*Astronomy in Focus, Volume 1*, Focus Meeting 8
*XXIXth IAU General Assembly, August 2015*
*Piero Benvenuti, ed.*

doi:10.1017/S1743921316002817

# Bayesian Planet Searches for the 10 cm/s Radial Velocity Era

**Philip C. Gregory**

Physics and Astronomy, University of British Columbia, 6224 Agricultural Rd, Vancouver, British Columbia V6T 1Z1, Canada
email: gregory@phas.ubc.ca

**Abstract.** A new apodized Keplerian model is proposed for the analysis of precision radial velocity (RV) data to model both planetary and stellar activity (SA) induced RV signals. A symmetrical Gaussian apodization function with unknown width and center can distinguish planetary signals from SA signals on the basis of the width of the apodization function. The general model for $m$ apodized Keplerian signals also includes a linear regression term between RV and the stellar activity diagnostic $\ln(R'hk)$, as well as an extra Gaussian noise term with unknown standard deviation. The model parameters are explored using a Bayesian fusion MCMC code. A differential version of the Generalized Lomb-Scargle periodogram provides an additional way of distinguishing SA signals and helps guide the choice of new periods. Sample results are reported for a recent international RV blind challenge which included multiple state of the art simulated data sets supported by a variety of stellar activity diagnostics.

**Keywords.** stars: planetary systems; methods: statistical; methods: data analysis; techniques: radial velocities.

---

At the current m/s RV precision level, intrinsic stellar activity (SA) has become the main limiting factor. New spectrographs are under development like ESPRESSO and EXPRES that aim to improve RV precision by a factor of approximately 100 over the current best spectrographs, HARPS and HARPS-N. Clearly, the success of these developments hinges on our ability to distinguish true planetary signals from SA induced signals. At the 'Towards Other Earths II" meeting held in Porto Portugal in September 2014, Xavier Dumusque challenged the community to a large scale blind test using simulated RV data at the 0.7 m/s level of precision, to understand the limitations of present solutions to deal with SA signals and to select the best approach. This paper describes a new approach using apodized Keplerian models which was tested on the first five of the RV Challenge data sets.

For the apodized Keplerian (AK) models, the semi-amplitude of the Kepler RV model is multiplied by a symmetrical Gaussian of unknown width, $\tau$, and with an unknown center of the apodizing window, $t_a$. Since a true planetary signal spans the duration of the data $\tau$ will be large while SA induced RV signals generally vary on shorter time scales. The general model for $m$ apodized Keplerian signals also included a linear regression term between RV and the stellar activity diagnostic $\ln(R'hk)$, as well as an extra Gaussian noise term with unknown standard deviation. The correlation term was particularly useful in removing long term SA signals associated with the stars magnetic cycle.

In addition to the RV measurements, the challenge data sets includes simultaneous observations of three stellar activity diagnostics. Two of these come from additional information on the spectral line shape that are extracted from the cross correlation function (CCF), the average shape of all spectral lines of the star. These two shape parameters are the CCF width (FWHM) and bisector span (BIS). The third diagnostic, $\ln(R'hk)$, is based on the Ca II H & K line flux that is sensitive to active regions on the stellar surface. A preliminary analysis of the first 5 data sets indicated a strong correlation between RV and the $\ln(R'hk)$ diagnostic and slighlty reduced correlation with the FWHM diagnostic.

The AK models were explored using an automated fusion MCMC algorithm (FMCMC; Gregory 2013), a general purpose tool for nonlinear model fitting and regression analysis. The AK models combined with the FMCMC algorithm† constitute a multi-signal apodized Keplerian periodogram.

The primary role of the AK models is to distinguish planetary signal candidates from SA signals. Suppose the results indicate that $k$ of the signals are planetary and $m-k$ are SA signals. Final model parameter estimates and model comparisons are based on subsequent runs using a model of $k$ Keplerians and $m-k$ apodized Keplerians.

The methodology is partially illustrated for the second challenge data set. The raw RV 2 data had a standard deviation of 8.58 m/s. After removing the best linear regression fit to $\ln(R'hk)$ as the independent variable, the standard deviation was reduced to 3.95 m/s. The top two rows of Figure 1 show the RV data and FWHM after removing the $\ln(R'hk)$ diagnostics together with their GLS periodogram on the right. In this analysis the FWHM (rhk corrected) is treated as a control.

It proved useful to construct a differential form of the Zechmeister & Kürster (2009) Generalized Lomb-Scargle (GLS) periodogram of selected period regions for the RV residuals. Two examples of this are shown in the bottom two panels. The black trace is from the upper right GLS periodogram in Figure 1. The dark gray trace is the negative of the GLS periodogram of the FWHM control (middle right panel) and the light gray trace shows the difference, the black trace plus the gray trace. The strongest signals at $P=3.77$ and 10.63 d have no significant counterpart in the control. Strong yearly aliases signal are also evident. For both periods the light gray trace and the black trace coincide closely near the positive peaks, consistent with a planetary signal or its alias. Signals in common to both black and dark gray traces (e.g., near $P=12.5$ d) indicates SA.

Figure 2 shows sample FMCMC results for an eight AK signal model. It is clear from this that there are three planetary candidates with periods of 3.77, 10.63, and 75.56 d whose apodization windows span the duration of the data. The apodization time constant of $\sim 800$ d for the 10.63 d signal make it a borderline P candidate. Also, the presence of SA activity at a period of 11.06 d in the close vicinity of the 10.63 d signal called into question a planetary interpretation of the 10.63 d signal. For the competition, the 10.63 d signal was reported as a probable planetary signal. The true planetary signals injected into the RV 2 data set included three with $K \geqslant 1$ m/s at $P=3.77, 10.64, 75.28$ and two with $K<1$ m/s at $P=5.79, 20.16$ d. Starting from the raw data (standard deviation of 8.58 m/s), which was dominated by SA, we have been able to achieve residuals of 1.42 m/s which is a factor of six lower but still a factor of two higher than the mean measurement uncertainty of 0.7 m/s. Further studies including consideration of other apodization functions (e.g., an asymmetrical Gaussian) are warranted.

† A more detailed description of FMCMC is available in Chapter 1 of "Supplement to 'Bayesian Logical Data Analysis for the Physical Sciences'," a free supplement available in the resources section of the Cambridge University Press website for my Textbook "Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with *Mathematica* Support." A *Mathematica* implementation of fusion MCMC is also available from the resource section. The supplement includes a detailed discussion of the priors adopted by this author for exoplanet RV analysis. Chapter 1 also includes a comparison of three marginal likelihood estimators used for Bayesian model comparison and concludes in favor of the Nested Restricted Monte Carlo (NRMC) estimator which is is used in this work.

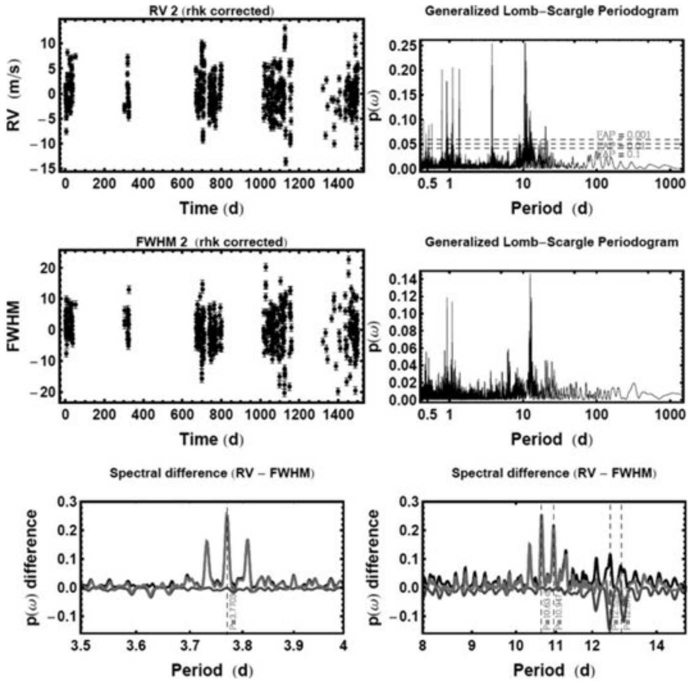


**Figure 1.** The RV 2 data and FWHM (control) after removing the $\ln(R'hk)$ diagnostics (rhk corrected) together with their GLS periodograms on the right. The differential GLS periodogram for two selected period ranges is shown in the bottom two panels.

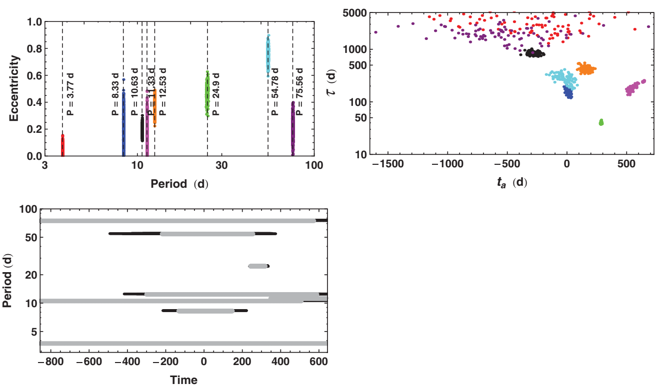


**Figure 2.** The upper right panel shows the eccentricity versus period parameters for sample MCMC iterations for the 8 signal apodized Kepler periodogram of the RV 2 data. The upper left panel is a plot of the apodization time constant, $\tau$, versus apodization window center time, $t_a$. The lower panel shows the apodization interval for each signal (gray trace for MAP values of $\tau$ and $t_a$, black for a representative set of samples which is mainly hidden below the gray).

## References

Gregory, P. C., 2007, *MNRAS*, 381, 1607-1616

Gregory, P. C., 2013, in *Astrostatistical Challenges for the New astronomy*, Joseph M. Hilbe (ed.), Springer Series in Astrostatistics 1,121

Zechmeister, M. & Kürster, M., 2009, *A&A*, 496, 577

https://doi.org/10.1017/S1743921316002817 Published online by Cambridge University Press