A SEARCH FOR VARIABILITY AND TRANSIT SIGNATURES IN HIPPARCOS PHOTOMETRIC DATA

A thesis presented to the faculty of
San Francisco State University
In partial fulfilment of
The Requirements for
The Degree

Master of Science
In
Physics: Astronomy

by

Badrinath Thirumalachari

San Francisco, California

December 2018

Copyright by
Badrinath Thirumalachari
2018

CERTIFICATION OF APPROVAL

I certify that I have read *A SEARCH FOR VARIABILITY AND TRANSIT SIGNATURES IN HIPPARCOS PHOTOMETRIC DATA* by Badrinath Thirumalachari and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirements for the degree: Master of Science in Physics: Astronomy at San Francisco State University.

Dr. Stephen Kane, Ph.D. Astrophysics
Associate Professor of Planetary Astrophysics

Dr. Joseph Barranco, Ph.D. Astrophysics
Chair & Associate Professor of Physics

Dr. Ron Marzke, Ph.D. Astronomy
Assoc. Dean of College of Science & Engineering

A SEARCH FOR VARIABILITY AND TRANSIT SIGNATURES IN HIPPARCOS PHOTOMETRIC DATA

Badrinath Thirumalachari
San Francisco State University
2018

The study and characterization of exoplanets has picked up pace rapidly over the past few decades with the invention of newer techniques and instruments. Detecting transits in stellar photometric data around stars already known to harbor exoplanets is crucial for exoplanet characterization. Due to these advancements we now have oceans of data and coming up with an automated way of performing exoplanet characterization is a challenge. In this thesis I describe one such method to search for transits in Hipparcos dataset containing photometric data for over 118000 stars. The radial velocity method has discovered a lot of planets around bright host stars and a follow up transit detection will give us the density of the exoplanet. The Transit Ephemeris Refinement and Monitoring Survey (TERMS) is a project that calculates predicted transit times of the known radial velocity planets. A method was developed where we used the TERMS data to search for transits in photometric data collected by the Hipparcos mission. Our method looks for photometric data in the Hipparcos mission that falls under the TERMS predicted transit duration and window. A statistical test is performed to validate the transit points in the transit window and duration. Our method was tested for the known transiting exoplanets orbiting HD 209458 and HD 189733. The method is working for the test host stars, after which it was applied to the selected 97 targets from the catalog of over 118000 stars. The results from such an analysis helps us to identify potential transit points or to rule our transits entirely in a data set. Targets of interest can also be identified and a follow up radial velocity or transit observation can be carried on. Our future goal is to apply this technique to other datasets containing stellar photometric data and multi-planet systems.

I certify that the Abstract is a correct representation of the content of this thesis.

Chair, Thesis Committee   Date

ACKNOWLEDGMENTS

A special thank you to my research advisor, Stephen Kane, for inspiring and supporting me to pursue this endeavor and the entire SFSU Physics & Astronomy Department. I would also like to thank my wife Abhishikta Pandit for supporting me through this process, she never stopped believing in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Hipparcos Mission and Hipparcos Catalog

The Hipparcos Catalog is the primary products of the European Space Agency's astrometric mission, Hipparcos. The satellite, which operated for four years, returned high quality scientific data from November 1989 to March 1993.

The products of the Hipparcos mission are two major astrometric catalogues, the Hipparcos Catalogue (of 118218 stars) and the Tycho Catalogue (of more than one million stars), both derived from instruments on board the Hipparcos satellite. The global data analysis tasks was undertaken by the NDAC and FAST Consortia, together responsible for the production of the Hipparcos Catalogue, and the Tycho Consortium, responsible for the production of the Tycho Catalogue. Each of the catalogues includes a large quantity of very high quality astrometric and photometric data. Only the final reduced and calibrated astrometric and photometric data are provided in a machine readable format for data analysis. These files can be mined from the NASA exoplanet archive online.

The photometry part of the mission provided an average of some 110 epochs acquired for all objects in the Hipparcos Catalogue. The epochs of the measurements are given in Terrestrial Time corresponding to the Solar System Barycentric Julian Date (BJD). The differences between BJD and Heliocentric Julian Date (HJD) is negligible for our study. The Hipparcos measurements were performed in the specific Hp band, which is centered near 450nm and has a width of 240nm. The precision of median Hipparcos magnitudes, for constant stars, is in the range 0.0004-0.007 mag (over the interval 2-12 mag) with corresponding individual transit errors in the range 0.003- 0.05 mag. The typical precisions, at around 8 mag, are around 0.0015 mag on the median, and 0.011 mag on the individual errors. These measurements can be used to detect possible transit points of a know exoplanets in collabration with the Transit Ephemeris Refinement and Monitoring Survey (TERMS).

1.2 Theory of Exoplanet Detection

There are several methods that have successfully detected exoplanets. Here I focus on the usage of the transit method and the radial velocity method. I describe these two methods in some detail.

1.2.1 Overview of detection methods

Pondering on the existence of worlds had led us to search worlds beyond our own solar system. However, unlike the planets in our Solar System which are close to us and thus appear very bright, exoplanets are very difficult to observe directly. The light of a planet is millions of times fainter than the light of its star and when seen from many light years away the planet appears close to the star and light from the planet is lost over by the glare of the star.

Several claims of exoplanet detections have been documented since the 19th century, but the first confirmed detections were made by Wolszczan & Frail (1992) who monitored the irregularities in the timing of pulsars. These exoplanets, PSR 1257+12 b and c, are a few Earth masses and orbit a pulsar. The first exoplanet found around a sun like star was in 1995 using the radial velocity technique by Mayor & Queloz around the star 51 Pegasi. This exoplanet, 51 Pegasi b, is classified as a Hot Jupiter with a mass of $.46M_J$ and it orbits close to its star about .052AU.

In the past three decades, the number of exoplanets discovered has gone up to more than 3800. This progress is the result of several improvements in instrumentation and observing techniques, such as the development of CCD cameras, the development of stable high resolution spectroscopy, and the introduction of computer-based image processing. It is also due to the space based missions like *Kepler* and ground-based transit searches like *OGLE*.

1.2.2 Radial Velocity Method

This method detects the oscillating Doppler shift in the stellar spectrum due to the periodic radial velocity motion (motion along the line of sight) of a star gravitationally tugged back and forth by an orbiting planet. This periodic stellar motion towards and away from us results in a Doppler shift in the light received from the star. Using spectroscopy, we measure these blue and red shifts in the star's spectrum, and derive its radial velocity data over time. A hot Jupiter like exoplanet, such as 51 Pegasi b, induces a radial velocity variation on its host star of about 50 ms^{-1}.

The following equation relates the star's radial velocity and the mass of its planetary companion (Lovis & Fischer et al. 2011):

$$K_* = \sqrt{\frac{G}{1-e^2}} M_p \sin i M_*^{\frac{1}{2}} a^{\frac{1}{2}} \tag{1.1}$$

where the radial velocity semi amplitude K_* is the average of the minimum and maximum radial velocities, G is the gravitational constant, e is the exoplanet's orbital eccentricity, M_p and M_* are the exoplanet's and the star's masses, and a is the semi-major axis of the exoplanet's orbit. Thus, in addition to characterizing the basic shape of an exoplanet's orbit (e and a), the radial velocity technique gives the minimum mass of the planet $M_p \sin i$. For the true mass to be measured, the inclination angle, i is needed, which can be obtained from transit observations if the planet transits its host star. The radial velocity method is most sensitive when the system is seen edge-on as the radial gravitational tug then appears stronger.

1.2.3 Transit Method

This method detects the passage of a planet in front of its host star. This event is called a transit. The passage of the planet behind its host star is called an occultation or a secondary eclipse. Jupiter creates a transit of 1% depth in front of the Sun, and the Earth create a depth of 0.08%. Analysis of the shape, duration and depth of the transit characterizes several exoplanetary parameters. Assuming the stellar radius is known, the exoplanet's radius can be determined from δ, the depth of the transit, using the following equation:

$$\delta = \frac{F_* - F_{transit}}{F_*} = (\frac{R_P}{R_*})^2 \tag{1.2}$$

where $F_{transit}$ and F_* are the in and out-of-transit fluxes, and R_p and R_* are the radii of the planet and star respectively (Seager & Mallen-Ornelas et al. 2003). The exoplanet's orbital eccentricity e, exoplanet's orbital radius a and exoplanet's orbital inclination i are characterized from the duration and shape of the transit (cf. Seager & Mallen-Ornelas 2003; Winn 2011).The transit method is particularly important because it is the only one providing a measurement of the dimension of the planet. Moreover, combined with the radial velocity measurements that provide the mass of the planet, it is possible to deduce the planetary mean density.

One disadvantage of the transit discovery method is that planets can only be detected if a chance alignment of the orbit causes the planet to cross the face of its host star from the perspective of the observer. For a circular orbit with random orientation and inclination the transit probability p is given by:

$$p = \frac{R_*}{a} \tag{1.3}$$

Where a is the semimajor axis of the planet's orbit, and R_* is the radius of the star. A general formula for the transit probability of a planet in an eccentric orbit is (Winn, 2014; Kipping, 2014), where e is the eccentricity.

$$p = \frac{R_*}{a} \times \frac{1}{1 - e^2} \tag{1.4}$$

The probability to observe a planet like the earth is about 0.47%. Even assuming that the alignment is perfect for the planets like Jupiter the transit probability starts to fall 0.086% in this case.

1.3 Basic Properties of Exoplanet Systems

Figure 1.1 and 1.2 displays the ranges in planet mass, radius and orbital periods of the exoplanets detected by Nov 29th 2018, for each of the detection methods. Some features in the plots of the detected exoplanets stand out. The planets detected via the transit method are clustered around shorter orbital periods(<100 days); this is actually a bias in the transit method as planets with shorter orbital periods are closer to the star and hence have a higher probability to transit. Most of the exoplanets detected by radial velocity measurements have masses and radii similar to Jupiter; this is because the perturbations caused by massive planets on their host stars are larger and easier to detect. With the first determination of planetary masses and radii, a discrepancy between the observation and theory arose. A significant fraction of the transiting exoplanets has larger radii than predicted by the theoretical models. Figure 1.3 displays the relationship between density and mass and is a very useful because from the radius and mass of the planet we can derive density and study of the density further lets us model interior structure of the planets. Figure 1.4 displays eccentricities of exoplanets discovered till date ranging from 0 to 0.97 (HD 20782b), while in our Solar System the maximum planet orbital eccentricity is 0.2 (Mercury).

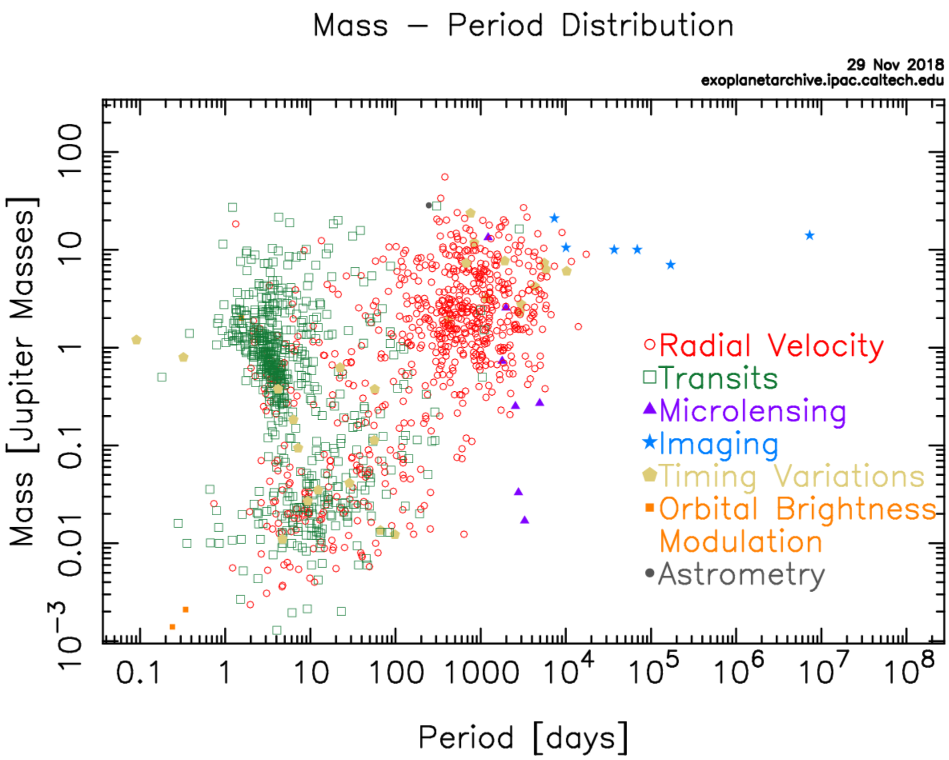


Figure 1.1: Plot showing mass vs orbital period relationship for exoplanets discovered till date. The colors/shape correspond to the method used to discover them. Most of the data points are clustered at around $1M_J$ but the red circles are clustered at longer orbital periods and the green boxes are centered at shorter orbital periods. Credit : NASA Exoplanet Archive

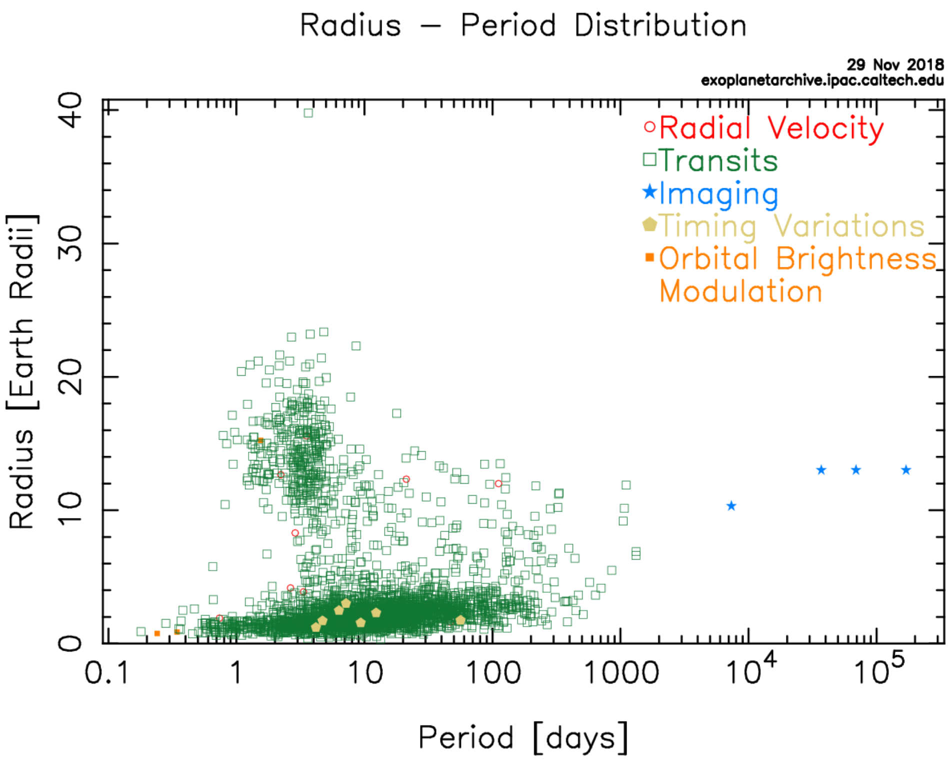


Figure 1.2: Plot showing mass vs radius relationship for exoplanets discovered till date. The colors/shape correspond to the method used to discover them.
Credit : NASA Exoplanet Archive

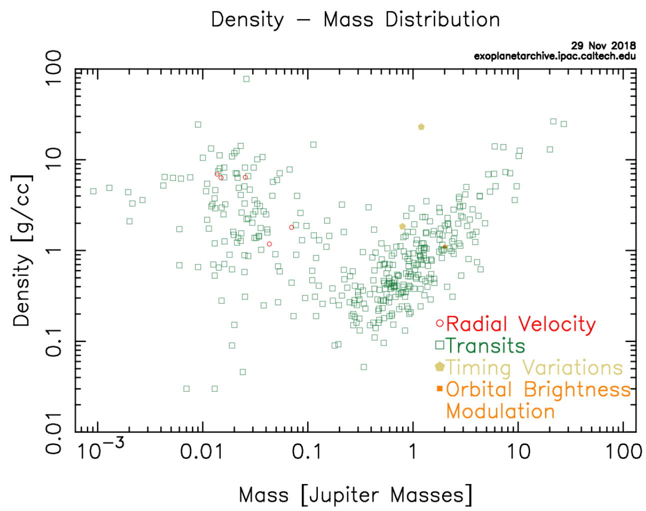


Figure 1.3: Plot showing density vs mass relationship for exoplanets discovered till date. The colors/shape correspond to the method used to discover them.
Credit : NASA Exoplanet Archive

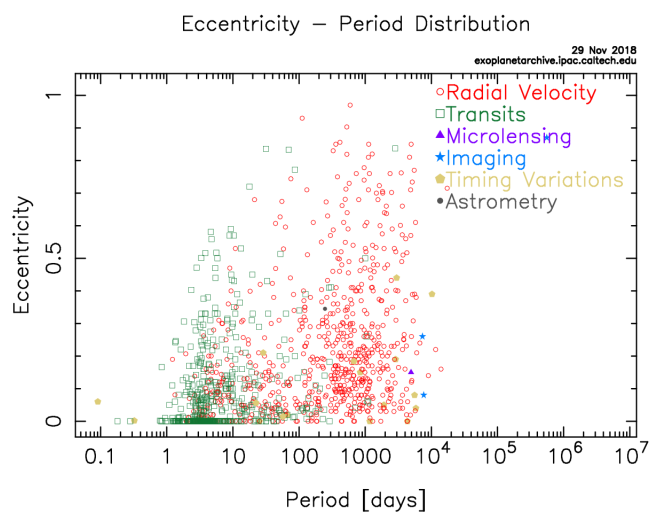


Figure 1.4: Plot showing eccentricity vs orbital period relationship for exoplanets discovered till date. The colors/shape correspond to the method used to discover them.
Credit : NASA Exoplanet Archive

Chapter 2

Transit Ephemeris Refinement and Monitoring Survey

2.1 The exoplanet orbit

In this section, parameters which are used to describe the orientation of an exoplanet orbit relative to the observer and the position of the planet are explained. A representation is given in Figure 2.1. The sky is the plane perpendicular to the line-of-sight to where the star lies. The two points where the planet's orbit crosses the sky are called nodes. In particular, the ascending node is where the planet is moving to the observer and the descending node is where the planet is moving away from the observer. The longitude of the ascending node, Ω, is the angle between the ascending node and a reference line in the sky. The inclination, i, is the smaller of the two angles between the orbital plane and the sky ($0 \leq i \leq \frac{\pi}{2}$). The argument of periapsis, ω, is the angle, in the orbital plane, between the ascending node and the periapsis. The three angles, Ω, i and ω, are used to fully describe the orientation of the orbit. The true anomaly, f, is the angular coordinate of the planet along the orbit, referred to the periapsis. Inferior conjunction is the instant when $f = \frac{\pi}{2} - \omega$. Superior conjunction is the opposite point, i.e. when $f = \frac{3\pi}{2} - \omega$.

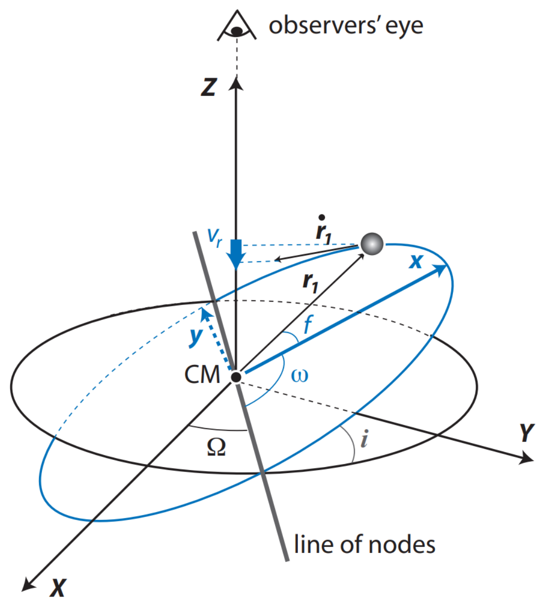


Figure 2.1: Orbital elements of an exoplanet orbiting a star. Figure from Murray and Correia (2011).

2.2 TERMS

The Transit Ephemeris Refinement and Monitoring Survey (TERMS) conducts radial velocity and photometric monitoring of known exoplanets in order to refine

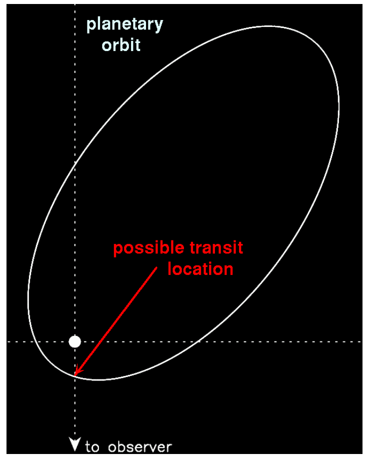


Figure 2.2: Figure depicting possible transit location from the line of sight. Credit : http://stephenkane.net/terms/ephemerides.html

planetary orbits and predictions of possible transit times. This effort is primarily directed towards planets not known to transit, but a small sample of our targets consist of known transiting systems.

In order to attempt to observe a transit of a known exoplanet, one needs to know precisely where and when to look. The limiting factor for successfully observing a known exoplanet host star during the predicted transit time is often the precision of the transit ephemeris. Particularly for intermediate-long period planets, this predicted time is often only roughly known. The quality of a calculated transit ephemeris is primarily determined by the uncertainties associated with the fitted orbital parameters, and the time elapsed since the most recent radial velocity data was acquired. Acquiring new high precision radial velocity data can easily mitigate both effects. With a prompt photometric observing strategy after the orbital parameters have been revised, one can maximize the chances of being able to obtain complete coverage of the transit observing window and thus either confirm or rule out the transiting nature of the planet.

2.3 Transit windows

The transit window is described here is defined as a specific time period during which a complete transit (including ingress and egress) could occur for a specified planet. The orbital parameters measured from fitting the radial velocity data of a planet are sufficient for calculating a transit ephemeris. The predicted time of midtransit can be calculated by using Kepler's equations. Firstly, the eccentric anomaly is calculated from the following

$$E = 2\tan^{-1}\left(\sqrt{\frac{1-e}{1+e}}\tan\frac{f}{2}\right) \tag{2.1}$$

where e is the orbital eccentricity and f is the true anomaly. As described in Kane (2007), the time of transit midpoint will occur when $\omega + f = \frac{\pi}{2}$, where ω is the argument of periastron. Substituting this for the true anomaly in equation (2.1) thus yields the eccentric anomaly at the point of predicted transit. The mean anomaly, M, which defines the time since last periapsis in units of radians, is then computed by

$$M = E - e\sin E \tag{2.2}$$

which can be converted to regular time units using

$$t_M = \frac{PM}{2\pi} \tag{2.3}$$

where P is the orbital period. The predicted midpoint of primary transit can then be calculated using

$$t_{mid} = t_p + \frac{PM}{2\pi} + nP \tag{2.4}$$

where t_p is the time of periastron passage and the term of nP incorporates the number of complete orbits which have transpired since t_p. The uncertainties in the orbital parameters (assuming they are symmetrical) can be propagated through these equations to determine the uncertainty in the predicted transit midpoint δt_{mid}, and the size of the transit window, t_{win}. The size of a transit window is mostly dependent upon the uncertainty in the period and the time elapsed since last observations were acquired. Thus, the beginning and end of a transit window are calculated by subtracting and adding (respectively) the uncertainties in t_p and P with respect to the transit midpoint, taking into account the number of orbits since periastron passage and the transit duration. The beginning of a particular transit window can be approximated by

$$t_{begin} = (t_p - \delta t_p) + (P - \delta P)\frac{M}{2\pi} + n(P - \delta P) - \frac{t_d}{2} \tag{2.5}$$

Where δt_p and δP are the uncertainties in t_p and P respectively, and t_d is the transit duration. Conversely, the end of the transit window is approximated by

$$t_{end} = (t_p + \delta t_p) + (P + \delta P)\frac{M}{2\pi} + n(P + \delta P) - \frac{t_d}{2} \tag{2.6}$$

Hence, the size of a given transit window is defined by subtracting equation (2.5) from equation (2.6), resulting in

$$t_{win} = 2(\delta t + \delta P\frac{M}{2\pi} + n\delta P) + t_d \tag{2.7}$$

which reduces to simply the transit duration as the uncertainties in P and t_p approach zero. From the equation we can see that δP has the potential to rapidly dominate the size of transit window, if n the number of orbits since discovery is large. We can rewrite equation 2.7 and determine the period uncertainty needed to achieve a certain transit window for a fixed δt_p and t_d.

$$\delta P = \frac{\pi(t_{win} - t_d - 2\delta t_p)}{M + 2\pi n} \tag{2.8}$$

These equations serve as first-order approximations which ignore the uncertainties in the orbital parameters of eccentricity and argument of periastron and instead focus on the time domain parameters of period and time of periastron passage. However, the equations also serve to overestimate the size of the transit window (the conservative approach) by assuming that the orbital inclination is edge-on compared with the line of sight. The consequence of this is that the maximum transit duration is allowed for (Kane et al. 2009).

2.4 Why Hipparcos photometry data?

The TERMS project aims at calculating the transit ephemeris of known RV planets. If we plot the histogram of the magnitudes of the exoplanet host stars discovered by the RV and Transit method it becomes clear.

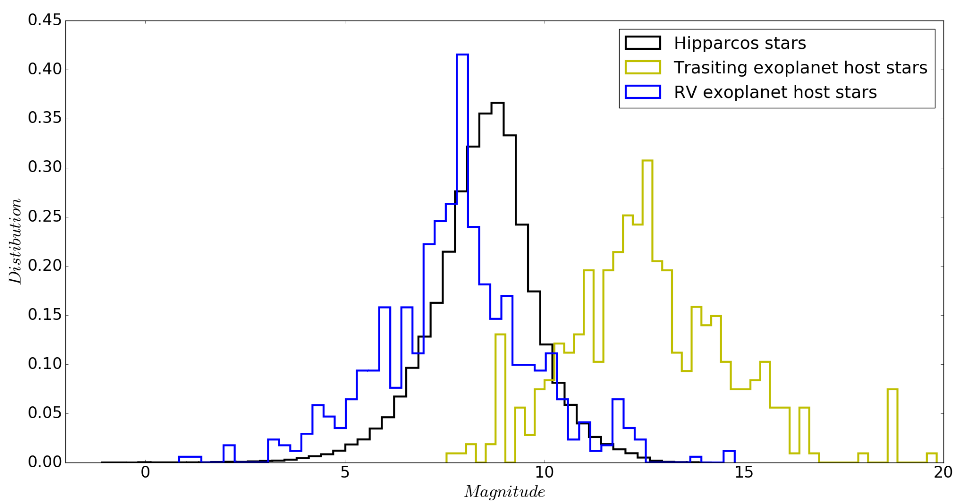


Figure 2.3: Distribution of V mag for Transiting and RV exoplanets plotted over the H_p mag distribution of Hipparcos stars

The peak of the Hipparcos distribution is 8.47 mag, while the RV technique is biased towards brighter stars with a peak V mag value of 7.80 and the transit method is biased towards fainter stars with a peak V mag value of 12.65. The Hipparcos distribution falls somewhere inbetween the RV and Transiting exoplanet host stars and is a good data set to investigate using the TERMS data.

Chapter 3

Data Analysis

3.1 Hipparcos data collection and description

All the Hipparcos photometry data is hosted in the NASA exoplanet archive, to extract them one by one, a script was written and it was able to extract all the Hipparcos photometry files from the archive. Once all the files were extracted it was sorted into planet hosting and non planet hosting Hipparcos star. The extracted data contained photometry files for 118170 Hipparcos stars of which 441 are planet hosting stars.

A sample Hipparcos data file for HIP 98505/HD 189733 is shown below:

```
NSTED INTERNAL KEYWORDS (NEEDED FOR TIME SERIES VIEWER)
X_AXIS                             = "BJD"
X_AXIS_UNITS                       = "days"
Y_AXIS                             = "Magnitude"
Y_AXIS_UNITS                       = "mag"
Y_AXIS_UNCERTAINTY                 = "Magnitude_Uncertainty"
Y_AXIS_UNCERTAINTY_UNITS           = "mag"
```

```
DESCRIPTION OF COLUMN HEADERS
COLUMN_BJD                          = "Barycentric Julian date of observation"
COLUMN_MAGNITUDE                    = "Magnitude in TIME_SERIES_DATA_FILTER"
COLUMN_MAGNITUDE_UNCERTAINTY        = "Uncertainty in MAGNITUDE"
COLUMN_DATA_QUALITY_FLAG            = "Data quality flag provided by HIPPARCOS project (9-bit integer, bit-packed)"
COLUMN_ACCEPTED                     = "1 if value was used by NStED in computing statistics; 0 otherwise"
|           BJD | Magnitude | Magnitude_Uncertainty | Data_Quality_Flag | Accepted |
|        double |    double |                double |               int |      int |
|          days |       mag |                   mag |                   |          |
  2447960.58338      7.8147                   0.008                   0          1
  2447960.65795      7.8252                   0.013                   2          1
  2447960.67227      7.8243                   0.012                   0          1
  2447960.74682      7.8177                   0.009                   0          1
  2447960.76117      7.7904                   0.010                   0          1
  2447960.83571      7.8269                   0.010                   0          1
```

From the sample data we can extract a lot of information of interest. The time stamp and the magnitude are especially useful while searching for transits. We only make use of data points that are flagged as accepted and do not consider any data row that have been flagged as corrupt or unaccepted.

3.2 Hipparcos data analysis and plots

To get a broad idea of how the data is distributed in the extracted files, multiples plots were created to understand the data better.

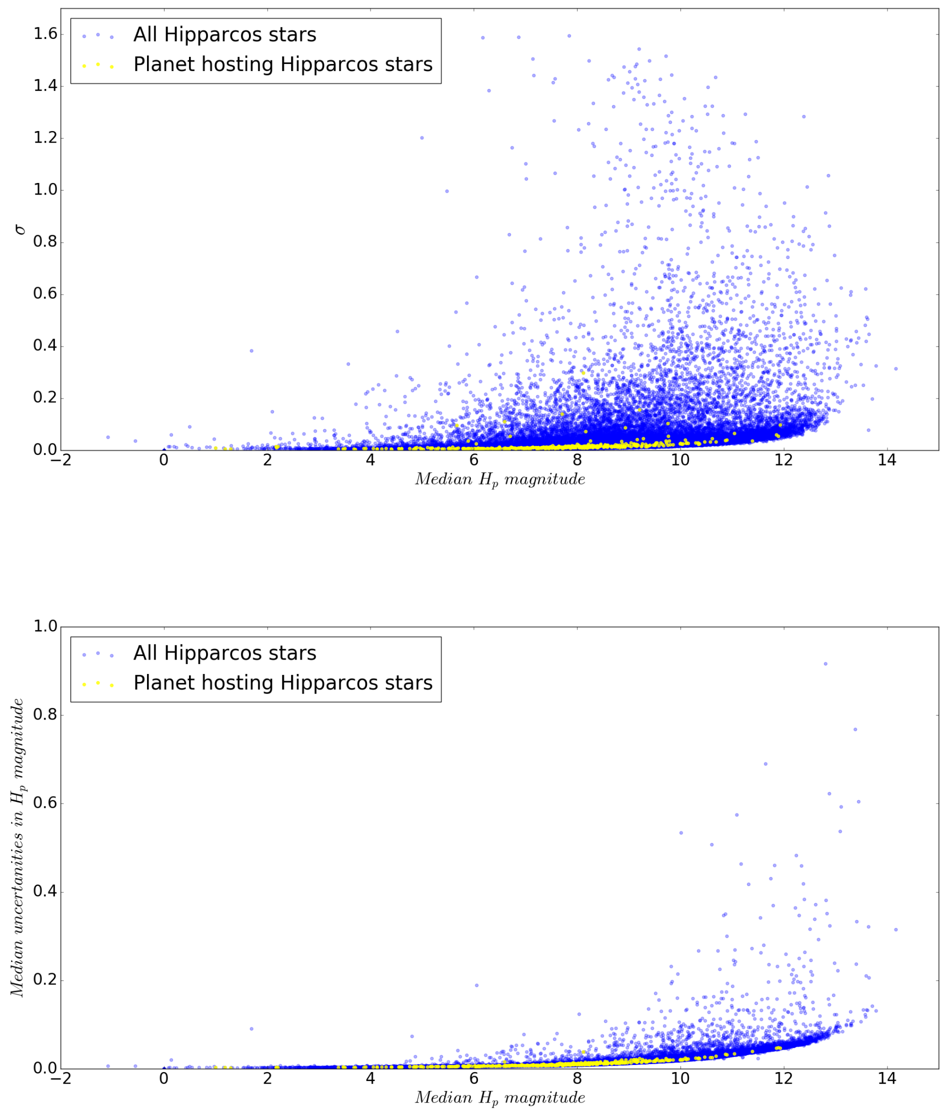


Figure 3.1: *Top* : The standard deviation and *Bottom* : variation of median uncertainty in each of the Hipparcos stellar photometry data with respect to the median Hipparcos magnitude. Blue points are all the Hipparcos stars and yellow are planet hosting Hipparcos stars

The one main thing noticeable in both the plots is that the standard deviation and the median uncertainty starts to increase with the increase in magnitude of the star. This shows that the instrument is less sensitive towards fainter stars in the catalog. Fortunately in our case we are interested in exoplanets discovered using the radial velocity technique which itself works well for bright stars. Most of our targets we are interested in are below 10th magnitude, the cutoff where the data stars to become more unreliable. To get a better idea of how the magnitude data is distributed for planet hosting Hipparcos stars as compared to the entire catalog of Hipparcos stars will give us a better insight.

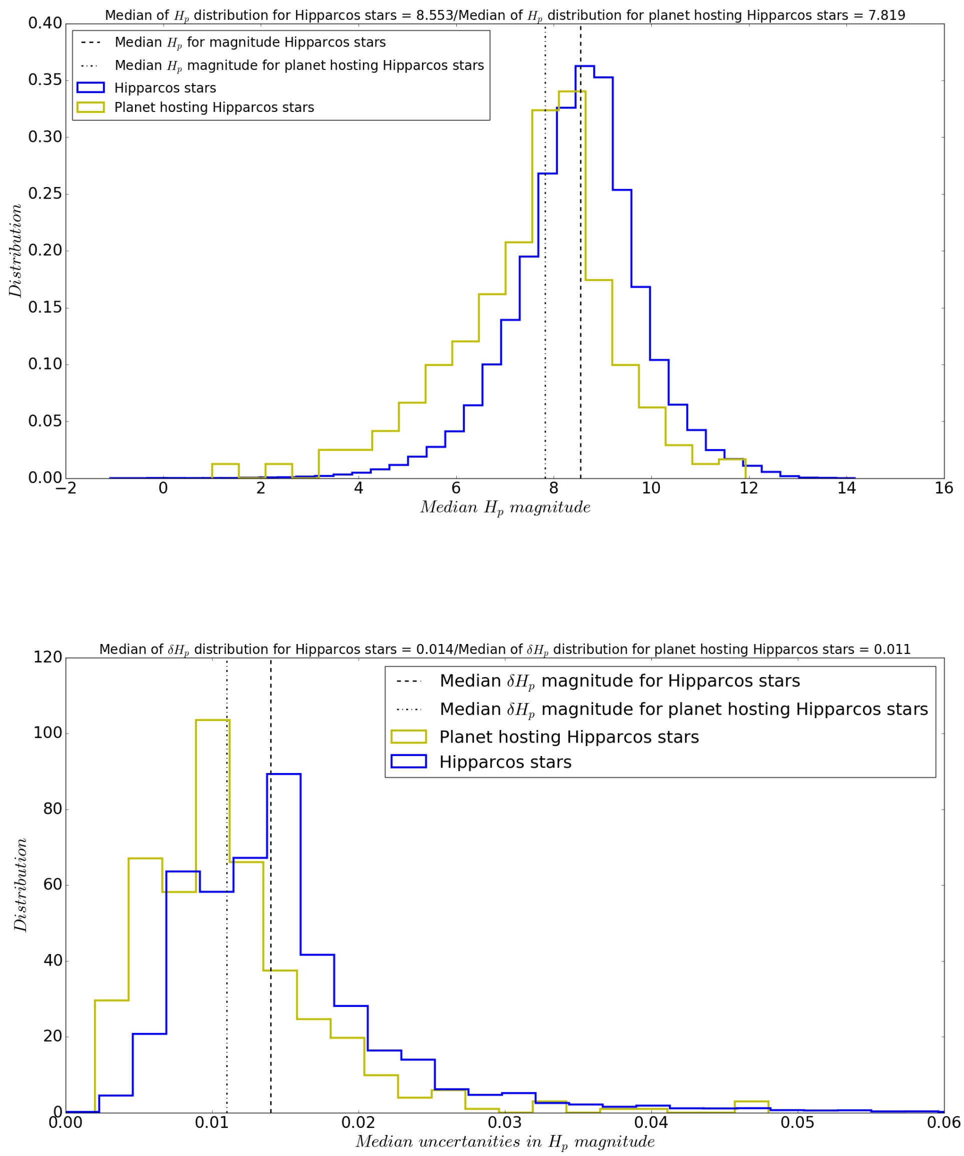


Figure 3.2: *Top* : the distribution of median Hipparcos magnitude. *Bottom* : distribution of median uncertainties in the Hipparcos magnitude data. All the Hipparcos stars are blue and planet hosting Hipparcos stars are yellow. We can clearly see that the planet hosting Hipparcos stars are brighter and have a less median uncertainty overall.

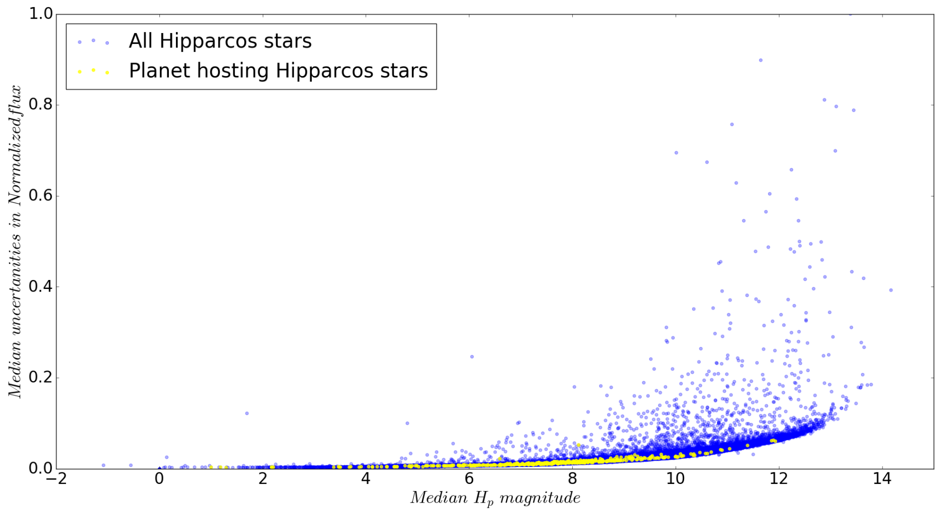
The median of the magnitude and uncertainty distribution of all the Hipparcos stars is 8.552 and 0.014 receptively, where as the median of the magnitude and uncertainty distribution of planet hosting Hipparcos stars is 7.819 and 0.011 respectively. Hence we can conclude that the planets found around stars using the radial velocity technique in the Hipparcos catalog tend to be brighter compared to the entire catalog of Hipparcos stars.

For the purpose of our analysis we are converting the magnitude data into a normalized flux using the following equation:

$$F = 10^{\frac{H_p}{-2.5}} \tag{3.1}$$

$$\Delta F_i = \frac{F_{Median} - F_i}{F_{Median}} \tag{3.2}$$

where F is the flux, ΔF is the normalized flux, H_p is the Hipparcos magnitude, F_{Median} is the median flux of a specific stellar target from the data and F_i is an individual data measurement.



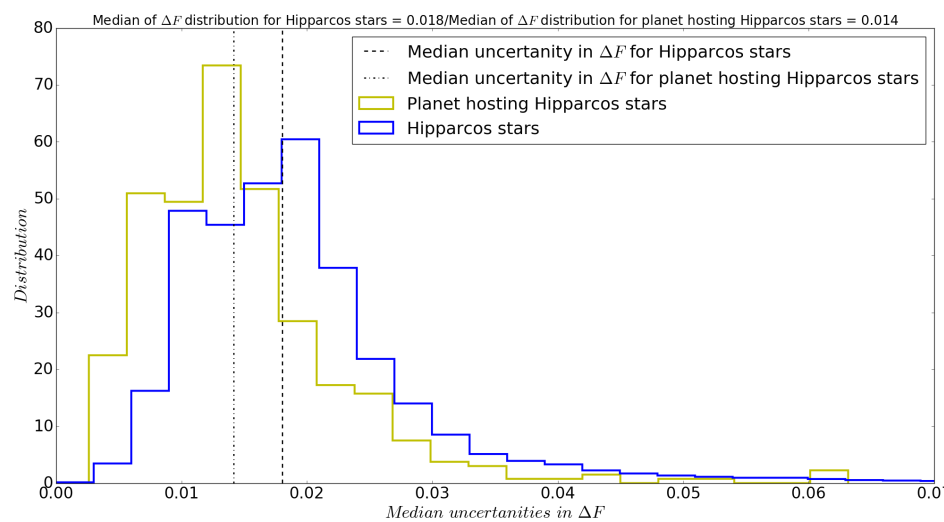


Figure 3.3: *Top* : the distribution of median normalized flux. *Bottom* : distribution of median uncertainties in the normalized flux data. All the Hipparcos stars are blue and planet hosting Hipparcos stars are yellow. We can clearly see that the planet hosting Hipparcos stars have less uncertainties in the normalized flux.

3.3 TERMS data

The TERMS data is hosted on a private server. It is a collection of transit epochs with the set transit window and also contains information on the parameters measured from fitting the radial velocity data . A sample TERMS data file for HIP 98505/HD 189733 is shown below:

```
PlanetName      > HD189733_b        Period          > 2.218576
dPeriod         > 0.000000         TimePeriastron  > 2454279.436714
dTimePeriastron > 0.000015         Eccentricity    > 0.000000
dEccentricity   > 0.000000         ArgPeriastron   > 90.000000
dArgPeriastron  > 0.000000         SemiMajorAxis   > 0.030995
dSemiMajorAxis  > 0.000615         PlanetMass      > 1.140390
dPlanetMass     > 0.056249         PlanetRadius    > 1.138000
dPlanetRadius   > 0.027000         NumberOfPlanets > 1
Transit         > 1                OrbitSource     > Bouchy 2005
StarMass        > 0.806000         StarRadius      > 0.756000
Vmag            > 7.670000         Teff            > 5040.000000
logg            > 4.587000         RA              > 20 00 43.71
Dec             > +22 42 41.26

 Predicted Transit Epochs: HD189733_ b
 Predicted Duration of a Central Transit:   130. Minutes
 Begin Transit Window           PREDICTED CENTRAL TRANSIT     End Transit Window
                                     All Times UT
                               HJD        Year M  D  H  M
 2454281.61022 2007  6 30  2 38   2454281.65529 2007  6 30  3 43   2454281.70036 2007  6 30  4 48
 2454283.82880 2007  7  2  7 53   2454283.87387 2007  7  2  8 58   2454283.91894 2007  7  2 10  3
 2454286.04737 2007  7  4 13  8   2454286.09244 2007  7  4 14 13   2454286.13751 2007  7  4 15 18
 2454288.26595 2007  7  6 18 22   2454288.31102 2007  7  6 19 27   2454288.35609 2007  7  6 20 32
 2454290.48452 2007  7  8 23 37   2454290.52959 2007  7  9  0 42   2454290.57466 2007  7  9  1 47
 2454292.70310 2007  7 11  4 52   2454292.74817 2007  7 11  5 57   2454292.79324 2007  7 11  7  2
```

These transit points are calculated from the parameters measured from fitting the radial velocity data. The equations used to calculate the transits points have been discussed in section 2.3.

Chapter 4

Transit point detection technique

The key aim of this project was to come up with a technique and automate the transit search process. The flowchart shown in the next page describes this process in detail. The key purpose of coming up with such technique is that we can automate it for many different photometric datasets that have been collected for not only transit search purposes but also for non-transit search purposes. Hipparcos photometric data set is one such example, because the main purpose of the measurements was not to look for transit but it was for collection of astrometric data.

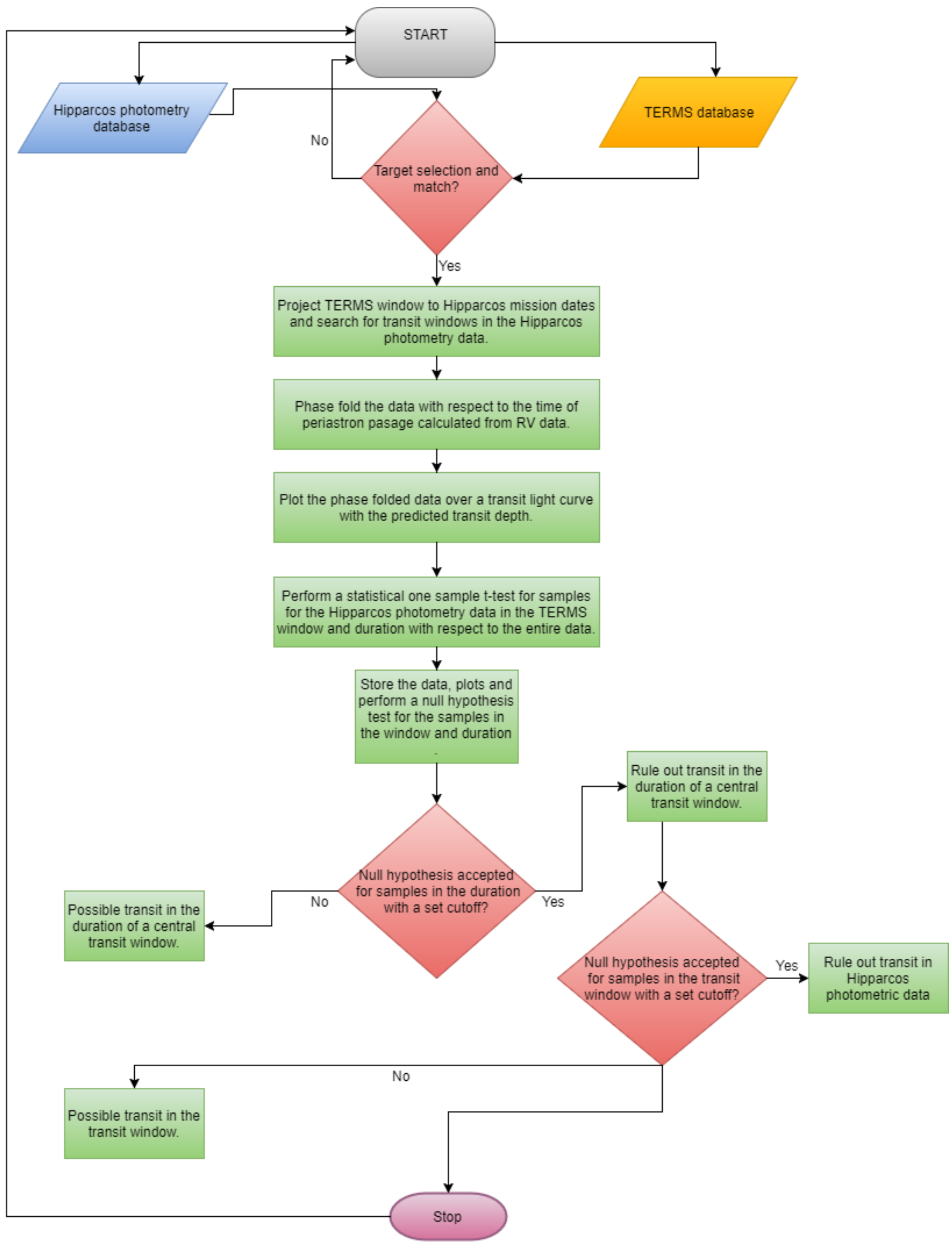


Figure 4.1: Flow chart describing the transit search method in the Hipparcos photometric data using the TERMS data

4.1 Target selection

For selecting targets to perform the analysis, we picked stellar targets belonging to the Henry Draper Catalogue (HD) of stars hosting only one planet. This decision was made because it was easy to match up Hipparcos star names with the HD star names from the TERMS data with the master catalog containing both the Hipparcos and HD names of the stellar targets.

Another criteria for target selection was the detectable predicted transit depth in the Hipparcos dataset. We set the cutoff for the transit depth at 1% because the mean uncertainty in the normalized flux is about 1.4%. We set the cutoff slightly below the mean uncertainty because a statistical analysis such as we are performing can reveal and exclude transits depending on the number of data points in the transit duration and window.

Using these criteria for filtering of the 441 planet hosting Hipparcos stars we end up with 97 potential targets for further analysis.

4.2 Window calculation and search

Once we know the TERMS window from the TERMS data, we can propagate the window backwards to the Hipparcos mission start date and then search for data points in that window.

4.3 Phase fold

To improve our data it is important to phase fold the the entire lightcurve to fit one cycle. For each photometry file the time between samples is irregular but we can fill in the gaps of data by folding the cycles of the planet orbiting the star over one another. This will overall improve the Signal to Noise ratio of our observation and define the peak of our primary transit more clearly.

$$\phi = \frac{t - t_p}{P} \tag{4.1}$$

where t is the time of each data point in the Hipparcos photometry file, t_p is the time of central transit and P is the period of the exoplanet which can be extracted from the TERMS data. Phase ϕ is the new coordinate that would be assigned to each data point. This way we can fit each data point in one period or one phase. With each cycle fit to the same phase our long light curve will turn into a phase-folded light curve. The phase folded light curve had been corrected to take in to account that at the time of central transit the planet is perfectly at inferior conjuncture with a phase of 0.

4.4 Hypothesis test

Going one step further after detecting the possible transit points, we perform hypothesis test with the samples in the TERMS transit duration and window with the entire Hipparcos photometric observation for that specific stellar target.

Lets define $\mu_{duration}$ as the sample mean of data points in the transit duration, μ_{window} as the sample mean of data points in the transit window and μ as the mean of the photometic measurements of a specific target.

- The null hypothesis (H_0) would be that the sample mean of data points in the transit duration or transit window will be equal to the mean of the photometic measurements of a specific target.

$$H_0 : \mu_{duration} = \mu$$

$$H_0 : \mu_{window} = \mu$$

- The alternate hypothesis (H_A) would be that the sample mean of data points in the transit duration or transit window will not be equal to the mean of the photometic measurements of a specific target.

$$H_A : \mu_{duration} \neq \mu$$

$$H_A : \mu_{window} \neq \mu$$

The t-test will signify how statistically different is a sample if picked from a set of observations. In our case its very useful because the t-statistic value and the p-value calculated from this test will tell us for sure if the transit points are statistically different. This technique is also termed as null hypothesis testing and the cutoff value (α) is set to 0.2 for validating the hypothesis. Using this powerful tool we can run hypothesis tests on all our selected targets and pick the most interesting ones based on the t-statistic value, p-value and the cutoff value. Depending on the results from the t-test we can classify the targets as interesting and not interesting.

4.5 Validation of algorithm

To validate if the method described above worked, we tested it on two test cases. The test cases were HIP 98505/HD 189733 and HIP 108859/HD 209458, both hosting transiting exoplanets with confirmed detection of transits in the Hipparcos photometric data.

4.5.1 HIP 98505/HD 189733

Bouchy et al. (2005) announced the detection of a 2.219$\pm$0.0005 day orbital period extra-solar planet that transits the disk of its parent star, the dwarf HD 189733. This detection was performed thanks to spectroscopic and photometric data collected at the Haute-Provence Observatory, France, as part of the ELODIE metallicity-biased search for transiting hot Jupiters (Da Silva et al.2005). HD 189733b is only the ninth known transiting extra-solar planet (Bouchy et al.2005), and the third transiting a star bright enough to be in the Hipparcos Catalogue (Perryman et al.1997). The Hipparcos Catalogue includes HD 189733 photometry for 176 accepted epochs with a median value of 7.82745±0.01462 and median uncertainty of 0.011 mag. Transits of HD 189733b in the Hipparcos photometric data was successfully reported in Bouchy et al. (2005) and G. Hebrard et al. (2005). The median value of 5 data points in the transit duration was 7.8576±0.02079. To see if there is a statistical significance of the data points in the transit duration the proposed analysis in figure 4.1 can be used. From the analysis we can see that the $t-value$ of the samples is 2.472, which means the mean of the sample from the population is fainter compared to the mean of the photometric data for HD 189733 from Hipparcos. With a $p-value$ of 0.068 its highly unlikely these samples are perfectly random. Results are summarized in table 4.1. The results of my analysis show that the transit points are in good agreement with the transit points detected in Bouchy et al. (2005) and G. Hebrard et al. (2005).

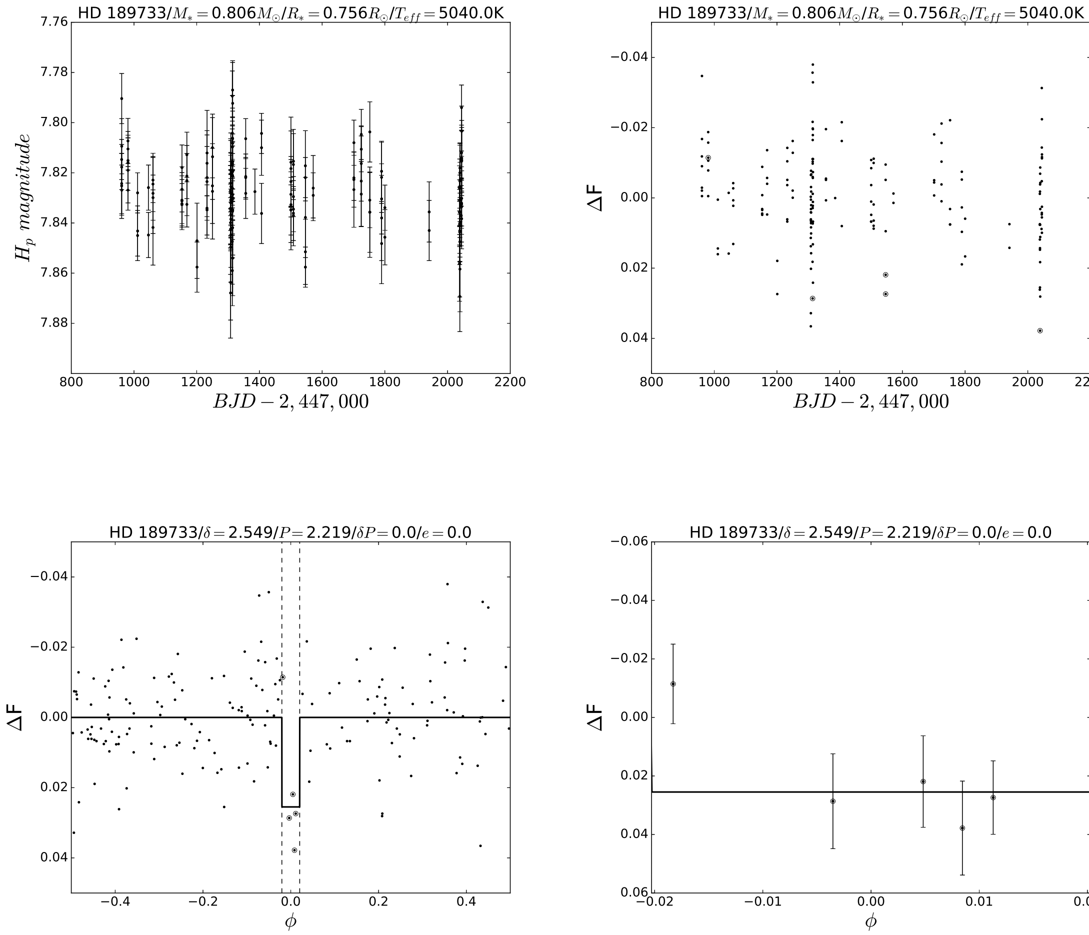


Figure 4.2: *Top left: Hipparcos* photometric measurements of HD 189733. *Top right:* Normalized *Hipparcos* photometric measurements of HD 189733. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 189733. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

4.5.2 HIP 108859/ HD 209458

The detection of HD 209458 b was simultaneously announced by Mazeh et al. (2000) and Henry et al. (2000) with an orbital period of 3.5250±0.003 days. HD 209458b is known to be a transiting extra-solar planet (Charbonneau et al. 2000). Spectroscopically, HD 209458 is similar to the Sun, with a $\log(R'_{HK}) = -5.0$ indicating low chromospheric activity (Henry et al. 2000). Hipparcos observed HD 209458 (HIP 108859) on 89 occasions, with a median value of 7.7719±0.01481 and an individual median standard error of 0.011 mag. Transits of HD 209458b in the Hipparcos photometric data was successfully reported in N. Robichon et al. (2000). The median value of 5 data points in the transit duration was 7.7979±0.0083. Performing the analysis the $t - value$ of the sample is 6.766, which is a lot of difference relative to the variation in the data and the samples are fainter then the star itself in that time window with a $p - value$ of 0.002 . The results of my analysis proposed in figure 4.1 shows that the transit points are in good agreement with the transit points detected in N. Robichon et al. (2000). The results for HD 209458b have been summarized in table 4.1.

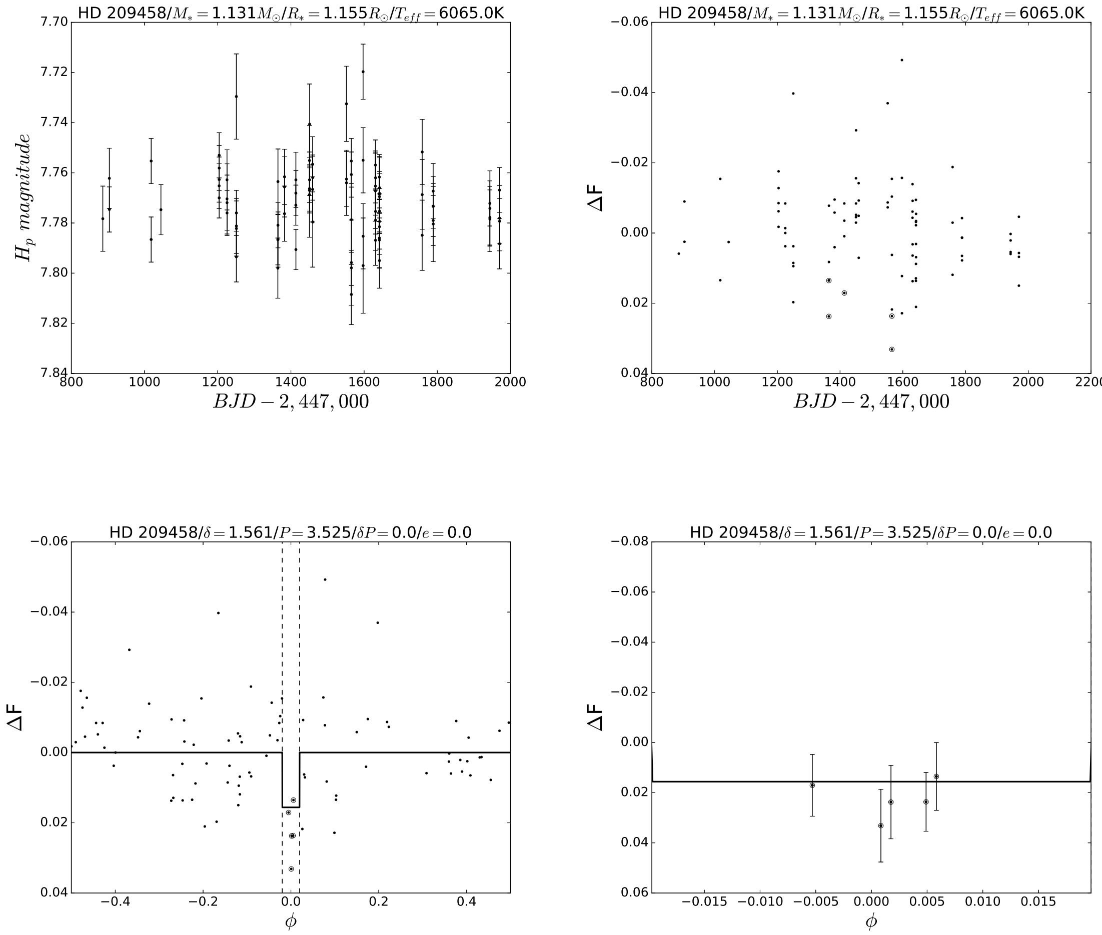


Figure 4.3: *Top left: Hipparcos* photometric measurements of HD 209458. *Top right:* Normalized *Hipparcos* photometric measurements of HD 209458. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 209458. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

4.6 Null hypothesis test for validation targets

The table below shows the results of the t-test for our validation/test cases HD 189733 and HD 209458.

Table 4.1: Data summary from analysis of HD 189733 and HD 209458

Target	t_d	t_w	N	N_d	N_w	t_{sd}	t_{sw}	p_d	p_w
HIP 98505/HD 189733	0.09	0.09	176	5	5	2.472	2.472	0.068	0.068
HIP 108859/ HD 209458	0.139	0.139	89	5	5	6.766	6.766	0.002	0.002

Where t_d is the transit duration, t_w is the transit window, N is the total number of observations, N_d is the number of observations is the transit duration, N_w is the number of observations is the transit window, t_{sd} is the t-statistic value calculated with the samples in the transit duration, t_{sw} is the t-statistic value calculated with the samples in the transit window, p_d is the p-value calculated with the samples in the transit duration and p_w is the p-value calculated with the samples in the transit window. Looking at the table we can clearly see that the t-statistic value is positive and the p-value is very low for both our validation targets, less then the cutoff $\alpha = .2$. This means that the samples that were marked as transit points do indeed differ from the entire photometric observations. Also an important feature to note here is that $t_{sd} = t_{sw}$ and $p_d = p_w$ because for these targets the δP is 0 and $t_d = t_w$.

4.7 Testing on the dataset

Running the analysis described above for the 97 targets, we can look at the data and observe some interesting behaviors. Of the 97 targets 51 targets have data points in the TERMS duration or TERMS window.

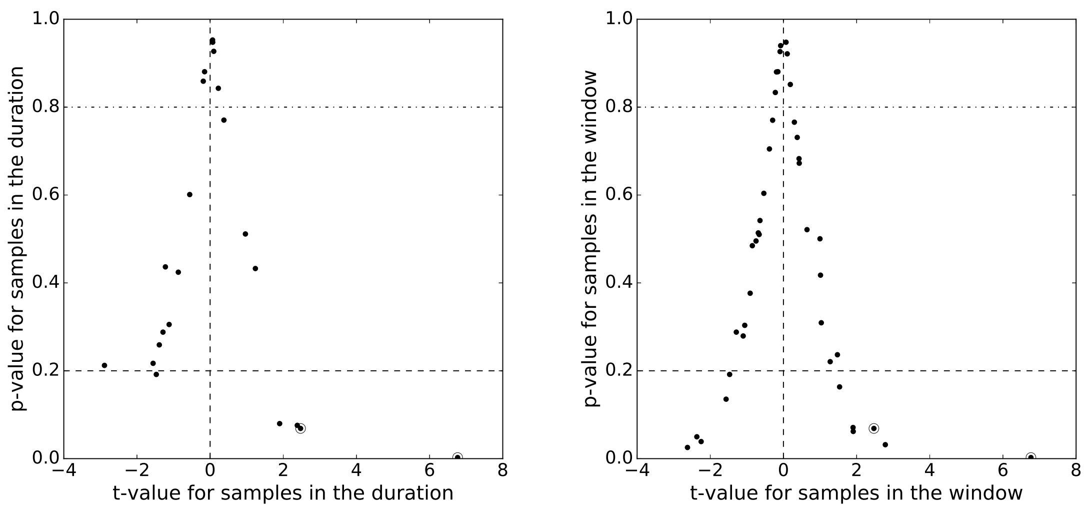


Figure 4.4: *Left* : The distribution of t-value and p-value calculated in the TERMS duration for the 51 targets. *Right* : The distribution of t-value and p-value calculated in the TERMS window for the 51 targets. The validation targets that are proven to transit in the Hipparcos data are marked with a circle. The horizontal dotted lines represent the cut off values we have set and the vertical dotted line is at $t - value = 0$.

Figure 4.4 makes it easier to come up with a set of criteria to classify targets based on the t-value, p-value and cutoff:

- If the targets have a $t - value > 0$ and $p - value < .2$ we can classify these targets as potential transit in the Hipparcos data during the TERMS duration/window.

- If the targets have a $t - value < 0$ and $p - value < .2$ we can classify these targets as potential transit rule out in the Hipparcos data during the TERMS duration/window.

- If the targets have a $p - value > .8$ we can classify these targets as statistically insignificant during the TERMS duration/window.

Chapter 5

Results and case studies

5.1 Potential transit in the Hipparcos data during the TERMS transit duration

These are cases classified as targets with potential transits in the Hipparcos data during the TERMS duration using the criteria we defined in section 4.7.

5.1.1 HIP 19207/ HD 285507

The detection of the first hot Jupiter in the Hyades open cluster, HD 285507b was reported by Quinn et al.(2014) with an orbital period of 6.0881±0.0018 days. The photometric monitoring of HD 285507 with KeplerCam on the FLWO 1.2 m telescope at the predicted time of conjunction on BJD 2456238.5 (UT 2012 November 7) and the resulting light curve showed no sign of a transit (Quinn et al.2014). The transit depth can be predicted by calculating the derived radius using the mass-radius-flux relation from Weiss et al.(2013), which is about 2.61% and $0.95R_J$ respectively. Hipparcos data contains 54 acceptable epochs from the 56 observations with a median value of 10.6178±0.09765 and a median uncertainty of 0.026. There are 5 data points in the Hipparcos transit duration with a median value of 10.6189±0.01498 calculated using the TERMS data. The Based on the results from the analysis summarized in table 5.1 these epochs are statistically significant with a $t_{sd} = 2.383$ and $p_d = 0.075$ which is less then out $\alpha = .2$. This could be transit signature in the Hipparcos data, but HD 285507 is a KV dwarf with $V = 10.47$ which is faint compared to the population of stars Hipparcos was observing and instrument detection limits can cause such variations in the data.

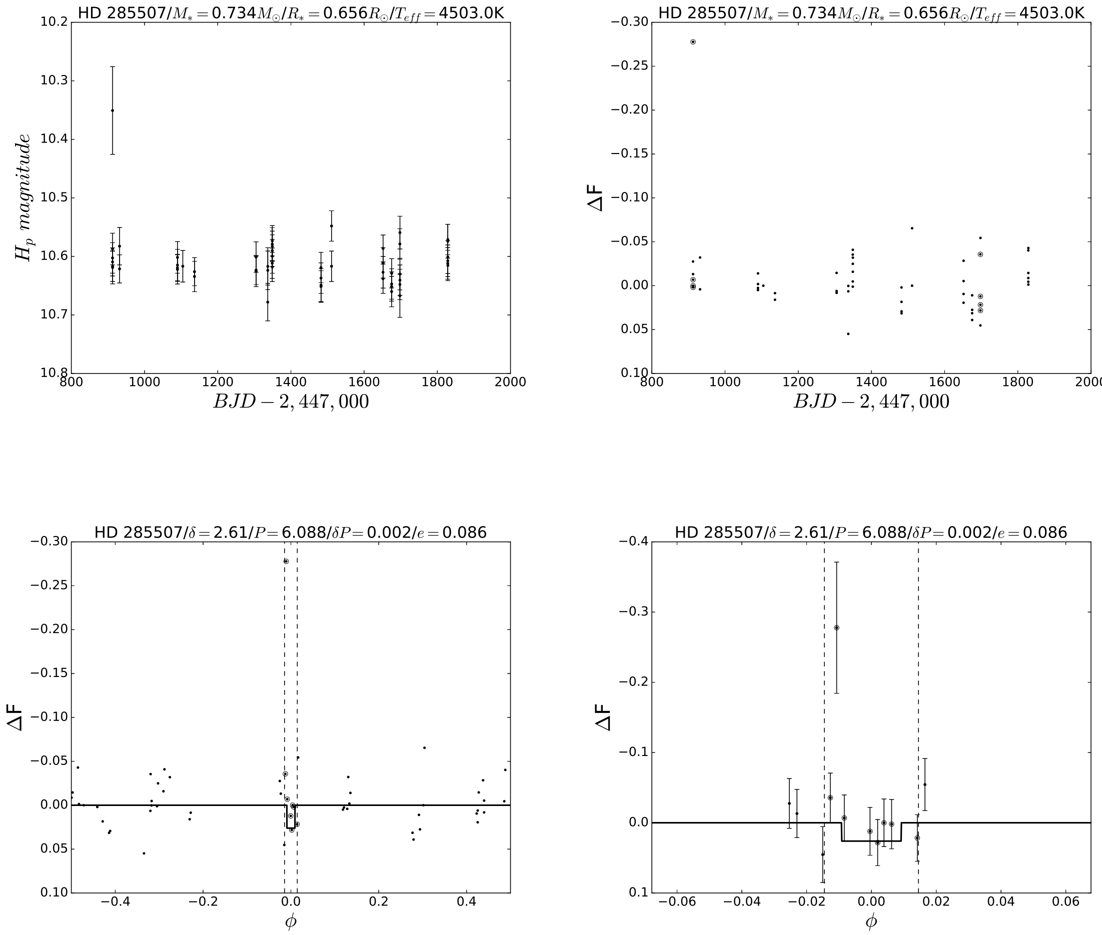


Figure 5.1: *Top left: Hipparcos* photometric measurements of HD 285507. *Top right:* Normalized *Hipparcos* photometric measurements of HD 285507. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 285507. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

5.1.2 HIP 6511/ HD 8535

The planet HD 8535b was discovered as a part of the High Accuracy Radial Velocity Planet Searcher (HARPS) for low-activity solar-type stars reported by Naef et al.(2010). HARPS is a high-resolution echelle spectrograph mounted on the ESO–3.6-m telescope in La Silla (Chile). HD 8535b has an orbital period of 1313±28 with a minimum mass of $0.68M_J$ and a semi-major axis of 2.45 AU. No follow up transit searches have been reported due to the planet's large orbital period. The host star exhibits a very low level of activity with a $\log(R'_{HK}) = -5.0$ (Naef et al.2010). This large P and δP has an impact on the TERMS transit duration and window calculations. The TERMS transit duration (t_d) for HD 8535b is about 0.767 days while the window (t_w) is about 365.1 days. The transit depth we could potentially observe is about 1%. We can use the already existing Hipparcos transit data and look for potential transits. Hipparcos mission has recorded brightness data for 187 epochs with median value 7.8192±0.01578 of which 14 data points are within t_d and 53 data points within t_w. As the window for the current target is very large, it makes sense to look at only the transit duration. The results of the analysis are summarized in table 5.1. The 14 epochs measured by Hipparcos are statistically significant with a median value of 7.82235±0.01053, $t_{sd} = 1.9$ and $p_d = 0.079$ which is less then out $\alpha = .2$. This could be transit signature in the Hipparcos data.

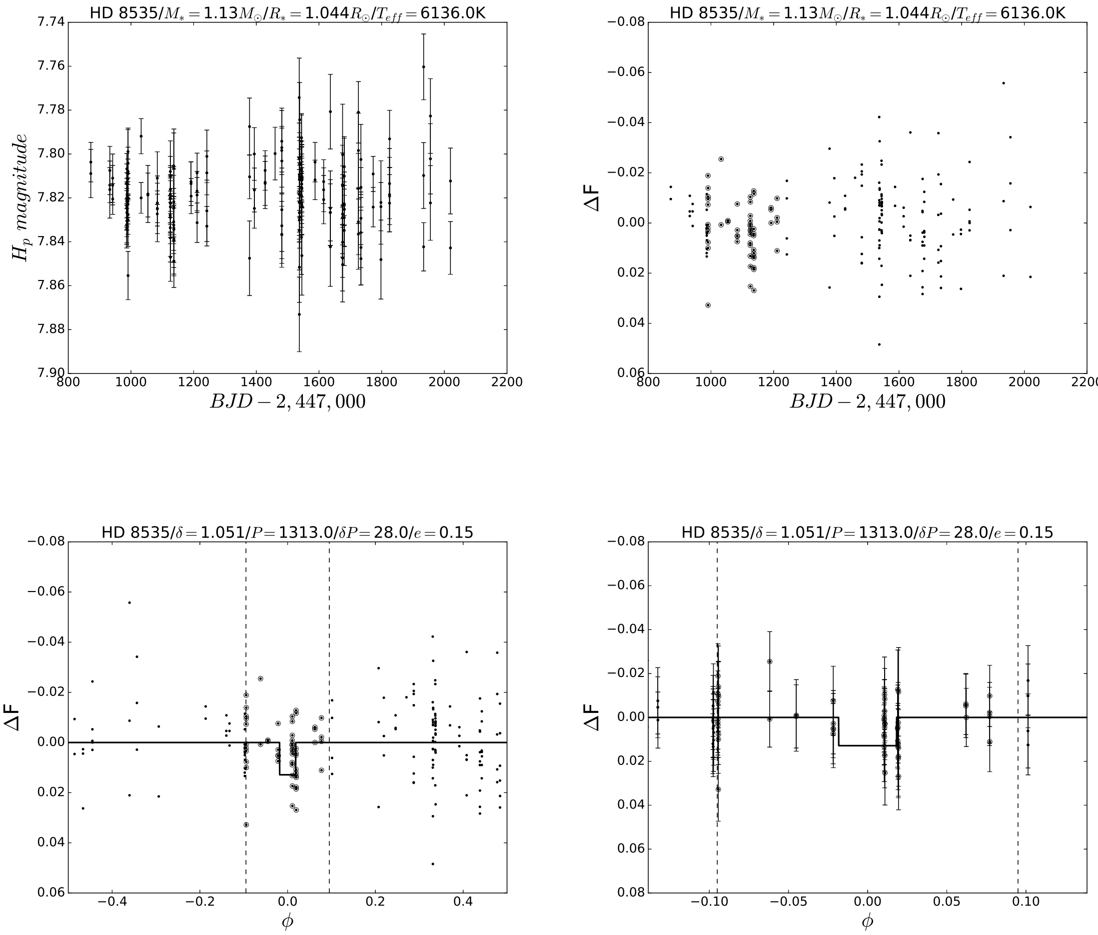


Figure 5.2: *Top left: Hipparcos* photometric measurements of HD 8535. *Top right:* Normalized *Hipparcos* photometric measurements of HD 8535. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 8535. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

The table below summarizes the results from our analysis.

Table 5.1: Data summary from analysis of HD 285507 and HD 8535

Target	t_d	t_w	N	N_d	N_w	t_{sd}	t_{sw}	p_d	p_w
HIP 19207/ HD 285507	0.112	0.824	54	5	8	2.383	-0.687	0.075	0.513
HIP 6511/ HD 8535	0.767	365.1	187	14	53	1.900	1.907	0.079	0.061

5.2 Potential transit rule out in the Hipparcos data during the TERMS transit duration

These are cases classified as targets with potential transit rule in the Hipparcos data during the TERMS duration using the criteria we defined in section 4.7.

5.2.1 HIP 38041/ HD 63765

The planet HD 63765b was discovered with the HARPS Echelle spectrograph mounted on the ESO 3.6-m telescope at La Silla as part of search for planets in orbit around the moderately active dwarfs. HD 63765b orbits a G9 dwarf at 0.94AU with an orbital period of 358.0±1.0 in a slightly eccentric orbit with eccentricity of 0.24, it has a minimum mass of $0.64 M_J$ reported by Ségransan et al. (2011). No transit searches were reported for this planet because of its long orbital period and low transit probability similar to that of the earth 0.43%. This one year Saturn mass like planet could have a theatrical predicted transit depth of 1.208%. The Hipparcos mission recorded stellar brightness data for 109 epochs and 7 data points are in the t_d and epochs extracted from the TERMS data. The median value of the data observed by Hipparcos is 8.2473±0.0141 with median value of the data points in the window 8.2406±0.0092. Performing the analysis described in figure 4.1 we can see that the 7 epochs observed are indeed statistically significant with a $t_{sd} = -1.471$ and $p_d = 0.191$ which is less then cutoff value $\alpha = .2$. The value of $t_{sd} < 0$ means that the difference relative to the variation of the sample data is negative, the star is getting brighter. In this unique case there are no data points in the transit window of this long period planet and $t_{sd} = t_{sw}$. As the star is getting brighter we can rule of the transit in the Hipparcos data. The reason for this might be due to the fact that HD 63765 is a moderately active G9V star with $\log(R'_{HK}) = -4.736 \pm 0.037$ and shows a large Ca II re-emission lines, which is an indication of significant chromospheric activity possibly induced by stellar spots or plagues (Ségransan et al. 2011).

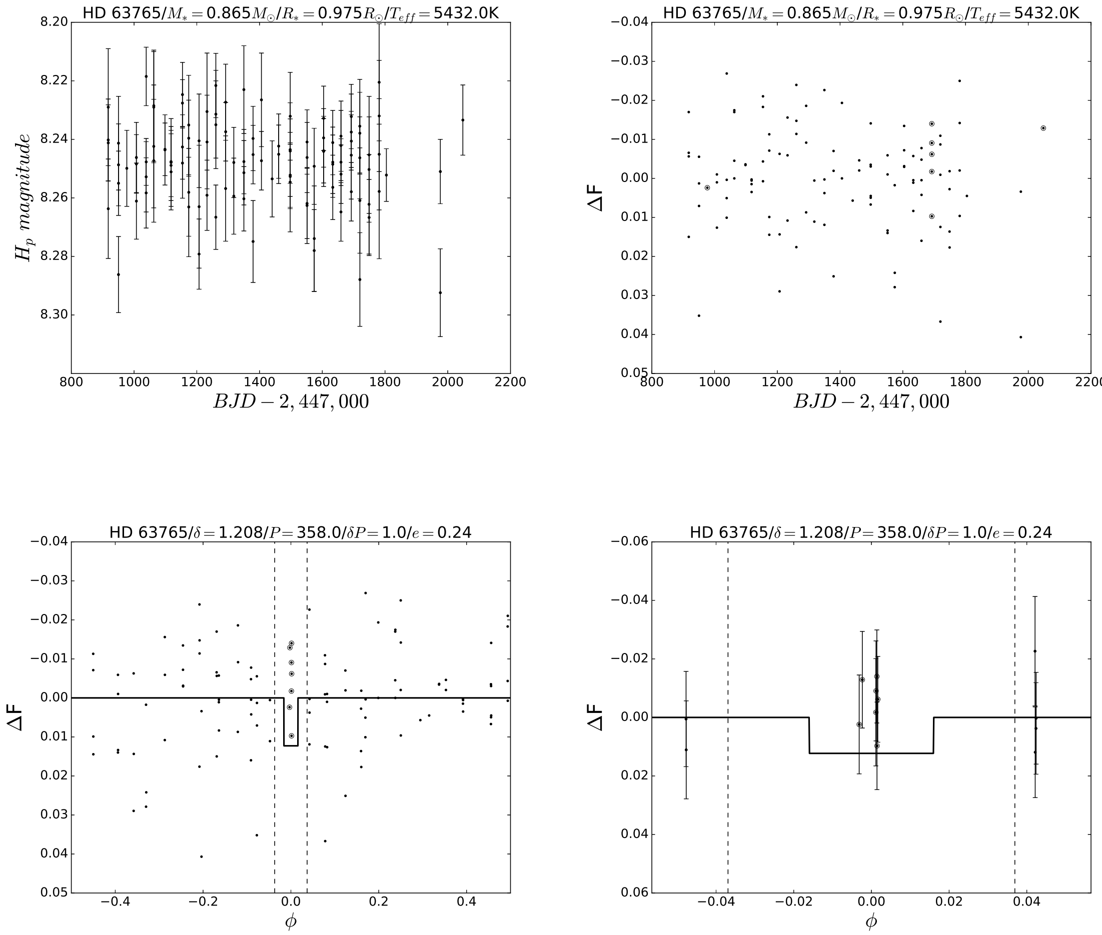


Figure 5.3: *Top left: Hipparcos* photometric measurements of HD 63765. *Top right:* Normalized *Hipparcos* photometric measurements of HD 63765. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 63765. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

The table below summarizes the results from our analysis.

Table 5.2: Data summary from analysis of HD 63765

Target	t_d	t_w	N	N_d	N_w	t_{sd}	t_{sw}	p_d	p_w
HIP 38041/ HD 63765	0.417	40.363	109	7	7	-1.471	-1.471	0.191	0.191

5.3 No significant change in TERMS duration

These are cases classified as targets with no statistical significance in the Hipparcos data during the TERMS duration using the criteria we defined in section 4.7.

5.3.1 HIP 97769/ HD 188015

The discovery of HD 188015b was reported in Marcy et al. (2005) from the radial velocity data of 975 FGKM dwarfs using the Keck 1 telescope over 6-7 year period. Later on the discovery of the planet and its derived properties were confirmed and corrected by Butler et al. (2006) three sources: observations at Lick Observatory using the Hamilton spectrograph (Vogt 1987), at Keck Observatory using HIRES (Vogt et al.1994), and at the 3.9 m Anglo-Australian Telescope (AAT) using UCLES (Diego et al. 1990). The planet has an orbital period of 461.2±1.7 with an eccentricity of 0.137. HD 188015b orbits the host star at a distance of 1.203 AU with a minimum mass of $1.50M_J$. HD 188015 is a G5IV star with $V = 8.24$. It is chromospherically quiet, with a $\log(R'_{HK}) = -5.05$ (Wright et al. 2004). No transit searches were reported due to the planet's large orbital period, but potentially it could have a transit depth of 1.07%. Hipparcos recorded data on this star for 210 epochs with median value of 8.3795±0.0167 of which 3 data points falling in the TERMS duration with a median value of 8.3665±0.0227. Performing the analysis shown in figure 4.1, we find that the data points in t_d have a $t_{sd} = 0.066$ and $p_d = .952$ which is above the $\alpha = 0.8$ cutoff for statistically insignificant. Interesting thing to note here is that the $t - value$ which signifies the sample variation from the data is close to 0. Hence we can classify this target as no change in TERMS duration.

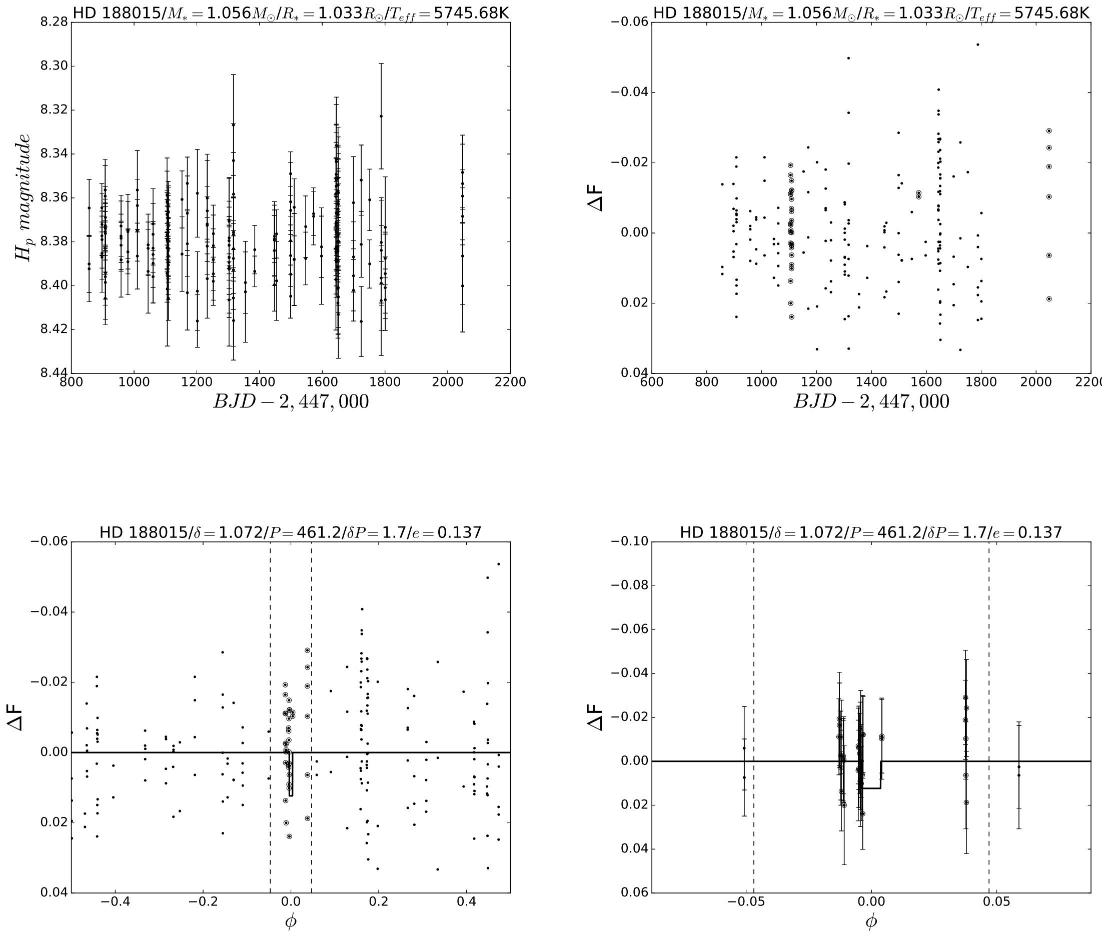


Figure 5.4: *Top left: Hipparcos* photometric measurements of HD 188015. *Top right:* Normalized *Hipparcos* photometric measurements of HD 188015. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 188015. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

5.3.2 HIP 25191/ HD 290327

HD 290327b was discovered as a part of HARPS search for planetary companions around southern low-activity solar-type stars. HD 290327 is a G8V star with a $\log(R'_{HK}) = -4.96$. The companion has an orbital period of 2443$\pm$161, an extreme long period planet. The minimum mass of the planet obtained from the RV data is $2.43M_J$ and orbits with a semi-major axis of 3.43 AU (Naef et al.2010). The potential transit depth we can observe for HD 290327b is 1.137%. No transit searches are reported for this planet due to its very large orbital period and the uncertainty in the calculated orbital period. Due the very large P and δP values we can expect the ratio of $\frac{t_d}{t_w} \approx 0$. In this case the calculated $t_w = 1128.301$ days, its just not practical to do any kind of followup observation. Hipparcos observed HD 290327 for 89 epochs with median value of 9.129$\pm$0.0202 of which 4 data points fall in the target duration with a median value of 9.1184$\pm$0.02998 with a $t_{sd} = 0.099$ and $p_d = 0.926$. The calculated $t-value$ from the analysis is close to 0 and the $p-value$ is above the cutoff value $\alpha = 0.8$. Hence we can classify the target as statistically insignificant in the transit duration.

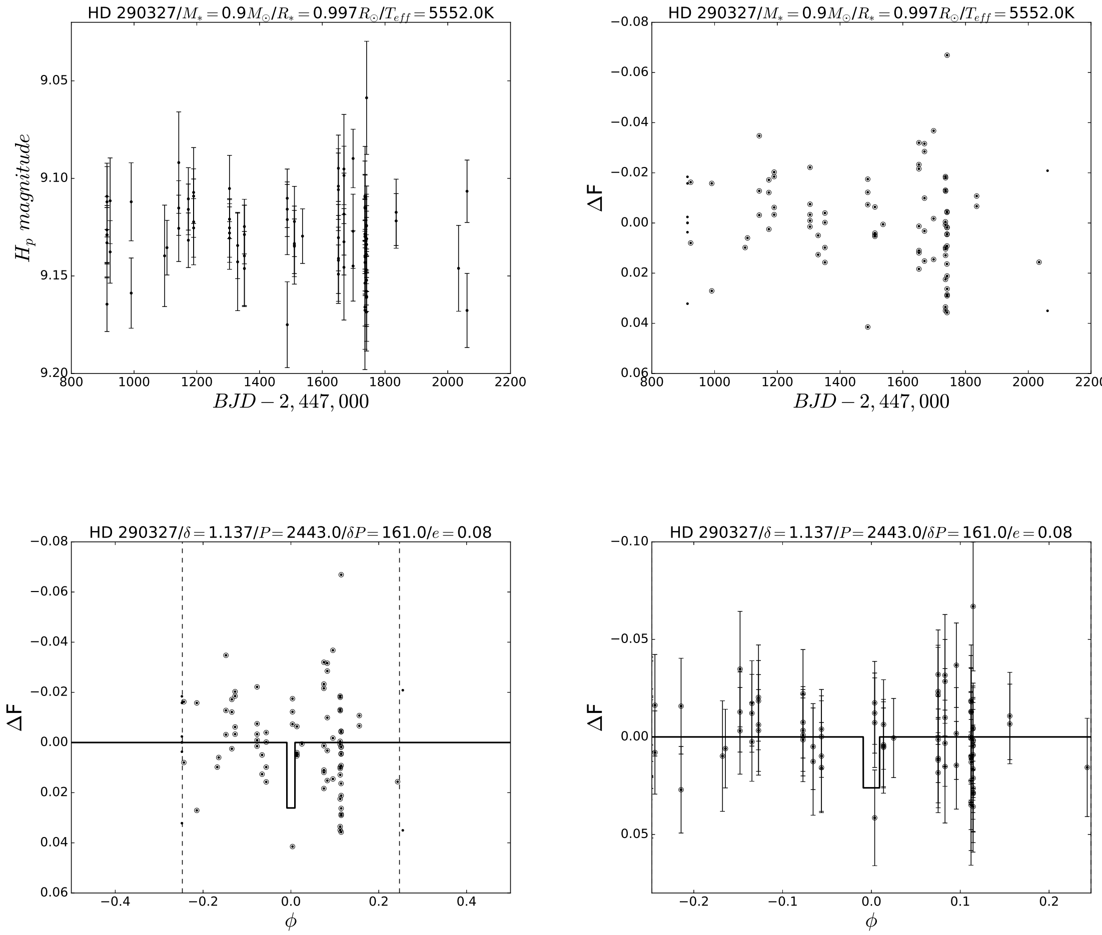


Figure 5.5: *Top left: Hipparcos* photometric measurements of HD 290327. *Top right:* Normalized *Hipparcos* photometric measurements of HD 290327. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 290327. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

5.3.3 HIP 39417/ HD 66428

The discovery of HD66428b was reported as part of an exoplanet catalog of 172 known low-mass companions with orbits established through radial velocity and transit measurements around stars within 200 pc. The radial velocity data here come from three sources: observations at Lick Observatory using the Hamilton spectrograph (Vogt 1987), at Keck Observatory using HIRES (Vogt et al. 1994), and at the 3.9 m Anglo-Australian Telescope (AAT) using UCLES (Diego et al. 1990). The planet orbit's its host star with a orbital period of 1973±31 days at 3.14 AU in an eccentric orbit with $e = 0.465$. This is a long period planet with a minimum mass of $2.64M_J$ (Butler et al. 2006). HD 66428 is chromospherically quite with $\log(R'_{HK}) = -5.08$. The potential transit depth we can observe is 1.184%. Hipparcos observed the star for 86 epochs with a median value of 8.3909±0.0111 of which 36 data points fall in the calculated t_d with a median value of 8.3998±0.0100. Performing the analysis shown in figure 4.1 the calculated $t_{sd} = -0.151$ with a $p_d = 0.88$ which is above the cut off value $\alpha = 0.8$. An interesting feature to notice is that even though the $t_w = 215.662$ days, there are no data points in the transit window apart from the transit duration itself, but the significance of these observations is not high enough and hence this target is classified as statistically insignificant.

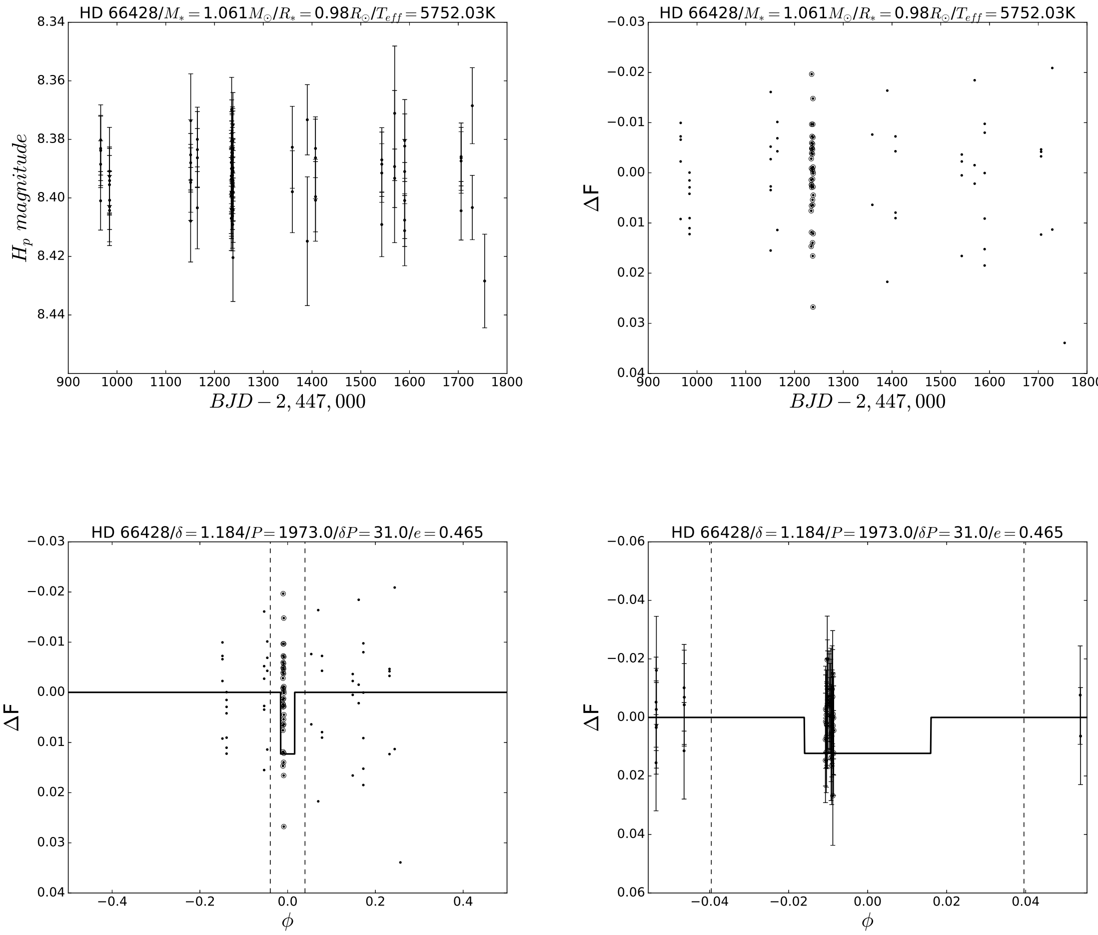


Figure 5.6: *Top left: Hipparcos* photometric measurements of HD 66428. *Top right:* Normalized *Hipparcos* photometric measurements of HD 66428. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 66428. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

5.3.4 HIP 113238/ HD 216770

HD 216770b was discovered as a part of the CORALIE survey for southern extra-solar planets. CORALIE planet-search program in the southern hemisphere (Udry et al. 2000) has been ongoing at the 1.2-m Euler Swiss telescope designed, built and operated by the Geneva Observatory at La Silla Observatory (ESO/Chile). The reported orbital period from the RV data is 118.45±0.55 days with a minimum mass of $0.65M_J$ on an slightly eccentric orbit with $e = 0.37$ at 0.46 AU from the parent star. HD 216770 is K0V star with a reported activity index of $\log(R'_{HK}) = -4.84$, a moderate jitter in the data can be expected (Mayor et al. 2004). Potentially we can observe a transit depth of 1.05%. Hipparcos observed HD 216770 for 99 epochs with a median value of 8.2709±0.0144 of which 6 data points fall in the transit duration with a median value of 8.2657±0.0100, calculated from the TERMS data. The calculated t_{sd} and p_d from the analysis shown in figure 4.1 are -0.187 and 0.859 respectively. The p_d value is above the cut of $\alpha = 0.8$ and hence the variation signified by the t_{sd} is not significant enough. Hence this target is classified as statistically insignificant.

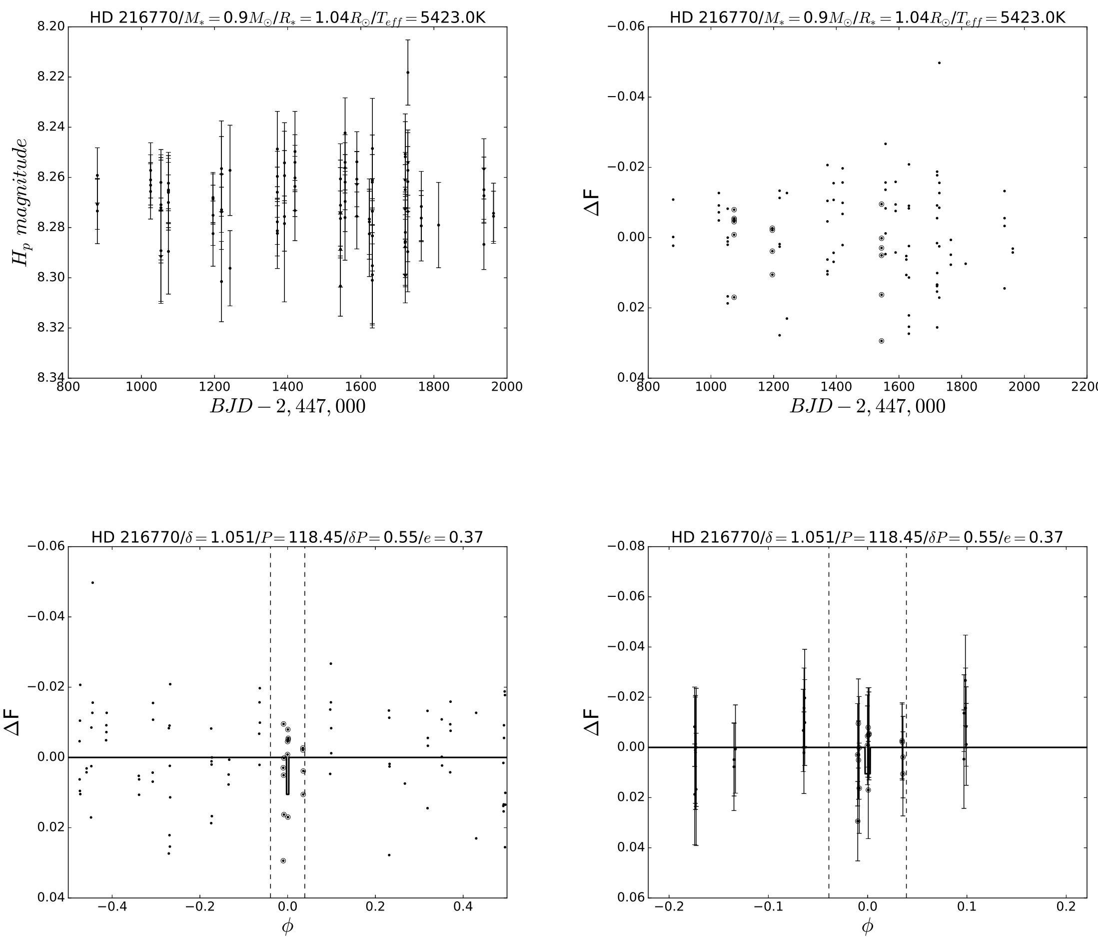


Figure 5.7: *Top left: Hipparcos* photometric measurements of HD 216770. *Top right:* Normalized *Hipparcos* photometric measurements of HD 216770. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 216770. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

5.3.5 HIP 61177/ HD 109246

The discovery of a Jupiter-mass companion around HD 109246 was reported as a part of SOPHIE search for northern extrasolar planets. SOPHIE is a spectrograph mounted on the 1.93-m telescope at the Haute-Provence Observatory. HD 109246 is a G0V star with an activity index of $\log(R'_{HK}) = -5.05$. HD 109246b has an orbital period of 68.27±0.13 days at 0.33 AU with an e of 0.12. The calculated minimum mass of the planet from the data is $0.77 M_J$ (Boisse et al. 2010). No photometric search for transits have been reported until now, even though the transit probability is about 1%. Hipparcos observed HD 109246 for 116 epochs with a median value of 8.8796±0.0131 of which 3 data points fall in the transit duration calculated from TERMS data with a median value of 8.8826±0.0146. The associated t_{sd} and p_d with the sample in the transit duration are 0.225 and 0.843 respectively. As the p_d value is greater then the cutoff $\alpha = 0.8$ the small sample variation compared to the data is insignificant and we can rule out these samples as statistically insignificant.

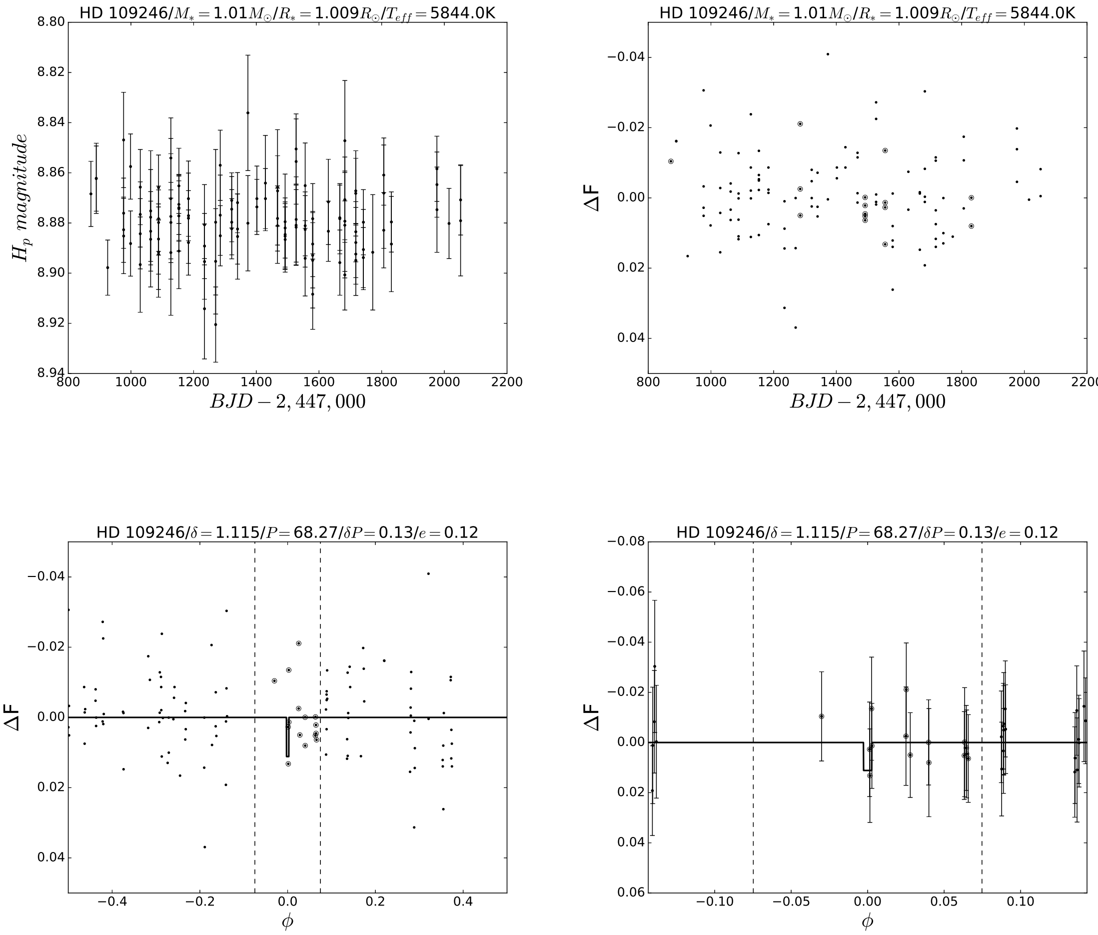


Figure 5.8: *Top left: Hipparcos* photometric measurements of HD 109246. *Top right:* Normalized *Hipparcos* photometric measurements of HD 109246. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 109246. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

The table below summarizes the results from our analysis.

Table 5.3: Data summary from analysis of HD 188015, HD 290327, HD 66428, HD 216770 and HD 109246

Target	t_d	t_w	N	N_d	N_w	t_{sd}	t_{sw}	p_d	p_w
HIP 97769/ HD 188015	0.684	80.527	210	3	34	0.066	-1.1	0.952	0.279
HIP 25191/HD 290327	1.127	1128.301	89	4	80	0.099	-0.092	0.926	0.927
HIP 39417/HD 66428	0.71	215.662	86	36	36	-0.151	-0.151	0.88	0.88
HIP 113238/HD 216770	0.62	52.103	99	6	16	-0.187	1.277	0.859	0.22
HIP 61177/HD 109246	0.364	19.545	116	3	15	0.225	0.432	0.843	0.672

5.4 Potential transit in the Hipparcos data during the TERMS transit window

These are cases classified as targets with potential transits in the Hipparcos data during the TERMS window using the criteria we defined in section 4.7. The table below summarizes the results from our analysis. For this case we only pick targets with small windows, because if the windows are large then the sample starts to resemble the distribution.

5.4.1 HIP 31246/HD 46375

HD 46375b was discovered using the Precision Doppler measurements from the Keck High-Resolution Echelle Spectrograph. The planet orbits a K1V star with an activity index of $\log(R'_{HK}) = -4.94$ typical for chromospherically quiet stars (Noyeset al. 1984). HD 46375b has an orbital period of 3.0240±.0005, a short period sub-saturn planet with a minimum mass of $0.249 M_J$. The semi-major axis is at 0.041 AU with an eccentricity of 0.063 (Marcy et al. 2000). The potential transit depth we can observe is 1.228%, but no transit searches or detections have been reported. Hipparcos collected brightness data on the target for 63 epochs with a median value of 8.0512±0.0133 of which only 2 data points fall in the calculated t_d. An interesting feature to notice here is that the t_d and t_w for this target are 0.097 days and 0.753 days respectively. As the window is pretty small we can do the analysis shown in figure 4.1 in the TERMS transit window. There are 7 data points in the TERMS transit window with a median value of 8.0641±0.008. The calculated t_{sw} and p_w are 2.785 and 0.128 receptively. From the calculated $t-value$ in the transit window t_{sw} we can see that the target is significantly dimmer, below the cut off $\alpha = 0.2$. We can classify this target as a potential transit signature in the TERMS window for Hipparcos data. Results are summarized in table 5.4.

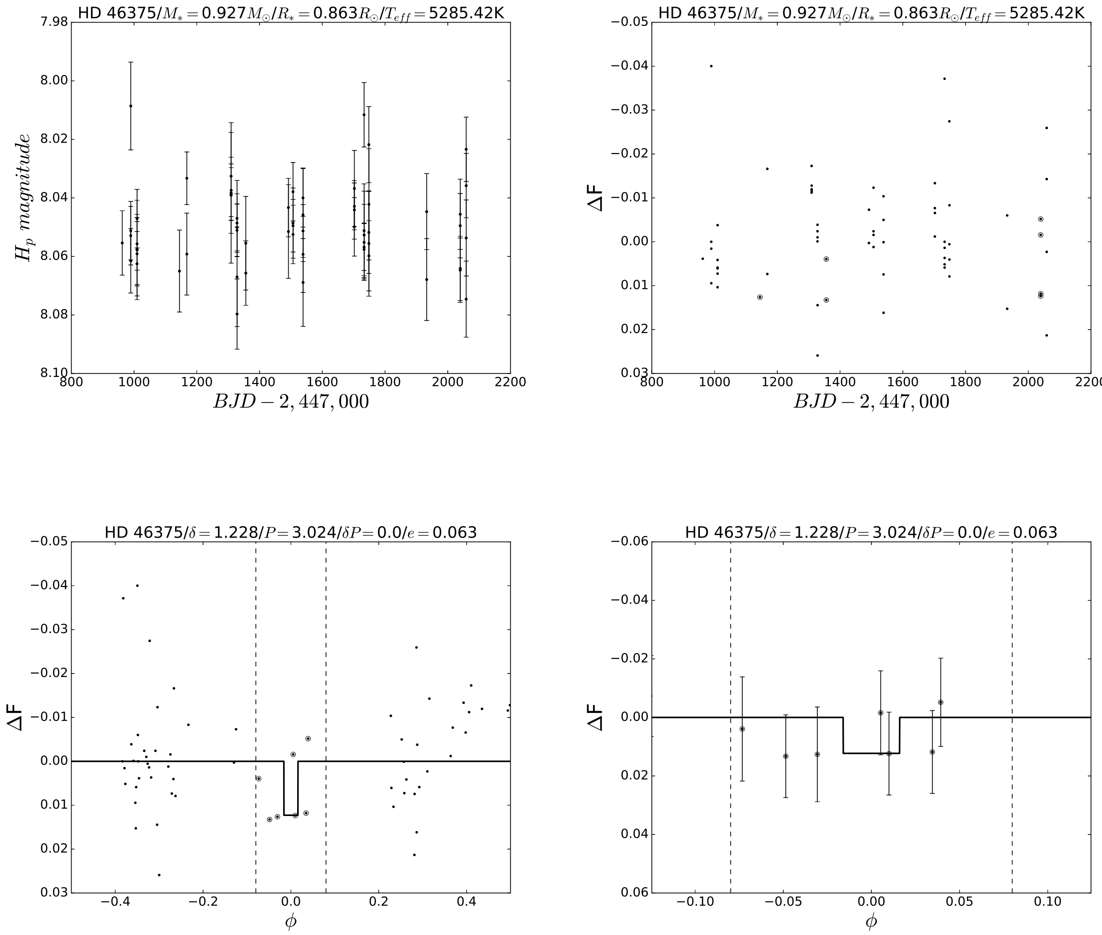


Figure 5.9: *Top left: Hipparcos* photometric measurements of HD 46375. *Top right:* Normalized *Hipparcos* photometric measurements of HD 46375. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 46375. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

Table 5.4: Data summary from analysis of HD 46375

Target	t_d	t_w	N	N_d	N_w	t_{sd}	t_{sw}	p_d	p_w
HIP 31246/HD 46375	0.097	0.753	63	2	7	0.965	2.785	0.511	0.128

5.5 Potential transit rule out in the Hipparcos data during the TERMS transit window

These are cases classified as targets with potential transit rule out in the Hipparcos data during the TERMS window using the criteria we defined in section 4.7. The table below summarizes the results from our analysis. For this case we only pick targets with small windows, because if the windows are large then the sample starts to resemble the distribution.

5.5.1 HIP 3093/HD 3651

HD 3651b was discovered in 112 observations of HD 3651 at Lick Observatory over the past 15 years and 26 observations at Keck Observatory over the past 6 years. This K0V star is chromospherically inactive with an activity index of $\log(R'_{HK}) = -5.01$. 10 years of observations at Fairborn Observatory show it to be photometrically stable to better than 0.001 mag. The planet has an orbital period of 62.23±0.03 days orbiting the star at 0.284 AU in an eccentric orbit with $e = 0.63$.

The minimum mass derived from the RV observations is $0.2M_J$ (Fischer et al. 2003). There was a transit search for this planet using the observations at Fairborn Observatory and no transits in the data were reported (Fischer et al. 2003). The potential transit depth that we can observe is 1.059%, which is very close to our limit set for the 1% transit depth cutoff. Hipparcos observed the target for 58 epochs with a median value of 6.030±0.006 of which 4 data points fall in the TERMS transit duration with a median value of 6.02785±0.006 and 8 data points fall in the TERMS transit window with a median value of 6.02785±0.004. The t_d and t_w values from the TERMS data are small enough that we can perform the analysis shown in figure 4.1 for samples in the transit duration and transit window. From the calculated t_{sd}, t_{sw}, p_d and p_w which are 0.538, 3.409, 0.259 and 0.05 respectively, the variation of the samples compared to the data is in the same direction. But the significance with which we are certain about the variation ($p - value$) is different and also by how much ($t - value$). In the transit window the p_w is less then the cut off value $\alpha = .2$. We can rule of transits in the TERMS transit window with high confidence as the target is getting brighter.

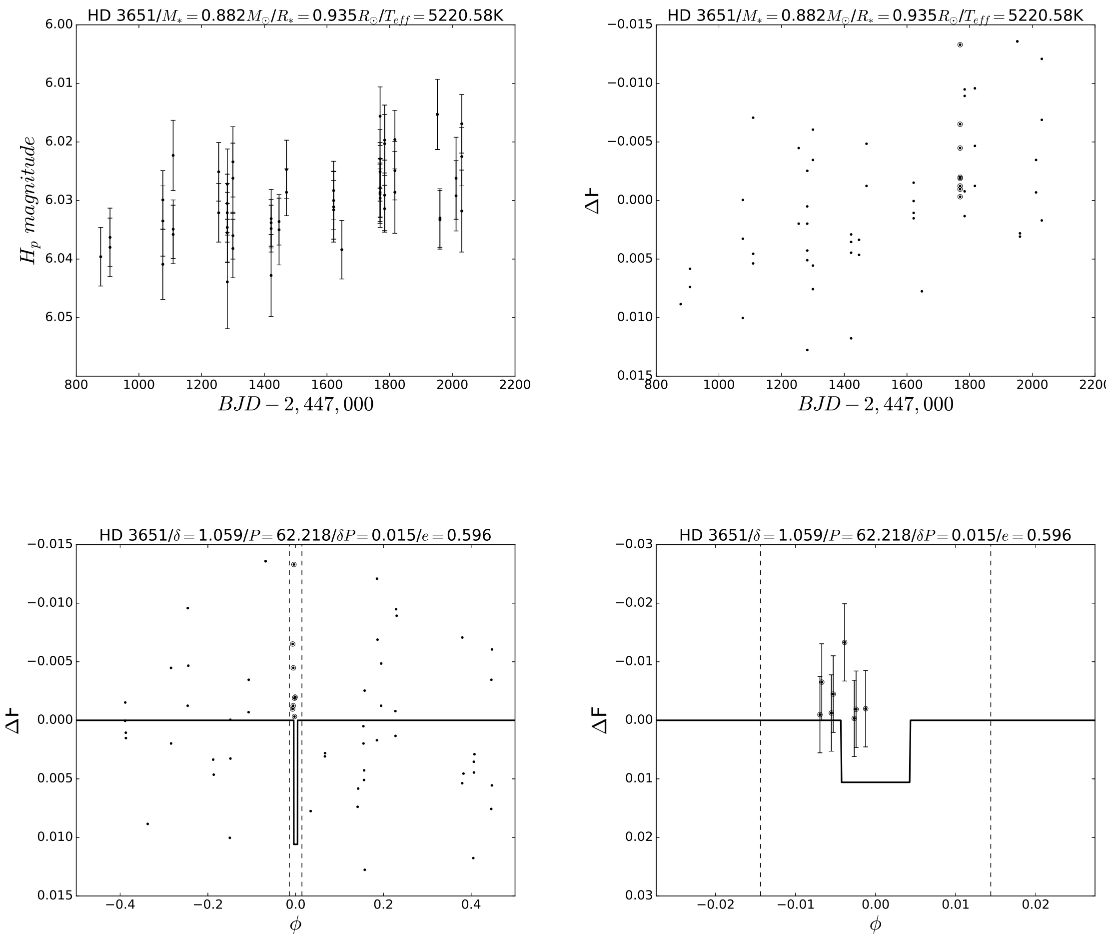


Figure 5.10: *Top left: Hipparcos* photometric measurements of HD 3651. *Top right:* Normalized *Hipparcos* photometric measurements of HD 3651. The photometric measurements in the TERMS window are surrounded by circles. *Bottom left:* Phase folded *Hipparcos* photometric measurements of HD 3651. The dotted line in the size of the TERMS window (t_w) and the solid line is a transit curve with a set transit depth (δ) and duration (t_d). *Bottom right:* Zoom onto the measurements in the transit window (t_w).

Table 5.5: Data summary from analysis of HD 3651

Target	t_d	t_w	N	N_d	N_w	t_{sd}	t_{sw}	p_d	p_w
HIP 3093/HD 3651	0.538	3.409	58	4	8	-1.389	-2.367	0.259	0.05

Chapter 6

Conclusions and future work

We have already proved that the method developed here works using the test/validation cases HD 189733 and HD 209458. A code was written in python to automate the same. Once we applied this technique to the Hipparcos dataset we did see some interesting results. The targets that we selected were a small subset of the TERMS data because we only looked at the targets that were present in the Hipparcos catalog, HD catalog and the TERMS dataset. Hence we could only summarize a few interesting cases in this thesis. With missions like GAIA (launched on December 19, 2013) that is already generation a very large data set filled with star brightness data of over one billion stars in our Milky Way galaxy and TESS (launched on April 18, 2018), we can use the tool proposed in this thesis to quickly filter out the interesting cases for further study and observations.

The different cases classified based on the criteria described in section 4.7 can be summarized and interpreted as follows:

- If the output of the analysis indicates a positive t-value for a target with a p-value lower then the cutoff in the TERMS duration window, it should be further investigated either by follow up observations or other photometric data on the target if available. For example the target HIP 19207/ HD 285507, upon further investigating the target we found that it does not transit N. Quinn et al. (2014). This can also be classified as a false positive in our analysis.

- If the output of the analysis indicates a negative t-value for a target with a p-value lower then the cutoff we can rule of transits in the TERMS duration/window for that specific set of photometric data. For example the target HIP 38041/ HD 63765.

- If the output of the analysis indicates a positive t-value for a target with a p-value lower then the cutoff in the transit duration window and if the TERMS data has a large window due to poor RV data, more accurate RV measurements will help to detect potential transit points. For example the target HIP 6511/ HD 8535.

- If the output of the analysis indicates a high p-value in the transit duration window we can with certainty conclude that the data points in the TERMS duration window do no differ significantly compared to the entire photometric measurements. Hence we can just conclude that the photometry dataset is not sufficient to detect or rule out transits in the TERMS duration. For example the target HIP 6511/ HD 8535.

- For interpreting the results for the data points in the transit window we have to be a bit careful because a large TERMS window would include a large number of data points from the observation and the sample will start to resemble the distribution. Hence we need to look for planets with comparatively small TERMS window and then classify them based on the t-value and p-value. For example the target HIP 3093/HD 3651 we can rule out potential transit in the TERMS window, but for the target HIP 31246/ HD 46375 we can see something interesting happening in the TERMS window and potential transits in the TERMS window is a possibility here.

Our future goal is to use many other photometric data collected over various different catalog of stars and look for potential transit points in the data using the TERMS data. Here we have only considered systems hosting only one planet but in the future we are planning to include multi-planet systems, which will give us even more targets to look for.

Chapter 7

Appendices

Appendix I: Tables

Table 7.1: Properties of case study systems

Target	Mass($M_\odot$)	Radius($R_\odot$)	T_{eff}(K)	Period(d)	δP	e	δ
HD 102195	0.87	0.938	5291.0	4.114	0.001	0.0	1.287
HD 104067	0.791	0.858	4969.0	55.806	0.049	0.0	1.0
HD 108341	0.843	0.79	5122.0	1129.0	7.0	0.85	1.822
HD 109246	1.01	1.009	5844.0	68.27	0.13	0.12	1.115
HD 111232	0.78	0.875	5494.0	1143.0	14.0	0.2	1.502
HD 114386	0.78	0.637	4819.64	937.7	15.6	0.23	2.776
HD 114762	0.895	0.859	5952.51	83.915	0.003	0.335	1.537
HD 114783	0.853	0.838	5135.09	493.7	1.8	0.144	1.611
HD 117207	1.031	0.952	5723.69	2597.0	41.0	0.144	1.26
HD 121504	1.18	0.844	6075.0	63.33	0.03	0.03	1.611

GJ 3021	0.9	0.858	5536.0	133.71	0.2	0.511	1.537
HD 130322	0.92	0.85	5387.0	10.709	0.0	0.029	1.574
HD 131664	1.1	1.062	5886.0	1951.0	41.0	0.638	1.012
HD 141937	1.048	1.064	5846.82	653.22	1.21	0.41	1.012
HD 145377	1.12	1.001	6046.0	103.95	0.13	0.307	1.137
HD 147513	1.072	0.85	5929.74	528.4	6.3	0.26	1.574
HD 148156	1.22	0.966	6308.0	1027.0	28.0	0.52	1.234
HD 154345	0.893	1.0	5468.12	3341.559	92.68	0.044	1.137
HD 156279	0.93	0.958	5453.0	131.05	0.54	0.708	1.234
HD 162020	0.8	0.526	4844.67	8.428	0.0	0.277	4.048
HD 164509	1.13	1.056	5922.0	282.4	3.8	0.26	1.012
HD 164922	0.927	0.899	5385.36	1155.0	23.0	0.05	1.404
HD 168746	0.916	0.888	5563.62	6.404	0.001	0.107	1.24
HD 171238	0.943	1.051	5467.0	1523.0	43.0	0.4	1.031
HD 188015	1.056	1.033	5745.68	461.2	1.7	0.137	1.072
HD 189733	0.806	0.756	5040.0	2.219	0.0	0.0	2.549
HD 192263	0.804	0.74	4975.16	24.356	0.005	0.055	2.076
HD 204941	0.74	0.85	5056.0	1733.0	74.0	0.37	1.483
HD 209458	1.131	1.155	6065.0	3.525	0.0	0.0	1.561
HD 210277	0.986	0.939	5555.23	442.19	0.5	0.476	1.287
HD 216770	0.9	1.04	5423.0	118.45	0.55	0.37	1.051
HD 218566	0.85	0.591	4820.0	225.7	0.4	0.3	2.441
HD 22781	0.75	0.72	5027.0	528.07	0.14	0.819	2.193

HD 28185	0.99	1.007	5656.0	379.0	2.0	0.05	1.115
HD 285507	0.734	0.656	4503.0	6.088	0.002	0.086	2.61
HD 290327	0.9	0.997	5552.0	2443.0	161.0	0.08	1.137
HD 3651	0.882	0.935	5220.58	62.218	0.015	0.596	1.059
HD 4208	0.879	0.877	5600.17	828.0	8.1	0.052	1.468
HD 46375	0.927	0.863	5285.42	3.024	0.0	0.063	1.228
HD 63765	0.865	0.975	5432.0	358.0	1.0	0.24	1.208
HD 66428	1.061	0.98	5752.03	1973.0	31.0	0.465	1.184
HD 6718	0.96	0.957	5746.0	2496.0	176.0	0.1	1.234
HD 70642	1.002	1.031	5705.54	2068.0	39.0	0.034	1.072
HD 7449	1.05	0.953	6024.0	1275.0	13.0	0.82	1.26
HD 80606	0.958	0.98	5572.51	111.437	0.0	0.934	1.253
HD 81040	0.96	0.919	5700.0	1001.7	7.0	0.526	1.343
HD 83443	0.991	0.943	5453.4	2.986	0.0	0.013	1.287
HD 8535	1.13	1.044	6136.0	1313.0	28.0	0.15	1.051
HD 87883	0.803	0.778	4958.0	2754.0	87.0	0.53	1.869
HD 92788	1.078	0.814	5836.08	325.81	0.26	0.334	1.733
HD 99109	0.94	0.973	5272.39	439.3	5.6	0.09	1.208

N.B. The orbital period, uncertainty in orbital period and the eccentricity are the properties of the first planet of the system in this case planet b.

Table 7.2: Derived properties of case study systems

Target	t_{sw}	p_w	t_w	t_d	t_{sd}	p_d	$\frac{t_d}{t_w}$	$\frac{p_d}{p_w}$
HD 102195	-0.641	0.542	1.129	0.128	-1.121	0.305	0.113	0.563
HD 104067	-	-	7.285	0.285	-	-	-	-
HD 108341	-0.664	0.51	392.776	0.441	-0.558	0.601	0.001	1.178
HD 109246	0.432	0.672	19.545	0.364	0.225	0.843	0.019	1.254
HD 111232	-2.623	0.025	208.252	0.659	-1.223	0.436	0.003	17.141
HD 114386	-0.222	0.833	351.526	0.730	-	-	0.002	-
HD 114762	-	-	1.251	0.340	-	-	-	-
HD 114783	1.013	0.418	57.326	0.497	-	-	0.009	-
HD 117207	-0.383	0.705	505.728	0.910	-0.87	0.424	0.002	0.602
HD 121504	-0.536	0.604	9.879	0.268	-1.559	0.217	0.027	0.359
GJ 3021	0.104	0.921	19.636	0.614	-	-	0.031	-
HD 130322	0.426	0.683	2.394	0.159	-	-	0.066	-
HD 131664	1.034	0.309	325.905	0.595	-	-	0.002	-
HD 141937	0.377	0.731	26.077	0.697	-	-	0.027	-
HD 145377	-	-	8.488	0.287	-	-	-	-
HD 147513	1.535	0.163	197.525	0.712	-	-	0.004	-
HD 148156	-0.749	0.495	211.058	0.483	-	-	0.002	-
HD 154345	-	-	1050.422	1.200	-	-	-	-
HD 156279	-1.289	0.288	17.052	0.961	-1.289	0.288	0.056	1
HD 162020	-	-	0.168	0.087	-	-	-	-

HD 164509	-0.077	0.94	108.301	0.610	-	-	0.006	-
HD 164922	0.188	0.852	836.171	0.798	-2.888	0.212	0.001	0.249
HD 168746	1.476	0.236	3.519	0.134	-	-	0.038	-
HD 171238	-	-	216.962	0.700	-	-	-	-
HD 188015	-1.1	0.279	80.527	0.684	0.067	0.953	0.008	3.413
HD 189733	2.473	0.069	0.090	0.090	2.473	0.069	1	1
HD 192263	-0.91	0.376	9.636	0.195	-	-	0.02	-
HD 204941	-	-	438.919	1.181	-	-	-	-
HD 209458	6.766	0.002	0.139	0.139	6.766	0.002	1	1
HD 210277	0.071	0.948	22.246	0.362	0.071	0.948	0.016	1
HD 216770	1.278	0.221	52.103	0.620	-0.187	0.859	0.012	3.892
HD 218566	0.299	0.766	333.925	0.259	-	-	0.001	-
HD 22781	-	-	2.376	0.714	-	-	-	-
HD 28185	-	-	41.035	0.569	-	-	-	-
HD 285507	-0.688	0.514	0.825	0.112	2.383	0.076	0.136	0.147
HD 290327	-0.093	0.926	1128.310	1.217	0.1	0.927	0.001	1.001
HD 3651	-2.368	0.05	3.409	0.538	-1.389	0.259	0.158	5.202
HD 4208	-2.25	0.039	373.965	0.713	-	-	0.002	-
HD 46375	2.785	0.032	0.753	0.097	0.965	0.511	0.129	16.093
HD 63765	-1.471	0.192	40.363	0.471	-1.471	0.192	0.012	1
HD 66428	-0.151	0.881	215.662	0.710	-0.151	0.881	0.003	1
HD 6718	-0.191	0.88	1012.204	1.175	-	-	0.001	-
HD 70642	-1.568	0.135	1047.152	1.068	-	-	0.001	-

HD 7449	1.903	0.071	105.652	0.668	0.377	0.77	0.006	10.871
HD 80606	-0.85	0.485	0.856	0.734	-	-	0.857	-
HD 81040	0.999	0.5	94.513	0.423	-	-	0.004	-
HD 83443	0.644	0.521	1.986	0.109	1.237	0.433	0.055	0.83
HD 8535	1.908	0.062	356.100	0.767	1.901	0.08	0.002	1.288
HD 87883	-1.058	0.303	763.931	1.633	-	-	0.002	-
HD 92788	-	-	16.162	0.631	-	-	-	-
HD 99109	-0.295	0.77	333.933	0.664	-	-	0.002	-

N.B. Some properties have been assigned to $-$, because the values could not be computed.

Appendix II: Code

Code to extract Hipparcos data from NASA exoplanet archive

```
format long
clc
clear

%For this code to work the list obtained from the archive must be ordered by HIP numbers
%Getting the Information/File of the confirmed planets from Nasa Exo-planet Archive

Planet_info_1 = urlread('http://exoplanetarchive.ipac.caltech.edu/cgi-bin/nstedAPI/nph-nstedAPI?table=exoplanets&select=pl_hostname,hd_name,hip_name,pl_pnum,pl_letter&order=hip_name&format=csv');

%writing the data to a file and formatting
dlmwrite('Planet_info_1.csv',Planet_info_1,''); %writing the extracted data in csv format
fid=fopen('Planet_info_1.csv');
```

```
temp_cell = textscan(Planet_info_1,'%s %s %s %s %s','HeaderLines',1,'Delimiter',',', 'EndOfLine','\n'); % scanning the text file using a specific Delimiter, here its ','
temp_structure = cell2struct(temp_cell,{'pl_hostname','hd_name','hip_name','pl_pnum','pl_letter'},2);  % Converting the extracted cells into a structure and storing the information as required
fclose(fid);

%Extracting HIP numbers from the file and formating it to append naming for the files to be extracted from the exo-planet archive%

hip_name = getfield(temp_structure,'hip_name'); %extracting and storing hip names of the stars
pl_hostname = getfield(temp_structure,'pl_hostname'); % extracting and storing host names of the stars

%Logic to count the number of HIP stars data
for i=1:length(pl_hostname)
```

```
    A = hip_name(i,1);
    B = strcmp(A,''); %string comparision very usefull
    if B == 0
        j=j+1;
    else
        j;
    end
end

%Display the number of HIP stars data we can extract
hip_stars_number = real(j);
temp_hipstars_number_data = sprintf('The total number of Hipparcos stars data we can extract are %d', hip_stars_number);
disp(temp_hipstars_number_data);

%Renaming the stars as per our need, here to generate the url link with proper string formatting
hip_number_formated = cell(hip_stars_number,1);
for i=1:hip_stars_number

```

```
A=hip_name(i,1);
B=char(A);
 if length(B) == 7
 C=B(5:7);
 D=strcat('0000',C);
 E=cellstr(D);
 hip_number_formated(i)=E;

elseif length(B) == 8
 C=B(5:8);
 D=strcat('000',C);
 E=cellstr(D);
 hip_number_formated(i)=E;

elseif length(B) == 9
  C=B(5:9);
  D=strcat('00',C);
  E=cellstr(D);
  hip_number_formated(i)=E;

elseif length(B) == 10
```

```
      C=B(5:10);
      D=strcat('0',C);
      E=cellstr(D);
      hip_number_formated(i)=E;

    elseif length(B) == 11
      C=B(5:9);
      D=strcat('00',C);
      E=cellstr(D);
      hip_number_formated(i)=E;

  end

end

%Generating URLs for collecting the data
%example URL %http://exoplanetarchive.ipac.caltech.edu/data/ExoData/0099/0099894/data/UID_0099894_PLC_001.tbl

url_links = cell(hip_stars_number,1); % intializing cell for saving url links of HIP data
```

```
hip_info = cell(hip_stars_number,1); % intialzing cell for saving the HIP data obtained
planet_number = cell(hip_stars_number,1); % intialzing cell for saving the HIP star data for the number of planets in the system

planet_number = getfield(temp_structure,'pl_pnum'); % extracting planet numbers from the intial data

%Logic to creat the url link and extract the HIP photometry data for all HIP stars
for i=1:hip_stars_number
 temp = sprintf('http://exoplanetarchive.ipac.caltech.edu/data/ExoData/%.4s/%s/data/UID_%s_PLC_001.tbl',char(hip_number_formated(i,1)),char(hip_number_formated(i,1)),char(hip_number_formated(i,1)));
 %disp(temp);
 url_links(i) = cellstr(temp);
 temp1 = urlread(char(url_links(i)));
 hip_info(i) = cellstr(temp1);
 %writing  the files to the Directory
```

```
 X = sprintf( 'HIP%s_%s.tbl',char(hip_number_formated(i)),
    char(planet_number(i)));
 dlmwrite(X,hip_info(i),'');
end

%Total number of HIP stars, because the data files are same
    for multiple planetary system
hip_stars_actual_number = length(dir('*.tbl'));
temp_hipstars_number = sprintf('The total number of
   Hipparcos stars data actually is %d',
   hip_stars_actual_number);
disp(temp_hipstars_number);

hip_number_planet = cell(hip_stars_actual_number,1);
```

Python function to load the Hipparcos data

```
# -*- coding: utf-8 -*-
"""
@author: Badrinath
"""
import numpy as np
import itertools

def hip_data_load(loc):
    t=[];mag=[];uncern=[]
    with open(loc) as f:
      for line in itertools.islice(f, 123, None):
        line = line.strip()
        columns = line.split()
        t.append(float(columns[0]))
        mag.append(float(columns[1]))
        uncern.append(float(columns[2]))
    f.close()
    t=np.array(t)
    mag=np.array(mag)
    uncern=np.array(uncern)
    return list(t),list(mag),list(uncern)


def hip_data_load_flag(loc):
    t=[];mag=[];uncern=[];acc=[];quality=[];t1=[];mag1=[];
       uncern1=[];
    with open(loc) as f:
      for line in itertools.islice(f, 123, None):
        line = line.strip()
        columns = line.split()
        t.append(float(columns[0]))
        mag.append(float(columns[1]))
        uncern.append(float(columns[2]))
        acc.append(float(columns[4]))
        quality.append(float(columns[3]))
```

```
      for i in range(len(acc)):
        if (acc[i] == 1) :
         t1.append(float(t[i]))
         mag1.append(float(mag[i]))
         uncern1.append(float(uncern[i]))
    f.close()
    t=np.array(t1)
    mag=np.array(mag1)
    uncern=np.array(uncern1)
    return list(t),list(mag),list(uncern)
```

Python function to search for TERMS window data points in Hipparcos data

```
# -*- coding: utf-8 -*-
"""
@author: Badrinath
"""
import itertools
import numpy as np
from hip_data_load import hip_data_load_flag
from matplotlib import pyplot as plt
import pandas as pd
import statistics
from astroML.plotting import hist
from sklearn.metrics import mean_squared_error
from math import sqrt

def transit_detect(loc1, loc2, loc3):

    punctuation_1 = "
       ABCDEFGHIJKMNOPQRSTUVWXYZabcdefgijklmnopqrstuvwxyz/\_:.h
       "

    def remove_punctuation_1(s):

```

```
        s_sans_punct = ""
        for letter in s:

            if letter not in punctuation_1:
                s_sans_punct += letter
        return s_sans_punct


    time_hip=[];temp_time=[];acc=[];#sig_figs=5;

    with open(loc1) as f1:
        for line1 in itertools.islice(f1, 123, None):
            line1 = line1.strip()
            columns1 = line1.split()
            temp_time.append(columns1[0])
            acc.append(columns1[4])

    for j in range(len(acc)):
        if (acc[j] == '1'):
            time_hip.append(temp_time[j])

    time_start_terms=[];time_mid_terms=[];time_end_terms=[];

    with open(loc2) as f1:
        for line1 in itertools.islice(f1, 8, None):
            line1 = line1.strip()
            columns1 = line1.split()
            time_start_terms.append(float(columns1[0]))
            time_mid_terms.append(float(columns1[6]))
            time_end_terms.append(float(columns1[12]))

    transit_time = time_end_terms[0] - time_start_terms[0]

    orbital_period_1 = round(time_start_terms[1] -
       time_start_terms[0],5)

    mid_transit_time = float(float(transit_time)/(float(2)))
```

```
    periastron = []
    with open(loc3) as f1:
        for line1 in itertools.islice(f1, 3, 4):
            periastron.append(line1.split('>')[1])
    stellar_mass = []
    with open(loc3) as f1:
        for line1 in itertools.islice(f1, 18, 19):
            stellar_mass.append(line1.split('>')[1])

    stellar_radius = []
    with open(loc3) as f1:
        for line1 in itertools.islice(f1, 19, 20):
            stellar_radius.append(line1.split('>')[1])

    period = []
    with open(loc3) as f1:
        for line1 in itertools.islice(f1, 1, 2):
            period.append(line1.split('>')[1])

    dperiod = []
    with open(loc3) as f1:
        for line1 in itertools.islice(f1, 2, 3):
            dperiod.append(line1.split('>')[1])

    ecc = []
    with open(loc3) as f1:
        for line1 in itertools.islice(f1, 5, 6):
            ecc.append(line1.split('>')[1])

    teff = []
    with open(loc3) as f1:
        for line1 in itertools.islice(f1, 21, 22):
            teff.append(line1.split('>')[1])

    hip_mission_time_start = 2447831.50000
```

```
    discovery_mid = float(periastron[0])


    terms_data_frame = pd.DataFrame()
    terms_data_frame['Transit_start'] = time_start_terms
    terms_data_frame['Transit_mid'] = time_mid_terms
    terms_data_frame['Transit_end'] = time_end_terms


    terms_past_data_start_transit= [];
    terms_past_data_mid_transit= [];
    terms_past_data_end_transit= [];

    temp_start = discovery_mid - float(time_end_terms[0] -
       time_start_terms[0])/2
    time_start_terms = [temp_start] + time_start_terms


    temp1 = time_start_terms[0];

    i=0;
    while time_start_terms[0] < time_hip[0]:
        temp_start = temp1 - orbital_period_1
        terms_past_data_start_transit.append(temp_start)


        temp_mid = temp_start + mid_transit_time
        terms_past_data_mid_transit.append(temp_mid)

        temp_end = temp_start + transit_time
        terms_past_data_end_transit.append(temp_end)

        temp1 = temp_start

        i=i+1
        if terms_past_data_start_transit[i-1] <
           hip_mission_time_start:
```

```
                break

        nono_2=[]
        nono_2_round_2=[]

        i=j=0;
        for i in range(len(time_hip)):
            for j in range (len(terms_past_data_start_transit)):
                if float(round(terms_past_data_start_transit[j],2))
                    <= round(float(time_hip[i]),3) <= float(round(
                    terms_past_data_mid_transit[j],2)):

                    nono_2.append(float(time_hip[i]))
                    nono_2_round_2.append(round(float(time_hip[i])
                        ,5))

                if float(round(terms_past_data_mid_transit[j],2)) <=
                     round(float(time_hip[i]),3) <= float(round(
                    terms_past_data_end_transit[j],2)):

                    nono_2.append(float(time_hip[i]))
                    nono_2_round_2.append(round(float(time_hip[i])
                        ,5))


        time1,magnitude1,uncertanity1 = hip_data_load_flag(loc1)
        time1=np.round(time1,5)
        magnitude1=np.round(magnitude1,5)

#storing results and numbers in dataframes
        terms_data_frame_hip = pd.DataFrame()
        terms_data_frame_hip['Transit_start'] =
            terms_past_data_start_transit
        terms_data_frame_hip['Transit_mid'] =
            terms_past_data_mid_transit
        terms_data_frame_hip['Transit_end'] =
            terms_past_data_end_transit
```

```


    transit_time_mag_df = pd.DataFrame()
    transit_time_mag_df['time'] = nono_2


    hip_data_frame = pd.DataFrame()
    hip_data_frame['time'] = time1
    hip_data_frame['magnitude'] = magnitude1
    hip_data_frame['uncertanity'] = uncertanity1
    hip_data_frame['d'] =  hip_data_frame['magnitude']/(-2.5)
    hip_data_frame['c'] =  hip_data_frame['magnitude'].median()/(-2.5)
    hip_data_frame['delta_d'] =  hip_data_frame['uncertanity']/(-2.5)
    hip_data_frame['delta_c'] =  hip_data_frame['uncertanity'].median()/(-2.5)
    hip_data_frame['transit_depth'] = 1 - 10**(hip_data_frame['d'] - hip_data_frame['c'])
    hip_data_frame['delta_transit_depth'] = np.sqrt(((-2.30259*10**(hip_data_frame['d'] - hip_data_frame['c']))*(hip_data_frame['delta_c']))**2 + ((2.30259*10**(hip_data_frame['d'] - hip_data_frame['c']))*(hip_data_frame['delta_d']))**2)


    common_df=pd.merge(hip_data_frame, transit_time_mag_df, how='inner')

    sigma_away = []
    #i=0;
    for i in range(len(common_df.magnitude)):
            temp_away_sigma = float(common_df.magnitude[i] - statistics.mean(hip_data_frame.magnitude))/float(statistics.stdev(hip_data_frame.magnitude))
            sigma_away.append(temp_away_sigma)

```

```
    mean_list = [statistics.mean(hip_data_frame.magnitude)] * len(common_df['magnitude'])

    common_df['mean of hip data'] = pd.Series(list(mean_list), index=common_df.index)
    common_df['sigmas away'] = pd.Series(list(sigma_away), index=common_df.index)
    common_df['time_2_sig'] =  pd.Series(list(np.round(np.array(common_df.time),2)), index=common_df.index)
    return (common_df,time1, magnitude1, uncertanity1, hip_data_frame,orbital_period_1,terms_data_frame_hip, terms_data_frame, periastron,stellar_radius,stellar_mass,period,dperiod,ecc,teff)
```

Python code to select targets and perform the hypothesis test

```
# -*- coding: utf-8 -*-
"""
@author: Badrinath

"""
import pandas as pd
from transit_detect_func import transit_detect
from hip_data_load import hip_data_load_flag
from astroML.plotting import hist
from matplotlib import pyplot as plt
import numpy as np
from PyAstronomy.pyasl import foldAt
from scipy import stats
import scipy.stats as stats

punctuation = "abcdefgijklmnopqrstuvwxyz_ "
def remove_punctuation(s):
    s_sans_punct = ""
    for letter in s:
        if letter not in punctuation:
            s_sans_punct += letter
```

```
    return s_sans_punct

def phase_correction(x):
    if x > .5:
        x=x-1
    return x

myfile = 'D:/Research/Research_python/terms_pyhon/
    transit_points_analysis_for_all_targets/terms_web_data/
    ephemerides.html'
str1 = open(myfile).read()
df_terms = pd.read_html(str1,header=0)
df_terms = df_terms[1]
df_terms = df_terms.drop([0])
df_terms = df_terms.sort('Planet').reset_index(drop=True)
df_terms.columns = ['Planet','Period','Window','Mid-p Err','
    Duration','Prob(%)','Depth(%)','Ephemeris','Status']



myfile = 'D:/Research/Research_python/terms_pyhon/
    transit_points_analysis_for_all_targets/terms_web_data/
    ephemerides_fromweb.html'
str1 = open(myfile).read()
df_terms_web = pd.read_html(str1,header=0)
df_terms_web = df_terms_web[1]
df_terms_web = df_terms_web.drop([0])
df_terms_web = df_terms_web.sort('Planet').reset_index(drop=
    True)
df_terms_web.columns = ['Planet','Period','Window','Mid-p Err
    ','Duration','Prob(%)','Depth(%)','Ephemeris','Status']


df_final = pd.concat([df_terms,df_terms_web]).drop_duplicates('
    Planet').reset_index(drop=True)
df_final = df_final.sort('Planet').reset_index(drop=True)

```

```
depth_cut_off = 1
df_transit_depth_1 = df_final[df_final["Depth(%)"]>=
    depth_cut_off]
df_transit_depth_1["Planet_clean"]= df_transit_depth_1["Planet
    "].apply(remove_punctuation)

df_hip_terms_common = pd.read_csv(r"D:/Research/Research_python
    /terms_pyhon/common_terms_targets_with_hipparcos.csv")
df_hip_terms_common["Planet_clean"] = df_hip_terms_common["
    hd_name_clean"]


df_targets_common=pd.merge(df_hip_terms_common,
    df_transit_depth_1, how='inner')
df_targets_common = df_targets_common.dropna()

punctuation = " HIPA"
def remove_punctuation(s):
    s_sans_punct = ""
    for letter in s:
        if letter not in punctuation:
            s_sans_punct += letter
    return s_sans_punct

punctuation_1 = "
    ABCDEFGHIJKMNOPQRSTUVWXYZabcdefgijklmnopqrstuvwxyz/\_:.h"
def remove_punctuation_1(s):
    s_sans_punct = ""
    for letter in s:
        if letter not in punctuation_1:
            s_sans_punct += letter
    return s_sans_punct
df_targets_common.to_csv("final_list_of_targets_1.csv")
df_all_targets = pd.read_csv(r'final_list_of_targets_1.csv')
del df_all_targets['Unnamed: 0']
df_hd_targets_only = df_all_targets.query('terms_targets > 0 ')
str1 = r'D:/Research/Research_python/terms_pyhon/
```

```
    transit_points_analysis_for_all_targets/
    backup_hip_planet_host'
str2 = r'D:/Research/Research_python/terms_pyhon/
    transit_points_analysis_for_all_targets/new_terms_data'
str3 = r'D:/Research/Research_python/terms_pyhon/
    transit_points_analysis_for_all_targets/new_mid_point_data'

hip_file_locations=[];terms_hip_targets_file_locations=[];
    terms_periastron_locations=[];

for i in range(len(df_hd_targets_only.pl_hostname)):
    terms_hip_targets_file_locations.append(str(str2+'/'+str(
        df_hd_targets_only.terms_targets[i])))
    terms_periastron_locations.append(str(str3+'/'+str(
        df_hd_targets_only.terms_targets[i])).replace("transits.
        txt","planet"))
    temp = remove_punctuation(df_hd_targets_only.hip_name[i])
    if len(temp)==1:
         temp1 = str1 + '/HIP'+'000000'+ temp +'_'+ str(int(
            df_hd_targets_only.pl_pnum[i])) +'.tbl'
    elif len(temp)==2:
         temp1 = str1 +'/HIP'+'00000'+temp +'_'+ str(int(
            df_hd_targets_only.pl_pnum[i]))+'.tbl'
    elif len(temp)==3:
         temp1 = str1 +'/HIP'+'0000'+temp +'_'+ str(int(
            df_hd_targets_only.pl_pnum[i]))+'.tbl'
    elif len(temp)==4:
         temp1 = str1 +'/HIP'+'000'+temp +'_'+ str(int(
            df_hd_targets_only.pl_pnum[i]))+'.tbl'
    elif len(temp)==5:
         temp1 = str1 +'/HIP'+'00'+temp +'_'+ str(int(
            df_hd_targets_only.pl_pnum[i]))+'.tbl'
    elif len(temp)==6:
         temp1 = str1 +'/HIP'+'0'+temp +'_'+ str(int(
            df_hd_targets_only.pl_pnum[i]))+'.tbl'
    hip_file_locations.append(str(temp1))

```

```

med_in_window=[]
std_in_window=[]
med_all=[]
std_all=[]
name=[]; p_value=[] ; t_value = []
p_value_scipy=[] ; t_value_scipy = []
p_value_w=[] ; t_value_w = []; name_w=[]
p_value_scipy_w=[] ; t_value_scipy_w = []
window=[];duration=[]
number_duration=[];number_window=[];
#custom target testing

master_df = []
i=0;
for i in range(len(hip_file_locations)):#len(hip_file_locations)
    print i
    #print hip_file_locations[i]
    if df_targets_common["pl_pnum"][i]==1:
        try:
            analysis = transit_detect(hip_file_locations[i], terms_hip_targets_file_locations[i], terms_periastron_locations[i])
            master_df.append(analysis)
            analysis[4]["corrected_phases"]= (analysis[4]["time"] - (float(analysis[8][0]))) / float(analysis[5]);
            analysis[4]["corrected_phases"]=(analysis[4]["corrected_phases"] % 1).astype(float)
            analysis[4]["corrected_phases"]=analysis[4]["corrected_phases"].apply(phase_correction)

            df_c=pd.merge(analysis[4], analysis[0], how='inner', on=['time'])

            print ("median and std of points in window: ",df_c
```

```
            ["magnitude_x"].median(), df_c["magnitude_x"].std() )
            print ("median and std of entire data: ",analysis[4]["magnitude"].median(), analysis[4]["magnitude"].std() )


            med_in_window.append(df_c["magnitude_x"].median())
            std_in_window.append(df_c["magnitude_x"].std())
            med_all.append(analysis[4]["magnitude"].median())
            std_all.append(analysis[4]["magnitude"].std())


            df_depth = pd.DataFrame()
            df_depth["Phase"]= np.linspace(-(df_targets_common["Duration"][i])/(analysis[5]*2),(df_targets_common["Duration"][i])/(analysis[5]*2),len(df_c))
            df_depth["Depth"] = [df_targets_common["Depth(%)"][i]/100] * len(df_c)

            df_t = pd.DataFrame()
            df_t["Phase"]=analysis[4]["corrected_phases"]
            df_t["Depth"]=np.zeros(len(analysis[4]["corrected_phases"]))
            for j in range(len(df_t["Depth"])):
                if df_t["Phase"][j] > -(df_targets_common["Duration"][i])/(analysis[5]*2) and df_t["Phase"][j] < (df_targets_common["Duration"][i])/(analysis[5]*2):
                    df_t["Depth"][j] = df_targets_common["Depth(%)"][i]/100

                    df_t1 = pd.DataFrame()
                    df_t1["Phase"]=np.linspace(-.5,.5,10000)
                    df_t1["Depth"]=np.zeros(len(df_t1["Phase"]))
```

```
                    for j in range(len(df_t1["Depth"])):
                        if df_t1["Phase"][j] > -(df_targets_common["Duration"][i])/(analysis[5]*2) and df_t1["Phase"][j] < (df_targets_common["Duration"][i])/(analysis[5]*2):
                            df_t1["Depth"][j] = df_targets_common["Depth(%)"][i]/100
                            df_t=df_t.sort("Phase")




        # T statistic analysis
            first_phase = np.array(df_t1["Phase"][df_t1["Depth"]!=0])
            data_in_transit_duration = df_c[(df_c.corrected_phases >= first_phase[0]) & (df_c.corrected_phases <= first_phase[-1])]
            data_in_transit_window = df_c[(df_c.corrected_phases >= float(analysis[6]['Transit_start'][0]-analysis[6]['Transit_end'][0])/(analysis[5]*2)*(1)) & (df_c.corrected_phases <= float(analysis[6]['Transit_start'][0]-analysis[6]['Transit_end'][0])/(analysis[5]*2)*(-1))]

            ttest=stats.ttest_ind(data_in_transit_duration["transit_depth_y"],analysis[4]["transit_depth"])
            name.append(df_targets_common["hip_name"][i]+'/'+df_targets_common["pl_hostname"][i])
            t_value.append(ttest[0])
            p_value.append(ttest[1])

            ttest_scipy = stats.ttest_1samp(
```

```
                data_in_transit_duration["transit_depth_y"],
                analysis[4]["transit_depth"].mean())
            t_value_scipy.append(ttest_scipy[0])
            p_value_scipy.append(ttest_scipy[1])


            ttest_w=stats.ttest_ind(data_in_transit_window["
                transit_depth_y"],analysis[4]["transit_depth"])
            name_w.append(df_targets_common["hip_name"][i]+'/'+
                df_targets_common["pl_hostname"][i])
            t_value_w.append(ttest_w[0])
            p_value_w.append(ttest_w[1])

            ttest_scipy_w = stats.ttest_1samp(
                data_in_transit_window["transit_depth_y"],
                analysis[4]["transit_depth"].mean())
            t_value_scipy_w.append(ttest_scipy_w[0])
            p_value_scipy_w.append(ttest_scipy_w[1])

            number_duration.append(len(data_in_transit_duration
                ["transit_depth_y"]))
            number_window.append(len(data_in_transit_window["
                transit_depth_y"]))

            if ttest[1] < .2:
                 print "Reject null hypothesis"
                 print "P value is : ", ttest[1]
            else:
                 print "Accept null hypothesis"
                 print "P value is : ", ttest[1]

            plt.figure(figsize=(10,8))
            plt.title(df_targets_common["hip_name"][i]+'/'+
                df_targets_common["pl_hostname"][i]+'/'+"$M_{*}
                = $"+str(round(float(analysis[10][0]),3))+'$M_{\
                odot}$'+'/'+"$R_{*} = $"+str(round(float(
                analysis[9][0]),3))+'$R_{\odot}$'+"/"+"$T_{eff}
```

```
                 = $"+str(round(float(analysis[14][0]),3))+"K",
                 fontsize=16)
            #plt.plot(analysis[0]["time"],analysis[0]["
               transit_depth"],color='black', marker='o',
               markerfacecolor='None',markeredgecolor='black',
               label="TERMS Window",linestyle = 'None')
            #plt.scatter(analysis[4]["time"],analysis[4]["
               transit_depth"],color='k', marker='.', label="
               Hipparcos photometry data")
            plt.errorbar(analysis[1]-2447000,analysis[2],
               analysis[3],color='k', marker='.', label="
               Hipparcos photometry data",linestyle = 'None')
            plt.gca().invert_yaxis()
            plt.xlabel("$BJD - 2,447,000$",fontsize=26)
            plt.ylabel("$H_{p}$ $magnitude$",fontsize=26)
            #plt.legend(loc=2,prop={'size':6})
            #plt.legend(loc=2,prop={'size': 16})
            plt.xticks(fontsize=16)
            plt.yticks(fontsize=16)
            mng=plt.get_current_fig_manager()
            mng.full_screen_toggle()
            plt.savefig(r"D:/Research/Research_python/
               terms_pyhon/final_target_anaysis/
               phase_figures_1_planet_window_2/" +
               df_targets_common["hip_name"][i].replace(" ",
               "") +'_hip.eps', format='eps', dpi=1200)
            plt.cla()
            plt.clf()
            plt.close()

            plt.figure(figsize=(10,8))
            plt.title(df_targets_common["hip_name"][i]+'/'+
               df_targets_common["pl_hostname"][i]+'/'+"$M_{*}
               = $"+str(round(float(analysis[10][0]),3))+'$M_{\
               odot}$'+'/'+"$R_{*} = $"+str(round(float(
               analysis[9][0]),3))+'$R_{\odot}$'+"/"+"$T_{eff}
               = $"+str(round(float(analysis[14][0]),3))+"K",
```

```
            fontsize=16)
        plt.plot(analysis[0]["time"]-2447000,analysis[0]["
            transit_depth"],color='black', marker='o',
            markerfacecolor='None',markeredgecolor='black',
            label="TERMS Window",linestyle = 'None')
        plt.scatter(analysis[4]["time"]-2447000,analysis
            [4]["transit_depth"],color='k', marker='.',
            label="Hipparcos photometry data")
        #plt.errorbar(analysis[4]["time"],analysis[4]["
            transit_depth"],analysis[4]["delta_transit_depth
            "],color='k', marker='.', label="Hipparcos
            photometry data",linestyle = 'None')
        plt.gca().invert_yaxis()
        plt.xlabel("$BJD - 2,447,000$",fontsize=26)
        plt.ylabel("$\Delta$F",fontsize=26)
        #plt.legend(loc=2,prop={'size':6})
        #plt.legend(loc=2,prop={'size': 16})
        plt.xticks(fontsize=16)
        plt.yticks(fontsize=16)
        mng=plt.get_current_fig_manager()
        mng.full_screen_toggle()
        plt.savefig(r"D:/Research/Research_python/
            terms_pyhon/final_target_anaysis/
            phase_figures_1_planet_window_2/" +
            df_targets_common["hip_name"][i].replace(" ",
            "") +'_hip_window_terms.eps', format='eps', dpi
            =1200)
        plt.cla()
        plt.clf()
        plt.close()

        plt.figure(figsize=(10,8))
        plt.title(df_targets_common["hip_name"][i]+'/'+
            df_targets_common["pl_hostname"][i]+'/'+"$\delta
             = $"+str(df_targets_common["Depth(%)"][i])
            +"/"+"$P = $"+str(round(float(analysis[11][0])
            ,3))+"/"+"$\delta P = $"+str(round(float(
```

```
                analysis[12][0]),3))+"/"+"$e = $"+str(round(
                float(analysis[13][0]),3)),fontsize=18)
            plt.plot(df_c["corrected_phases"],analysis[0]["
                transit_depth"],color='black', marker='o',
                markerfacecolor='None',markeredgecolor='black',
                label="TERMS Window",linestyle='None')
            plt.scatter(analysis[4]["corrected_phases"],
                analysis[4]["transit_depth"],color='k', marker
                ='.',alpha=1, label="Hipparcos photometry data"
                )
            #plt.errorbar(analysis[4]["corrected_phases"],
                analysis[4]["transit_depth"],analysis[4]["
                delta_transit_depth"],color='k', marker='.',
                alpha=1, label="Hipparcos photometry data" ,
                linestyle = 'None')
            plt.gca().invert_yaxis()
            #plt.scatter(df_t["Phase"],df_t["Depth"],color="k",
                 label = "Transit Duration",marker='_')
            plt.plot(df_t1["Phase"],df_t1["Depth"],color="k",
                label = "Transit Duration",linestyle='-',lw=2)

            plt.axvline(float(analysis[6]['Transit_start'][0]-
                analysis[6]['Transit_end'][0])/(analysis[5]*2)
                *(-1), color='k', linestyle='--', label="Current
                 Transit Window")
            plt.axvline(float(analysis[6]['Transit_start'][0]-
                analysis[6]['Transit_end'][0])/(analysis[5]*2),
                color='k', linestyle='--')
            plt.xlabel("$\phi$",fontsize=26)
            plt.ylabel("$\Delta$F",fontsize=26)
            plt.xlim([-0.5,.5])
            #plt.legend(loc=2,prop={'size': 16})
            plt.xticks(fontsize=16)
            plt.yticks(fontsize=16)
            mng=plt.get_current_fig_manager()
            mng.full_screen_toggle()
            plt.savefig(r"D:/Research\Research_python/
```

```
                terms_pyhon/final_target_anaysis/
                phase_figures_1_planet_window_2/" +
                df_targets_common["hip_name"][i].replace(" ",
                "") +'_hip_window_terms_phase.eps', format='eps
                ', dpi=1200)
            plt.cla()
            plt.clf()
            plt.close()

            plt.figure(figsize=(10,8))
            plt.title(df_targets_common["hip_name"][i]+'/'+
                df_targets_common["pl_hostname"][i]+'/'+"$\delta
                 = $"+str(df_targets_common["Depth(%)"][i])
                +"/"+"$P = $"+str(round(float(analysis[11][0])
                ,3))+"/"+"$\delta P = $"+str(round(float(
                analysis[12][0]),3))+"/"+"$e = $"+str(round(
                float(analysis[13][0]),3)),fontsize=18)
            plt.plot(df_c["corrected_phases"],analysis[0]["
                transit_depth"],color='black', marker='o',
                markerfacecolor='None',markeredgecolor='black',
                label="TERMS Window",linestyle='None')
            #plt.scatter(analysis[4]["corrected_phases"],
                analysis[4]["transit_depth"],color='k', marker
                ='.',alpha=1, label="Hipparcos photometry data"
                )
            plt.errorbar(analysis[4]["corrected_phases"],
                analysis[4]["transit_depth"],analysis[4]["
                delta_transit_depth"],color='k', marker='.',
                alpha=1, label="Hipparcos photometry data",
                linestyle = 'None' )
            plt.gca().invert_yaxis()
            #plt.scatter(df_t["Phase"],df_t["Depth"],color="k",
                 label = "Transit Duration",marker='_')
            plt.plot(df_t1["Phase"],df_t1["Depth"], color="k",
                label = "Transit Duration",linestyle='-',lw=2)

            plt.axvline(float(analysis[6]['Transit_start'][0]-
```

```
                    analysis[6]['Transit_end'][0])/(analysis[5]*2)*(-1), color='k', linestyle='--',label="Current Transit Window")
            plt.axvline(float(analysis[6]['Transit_start'][0]-analysis[6]['Transit_end'][0])/(analysis[5]*2), color='k', linestyle='--')
            plt.xlabel("$\phi$",fontsize=26)
            plt.ylabel("$\Delta$F",fontsize=26)
            plt.xlim([-(df_targets_common["Window"][i])/(analysis[5]*2),(df_targets_common["Window"][i])/(analysis[5]*2)])
            #plt.legend(loc=4,prop={'size':16})
            plt.xticks(fontsize=16)
            plt.yticks(fontsize=16)
            mng=plt.get_current_fig_manager()
            mng.full_screen_toggle()
            plt.savefig(r"D:/Research/Research_python/terms_pyhon/final_target_anaysis/phase_figures_1_planet_window_2/" + df_targets_common["hip_name"][i].replace(" ", "") +'_hip_window_terms_phase_zoon.eps', format='eps', dpi=1200)
            plt.cla()
            plt.clf()
            plt.close()
            plt.show()
            window.append(df_targets_common["Window"][i])
            duration.append(df_targets_common["Duration"][i])
        except Exception:
            pass


t_test_df_scipy = pd.DataFrame()
t_test_df_scipy["Name"] = name
t_test_df_scipy["t_value"] = t_value_scipy
t_test_df_scipy["p_value"] = p_value_scipy

```

```
t_test_df = pd.DataFrame()
t_test_df["Name"] = name
t_test_df["t_value"] = t_value
t_test_df["p_value"] = p_value

t_test_df_scipy_w = pd.DataFrame()
t_test_df_scipy_w["Name"] = name_w
t_test_df_scipy_w["t_value"] = t_value_scipy_w
t_test_df_scipy_w["p_value"] = p_value_scipy_w

t_test_df_w = pd.DataFrame()
t_test_df_w["Name"] = name_w
t_test_df_w["t_value"] = t_value_w
t_test_df_w["p_value"] = p_value_w

t_test_df_scipy["window"]=window
t_test_df_scipy["duration"]=duration
t_test_df_scipy_w["window"]=window
t_test_df_scipy_w["duration"]=duration
t_test_df_scipy_w["t_value_d"] = t_value_scipy
t_test_df_scipy_w["p_value_d"] = p_value_scipy

t_test_df_scipy_w["duration_window_ratio"]=t_test_df_scipy_w.
    dropna(subset=["p_value"])["duration"]/t_test_df_scipy_w.
    dropna(subset=["p_value"])["window"]
t_test_df_scipy_w["pvalue_ratio_duration_window"]=
    t_test_df_scipy.dropna(subset=["p_value"])["p_value"]/
    t_test_df_scipy_w.dropna(subset=["p_value"])["p_value"]


#targets
df_target_transits_durations = t_test_df_scipy_w[(
    t_test_df_scipy_w["t_value_d"] > 0) & (t_test_df_scipy_w["
    p_value_d"] < .2 )]
df_target_no_transits_durations = t_test_df_scipy_w[(
    t_test_df_scipy_w["t_value_d"] < 0) & (t_test_df_scipy_w["
    p_value_d"] < .2 )]
```

```
df_target_no_significance_durations = t_test_df_scipy_w[(t_test_df_scipy_w["p_value_d"] > .8 )]


df_target_transits_window = t_test_df_scipy_w[(t_test_df_scipy_w["t_value"] > 0) & (t_test_df_scipy_w["p_value"] < .2 )]
df_target_no_transits_window =  t_test_df_scipy_w[(t_test_df_scipy_w["t_value"] < 0) & (t_test_df_scipy_w["p_value"] < .2 )]
df_target_no_significance_window = t_test_df_scipy_w[(t_test_df_scipy_w["p_value"] > .8 )]


plt.scatter(t_test_df_scipy_w.dropna(subset=["p_value"])["duration"]/t_test_df_scipy.dropna(subset=["p_value"])["window"],t_test_df_scipy.dropna(subset=["p_value"])["p_value"]/t_test_df_scipy_w.dropna(subset=["p_value"])["p_value"])
plt.axhline(float(1), color='k', linestyle='--')
plt.axvline(float(1), color='k', linestyle='--')
#plt.axis("equal")
#plt.xlim(0,2)
#plt.ylim(0,1)
plt.xticks(fontsize=18)
plt.yticks(fontsize=18)
plt.ylabel("$p-value_{duration}/p-value_{window}$",fontsize=20)
plt.xlabel("$t-value_{duration}/t-value_{window}$",fontsize=20)
plt.savefig(r"D:/Research/Research_python/terms_pyhon/final_target_anaysis/phase_figures_1_planet_window_2/" + 'p_vs_t_ratio.eps', format='eps', dpi=1200)


plt.figure()
plt.scatter(t_test_df_scipy_w.dropna(subset=["t_value_d"])["t_value_d"],t_test_df_scipy_w.dropna(subset=["p_value_d"])["p_value_d"], color = 'k')
plt.plot(df_target_transits_durations.iloc[0]["t_value_d"],
```

```
    df_target_transits_durations.iloc[0]["p_value_d"],color='
    black', marker='o', markerfacecolor='None',markeredgecolor='
    black',linestyle = 'None', markersize=8 )
plt.plot(df_target_transits_durations.iloc[1]["t_value_d"],
    df_target_transits_durations.iloc[1]["p_value_d"],color='
    black', marker='o', markerfacecolor='None',markeredgecolor='
    black',linestyle = 'None', markersize=8)
plt.axvline(float(0),color='k', linestyle='--')
plt.axhline(float(0.2),color='k', linestyle='--')
plt.axhline(float(0.8),color='k', linestyle='-.')
plt.xticks(fontsize=18)
plt.yticks(fontsize=18)
plt.ylabel("p-value for samples in the duration",fontsize=18)
plt.xlabel("t-value for samples in the duration",fontsize=18)
plt.ylim(0,1)
plt.savefig(r"D:/Research/Research_python/terms_pyhon/
    final_target_anaysis/phase_figures_1_planet_window_2/" + '
    p_vs_t_duration.eps', format='eps', dpi=1200)



plt.figure()
plt.scatter(t_test_df_scipy_w.dropna(subset=["t_value"])["
    t_value"],t_test_df_scipy_w.dropna(subset=["p_value"])["
    p_value"], color='k')
plt.plot(df_target_transits_window.iloc[2]["t_value_d"],
    df_target_transits_window.iloc[2]["p_value_d"],color='black
    ', marker='o', markerfacecolor='None',markeredgecolor='black
    ',linestyle = 'None', markersize=8)
plt.plot(df_target_transits_window.iloc[1]["t_value_d"],
    df_target_transits_window.iloc[1]["p_value_d"],color='black
    ', marker='o', markerfacecolor='None',markeredgecolor='black
    ',linestyle = 'None', markersize=8)
plt.axvline(float(0),color='k', linestyle='--')
plt.axhline(float(0.2),color='k', linestyle='--')
plt.axhline(float(0.8),color='k', linestyle='-.')
plt.xticks(fontsize=18)
```

```
plt.yticks(fontsize=18)
plt.ylabel("p-value for samples in the window",fontsize=18)
plt.xlabel("t-value for samples in the window",fontsize=18)
plt.ylim(0,1)
plt.savefig(r"D:/Research/Research_python/terms_pyhon/
    final_target_anaysis/phase_figures_1_planet_window_2/" + '
    p_vs_t_window.eps', format='eps', dpi=1200)
```

References

Boisse et al. 2010, A&A, 523, 88
Bouchy et al. 2005, A&A, 444, L15–L19
Butler et al. 2006, ApJ, 646, 1 , 505-522
Charbonneau et al. 2000, ApJ, 529, 1 , L45-L48
Fischer et al. 2003, ApJ, 590, 2 , 1081-1087
Hebrard G., et al. 2005, A&A, 445, 341-346
Henry et al. 2000, ApJ, 529, 1 , L41-L44
Kane S., et al. 2009, PASP, 121, 884, 1096
Kane S., et al. 2007, MNRAS, 380, 4, 1488–1496
Lovis, C & Fischer, D., 2011, University of Arizona Press, 2011, 526 pp, p.27-53.
Marcy et al. 2005, ApJ, 619, 1 , 570-584
Mayor et al. 2004, A&A, 415, 391-402
Naef et al. 2010, A&A, 523, 15
Quinn et al. 2014, ApJ, 787, 1 , 10
Robichon, N., et al. 2000, A&A, 355, 295-298
Seager, S., et al. 2003, ApJ, 585, 1038-1055
Ségransan et al. 2011, A&A, 535, 54
Weiss et al. 2014, ApJ, 768, 14
Winn., 2011, [25 pg] Chapter of the graduate-level textbook, EXOPLANETS, ed. S. Seager, University of Arizona Press
Wolszczan, Aleksander & A. Frail, D., 1992, Nature 355, 145-147.
"NASA Exoplanet Archive", https://exoplanetarchive.ipac.caltech.edu/
"TERMS Website", http://stephenkane.net/terms
"The Hipparcos space astrometry mission", https://www.cosmos.esa.int/web/hipparcos/