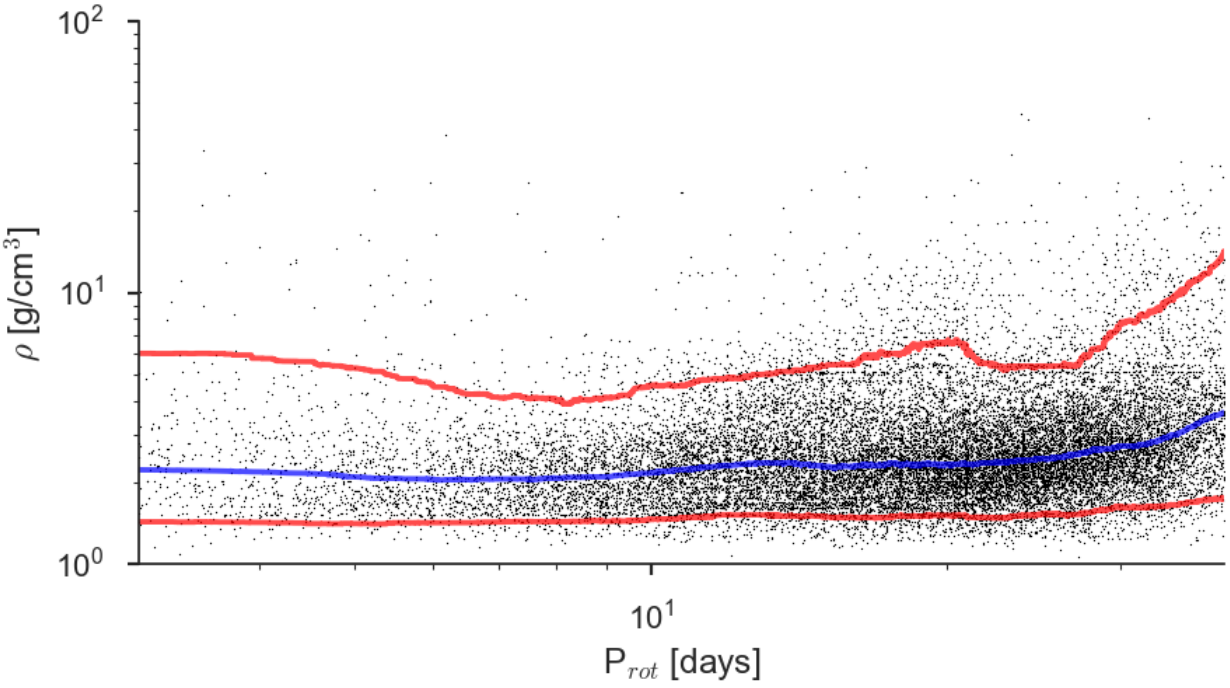

Fig. 7.— Density as a function of rotation period. The upper (95th) and lower (5th) percentiles are given in red, while the moving median is given in yellow.

Table 1. Mean χ^2 for upper and lower envelopes of $P_{\rm rot}$ vs. $T_{\rm eff}$.

Model	$\overline{\chi^2}_{\rm upper}$	$\overline{\chi^2}_{\rm lower}$
M_0	3.746×10^{-1}	9.358×10^{-1}
M_1	2.935×10^{-2}	4.165×10^{-1}
M_2	5.743×10^{-3}	4.421×10^{-1}
M_3	6.821×10^{-3}	4.772×10^{-1}