A Bayesian approach shows no correlation between transit-depth and stellar metallicity for confirmed and candidates Kepler gas giants planets

C Nehmé¹,² and P Sarkis³
¹Department of Physics & Astronomy, Notre Dame University- Louaize, PO Box 72, Zouk Mikael, Lebanon
²CEA/DRF/Irfu/SAp, F-91191 Gif-sur-Yvette, France
³Universität Bern Space Research & Planetary Sciences Division Bern, Switzerland
E-mail: cnehme@ndu.edu.lb

Abstract. Previous study to investigate the correlation between the transit depth and the stellar metallicity of Kepler’s (Q1-Q12) gas giant planets (radii of 5-20R_⊕) has led to a weakly significant negative correlation. We use the cumulative catalog of planets detected by the NASA Kepler mission Q1-Q17 catalog, as of April 2015, to perform a solid statistical analysis of this correlation. In the present work, we revise this correlation, within a Bayesian framework, for two large samples: sample A confirmed planets and sample B (confirmed + candidates). We expand a hierarchical method to account for false positives in the studied samples. Our statistical analysis reveals no correlation between the transit depth and the stellar metallicity. This has implications for planet formation theory and interior structure of giant planets.

1. Introduction
NASA’s Kepler Mission revolutionized the field of extra solar planets. We now can place statistical constraints on the (observed) properties of planets as well as on theories of planet formation. Clues on the nature of giant planet formation might be revealed from two correlations with stellar metallicity of main sequence stars hosting these planets. The first one is the correlation between the number of giant planets and stellar metallicity revealed by radial velocity surveys[1] and by transit surveys [2]. The second correlation is a positive trend between the mass of heavy-elements in giant planets and the stellar metallicity [3]. As Kepler is searching for transiting exoplanets around their host stars. It is primordial to study the correlation between the transit depth and the stellar metallicity of Kepler’s gas giant candidates. Earlier studies conducted on this correlation between has led to different results. For instance, Dodson-Robinson reported a negative correlation with a weak statistical significance. The author studied the transit depth of 213 giant planets from (Q1-Q12) catalog with estimated radii of 5-20R_⊕ with the values of [Fe/H] taken from the Kepler Input Catalog (KIC). Dodson-Robinson [4] interpreted the negative correlation as evidence that metal rich planets of a given mass are denser than their metal poor counterparts leading to small radii [6]. Here, we will use the latest available catalog Q1-Q17 [7] to study correlation between the transit depth and stellar metallicity.
In the previous study, the author have assumed that the measurement uncertainties are negligible. This assumption is not justified because these uncertainties, especially on radius measurements that depend on the stellar parameters, can be large compared to the scales of interest. Also, a very important component of any study of exoplanet populations is the treatment of detection efficiency. In a transit survey, small planets with long periods are much harder to detect than large planets orbiting close to their stars. Moreover, large-scale surveys, such as Kepler, are subject to selection effects and biases. Kepler will likely detect a bigger number of planets around bright stars than around faint stars. Thus, a sample of stars with transiting planets may not accurately represent the true intrinsic distribution of the discovered planets. Gaidos E and Mann A [8] reported the importance of including these effects since they can lead to biases in the properties of transiting planets and their host stars. We develop a framework to account for the uncertainties present on the observed parameters and we use Bayesian statistics to study the correlation. Our method is based on hierarchical Bayesian modeling introduced by [9]. We update the method to account for the false positive rate of Kepler giant planets.

2. Selection criteria and samples
We use the cumulative catalog of planets detected by the NASA Kepler mission which, as of April 2015, consisted of Q1-Q17 catalog [7]. Following [4] and [10] we define gas giant planets as planets having a radius between 5-20R⊙. The stellar parameters were taken from the Kepler stellar Q1-Q16 database [11]. We ended up by having a sample of 390 planets of which 84 are confirmed and 306 are candidates giant planets.

With the goal of performing a robust statistical method, we prepared two different samples. The first sample (sample A) consists of all the 84 confirmed giant planets. The second sample (sample B) contains the 390 confirmed + candidates. We then restrict sample B to stars with a Kepler magnitude Kp < 16 mag to avoid faint stars and we choose FGK Main sequence stars for Kepler’s high efficiency of detection giant transiting planets around those stars. In addition to the cuts made on the stellar sample we also require P < 90 days because of the low efficiency at detecting long period [12] and SNR > 18.8. After performing all the cuts (table 1), sample B will be left with 105 confirmed + candidates.

| Table 1. Summary of the cuts performed to obtain sample B |
|------------------------------------------------------------------------------------------|
| Parameter                        | Value                  |
|----------------------------------|------------------------|
| Stellar effective temperature, T_{eff} | 4000 - 7000 K          |
| Stellar gravity, log g (cm/s^{-2})    | 4.0 - 5.0              |
| Stellar Radius, R_⋆                 | 0.7 - 1.4 R⊙           |
| Planetary Radius, R_p              | 5 - 20 R⊙             |
| Orbital Period, P                  | < 90 days              |
| Detection Efficiency, SNR          | > 18.8                 |
| Kepler magnitude, K_p              | < 16 mag               |

3. Method
Hierarchical Bayesian Modeling (hereafter HBM) allows for intrinsic scatter and heteroscedastic measurement errors i.e the uncertainties for each data points are different. We followed an approach similar to that proposed by Kelly B C [9]. We constructed the likelihood function in a simple way in order to relate the parameters of interest to the observed data, taking into account
Figure 1. A graphical presentation of our HBM is given in this illustration. The gray nodes are the observed parameters. The true missing parameters are in the white nodes. The blue (upper left) nodes are the nuisance parameters and the red nodes (upper right) are the parameters of interest. \( \text{FeH}_i \) = stellar parameter of the \( i^{th} \) planet, \( \sigma_{\text{FeH},i} \) = uncertainty on the stellar metallicity of the \( i^{th} \) planet, \( \delta_i \) = transit depth of the \( i^{th} \) planet, \( \sigma_{\delta,i} \) = uncertainties on the transit depth of the \( i^{th} \) planet, \( \text{FeH}_{ti} \) = true stellar metallicity of the \( i^{th} \) planet, \( \delta_{ti} \) = true transit depth of the \( i^{th} \) planet, \( \mu \) and \( \tau \) = nuisance parameters, \( \alpha, \beta \) and \( \sigma \) = parameters of the linear model.

the measurements uncertainties. We extended the method in [9] to account for false positive by Fressin F [13]. This updated statistical method is used to study the correlation between the transit depth (\( \delta \)) and the metallicity (FeH) of host stars. A graphical illustration of our HBM is given in Figure 1. Markov chain Monte Carlo (hereafter MCMC) was performed using the Python package PySTAN, a package for Bayesian inference. We ran models with 4 Markov Chains, with 5000 iterations for the first sample and 10 000 iterations for the second one. The first 50 per cent of each chain was discarded as ”burn-in”. This work is the first to perform a full HBM to study correlations in general and the correlation between the transit depth and the metallicity of Kepler’s giant planets, in particular. Most importantly, the quantification of the intrinsic dispersion which has not been characterized before, is now defined.

4. Results
The equation of the ”best-fit” linear models are:

\[ \delta = (0.07 \pm 0.014) + (0.02 \pm 0.14) \text{FeH} \]

with an intrinsic scatter of \( \sigma = 0.03 \pm 0.005 \) for the confirmed planets sample (sample A). For sample B the best fit and the intrinsic scatter are \( \delta = (0.13 \pm 0.06) + (-0.18 \pm 0.06) \text{FeH} \) and \( \sigma = 0.05 \pm 0.02 \), respectively. The transit depth is plotted against the metallicity of the host star along with their uncertainties, in the right panel of the Figure 2. The ”best-fit” models (dotted lines) are shown along with 100 random samples from the MCMC chain. This shows clearly the existence of a large intrinsic scatter. A robust way to demonstrate the absence of correlation between the transit-depth and the stellar metallicity.

5. Discussion
We presented within a Bayesian framework, a study of the correlation between the transit depth of Kepler’s giant planets and the metallicity of the host star. Data from Kepler(Q1-Q17) [7] allowed us to characterize the intrinsic scatter in the relation with a robust statistical analysis. We expand the hierarchical model presented in [9], to account for the false positive rates. We also considered in the model the relevant selection effects. The use of Markov Chain Monte Carlo (MCMC) to fit models to observations is becoming a standard practice in astronomy. We
Figure 2. Transit depth ($\delta$) of Kepler’s giant planets vs. the metallicity of the host star ([Fe/H]). The dashed lines represent the best fit line and the light lines are samples from the MCMC chain. The dashed lines represent the best fit line and the light lines are samples from the MCMC chain.

performed MCMC using the package PyStan to estimate the parameters of the hierarchical linear model. We established that there is no correlation between the transit depth and the stellar metallicity of Kepler’s gas giant planets. Our result indicate that there is a relatively large intrinsic scatter in the relation. Hence, the previous results could probably be an artifact which shows the importance of accounting for uncertainties and for possible false positives. It also proves the importance of accounting for selection effects and biases within the transit surveys, such as Kepler [4] interpreted the negative correlation as evidence that metal-rich planets of a given mass are denser than their metal-poor counterparts, leading to smaller radii [6]. On the contrary, with our robust statistical method, we have proven the independence of transit depth and the stellar metallicity. It certainly warrants further investigations to check what planetary formation model can explain the outcome.

References
[1] Fischer D A and Valenti J 2005 ApJ 622 1102
[2] Wang J and Fischer D A 2013 arXiv:astro-ph/1310.7830
[3] Miller N and Fortney J J 2011 ApJ 736 L29
[4] Dodson-Robinson S E 2012 ApJ 752 72
[5] Borucki W J, Koch D G, Basri G, Bathalha N et al 2011 ApJ 736 19
[6] Fortney J J and Nettelmann N 2010 Space Sci. Rev. 152 423
[7] Mullaly F, Coughlin J L, Thompson S E, Rowe J et al 2015 ApJS 217 31
[8] Gaidos E and Mann A 2013 ApJ 762 81
[9] Kelly B C 2007 ApJ 665 1489
[10] Schaufman K C and Laughlin G 2011 ApJ 738 177
[11] Huber D et al 2014 ApJS 211 2
[12] Petigura E A, Howard A W and Marcy G W 2013 PNAS 110 19273
[13] Fressin F et al 2013 ApJ 766 81