Project Brief

Project Title: Red Herrings in Radial Velocity? Assessing the Authenticity of Planets Discovered Around Sun-Like Stars

Author: Garima Prabhakar

Date: 12/27/19

Background:

Introduction

The search to uncover exoplanets is one of the fastest expanding fields in astrophysics, allowing for the discovery of thousands of exoplanets in only a couple of decades, and revolutionizing the generic outlook on planetary system physics and habitability. The ultimate goal of this field is to find a planet that can and may already hold life. Moreover, in 2012, researchers at the University of Geneva claimed to have found an earth-mass planet orbiting Alpha Centauri B, the closest star system to the sun, a discovery that was received with widespread acclaim from scientists and ordinary people alike. However, in 2016 another team discovered that the supposed earth-mass planet orbiting the star wasn't a planet at all, but instead an artifact of noise being produced by the star itself. This problem of inherent noise in exoplanetary detection measurements is one of the most frequent and hardest to resolve causes of error in the detection of an exoplanet. This project attempts to check the current exoplanet population for cases like Alpha Centauri B b. The aim of this project is to find the likelihood that an exoplanet close to its host star that has already been detected and confirmed with the radial velocity exoplanetary detection method is not a planet at all, but a source of noise.

An Overview of Exoplanetary Research

Exoplanets are planets outside the solar system. The first exoplanet, a large and close-orbiting planet larger than Jupiter (called "Hot Jupiters") was discovered around a by a team of researchers in the mid 1990s. Since then, the field as exponentially increased as a result of newfound interest in research and funding (D. Hall et al 2018). Missions like the Hubble Space Telescope, Kepler, and TESS surveys have helped to increase the exoplanet population discovered from one exoplanet to 400 (Figure 1).

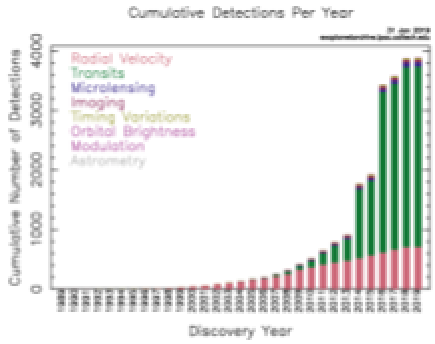

Figure 1: Statistics of exoplanet detection over time (NASA Exoplanet Archive)

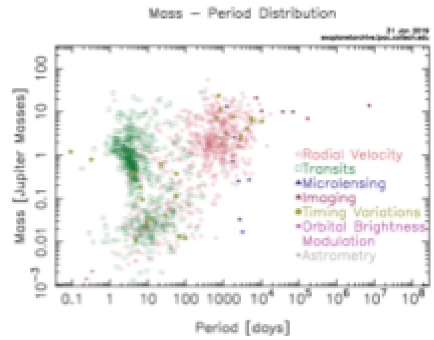

Figure 2: Mass vs. period of planets discovered until January 2019 using different detection methods (NASA Exoplanet Archive)

There are a variety of methods used to discover exoplanets, with the most popular being the radial velocity and transit method. The transit method aims to measure decreases in the light intensity coming from a star over time as a planet passes over the stellar disk. The radial velocity method is a more indirect method. The radial velocity method aims to detect movements of the star as a planet pulls on it. The planet and star are known to orbit around their common barycenter. The planet exerts the same gravitational force on the star that the star exerts on the planet. Because of this, the star gets pulled because of the gravitational influence of the star (Fischer & Lovis). This movement of the star because of the planet can be detected through doppler shifts in the star's spectra as the star moves towards or away from the Earth (Fischer & Lovis). Throughout this background brief, the radial velocity detection method will be focused on the most as it is the focus of this project.

Problems in Exoplanet Recovery

Even though the radial velocity method of exoplanet detection is extremely accurate when taking into consideration the sheer distance that these measurements are being taken, there are still key challenges that prevent the field from discovering earth-sized exoplanets in habitable zone regions more frequently.

A crucial problem in exoplanet detection is the presence of noise in the measurements (Fulton 2018). Often, noise sources such as the weather and the earth's revolution can be factored out through in-space surveys, but things like instrumental noise, photon noise, and especially stellar activity can prevent the detection of low-mass or high-period exoplanets because their signal masks out the fainter exoplanet signals (D. Hall 2018). Moreover, many times sources of noise such as stellar activity, known as jitter, pose a bigger problem because this source of noise can act exactly like an exoplanet signal and due to this many times is classified as a planet when it is in reality an alias (Dumusque 2014). Figure 4 introduces a flowchart that describes the major causes the stellar activity problem, and figure 5 presents a table of the problems with current exoplanet recovery systems.

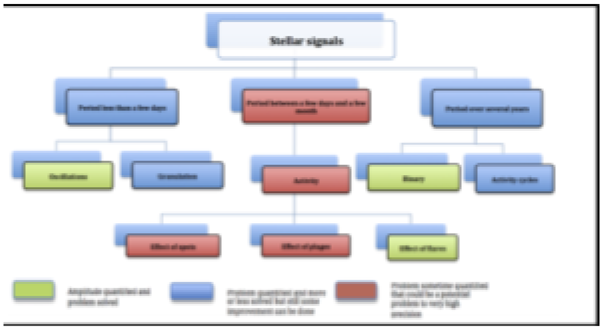

Figure 4: A flowchart of specific causes of stellar activity that hamper high-precision exoplanet detection (Plavchan 2015).

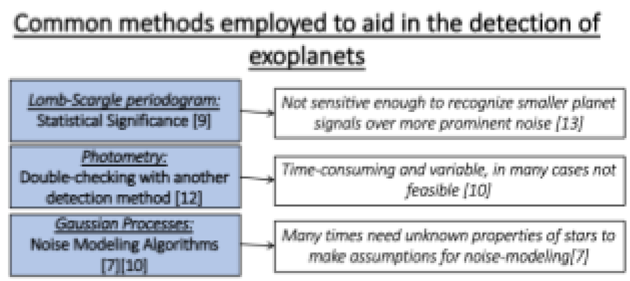

Figure 5: A table of current commonly used exoplanet detection methods and their drawbacks due to stellar activity.

Project Definition:

Problem Statement: Exoplanets, or planets outside of the solar system are often mis-detected due to the presence of astrophysical noise.

Goal: The goal of this project is to assess the false-positive rate of small-period exoplanets detected with the radial velocity method.

Experimental Design/Research Plan Goals:

Overall Experiment Materials

- Macbook Pro computer, late 2019 model, 8 GB RAM
- Access to NASA Exoplanet Archive
- Access to Terminal (command-line interface)
- MATLAB software
- Python software
- SOAP 2.0 python package
- Numpy and scipy python package
- Astropy python package

<u>Procedure</u>

Transit Procedure

Goal 1: Design a single pipeline that can be used to search for planets and characterize pipeline completeness

<u>Strategy:</u>

- Take the constrained sample of stars and take their planet signals out of them by running them through a planet detection algorithm.
- Then, inject trial planet signals and check to see what fraction of the signals the planet detection algorithm is able to detect.

I. **Stellar Sample Selection**
   A. <u>Constraining Factors</u>
      1. Constrain Effective Temperature (T_{eff}), Luminosity, Metallicity, Radius, and Mass of the Target Star (In my case, G Solar-Type stars)
      2. Take out stars with small number of measurements
      3. Take out stars with large amounts of "flagged" data
      4. Take out stars that are noise (as defined by the literature)
      5. Double-check the data and remove stars with high FAP for all of their signals as computed by
   B. <u>Preparing the Data</u>
      1. Obtaining all data of each selected star from
      2. Exclude all flagged for low-quality
      3. Detrend RV curves
      4. Look for anomalies and indicators of any stellar activity
         a) Stellar activity indicators

II. **Planet Detection Pipeline**
   A. <u>Predicting Planets</u>
      1. Predict expected orbital period based on mass and radius of star
      2. Decompose data into a periodogram
      3. Compute signal detection efficiency (SDE) (Kovács et al. (2002)
      4. Rerun the LSP with periods around all of the periods detected (sampling from normal distribution)
      5. To prevent choosing the harmonic of the true planetary period, search for periodogram peaks at 1/n trigger periods (where n= 2-7) as well
         a) Choose the period for which $\Delta\chi^2$ is minimized.

6. Repeat steps 1-6 two more times
    a) Stop planet searching if no SDE>6
    b) Stop planet searching if no FAP<20%
    c) "None of the transit models for the peaks were accepted"
    d) The code completed three iterations of searching for planets

B. Planetary Properties
    1. Use Radvel, assume circular orbits, as eccentricity is not well-constrained.

III. **Vetting Pipeline (**Cutting down transit events, confirming planets)

A. Accounting for noise I
    1. Look at stellar rotation as a function of stellar magnitude/luminosity and scrutinize periods at/at harmonics of the stellar rotational period
    2. Visually inspect remaining samples to see if they're from noise
    3. Again, check to see if the signals weren't just harmonics of the true planetary period (ranked with $\Delta\chi^2$)

B. Accounting for noise II
    1. Check visually and more thoroughly for common false positives such as eclipsing binaries (for transit), etc.
    2. Check whether previous studies had identified the signal as a false positive

C. Post-Vetting: Planet Confirmation
    1. Rerun pipeline on remaining stars to see if there was any planet left undetected
    2. Repeat vetting pipeline
    3. Repeat steps 1-2 one more time (3 iterations of detection and vetting pipeline in total)
    4. Check literature for any close stellar companions for each star
        a) If there is, implement a correction factor on the signal (PERIOD??)(see dressing 2015's correction factor, which is based on the magnitude difference and radius between the two stars)
        b) Can I use the mass-radius relation for my correction factor given its inherent bias?
    5. Apply radius (mass) dependent false-positive probability of a planet (TABLE 1 OF FRESSIN 2013, RESEARCH MORE ON THIS)
    6. Report any known planets missed by the pipeline

IV. **Estimation of Planetary Properties**

A. "Used MCMC analysis"
    1. I can use Radvel

V. **Planet Injection Pipeline** (In order measure planet occurrence rate you need to know the *completeness of the planet candidate list*)

A. Prologue: Detection process

1. Signal identification (LS periodogram, LSP)
    a) The period of the signal is identified as a peak in the periodogram
2. Signal confirmation
    a) The signal is accepted because it provides a 5σ improvement to a straight line fit

B. Planet Injection

1. For each star, generate 2000 “trial planets” (w/ 0.5-200 day period sampled from log uniform distribution)
2. Construct transit (radial velocity) models from stellar and planet parameters from each star used
    a) *Address the question: will the trial planets be accepted in stage V.A.2?*
3. Reject a trial planet if its $\Delta\chi^2$ indicates that its “not preferred”
4. Using a “straight-line fit method described in section 3.2”, superimpose each trial planet signal into the star’s data (along with all of the noise, etc.)
5. Repeat step 3 for the data in step 5
6. Rank the trial planets in increasing order of $\Delta\chi^2$
7. Randomly select planets that are closest to the $\Delta\chi^2$ threshold (thin sample by 50%)
8. Conduct up to 25 additional runs to recover at least 10 planets per star
9. Run the planet detection pipeline on the radial velocities) in #8. Additionally, halt the process as soon as the injected signal is found (otherwise terminate the program like you usually do).

Procedure for Radial Velocities

Goal 1: Design a single pipeline that can be used to search for planets and characterize pipeline completeness

Strategy:

- Take the constrained sample of stars and take their planet signals out of them by running them through a planet detection algorithm.
- Then, inject trial planet signals and check to see what fraction of the signals the planet detection algorithm is able to detect.

I. **I. Stellar Sample Selection**

A. Constraining Factors

1. Constrain Effective Temperature (T_{eff}), Luminosity, Metallicity, Radius, and Mass of the Target Star (In my case, G Solar-Type stars)
2. Only select stars that have been observed with both Kepler and HARPS
3. Take out stars with small number of measurements
4. Take out stars with large amounts of “flagged” data
5. Take out stars that are noise (as defined by the literature)
6. Double-check the data and remove stars with high FAP for all of their signals as computed by

B. Preparing the Data

1. Obtaining all data of each selected star from
2. Exclude all flagged for low-quality
3. Detrend RV curves
4. Look for anomalies and indicators of any stellar activity
   a) Stellar activity indicators

II. **Planet Detection and Vetting Pipeline**

A. Follow “Procedure for Identifying Variability and Trends in Radial Velocity Data” (Zechmeister et al. 2009, Hatzes et al. 2006)

B. Vetting

1. Check visually and more thoroughly for common false positives such as eclipsing binaries (for transit), etc.
2. Check whether previous studies had identified the signal as a false positive
3. Rerun pipeline on remaining stars to see if there was any planet left undetected
4. Repeat vetting pipeline
5. Repeat steps 1-2 one more time (3 iterations of detection and vetting pipeline in total)

6. Check literature for any close stellar companions for each star
    a) If there is, search periodogram for the signal corresponding to the stellar companion
    b) delete this signal
7. Report any known planets missed by the pipeline

III. Estimation of Planetary Properties

A. "Used MCMC analysis" (Zechmeister et al 2009)
    1. I can use Radvel

IV. Planet Injection Pipeline (In order measure planet occurrence rate you need to know the *completeness of the planet candidate list*)

a) Prologue: Detection process
    (1) Signal identification (LS periodogram, LSP)
        (a) The period of the signal is identified as a peak in the periodogram
    (2) Signal confirmation
        (a) The signal is accepted because it provides a 5σ improvement to a straight line fit

b) Planet Injection
    (1) For each star, generate 2000 "trial planets" (w/ 0.5-200 day period sampled from log uniform distribution)
    (2) Construct transit (radial velocity) models from stellar and planet parameters from each star used
        (a) Address the question: will the trial planets be accepted in stage V.A.2?
    (3) Reject a trial planet if its $\Delta\chi^2$ indicates that its "not preferred"
    (4) Using a "straight-line fit method described in section 3.2", superimpose each trial planet signal into the star's data (along with all of the noise, etc.)
    (5) Repeat step 3 for the data in step 5
    (6) Rank the trial planets in increasing order of $\Delta\chi^2$
    (7) Randomly select planets that are closest to the $\Delta\chi^2$ threshold (thin sample by 50%)
    (8) Conduct up to 25 additional runs to recover at least 10 planets per star
    (9) Run the planet detection pipeline on the radial velocities) in #8. Additionally, halt the process as soon as the injected signal is found (otherwise terminate the program like you usually do).

2. **Predicting Detectability**
    a) Find the fraction of BLS trial planets recovered as a function of the $\boldsymbol{\Delta}\chi^2$ "computed in the second round of model tests"
        (1) Rank BLS planets by $\boldsymbol{\Delta}\chi^2$ and find recovery fraction for each group of 1000 planets.
        (2) Smooth the resulting histogram and compute the likelihood of detection for the non-BLS runs using a cubic spline interpolation based on the smoothed histogram
        (3) Dressing et al limited the max detection likelihood to 91.5% (max value of CSI in the histogram)
    b) <u>Assessing Pipeline Performance</u> *Map recovery fraction to SNR, (remember to inject multiple trial planets per star)*
        (1) Generate radial velocity detectability maps by "summing individual maos"
            (a) Use smaller grids, and develop sensitivity maps

3.

Procedure for Identifying Variability and Trends in Radial Velocity Data (Zechmeister et al. 2009, Hatzes et al. 2006)

Perform "several statistical tests to identify variability and RV trends in our data"
Ask the question: *Is the observed variability or rms σ significantly higher than the mean measurement error σ_{RV}?*
<u>Strategy</u>

- F-test
- Goodness of fit for a constant model vs. non-constant model

1. F = σ^2/σ_{RV}
    a. If p(f) form the f-test is low, this means that measurement/instrument error is not enough to explain the variability.
    b. Else, the variability is not significant enough to be something other than photon noise, instrumentation error, and measurement error.
    c. Very high and significant rms may indicate a stellar companion/brown dwarf/hot jupiter
2. Drop any stars with insignificant F-value.
3. Calculate the χ^2 above the "weighted RV mean" (see equation in a.) and derive the probability from the χ^2-distribution.
    [3] When each measurement has the same error σ_{RV}, one gets
    a. $$\chi^2_{\mathrm{red}} = \frac{\chi^2_{\mathrm{const}}}{N-1} = \frac{1}{(N-1)\sigma^2_{RV}} \sum (RV_i - \overline{RV})^2 = \frac{\sigma^2}{\sigma^2_{RV}} = F.$$
4. If the probability sampled is significant, there is a good chance that the variability is not due to instrumentation, photon, or measurement noise.
5. To check for long-term variability, fit a "weighted linear slope" to the data.
6. A high probability of χ^2_{slope} means that the fit is acceptable, whereas low probability means that there is more variability than that
7. Calculate $\chi^2_{\mathrm{constant}}$ and χ^2_{slope}.
    a. $$F_{\mathrm{slope}} = (N-2)\frac{\chi^2_{\mathrm{constant}} - \chi^2_{\mathrm{slope}}}{\chi^2_{\mathrm{slope}}}.$$
8. If F_{Slope} is low, the slope fit is not enough and that there is more variability that is probably not due to noise, if it is high, the slope fit is an improvement.
9. Add the linear trend to the model if it is significant enough to explain the variability.

Periodogram Analysis *(look for significant signals in RV data for each star)*

1. Compute generalized lomb-scargle periodogram (GLS) for each star (Zechmeister & Kürster 2009), available through Astropy (fit_mean = True).
2. Calculate FAPs by bootstrap randomization (available through Astropy (method = "bootstrap")
   a. Generate 1000 random data sets (resolves FAP $>10^{-3}$)
3. If the FAP is below the threshold of 0.01, the signal is significant enough to be either stellar activity or another companion.
4. Drop all signals that do not pass the FAP threshold.

Upper Detection Limits (INCOMPLETE, REFER ZECHMEISTER & KURSTER 2009)
Answer the question: At what signal amplitude does the GLS for the dataset obtain a FAP of 0.01?

1. Use notation from Zechmeister & Kürster (2009):
   a. $p_y(\omega) = \frac{1}{YY} \cdot \frac{SS \cdot YC^2 + CC \cdot YS^2 - 2CS \cdot YC \cdot YS}{CC \cdot SS - CS^2}$
   b. $p_x(\omega) = \frac{1}{XX}\left[p_y YY + XX - YY\right] = 1 - \frac{(1-p_y)YY}{XX}.$
   c. $XX = YY\frac{1-p_y}{1-p_x}.$
   d. $$\begin{aligned} XX &= \sum w_i(x_i - \overline{x})^2 \\ &= \sum w_i(y_i - \overline{y} + a(\cos\omega_0 t_i - C_0) + b(\sin\omega_0 t_i - S_0))^2 \\ &= YY + a^2 C_0 C_0 + b^2 S_0 S_0 + 2aYC_0 + 2bY_0 S_0 \\ &\quad +2abC_0 S_0. \end{aligned}$$
   e. $A(\omega, \varphi) = -\frac{\beta}{\alpha} \underset{(-)}{+} \sqrt{\left(\frac{\beta}{\alpha}\right)^2 + 4\frac{\gamma}{\alpha}},$

2. Or, inject 1000 signals with amplitudes ranging from 0 m/s to 1 km/s into the data (randomly sampled from normal dist) for "12 different equidistant phases".
3. Record the amplitude at which FAP most nearly equals 0.01
4. Repeat steps 2-3 with 1000 injected signals at $\pm$ 25 m/s
5. Return the amplitude at which FAP = 0.01

Checking for Stellar Activity Presence (keep in mind that HiRES data only uses H-alpha/S-value stellar activity indicators, and the dataset reports -1 if they aren't available for the dataset)

H-alpha/S-value (Zechmeister et al (2009), Hatzes et al (2006))

1. Use the H-alpha vs. time data to generate GLS periodograms, and find peak with the lowest FAP (define threshold?).
2. If the RV signal at the peak found in step 1 has a FAP<0.01, subtract it from the RV data
3. Subtract the peak found in step 1 from the h-alpha periodogram

4. Repeat steps 1-3 for the next lowest FAP in the h-alpha periodogram

Timeline:

Project Schedule: Click here

- September 20 - September 27:
  - Finalize topic
  - Preliminary literature research (5 overview papers, 2 literature reviews)
- September 27 - October 31:
  - Full research
- November 1 - December 1:
  - Stage 3 Prototype 1 (procedure, design, implementation)
- December 2 - December 15:
  - Analysis, presentation
  - Research Review and direction
- December 16 - January 2:
  - Stage 3 Prototype 2 (Check-in)
  - Analysis and conclusion
- January 2 - January 15:
  - Literature review and resolution
- January 15 - February 1:
  - Stage 3 Prototype 3 (Check-in)
  - Analysis and conclusion
  - Literature review and resolution
- February 1 - March 1:
  - Construction of final design:
  - Analysis and conclusion
- March 1 - March 10:
  - Preparation of presentation
- March 10 - March 20:
  - Presentation practice

Background Knowledge Goals and Roadblocks:

Date	Topic	Completed Date
09/18/19	Problems with current exoplanet detection methods	10/9/19
09/19/19	very imbalanced datasets and	...

	neural networks	
09/19/19	Problems with my previous model	10/15/19
10/20/19		

Project Brainstorming Overview

What are you pursuing?	**Why?**
Making exoplanet detection systems more accurate	I have background knowledge, and a pretty good idea of where I can take the project-- it's more exciting when you can do more with project. I also (potentially) have help.
Making rogue asteroid detection systems more accurate	Sounds very exciting, and there really hasn't been much previous research.

Want to use	**Why**	**Assumptions**	**How assumptions can be challenged?**
Making exoplanet detection systems more accurate	I have background knowledge, and a pretty good idea of where I can take the project-- it's more exciting when you can do more with project. I also (potentially) have help.	● The results of my previous projects are accurate. ● Multiple exoplanets can be detected at the same time in RV signals (I don't have to subtract signals).	● Check with an expert to see if previous results were accurate. ● Experiment with this and do some literature review.
Making rogue asteroid detection systems more accurate	Sounds very exciting, and there really hasn't been much previous research.	● Using ATLAS data to detect rogue asteroids.	● Check for data on websites.

What aren't you pursuing?	**Why?**

Finding ways to more efficiently reforest America (mapping tree types and instructions to forest geography)	There might not be enough data, and the procedure (even an outline) to do this seems very unclear.
Detecting microplastics in the air	This also has an unclear methodology, and not sure how I would go about doing this (or what implications trying to move in this direction would have)

References

1. Plavchan P, Latham D, Gaudi S, et al. Radial velocity prospects current and future: A white paper report prepared by the study analysis group 8 for the exoplanet program analysis group (ExoPAG). *arXiv preprint arXiv:1503.01770*. 2015.

2. Dumusque X, Boisse I, Santos NC. SOAP 2.0: A tool to estimate the photometric and radial velocity variations induced by stellar spots and plages. *Astrophys J*. 2014;796(2):132.

3. Fulton BJ, Petigura EA, Blunt S, Sinukoff E. RadVel: The radial velocity modeling toolkit. *Publications of the Astronomical Society of the Pacific*. 2018;130(986):044504.

4. Beaugé C, Ferraz-Mello S, Michtchenko TA. Multi-planet extrasolar systems—detection and dynamics. *Research in Astronomy and Astrophysics*. 2012;12(8):1044.

5. Haywood RD, Collier Cameron A, Queloz D, et al. Planets and stellar activity: Hide and seek in the CoRoT-7 system. *Monthly notices of the royal astronomical society*. 2014;443(3):2517-2531.

6. Hall RD, Thompson SJ, Handley W, Queloz D. On the feasibility of intense radial velocity surveys for earth-twin discoveries. *Monthly Notices of the Royal Astronomical Society*. 2018;479(3):2968-2987.

7. Lovis C, Fischer D. Radial velocity techniques for exoplanets. *Exoplanets*. 2010:27-53.