# A Bayesian framework to improve exoplanets detection via the radial velocity technique

**Principal Investigator: Graça Rocha (326)**
**Andrew Howard (Caltech), Benjamin Fulton (IPAC, Caltech), Arpita Roy (Caltech)**

**Program: Topic Area**

## Project Objective:

The objective of this research work was to initialize the development of a more sophisticated method of analyzing Radial Velocity (RV) data, using a Bayesian approach to simultaneously and accurately characterize the statistical properties of the noise in the data, including systematic and astrophysical noise, compared to true Radial Velocity (RV) signals.

The radial velocity (RV) method is one of the most successful techniques for detecting exoplanets. As a planet orbits a star, the planet's gravitational force forces the star to orbit around the center of mass of the system. The RV technique looks for this velocity of a host star induced by the gravitational effect of an orbiting planet, along our line of sight, which is called *radial velocity* of the star. This is done spectroscopically, since starlight appears to be alternately shifted to longer (redder) and shorter (bluer) wavelengths as it moves away from or towards the observer, respectively, due to the Doppler effect caused by this motion. The Doppler effect due to a planet is very small (as a benchmark, Earth induces an RV shift of 9 cm/s on the Sun), and its detection is achieved by careful analysis of the spectra over a wide range of wavelengths covering thousands of spectral lines. To search for planets, candidate stars are chosen, and for each star a time series of RV spanning days to decades is collected.

With improvements in instrumentation, RV surveys have become sensitive to low-mass planets (RV≈1 m/s). High-resolution spectra are rich in Doppler information content, but in practice measurements are often limited by systematics that can add quasi-coherent noise that mimic low-mass planet signals. Examples systematics are: (1) variable point spread function of unstabilized instruments; (2) intrinsic stellar variability (e.g., star spots, convective motion, pulsations); (3) environmental conditions (temperature, humidity, telescope focus). These contaminants challenge our detection of low-mass planets (e.g., Fulton et al. 2015 (HD7924), Robertson et al. 2015 (Gl581).

This study investigated possible ways of implementing a novel Bayesian approach for the analysis of RV data. We assembled representative parameter correlations to help construct the covariance of the Multivariate Gaussian Likelihood; investigated several systematics and their characterization along with potential parameter degeneracies to help devise an adequate data model. Our results helped us to decide for specific strategies when implementing our new approach.
When focusing on accounting for the Point Spread function, PSF, inacuraccies to better constrain RV measurements, we adopted the Markov Chain Beam randomization , **MCBR,** approach developed by the PI of this proposal for CMB data analysis (Rocha et al…). The prototype module is being implemented in **RadVel.**

The (ongoing) full integration of **MCBR** in **RadVel** is guided by the findings of this proposal.

## FY18/19 Results:

## Procedure:

Our approach is devised within the Bayesian framework, this encompasses: (1) Defining the data model: $d=s+n_c+n_i$, where s=signal (here the true RV signal due to the presence of a planet), $n_c$=noise due to contaminant sources and systematics (e.g., barycentric motion, intrinsic stellar activity, instrument state), and $n_i$=instrument noise; (2) Setting the model priors: they are built on knowledge of the contaminant sources, for example: periodicity of the barycentric motion, the quasi-periodicity of stellar activity, instrument state; and the correlation information between the time series data and the contaminants. (3) Defining the Likelihood shape: the Likelihood is assumed a Multivariate Gaussian probability distribution function. (4) Apply one of the established sampling techniques to get the marginalized posterior distribution of the planet's orbital parameters and its best-fit model as well as its evidence, and ascertain whether a given detection is a real planet or not. (5) While most existing techniques can only be applied once a planet has already been detected (e.g., Rajpaul et al 2015, Jones et al 2017), our approach will perform blind Bayesian detection and characterization of exoplanet.

In this study we collected the information from current data and existing analysis to help implementing the new method:

(1) We investigated the parameter correlations observed in the data (as shown in Figure 1. middle panels) to construct the covariance matrix of the Multivariate Gaussian Likelihood.
(2) The intrinsic stellar activity, ISA, is one of the leading challenges for exoplanet detection efforts. ISA can mimic a Doppler shift signal which is usually modelled as an additional noise term, the RV jitter (eg. Ford 2006). Figure 1. ( right hand side panel) depicts the effect of stellar activity. We devised how to construct the data model – following the framework set in place by Rajpaul et al 2015 and Jones et al 2017, we adopt a data-driven method for constructing stellar activity indicators and resort to Doppler-constrained dimension reduction approaches, more specifically, the PCA method. Where the PCA method decomposes the stellar spectra into orthogonal vectors along which variation in the spectra is observed across time. This basis can be used to express the $n_c$ component due to ISA in the data model d, for example, and to construct the covariance matrix of the coefficients that express the contribution of the ISA to the observed data set.
(3) Currently the PSF is modelled as a superposition of Gaussians. An asymmetric PSF will cause spurious Doppler shifts in the stellar lines that are not compensated by an identical displacement in the wavelength scale derived from the iodine lines. We investigated how an inaccurate PSF model can mimic the signal of interest. Figure 1. (LHS) depicts the effect of small variations in the PSF. To model the effect we resorted to a technique called, Markov Chain Beam Randomization, **MCBR**, developed by the PI of this proposal for CMB data analysis. This approach allows a joint data analysis of both the PSF variations and RV signal: we sample jointly the planet Keplerian orbit parameters and (variable) PSF and marginalize over the PSF uncertainties. The PSF variation is sampled directly from the set of PSF fit parameters (function of time) obtained with the Doppler code.
(4) Figure 2. exemplifies the presence of residual correlations in the data after removal of three detected planets (last panel). Our approach will enable to statistically quantify, using a Bayesian framework, our confidence when categorizing these residual signals as spurious.

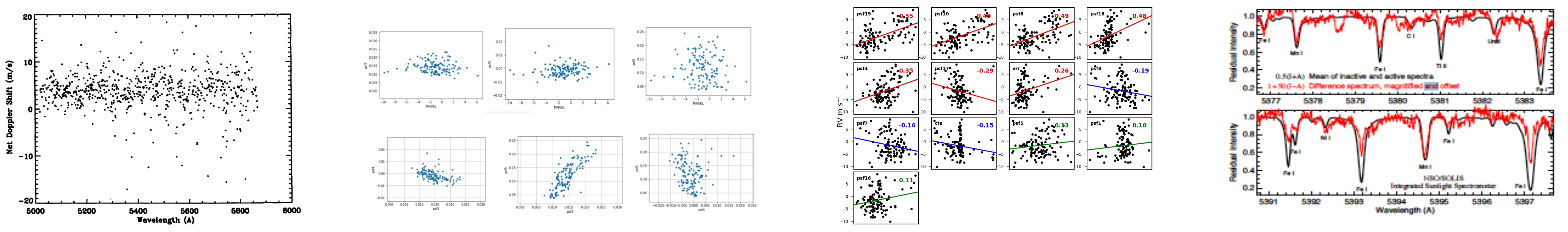

**Figure 1.** From Left hand side: A systematic error of 4 m $s^{-1}$ will accrue to Doppler measurements due to the relatively small variation in the PSF of 1%. - This is a variation smaller than can be easily stabilized at the telescope; Two middle panels from left to right: Example of parameter correlations observed in the data for a star without planets from our work with HIRES data; another example of parameter correlations (Fulton et al, 2018). These correlations can be used to construct the covariance matrix of the Multivariate Gaussian Likelihood. Right hand side: Full disk solar spectrum and the magnitude difference between active and inactive states. Individual lines respond differently to activity changes, providing a way to help distinguish surface phenomena from Keplerian velocity shifts (Fisher et al. 2016.);

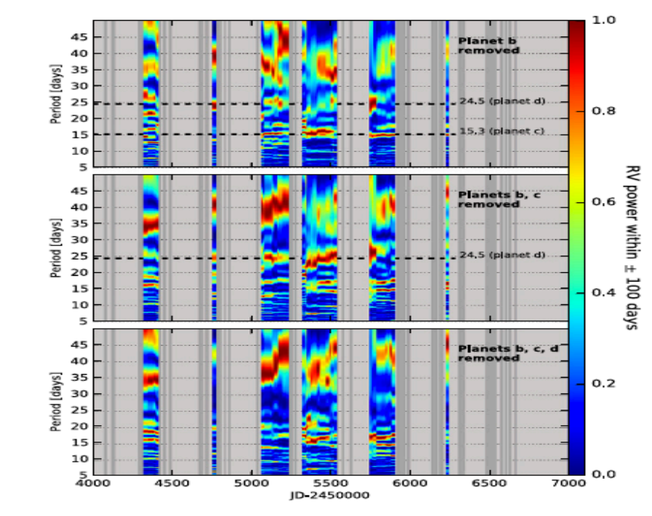

**Figure 2.** Top: periodogram of Keck HIRES RV measurements ±100 days versus time (JD-2450000) with the Keplerian model of planet b (P = 5.4 days) removedMiddle: same as top, with planets b and c removed. Bottom: same as top, with planets b, c, and d removed. This figure exemplifies the presence of residual correlations in the data after removal of the three detected planets.(Fulton et al 2015).

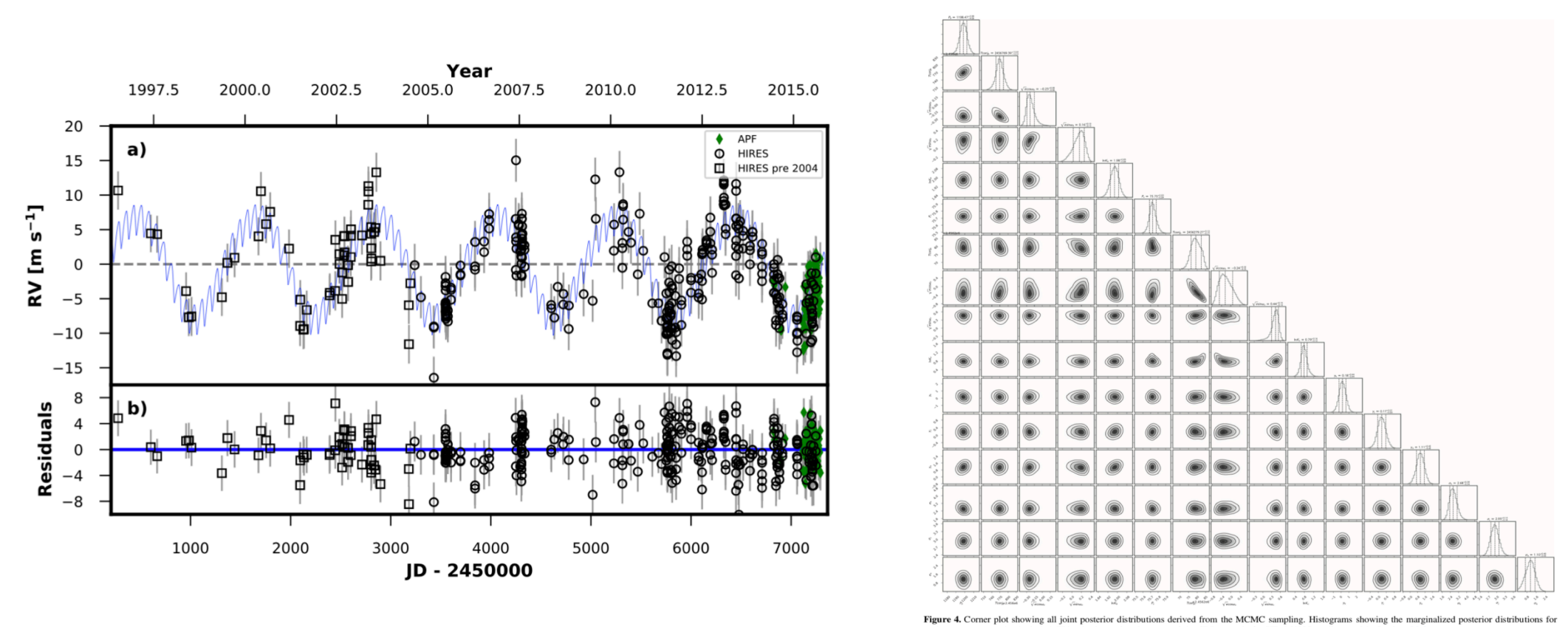

**Figure 3.** LHS: Example of RV measurement for a star with planets, HD 164922; RHS: Posterior distributions of the Keplerian orbit parameters when PSF variations are not accounted for. (Fulton et al, 2018)

## Discussion and Conclusions:

The findings of this proposal guided the full implementation of our method. Once the approach is fully implemented and integrated eg. with RadVel we will continue applying it to a subset of HIRES data from 2004-2018 from the Eta-Earth Survey of 166 nearby stars (Howard et al. 2010), that famously measured the occurrence rate of Earth-like planets. We expect our approach will help discover new planets, place better limits on non-detections, and measure the planet mass function in the 3-30 Earth-mass domain.

## Benefits to NASA and JPL (or significance of results):

This study along with upcoming research will set the stage for development of algorithms for use with the new KPF and PARVI instruments that are part of the emerging JPL/Caltech world-leading role in RV science and exoplanets.

## Publications:

1. Andrew Howard, et al., *Science* **330** (October 29, 2010): pp. 653-655.
2. Benjamin Fulton et al., *ApJ* **705** (June 1, 2015): 175 (17pp).
3. Paul Robertson et al., *Science* **345**, (July 25, 2014): pp 440-4.
4. V. Rajpaul et al., *MNRAS* **452, (**June 24, 2015**)**: pp 2269-2291.
5. David Jones et al,. eprint *arXiv*:1711.01318 (November, 2017): 39pp.
6. Debra Fischer et al., *PASP* 128:066001 (June 2016): 43pp

**National Aeronautics and Space Administration**
**Jet Propulsion Laboratory**
California Institute of Technology
Pasadena, California

**www.nasa.gov**

**Graca Rocha/R.19.023.108**
**Contact Information: 818-393-0095, Graca.M.Rocha@jpl.nasa.gov**

**Poster No. RPC-033**