STRESS - STEREO Transiting Exoplanet and Stellar Survey

by

Vinothini Sangaralingam

A thesis submitted to

The University of Birmingham

for the degree of

DOCTOR OF PHILOSOPHY

Astrophysics and Space Research Group

School of Physics and Astronomy

The University of Birmingham

August 2011

Abstract

The Solar TErrestrial RElations Observatory - *STEREO*, is a system of two identical spacecraft in Heliocentric Earth orbit. We use the two Heliospheric Imagers (HI) on board, which are wide angle imagers with multi-baffle systems to do high precision stellar photometry in order to search for exoplanetary transits and understand stellar variables. The cadence (40 min for HI-1 and 2 hrs for HI-2), high precision ($< 1\%$ scatter for $R < 8$ and a few per cent for fainter objects), wide magnitude range (R mag - 4 to 12) and broad sky coverage (nearly 20 percent just for HI-1A and 60 per cent of the sky in the zodiacal region for all the instruments combined) of this instrument marks this in a space left largely devoid by other current projects. This thesis describes the instrument and its capabilities, the semi-automated pipeline devised for the reduction of this data and the general characteristics of the data obtained. This thesis explains the data analysis methods used in identifying exoplanetary transits and stellar variables along with some of the early results. The thesis concludes with a description of the work in progress and also outlines some of the future prospects of this massive STRESS data set.

To my Parents

Madhu and Sathish

Acknowledgements

Statement of originality

The research work presented in this thesis was undertaken during the period 2007 to 2011, at University of Birmingham, UK under the guidance of Dr. Ian Stevens. The majority of work presented here has been done by the author and those portions which were done as collaboration are as explained below.

Chapter 1 introduces the two major fields of study in brief - extrasolar planets and stellar variability.

Chapter 2 summarises the *STEREO* spacecraft system, the Heliospheric Imager instruments and their capabilities.

Chapter 3 describes the basic image reduction technique, mainly devised by the *STEREO* HI instrument team. The later part of this chapter contains the analysis pertaining to the removal of the Solar F-corona. This is accepted for publication as part of the rest of the semi-automated pipeline described in the next few chapters.

In Chapter 4, the semi-automated photometric pipeline developed by the author for the *STEREO*-HI data is discussed in detail - the catalogue used and the aperture photometry performed to derive the stellar light curves.

Chapter 5 summarises the characteristics of the obtained data along with some of the trend removal methods which were attempted at this time. The work described in chapters 3, 4 and 5 are accepted for publication by MNRAS, entitled : STRESS - STEREO TRansiting Exoplanet and Stellar Survey I: Introduction and Data pipeline (Sangaralingam & Stevens, 2011).

In Chapter 6, the stellar variability analysis pipeline which was developed for this data is explained. It also contains brief introduction to the algorithms which were used in this pipeline. A subset of sample results are also included in this chapter. A more extensive set of examples is given in the appendix. The analysis and results obtained will be published in a forthcoming paper (Sangaralingam et al.).

Chapter 7 describes the transit analysis pipeline developed for this data. This work is done in collaboration with Ms. Gemma Whittaker, University of Birmingham, UK. She developed the basic candidate identification pipeline used in this analysis. This work is also due to be published in the future as Whittaker et al.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Exoplanets

From the time when Man formed his first questions, the question "Are we alone?" has been around. The quest to find a companion to Earth and to gather knowledge about them, has cost many a man's sleep and questioned his integrity. When Galileo invented the telescope, many became obsessed with finding life outside the Earth. Mars was considered a fair prospect. Percival Lowell in 1895, claimed the presence of canals in Mars, which led to a sudden influx of interest in the search for life elsewhere and consequently a systematic search for planets like Earth later. Struve (1952) proposed the first radial velocity program dedicated to the search of planet-like companions of stellar bodies. He proposed that planets at a distance of $1/50^{th}$ of astronomical unit, with a mass of Jupiter, would be observable from spectrograph present at that time. The earliest review of the current methodologies used for the identification of these systems is by Black (1980). However, the evidence of a planet itself came four decades after Struve's proposal. This launched the field of extrasolar planets or exoplanets with the discovery of the first planet in 1992 around the pulsar 1257+12 by Wolszczan & Frail (1992). This discovery brought the dreams of so many centuries to a reality and opened up immense prospects for the future of planetary studies. The search for planets around main sequence stars intensified and the first planet around a normal star, 51 Peg was discovered in 1995 by Mayor & Queloz (1995).

1.2 Current Planet Properties

Since this discovery, the field has grown by leaps and bounds, both in terms of the range of planetary parameter space covered and the technical capabilities of the searches. As of 04

August 2011, there are five different methods by which planets and planetary systems have been discovered. With a total count of 571 planets, 521 planets observed to have radial velocity measurements, 136 of them with transiting measurements and 41 of them with microlensing, imaging and transit timing variations together. There are 58 multiplanet systems in total. The growth of this field is portrayed in Fig. 1.1, which shows the number of planets discovered each year since 1989[1],[2]. There are many contributing factors to the identification of a planet. One of the major criteria is the number of planets searched. Many of the planetary search programs were a little biased, in terms of their magnitude search. Fig. 1.2 illustrates this in terms of the magnitude of the planet host. In many of these searches, there is a trade-off between the number of stars observed and the instrumental capability to observe milli-magnitude variation for transits and less than ms^{-1} precision for radial velocity, apart from other constraints. This results in observing fainter stars, since the number density is large, but this can also result in more confusion. For example *KEPLER* operates a specific group called '*KEPLER* Science Operations centre' (SOC) which is dedicated to the development of a photometric pipeline, taking into account the various aspects of the instrument and the field (cf. Jenkins et al., 2010a, Twicken et al., 2010a and Twicken et al., 2010b for detailed discussions).

One of the biggest surprises encountered by the community is the characteristics of the discovered planets. It was difficult initially to comprehend the structure, formation, evolution of these planets, since they defied all known models based on our Solar system. For example. the objects in the Solar system have a fairly low eccentricity, with maximum values being exhibited by long-period comets. Fig. 1.3 shows the known eccentricities of exoplanets plotted against their period. The range of eccentricities observed so far is very different from those observed within the Solar system, which are plotted in red. We find that there is no preferred eccentricity for planetary systems and at first looks to be very random as they vary from 0 - 1. This variation is explained as arising from planet-planet scattering (Ford & Rasio, 2008), i.e, during the early formative stages when the orbital periods were not stabilised, planetesimal collisions led to some bodies being flung into highly eccentric orbits and in such a model, there is no preferred

---

[1]Courtesy: http://exoplanet.eu

[2]Latham et al. (1989) reported the presence of a companion to the star HD 114762 with a mass of $10.98 M_{Jup}$. This was later confirmed as a planetary body in a highly eccentric orbit with a period of 83.915 days at a distance of 0.353 AU, hence its inclusion in Fig. 1.1
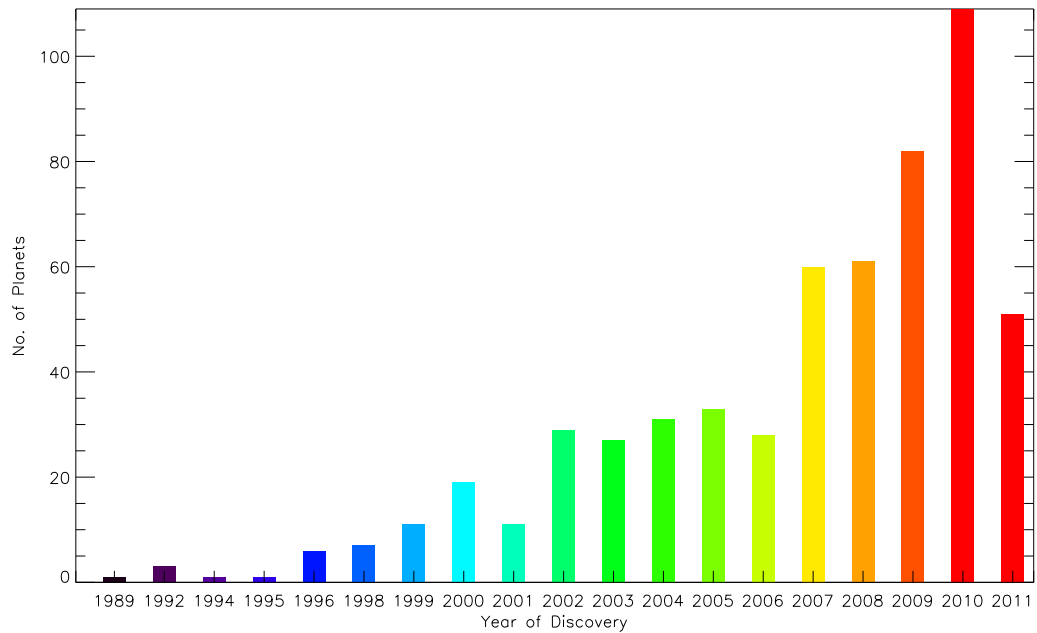
Figure 1.1: A plot of the number of planets discovered each year since 1989. Courtesy:http://exoplanet.eu.
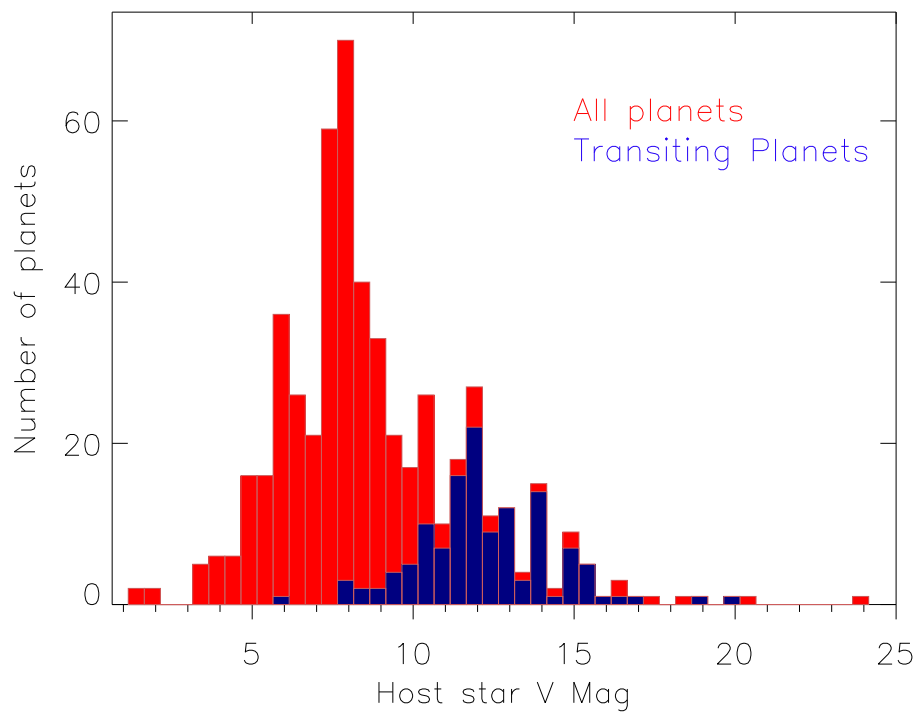
Figure 1.2: The distribution of the magnitude of exoplanetary host stars indicate a definite bias towards fainter end of the spectrum, specifically in the case of transit search. This can explained as the need to search for large number of stars at the same time have the instrumental capability to observe milli-magnitude variations in stellar intensity.
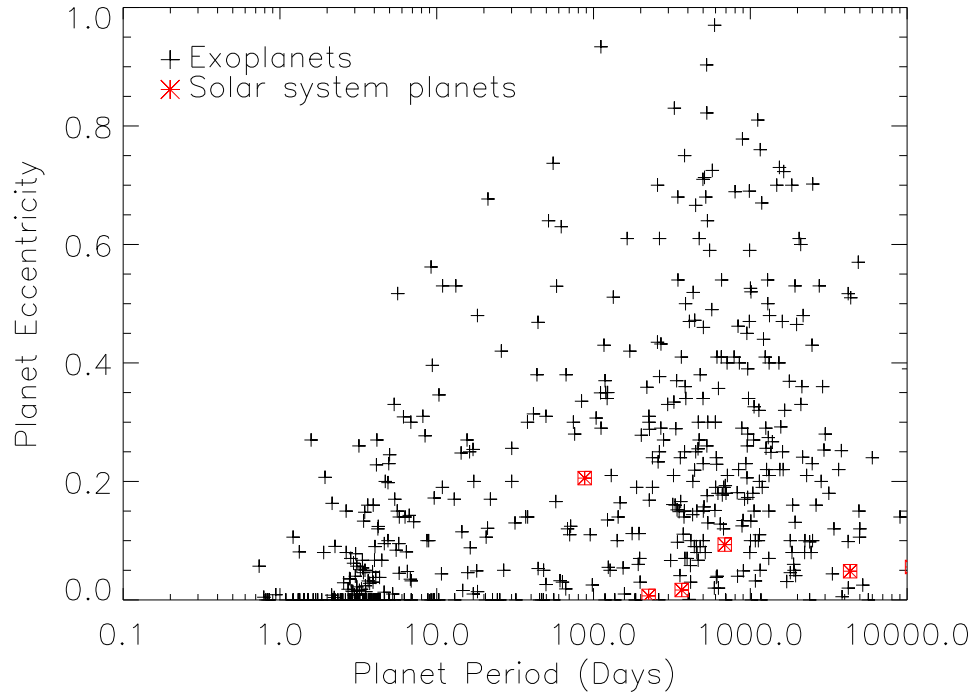
Figure 1.3: The distribution of the eccentricity of known planets against their orbital period illustrates the complexity of the problem, presented to the known planet formation models, arrived based on Solar system bodies. The position of the Solar system models are in red for comparison. The largest eccentricity is exhibited by Mercury with a value of 0.28. in our Solar system bodies.

eccentricity. The highest eccentricity of 0.927 is exhibited by HD 80606b which has an orbital period of 111 days and an $M\sin i$ value of 3.9 Jupiter-mass (M_{Jup}; Naef et al., 2001). Another anomaly is that some of these planets, even-though they are in the Jupiter mass range, they are discovered very close to their parent star. These planets are so close that their surfaces are being evaporated by the host star radiation. This new class of planets are called as 'Hot Jupiters' and 'Very Hot Jupiters'. As seen from this plot, our closest planet, Mercury is nearly in the middle of this distribution. To note is the fact, most of the transiting planets in this distribution are close-in because they are easy to find and hence some of the sample in this distribution is actually affected by observational bias. However, the present theoretical framework based on our Solar system bodies are unable to completely account for the formation of these Hot Jupiters and this necessitated the development of new formation models.

The classical theory explains the formation of planets as follows: the spinning and contracting proto-planetary nebula, in order to conserve its angular momentum, flattens and forms a

disk. The central part containing the maximum mass, initiates hydrogen burning in the core and forms the protostar. The rest of the disk contains a multitude of particles, which collide and stick together forming larger bodies. These smaller cores accrete the surrounding gas and dust until the disk disappears, giving rise to protoplanets and eventually planets (cf. Mordasini et al., 2008 for a detailed review). This formation model mainly depends on the accretion of material from the disk and the disk lifetimes, which is around 7Myr (Pollack et al., 1996). This in-situ formation model, however cannot explain the close Hot Jupiters. This led to a need for migration of these planets, where the cores are formed farther out in the disk but they slowly migrate towards the central star. The inclusion of the migration process in the core-accretion model, resulted in the formation of giant planets closer to the star at shorter time scales of the order of 1Myr (Alibert et al., 2005). This model also enabled continuous accretion of material by the core during its migration inwards, thus resulting in very massive planets, very close to the central body. However, all these models can be tested, verified or confirmed only with the discovery of more exoplanets which are non biased in terms of size as well as period. This is the aim of missions like *KEPLER* - "Determine the variety of orbit sizes and planet reflectivities, sizes, masses and densities of short-period giant planets"[3] and *CoRoT*.

At present, there are more than 100 active exoplanet search projects from both ground and space[4]. Some of the high contributors in this field are projects like Transatlantic Exoplanet Survey (TrES; Dunham et al., 2004), Microlensing Observations in Astrophysics (MOA; Bond et al., 2004), Optical Gravitational Lensing Experiment (OGLE; Udalski et al., 1992), Hungarian Automated Telescope (HATnet; Bakos et al., 2002), COnvection, ROtation and planetary Transits (*CoRoT*; Baglin et al., 2000), Wide-Angle Search for Planets (WASP/SuperWASP; Pollacco et al., 2006), *KEPLER* (Borucki et al., 2010), XO (McCullough et al., 2005), California & Carnegie Planet Search (N2K; Fischer et al., 2004), Micro-variability and Oscillations of STars (*MOST*; Matthews et al., 2000), etc. Table. 1.1 gives a rough outline of the various bits of parameter space covered by some of the above mentioned projects. One thing to note in these projects, is that, the ground based projects, though cover the whole sky and has a high precision, are day-night limited. Whereas, the space-based projects can have round the clock observations but

---

[3]http://kepler.nasa.gov/Mission/QuickGuide/
[4]http://www.exoplanet.eu

are either limited in their sky coverage or magnitude or precision.

The *STEREO* TRansiting Exoplanet and Stellar Survey (STRESS) project involves the use of the Heliospheric Imager instruments onboard the twin *STEREO* spacecraft for the study of exoplanetary transits and stellar variability. This project is described in detail in the rest of this thesis. Fig. 1.4 places STRESS in perspective with respect to the current exoplanet search projects, in terms of the position in the sky as well as magnitude. The plot contains all the known exoplanets plotted as circles, the known transiting planets plotted as stars and the position of the *STEREO* field of View (the FOV is ± 11 degrees of the *STEREO* spacecraft path, which is a little overestimation of the FOV). The magnitude of all the plotted objects are less than $V = 11$, thus is a conservative estimate of the *STEREO* observing capacities. Out of the known 144 transiting exoplanets, 33 of them are within the magnitude range of *STEREO*. However, the only system which is present in the FOV is the five planet system 55 Cnc (von Braun et al., 2011). Alas, the five planets are either too small or have too large an orbital period. Hence we couldn't identify them in our instrument. This example also stresses the point that the data from STRESS is very unique in many different ways.

1.3   Detection Methods

Since the time when Struve (1952) described radial velocity and transit as a probable method of discovering planetary bodies around stellar systems, many different new methods have been developed for this purpose. Some of the most yielding methods so far are described in brief below.

1.3.1   Radial Velocity

A massive planet will induce a Doppler wobble in the parent star due to gravitational interaction. This can be observed as radial velocity (RV) motion of the stellar body (line of sight motion). This small change in velocity can be measured very precisely by high resolution spectroscopy. At present this is the most successful method of detecting exoplanets with 523 planets in 52 multiple planetary systems to its credit. The 0.4 M_{Jup} planet around 51 Peg was

Project	Magnitude Range	CCD FOV	Sky area observed	Cadence	Search Method
TrES	$R = 10 - 13$	$6° \times 6°$	0.2% (STARE & PSST)	9min	Transits
MOA/OGLE[1]	$I < 18$	$6' \times 6'$	All sky	-	Microlensing
HATnet	$R = 9.5 - 12$	$10.5° \times 10.5°$	10% of sky	3.5min	Transits
SuperWASP	$V = 7 - 15$	$7.8° \times 7.8°$	All Sky	30sec	Transits
CoRoT	$V = 11 - 16$	$2.8° \times 2.8°$	0.03% of sky	512sec/32sec	Transits
KEPLER	$K_p = 9 - 16$	$5deg^2$	0.2% of sky	30min/1min	Transits
XO	$V = 9 - 12$	$7.2° \times 7.2°$	55% of the sky	10min	Transits
N2K	brighter than $V = 10.5$	n/a	2000 stars	$4ms^{-1}$	Radial Velocity
MOST[2]	fainter than $V = 6.$	$4.3°$	individual	< 1min	Transits
HARPS[3]	$V = 3 - 16$	aper $= 1''$	Southern Sky	$1ms^{-1}$	Radial Velocity
CORALIE[4]	$V < 9$	aper$= 2''$	Southern Sky	$3 - 4ms^{-1}$	Radial Velocity
STEREO HI-1	$R = 3 - 11$	$20° \times 20°$	each 20% of sky	40 min	Transits
STEREO HI-2	$R = 2 - 11$	$70° \times 70°$	each 60% of sky	2 hrs	Transits

Table 1.1: Different observation capabilities of some of the current exoplanet search programs.
1 - So far the host stars are in the magnitude range 15 - 23.
2 - Walker et al. (2003), 3 - Mayor et al. (2009), 4 - Queloz & Mayor (2001).
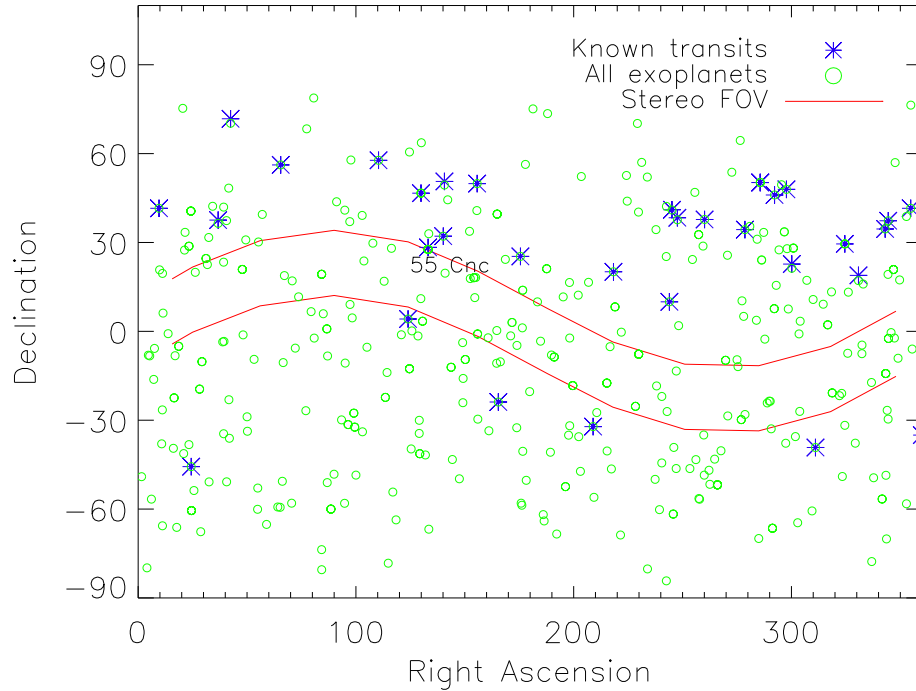
Figure 1.4: The plot is an illustration of the position of *STEREO* in today's exoplanet search programs. The red lines represent the boundary of *STEREO* FOV. This is an overestimation of the FOV by around 10% (i.e., 22° instead of 20°). The blue stars represent the currently known transiting exoplanets and the circles represent all the known exoplanets. The magnitude range of all the data is $V < 11$. There are 33 transiting planets in the magnitude range of *STEREO* but only one of them is in the FOV, 55 Cnc system. The data was obtained on 04 August, 2011.

discovered through this method whose RV curve is given in Fig. 1.5.

The expected velocity wobble K_1 for a star of mass M_1 orbited by a planet of mass M_2 can be shown to be,

$$K_1 = 212.9 \ \frac{M_1^{\frac{1}{3}}}{P} \ \frac{q}{(1+q)^{\frac{2}{3}}} \frac{\sin i}{\sqrt{1-e^2}} \ \ [\text{km s}^{-1}] \tag{1.1}$$

where, $q = M_2/M_1$, and i is the inclination of the orbital axis with respect to the line of sight direction. The masses are expressed in solar masses and the orbital period (P) in days. We can only use the projected mass of the system, $M \sin i$, as opposed to the real mass of the system in this calculation, since it is not possible to extract the orbital inclination information from just the radial velocity measurement. Thus, it is not possible to calculate the true mass of the components for the non-transiting systems. However, in the case of transit measurements, there is an implied inclination information, that is, only those systems whose orbital planes are in our
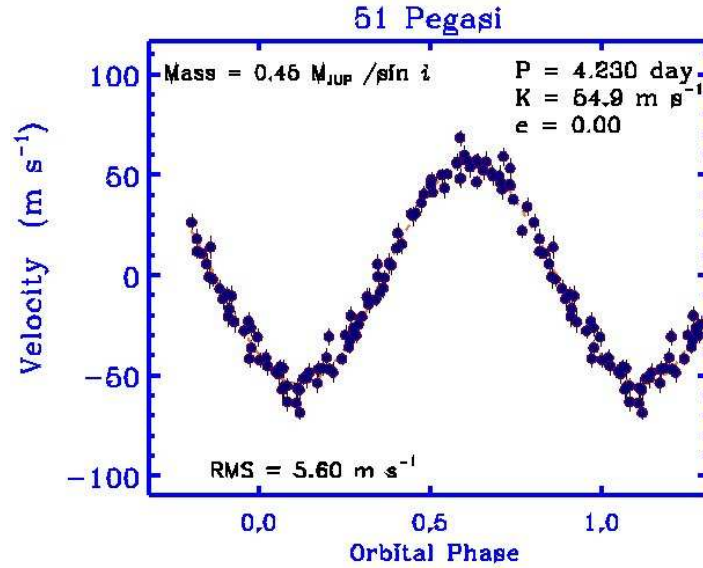
Figure 1.5: The Radial Velocity Curve as obtained by Mayor & Queloz (1995) for 51 Peg from the ELODIE spectrograph of the Haute-Provence Observatory.

line-of-sight can be observed to transit. From Eq. 1.1, we can find that this velocity wobble is only 13 m s^{-1} for a M_{Jup} planet with a period of 12 years and for Earth-mass planets ($M_{\oplus}$) it is only around 8 cm s^{-1}. Presently the most accurate RV instrument is HARPS spectrograph at La Silla in Chile, which can achieve a precision of 1 m s^{-1} (Mayor et al., 2003), which is sufficient to detect planets in a short-period orbit, with a few Earth-masses (Howard et al., 2009). Though this instrument can detect these objects, derivation of other physical parameters is difficult with this data alone. However, if supplementary high precision photometric data is available, it is possible to define many physical and orbital parameters (orbital period, eccentricity and planetary mass) to a high degree of accuracy. Recently, the field has branched from traditional visible wavelength measurements to Near-Infra Red (NIR) spectroscopy. These wavelengths are suitable for studying planets around low mass stars, as the RV values are quite high for them (see Reiners et al., 2010 for a detailed discussion).

1.3.2   Microlensing

One of the observable effects of the Theory of Relativity is gravitational lensing - magnification of light from a farther source by the presence of a massive body in the line of sight. This effect is used in discovering planetary bodies around stars which acts as lens in these events
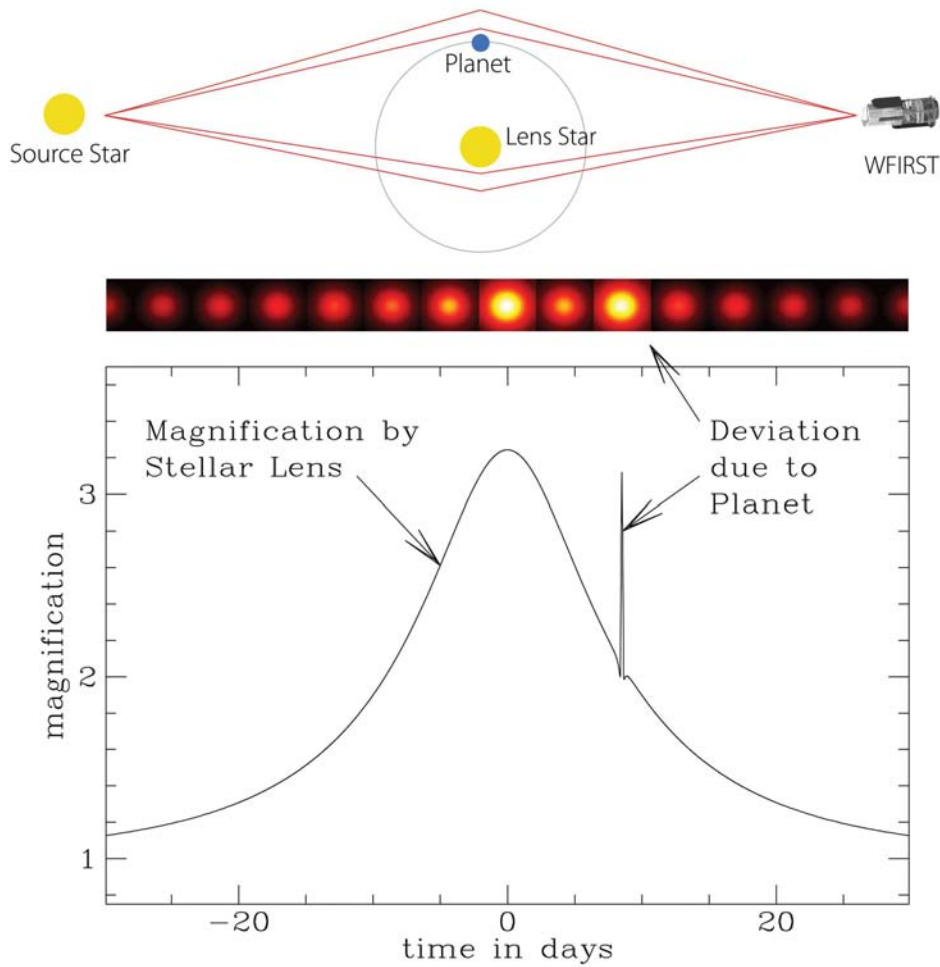
Figure 1.6: Graphical representation of a lensing event around a planetary system as observed by the proposed microlensing mission, WFIRST and a plot of magnification as a function of time. The additional magnification due to the presence of a planet can be clearly identified from the additional peak in the lensing light curve.

(Fig. 1.6). The first planet to be thus discovered is by OGLE group in 2004 around the star OGLE235-MOA53 (see Fig. 1.7 for the actual light curve). This is a 2.6 M_{Jup} planet at an orbital radius of 5.1 AU (Bond et al., 2004). The important point to note is that these events are observable only if the plane of the planetary system is parallel to our line of sight and hence these events are rare. On the other hand, this method is completely independent of planetary mass and thus, can be used in the detection of Earth mass ($M_\oplus$) planets. This method has resulted in the discovery of 13 planets so far, one of them with mass as low as $3.18 M_\oplus$ designated as MOA-2007-BLG-192-Lb (Bennett et al., 2008). Ultimately microlensing is capable of identifying Solar system analogs as well as free floating planets, thereby aiding in identifying planets which cannot be discovered by other methods, thus completing the statistical picture

for theoretical predictions (see Gaudi, 2010 for a review).
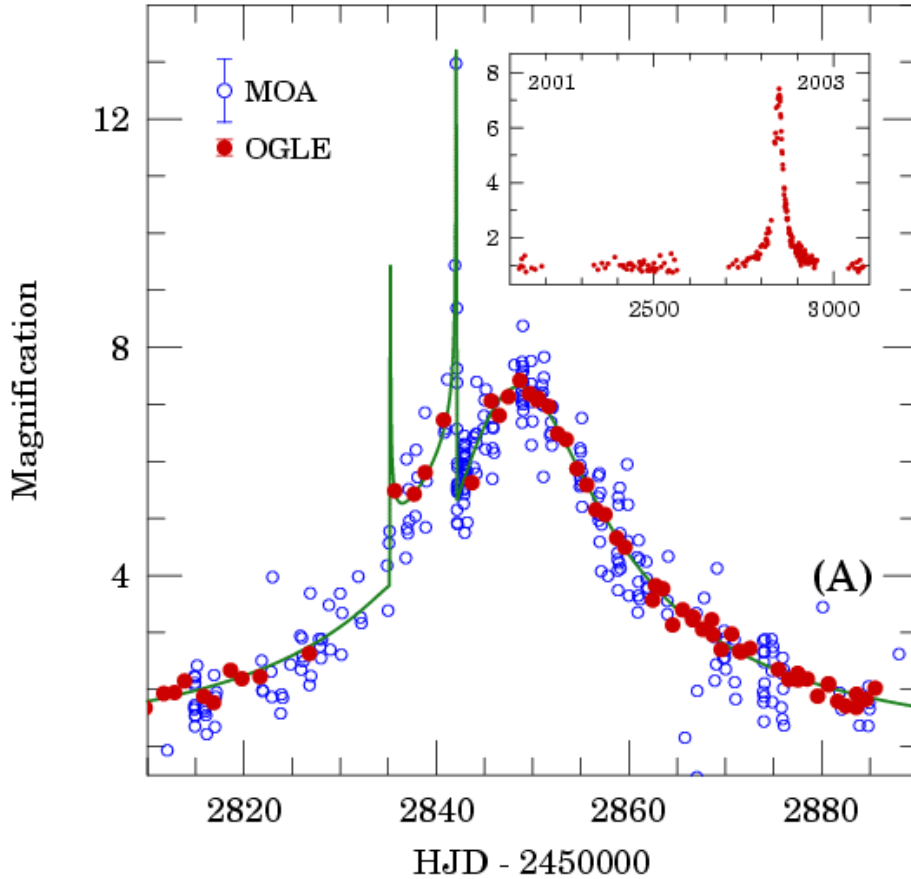
Figure 1.7: Light curve of OGLE 2003-BLG-235/MOA 2003-BLG-53, the first planet to be detected by gravitational microlensing. Insert shows the light curve from 2001 to early 2004 (Courtesy: http://ogle.astrouw.edu.pl).

1.3.3   Direct Imaging

This is one of the most straight forward methods of detecting planets. However, the very small contrast ratio between the star and its planet, provided the hurdle that needs to be crossed in order to discover planets in this method. This threshold was crossed in 2008, with the announcement of two systems by this method at the same time. One is the discovery of a planet around the bright star Fomalhaut, from the Hubble images taken in 2004 and 2006. The presence of a planet was offered as the explanation for the gap in the circumstellar disk around the parent star. The presence and orbital motion of the planet was observed in the two Hubble images taken 2 years apart (Kalas et al., 2008; Fig. 1.8). The other planetary system which was

announced at the same time, was the presence of three planetary bodies around the star HR
8799. Two of the planets were imaged using the Keck and Gemini ground telescopes and was
decoupled from noise using angular differential imaging technique (ADI; Marois et al., 2008).
Fig. 1.9 shows the discovery of the three planets by applying ADI technique to the Keck (upper
left) and Gemini (upper right) images.
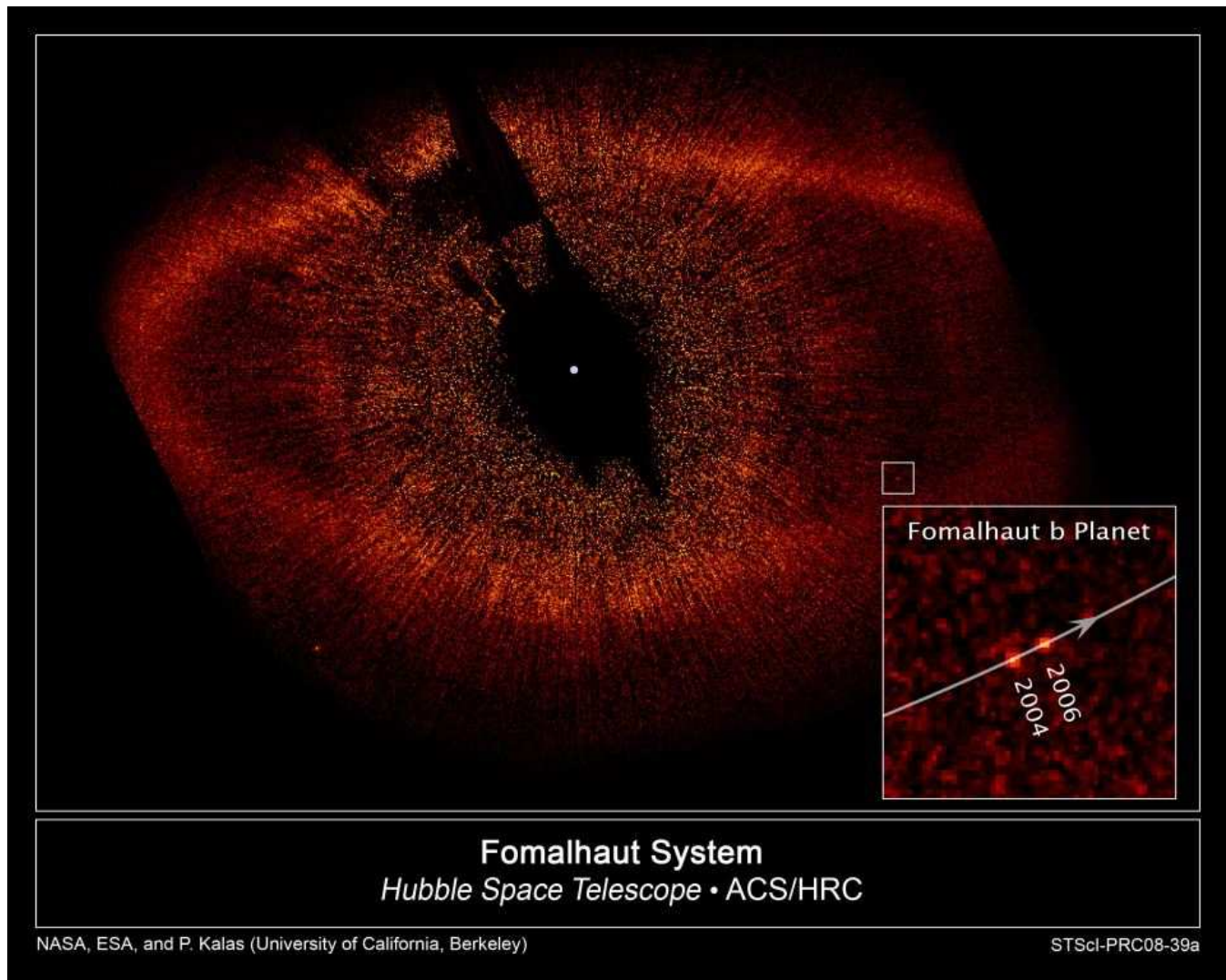
Figure 1.8: HST combination image of Fomalhaut with its debris disk and the position of the planet in 2004 and
2006. Courtesy: Kalas et al., 2008.

1.3.4   Astrometry

Astrometry is the exact measurement of the position of an object in the sky. A planet's
gravitational interaction on its parent star will induce a periodic wobble in the sky position of
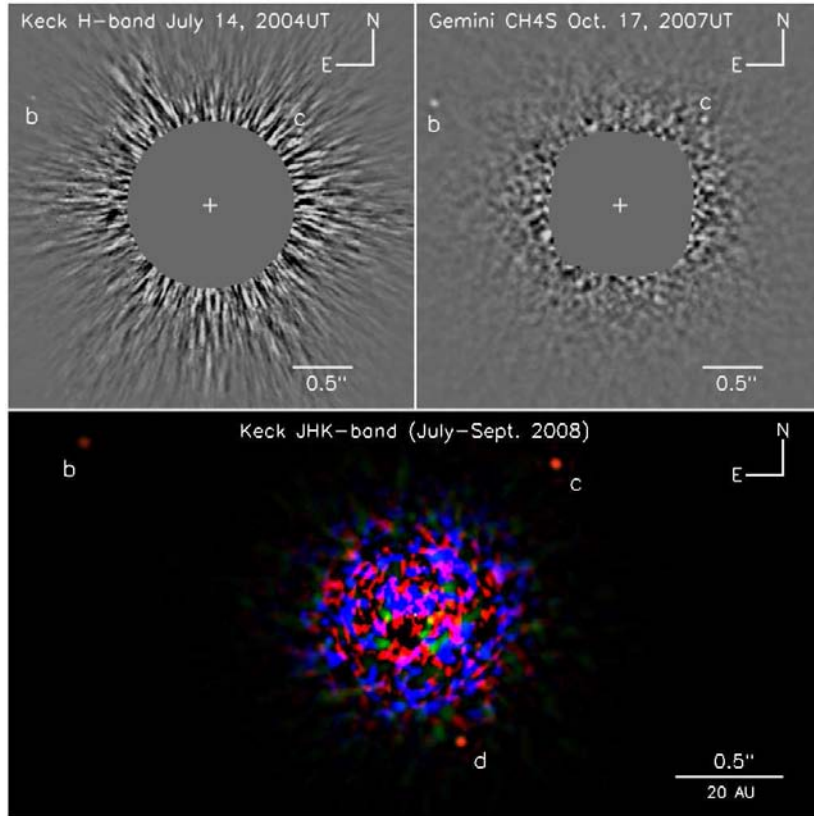
Figure 1.9: The three planet system directly imaged using the Keck and Gemini telescopes. The upper left image is taken from the Keck and the right from the Gemini telescopes. The host star is removed in both the images by ADI processing. Courtesy: Marois et al., 2008.

the star. This is used in predicting the presence of the exoplanet. Astrometry is sensitive to all orbital inclination and hence can be used in the determination of the inclination of the planet-star mass ratio. Owing to its sensitivity to large orbital radii, it can act as a complement to both RV transit techniques and transit measurements. The amplitude of astrometric signature, θ can be written as,

$$\theta = \frac{m}{M} \frac{r}{L} \tag{1.2}$$

where, θ is in arcsecs, m and M are the mass of the planet and the host star in Solar masses respectively with r being the planet's orbital radius in AU and L as the distance to the star in parsec.

This method has not been very successful so far, but this scenario will change with the launch of missions like *GAIA*, which will measure the position of one billion stars to a precision

level of 20 - 25 microarcseconds. This mission is projected to find thousands of exoplanets. At present only *Hubble* telescope is used in astrometric observation of exoplanets, for example, Gliese 876b at a period of 61 days with a mass of $2.27 R_{Jup}$ (Benedict et al., 2002). Like the radial velocity surveys, the information about the characteristics of the planet derived from this method is also limited.

1.3.5 Transits

Transits occur when star light is partially occulted by a planetary body. This phenomenon can be observed if the system is aligned with its orbital plane in the line of sight direction. For a Jupiter-size planet orbiting close to its parent star, the probability of a transit is 10%. However, in the case of an Earth-like planet at 1AU, this probability of transit is less then 1% (Borucki & Summers, 1984). The change in stellar intensity is a few percent during the transit. For a Jupiter sized planet around a Sun like star, it is about 1% and for Earth sized planets it is only around 0.01%. These two conditions posed a technical challenge, which had to be overcome before an exoplanetary transit can be observed. Since then, the number of planets discovered by this method has increased considerably. Currently, the roles are reversed, that we have the capability for very high precision photometers like the *KEPLER*, *CoRoT*, *MOST*, etc., but no supporting radial velocity techniques to confirm the claims made by these programs.

The first observation of a transiting planet was made by Charbonneau et al. (2000), around the star HD 209458 (Fig. 1.10). This planet was already discovered by radial velocity measurements and was later confirmed by transits. From a transit light curve, it is possible to determine the ratio of planetary and stellar radius as given by Eq. 1.3, where F is the flux, ΔF is the loss in flux during mid-transit, R_p is the radius of planet and $R_\star$ is the stellar radius.

$$\frac{\Delta F}{F_0} = \left(\frac{R_p}{R_\star} \right)^2 \qquad (1.3)$$

It is also possible to calculate the orbital period of the planet, P and the projected orbital inclination, i from transit measurements. Observing spectroscopically, we can determine the period (P) and $M \sin i$ values (projected mass of the planet) of the system. Stellar radius can
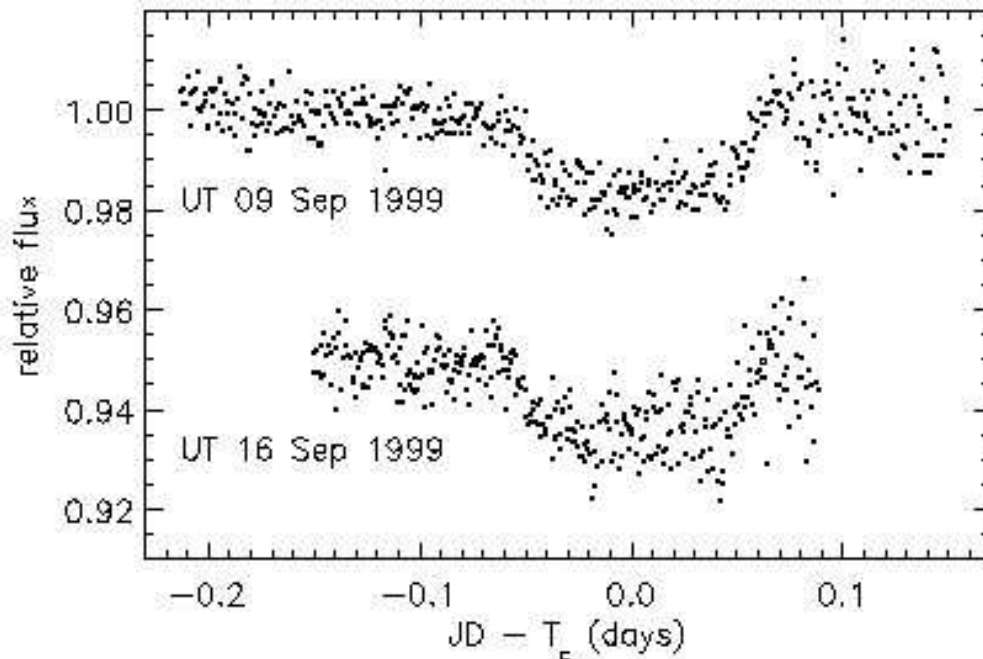
Figure 1.10: The transit light curves obtained by Charbonneau et al. (2000) for HD 209458 for two consecutive transits. The normalised flux intensity is plotted against normalised Julian date. This $0.6M_{Jup}$ planet with an orbital period of 3.5 days shows a dip of around 2% in the stellar intensity.

be calculated by various methods. It can more accurately calculated either in the presence of stellar spectra or through asteroseismological inferences. However, even without the presence of spectral information, the radius can be estimated to the first degree knowing the luminosity of the star. Combining all these measurements, it is possible to extract many parameters which describe the system, like the planetary and stellar radius, their mass, orbital period of the planet, planetary density, atmospheric composition etc. In many current transiting projects, the stellar parameters are constrained by the study of stellar interiors using asteroseismology. The idea of using the same high precision data for the two purposes, enables us to arrive at a better estimate of the planetary radius and thus is becoming the norm. It is important to note that without the transit measurements, these parameters can only be estimated and not accurately known (see Winn, 2009 and Kipping, 2010 for a detailed discussion on the parameters which can be derived from transit light curve). It is important to stress that at present, the most well studied exoplanetary systems are those which can transit, for example, HD 209458, HD 189733, etc.

It is possible to extract not only the physical parameters of the planet like the mass, radius (and thereby density), orbital parameters, but also the atmospheric composition, scale height

and profile of these transiting planets from spectrophotometric observation. Differential spectroscopy of the spectrum of the star and the planet together during the primary and secondary eclipses can give a wealth of information about the atmospheric parameters. The varying radius of the planet in different filters gives us information about the different atmospheric layers and thus helps us to create an atmospheric temperature profile as well as find its constituents (Seager, 2008). One such well studied object is HD 189733b with a $M \sin i$ of 1.15 M_J and an orbital period of 2.2 days around a K1 star (Bouchy et al., 2005). This exoplanet was observed by *Spitzer* and found to contain methane, water and traces of other molecules (Swain et al., 2008; Tinetti et al., 2007). Another planet whose atmosphere is well studied is HD 209458b around a G0V star, where traces of silicon, oxygen and carbon are found (Vidal-Madjar et al., 2004). This planet is so close to the star (0.045AU) that the planet is tidally locked with the star and has an extended evaporating atmosphere (Vidal-Madjar et al., 2003). An understanding of these parameters of the exoplanets, helps in building a stable model for the formation of the extreme exoplanets and also in arriving a statistical model for the different types of planetary bodies. It is important to stress at this point that all these very well studied systems are bright stars with Hot Jupiter companions, which are exactly the kind of systems that we are planning to identify, classify and study using STRESS data. Thus, the result from our study will add to current exoplanet scenario not in just quantity but it will be realitively easy (due to the host star magnitude ranges, period searched, radius of the planetary body detectable in our data) to undertake further in-depth spectroscropic and photometric study, thus adding to the quality of the known systems.

Transit Timing

A fairly recent off-shoot of the transit studies, is the identification of additional planetary systems through the study of variations in the predicted transit timings of the known systems (Holman & Murray, 2005; Agol et al., 2005). When we can predict the transit timing to an accuracy of a few minutes, it enables us to identify additional planets in the system which are gravitationally coupled with the transiting planet, by systematic variations in the start of the transit. There is no necessity for the additional bodies to actually transit the parent star. This

is a relatively new method where the first predictions where made based on the variations of WASP 3-b for the presence of WASP 3-c (Maciejewski et al., 2010). This planet is yet to be confirmed from radial velocity measurements. The timing variations depend on the exact geometry of the system as well as on the masses of the different bodies. This method has the capability of detecting Super-Earths from timing variations around Hot-Jupiters (see Payne, Veras & Ford, 2010 for a review of the current scenario.)

1.4   Future Prospects

The field of exoplanets is still very young and the prospects are endless - in terms of science to be learnt and technologies to be developed. It is essential that we first understand the frequency of the types of exoplanets before making valid models on their formation and evolution and thus predict their presence, in order to direct searches. Still, it is not a blind search either, since the goal of many of the current program is to identify Earth-like planets in the Habitable Zone (HZ) of the host stars. This led to searches around sun-like stars, typically, F, G and K spectroscopic types and also around M-dwarfs. Currently, the search for planets around M-dwarfs have intensified (Quirrenbach et al., 2011; Reiners et al., 2010). This trend stems from the fact that owing to the smaller stellar radius, the influence of a planet around it would be more prominent and hence it would be easy to identify the planet by both RV as well as transits as well as the fact that they are the dominant stellar population in our galaxy (Bonfils et al., 2011). Recently, Israelian et al. (2009) observed that planets are being increasingly found around stars which have a lithium abundance less than the Sun. However, this was also disputed by a further study of 117 stars by Baumann et al. (2010), who found no evidence of correlation between lithium depletion and presence of planet. They rather find that loss of lithium in the star increases with age and suspect that systematics present in the data led to the conclusion by Israelian et al. (2009). Also, we are searching for only mainstream planets which from theoritical constraint can form and sustain a planet in its HZ during its life time and in effect ignoring planets in extreme environment like pulsars, free-floating planets, etc. Methods are currently in place or being developed to identify planets in these systems, namely Gravitational microlensing, Pulsar timing, etc. Future missions like *GAIA*, *PLATO* and *JWST* have a bigger role to play in this field due
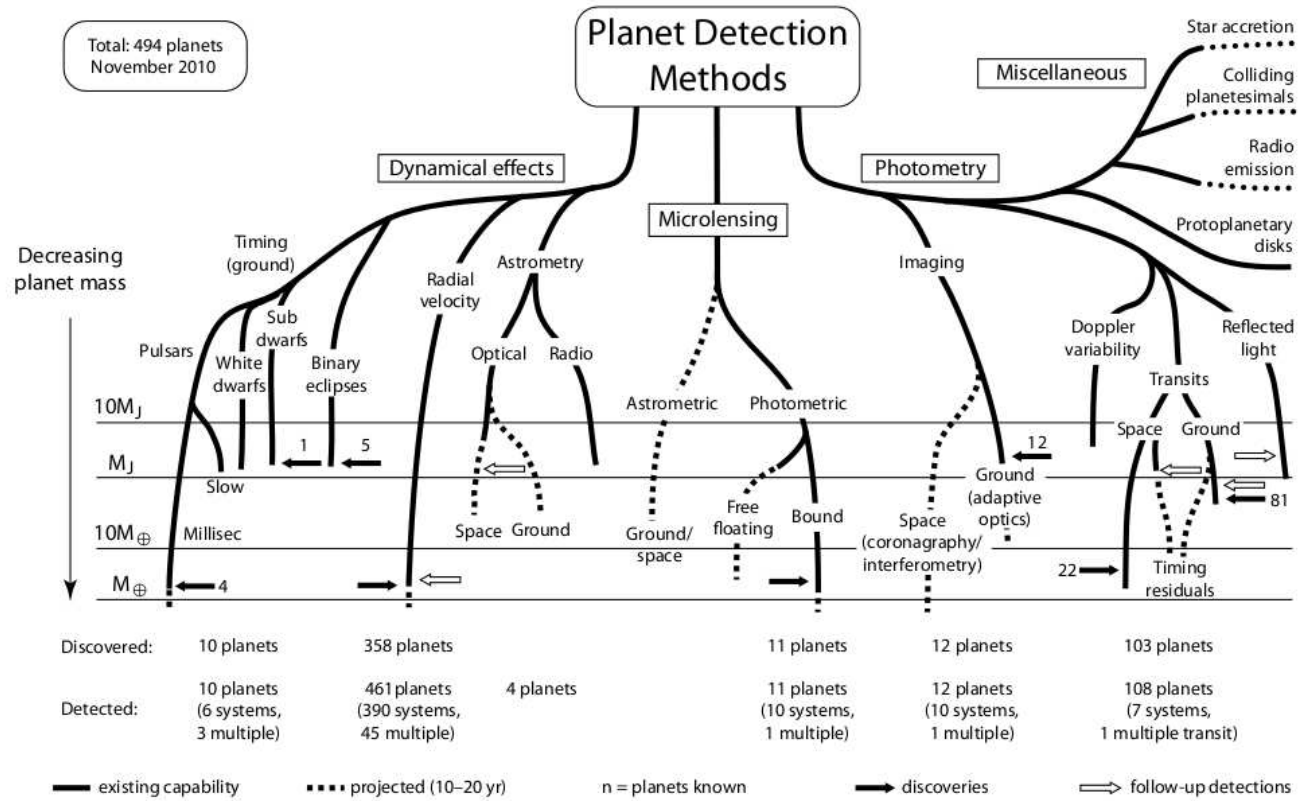
Figure 1.11: An illustration of the various known detection methods, their yields, updates as well as the probable new methods which can contribute to the field. Courtesy: Perryman et al., 2005 (original), 2010 (updated).

to each of its individual technical specifications. For example, *GAIA* will provide astrometry to nearly a billion stars down to $V = 20$ with an accuracy limit of $20\mu as$ and has a capability to detect tens of thousands of exoplanets[5]. The MIRI (Mid-Infrared Instrument) in *JWST* will play a role similar to that of the *Hubble* in terms of follow-up of potential planets and planetary disks. Fig. 1.11 gives an overview of the current and future prospects in this field.

Some of the projected values for the various parameters like the planet frequency, radius, host parameters, etc are already becoming outdated by missions like *KEPLER*, *COROT* and also by the modifications to methods and instruments from the ground, like novel reduction methods such as ADI in direct imaging, etc. All these emphasise the need and the important role these searches play in the current and future exoplanetary studies. In spite of the multitude of projects currently active, from Table 1.1 and Fig. 1.2, it is evident that the bright stars are still left off from these searches. Hence the need for projects like STRESS, which can play a very important role in this circumstance. When we compare *STEREO* with space-based mission like *COROT*, even with a higher cadence of about 512 seconds, it covers less than a percent of the sky and is limited to magnitudes fainter than 11. *STEREO* HI-1 can cover nearly 20% of the sky while HI-2 covers around 50% of the sky, for a magnitude range brighter than 12. Thus, it is evident that this project using *STEREO* plays a very significant role in today's scenario due to its large coverage in terms of magnitude space and area of sky. It becomes more efficient when the light curves from both the *STEREO* instruments are combined. This is also evident from Fig. 1.4 - the clear absence of any known exoplanet in most of *STEREO* FOV. The forthcoming chapters in this thesis describes the *STEREO*-HI instruments, their capabilities, characteristics of the data which enables these searches to be conducted as well as some of the early results from those data.

1.5   Stellar Variability

An added advantage of many of these photometric planet search programs is the high quality of data available - in terms of precision, sky coverage, cadence, magnitude range, etc., for

---

[5]http://sci.esa.int/science-e/www/object/index.cfm?fobjectid=31197

additional science goals. This is evident by the mammoth database of variable stars which is available to the scientific community at large, derived from the exoplanet search programs (Balona, 2011; Balona et al., 2011; Nataf, Stanek & Bakos, 2010; Norton et al., 2007; Debosscher et al., 2009; Lebzelter, 2011). The study of these periodic variables offer us an insight into the stellar properties, the different evolutionary stages of the objects, aids us in the calculation of stellar ages through seismology, derivation of stellar radii and masses through binary system studies, provided the first method of determining the distances to objects in the universe and many more. Similar to the amount of information obtained from them, the types of variables are in-numerous. The high quality data that is available in the past decade has resulted in the discovery previously unknown types of variables (see Eyer & Mowlavi, 2008 for a detailed review).

All stellar bodies are variable at some level. Those objects that exhibit either a periodic or semi-periodic variability in the observed intensity are known as variable stars. The photometric sky coverage and cadence of the *STEREO* HI instrument is well suited to the identification and analysis of different types of variable stars. One of the two main science goals of this project is to collect photometric information and classify these variable objects to the best of the available resources. Variables are broadly classified into *intrinsic* and *extrinsic* variables based on their physical characteristics (refer Table 1.2). According to the General Catalogue of Variable Stars (GCVS)[6], they are broadly grouped into eight different types with sub-types therein. All the variability discussed here in this section and in the rest of the thesis refer to photometric variation. However, variations in their stellar spectra are not included nor discussed here. Some of the most prominent well studied variability types are briefly explained below. It is to note that this list is non-exhaustive in all respects and also there can be objects which fulfill more than once criteria as well.

1.5.1 Intrinsic Variables

This group of objects vary in their intensity due to changes in their internal physical properties. These changes may cause sudden increase in their intensity like sudden explosions in their

---

[6]http://www.sai.msu.su/groups/cluster/gcvs/index.htm
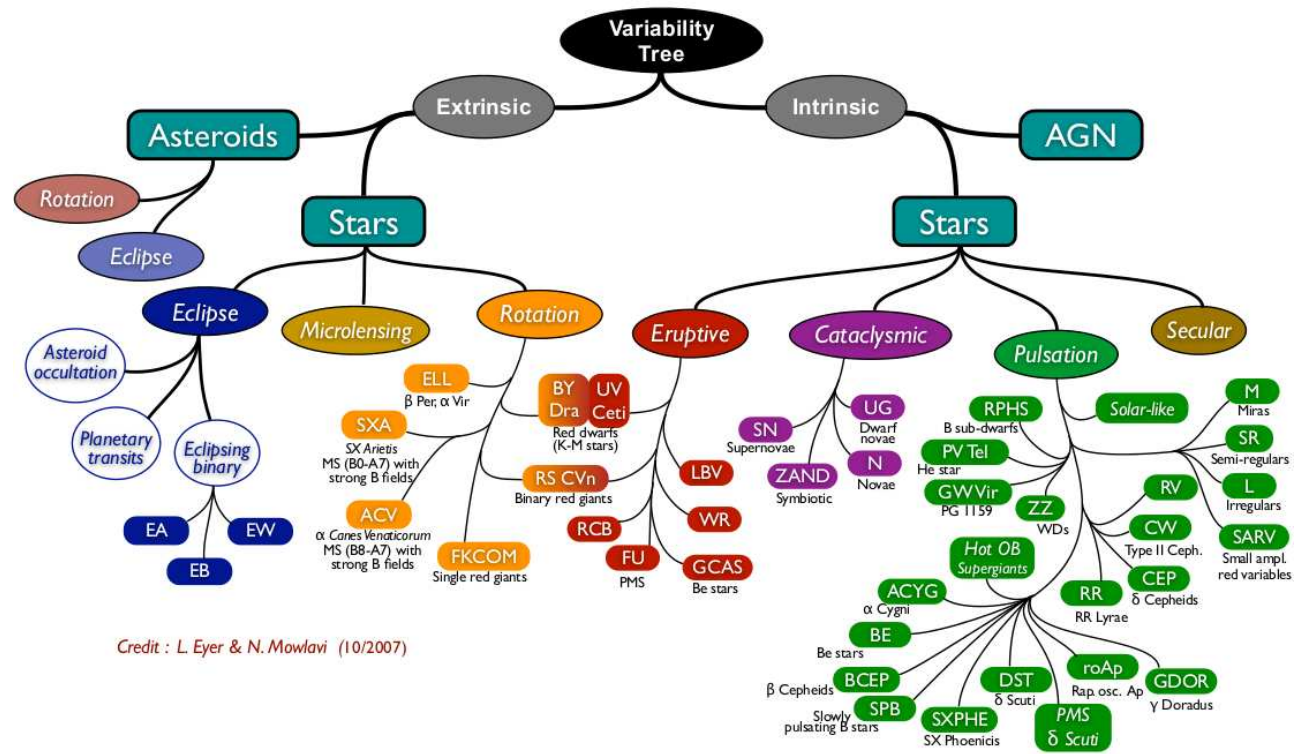
Figure 1.12: An organisation of the types of stellar variability. Courtesy: Eyer & Mowlavi (2008).

Types	Spectral Class	Periods	Change in magnitude (Δm)	Example
Eruptive				
Flare Stars	K to M	minutes to hours	3.0 to 4.0	UV Cet
FU Orions	A to G	years	several magnitudes	FU Ori
Pulsating				
RR Lyrae	A2 to F6	1 h to 24 h	0.0 to 0.5	RR Lyrae
Classical Cepheids	F6 to K2	1d to 50d	0.5 to 6.0	Polaris
W Virginis	F2 to G6	2d to 45d	0 to 3.0	W Vir
RV Tauri	G,K	20d to 150d	3.0 to 4.0	RV Tau
Semi-regular	M,R,C,N,S	20d to 2000d	1.0 to 3.0	Betelgeuse
Mira	M	months	2.5 to 11.0	Mira
β Cepheids	O,B	2h to 6h	0.01 to 0.3	β Ceph
δ Scuti	A0 to F5	1h to 3h	few hundredths to 2.0	δ Scu
γ Dors	F	0.5d to 3d	0.05 to 0.1	γ Dor
Eclipsing				
Algol	all	0.1 to 10000d	0.1 to 3.0	ε Auri
β Lyrae	O to A	0.5d to 200d	0.1 to 1.5	β Lyrae
W Ursae Majoris	F,G,K (Dwarfs)	0.2d to 0.8d	0.2 to 1.0	W UMa
Rotating				
α^2 CVn	B9 to A5	0.5d to 100d	0.02 to 0.8	CU Vir
BY Draconis	K - M	2d to 10d	few hundredths	BY Dra
RS CVn	F,G	1d to 10d	0.5 to 1.0	RS CVn
Solar - like Oscillations	Sun-like stars[1]	minutes to several hrs	n/a	Sun

Table 1.2: The different types of variable known, their general host spectral types, period ranges and examples are given in the above table.
1 - Sun-like stars refer to stars whose oscillations are stohastically excited in the convection zone of the stellar structure. See Bedding & Timothy (2011) for a review of this type of oscillations.

atmosphere. These variables also exhibit pulsations - alternative contraction and relaxation of the whole star. This class of stars are a wealth of information as they provide insight into the evolution of the star through its changing internal structure. Stars like Cepheids also fall into this category, which provides very accurate distance measurements. This class can be further sub-divided into *Eruptive* and *Pulsating* variable stars.

Eruptive Variables

These stars vary in intensity due to violent physical activity, such as eruptions or flares which occur in their chromosphere or corona. Some well known eruptive variables are T-Tauri, UV Cet and also some Wolf-Rayet stars. These changes are generally intense but short-lived. There are also another type of eruptive variables which are called as *Cataclysmic Variables*. They vary in brightness due to thermonuclear processes. Supernova and nova come under this category. Supernova are instances when the stellar core collapses, shedding the outer layers with very high velocities and this leaves behind a remnant Black hole or a Neutron star. The object at this instant increases in brightness many fold and outshines the parent galaxy itself. On the other hand a Nova causes a sudden increase in brightness of more than five or six magnitudes within a few days but slowly relaxes to its pre-nova intensity. For example, the RS Ophiuchi system, which is a recurrent nova, which changed its magnitude from the regular $V = 12.5$ to $V = 5$ during the outbursts. The changes in this system is due to the accumulation of matter by a white dwarf and the recent outburst happened in 2006 (Hernanz & José, 2008). A few other well studied cataclysmic variables are U Gem, SU Ursa Majoris, SS Cygni, etc (Mattei et al., 1987; Warner, 1995). Another type of Eruptive variables are the Wolf-Rayet stars. These are evolved massive stars, that exhibit strong stellar winds of upto 2000kms^{-1} and lose nearly 10^{-4} to $10^{-5} M_{\odot}$ (Cherepashchuk, 1992).

Pulsating Variables

These stars show periodic expansion and contraction of their surface layers (as the name suggests - pulses) and thus vary in their luminosity. The variations can be both periodic and semi-periodic. This class contains a large variety of pulsating stars starting from the very long period, high magnitude changing Miras to the very short period δ Scuti's to the Solar-like oscillations observed in red giants and other stellar type Christensen-Dalsgaard (2011); Hekker (2010b); Bedding & Timothy (2011). Some of the first known variables belong to this type, like the Cepheids, Miras, RR Lyrae stars, δ Scuti, γ Doradus, etc. Refer to Fig. 1.13 for a schematic of the various pulsating stars represented in the Hertzsprung-Russell (H-R) diagram. The Cepheid variables have a strict relationship between its period and luminosity, thus acting
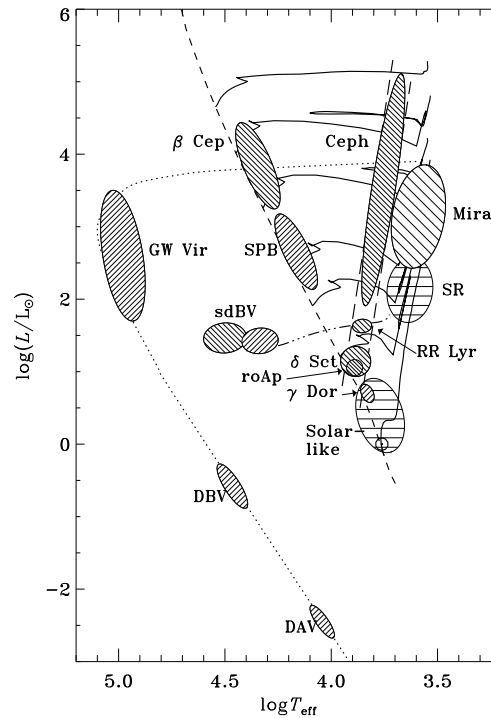
Figure 1.13: A schematic representation of different pulsating variables in the H-R diagram (Christensen-Dalsgaard, 1998).

as distance markers or *standard candles*. Fig. 1.14 shows the light curve of a classical Cepheid - HD 154365, which was observed by the HI-1A instrument. The left plot is a portion of the light curve and the insert contains the data available for three epochs with the HI instrument. A phase folded light curve of the star at a period of 4.05 days is shown on the right. The Mira type of variables are those that vary their magnitude to such large extent that some of them alternate between being visible and invisible to the naked eye. Another well known pulsating variable type is the RR Lyrae type stars. They are short periodic and hence are also known as short period Cepheids.

1.5.2 Extrinsic Variables

In this case, the variation in the observed intensity occurs due to the geometrical arrangement mainly between the observer and the target and also the brightness variation caused by features on the surface of the star due to rotation. The former are known as the Eclipsing binary stars, but can also sometimes contain more than two stellar bodies and the latter as rotational
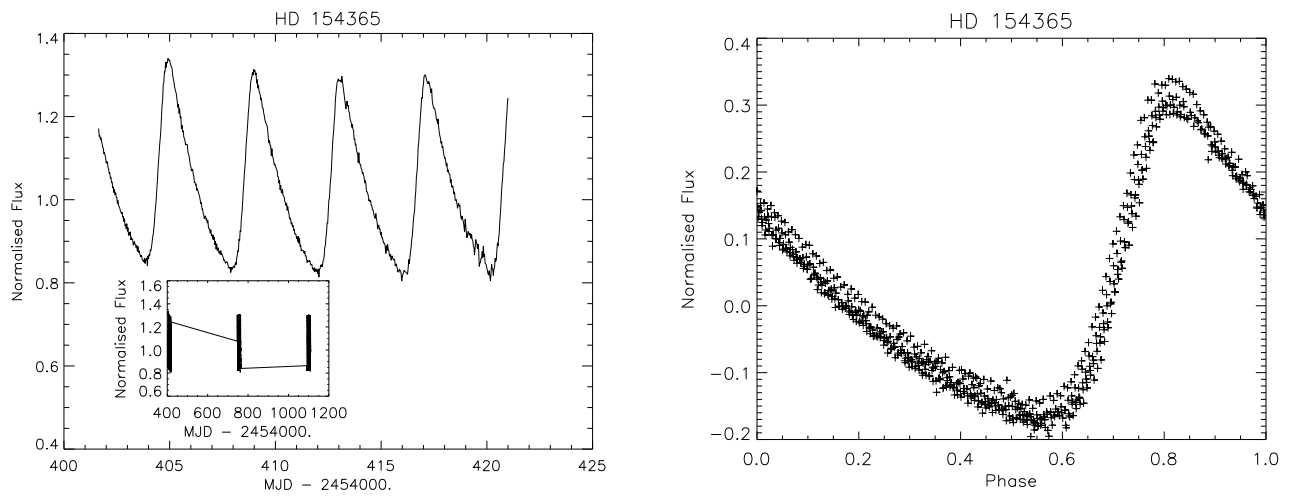
Figure 1.14: STRESS light curve of the classical Cepheid variable - HD 154365, a G0 star at $V = 7.37$. The insert on the left shows the data availabilty for three epochs. The plot on the right is the phase folded light curve with a period of 4.05 days.
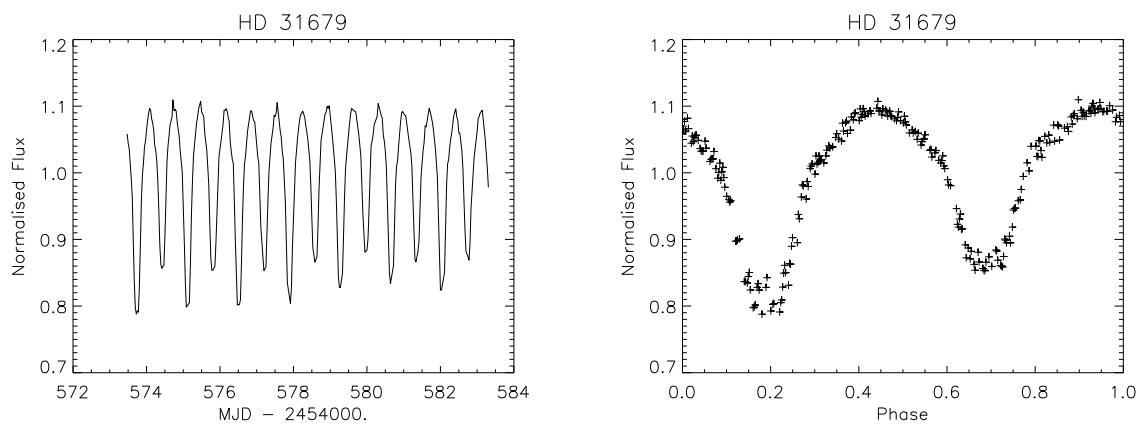
Figure 1.15: STRESS light curve of β Lyrae type eclipsing binary HD 31679 with a $V = 7.95$. The right plot is folded on a period of 1.38 days.

variables.

Eclipsing Binaries

These stars, as the name suggests, are binary stars which eclipse each other and thus a variation in light intensity is observed. The intensity of these objects vary depending on the two components of the system as well as the distance between them. A very well known type eclipsing binary is the Algol type. One of the stars in this configuration is more massive and belongs to the main sequence and the less massive is usually a sub-giant star. The two stars are

semi-detached [7] and has a characteristic light curve with prominent eclipses (Budding, 1986).
Another is the β Lyrae type where the components exactly fill their Roche lobe and hence appear ellipsoidal (Wilson, 1979). An example of this type of binary as observed by *STEREO* HI-1A instrument is shown in Fig. 1.15. This is a $V = 7.95$, B5 star at around 216 parsec with a binary period of 1.38 days. Some binaries like the W UMa have two stars which are in contact and hence are highly deformed and they fill their Roche lobes completely. Hence their intensity variation is very different from the other two types.

Rotating Variables

Rotating variables are stars that show variation in their intensity due to stellar rotation when the star spots or bright patches on the surface of the star changes its orientation with respect to the observer. Some of them change their intensity due to magnetic effects and they are known as magnetic variables. Some well known rotating variables are CU Vir, CM Tau, BY Dra, etc. There is another subclass of magnetic variables called the Magnetic Chemically Peculiar (MCP) stars. Another subtype is the RS CVn, which is also a binary with 2 sub-giant components. Hence these provide interesting platforms to study the stellar evolution of binary systems as well as the individual stellar evolution.

1.6   Summary and Final Thoughts

The study of stellar variabilities is in its prime at the present time. This includes the exciting discoveries made in the field of extra-solar planets as well as the progress in understanding of other types of stellar variables. The large number of ground and space based programs are increasing and we are observing these changes in an array of different wavelengths. However, the field is still in its infancy interms of a statistical analysis needed for generalisation of the various process and thus arrive at a conclusive formation and evolution model for the exoplanets as well as the variable stars. The contribution of *STEREO* instruments for this purpose is highly significant. The continuous monitoring of a large area of the sky covered in this specific wavelength

---

[7]Semi-detached are systems where one of the objects fills its Roche lobe and other object does not.

(R) is very useful in terms of the photometric points it can contribute to the field (as a rough estimate, this is around a billion data points). This thesis deals with the development of the *STEREO* HI instrument data for these two specific purposes. Chapter 2 describes the spacecraft and the instrument, Chapters 3 and 4 describes the photometric reduction pipeline developed for this process. The basic characteristics of the data will be discussed in Chapter 5, along with some of the systematic noise removal techniques used. Material covered in Chapters 3,4 and 5 has been accepted for publication in MNRAS, titled - STRESS - *STEREO* TRansiting Exoplanet and Stellar Survey: Introduction and Data reduction pipeline - I (Sangaralingam & Stevens, 2011). Chapter 6 defines the stellar variability analysis pipeline and the preliminary results derived from the data. The complete data is due to be published as a catalogue in the near future, a sample of which is given in the appendix. Chapter 7 explains the exoplanet transit analysis pipeline and first results obtained. This analysis is being prepared for publication in the near future as Whittaker et al. The thesis is summarised with further prospects for this data in the final chapter.

Chapter 2

STEREO - HI

2.1   *STEREO* as a spacecraft

The Solar Terrestrial Relations Observatory - *STEREO*, is a joint NASA and ESA mission. It was launched on October 25, 2006, aboard a Boeing Delta II rocket from Cape Canaveral Air Force station in Florida. It is a system of two identical spacecrafts located in the ecliptic plane, in a heliocentric orbit at a radii of 1 AU. The two spacecrafts, *STEREO* - Ahead (A) and *STEREO* - Behind (B), will be leading and lagging Earth in its own orbital path, respectively. Each of the spacecrafts separate from Earth at an angular rate of 22° per year. Fig. 2.1 gives the orbit of the two spacecraft since launch till date with respect to the Sun-Earth line (Kaiser et al., 2008). Each spacecraft weighs nearly 600kg and is 6.67m wide when the Solar arrays are deployed. It is capable of an attitude control within 7 arc-seconds, thus enabling very precise pointing compared to its CCD resolution.

This mission was designed with an aim to observe, study and understand Coronal Mass Ejections (CME) from the Sun, their motion and effects in the interplanetary medium and to enable a stereoscopic reconstruction of the Sun and its atmosphere. It offers in-situ measurements of the energetic particles in the region within 1 AU. Since its launch, it has also helped in the study of many comets and Solar system bodies which has crossed its Field-of-View (FOV), offering an insight into the impact of the various coronal components on these bodies as a function of distance, chemical composition of both the bodies and the energetic particles from the Sun, magnetic field interaction, etc. (Hess & Zhang, 2011, example.).

The two spacecrafts contain remote sensing and in-situ measuring instruments to provide the first-ever stereoscopic images of the Sun and the interplanetary medium in-between the Sun and Earth. It has helped in advancing our understanding of the physics of CMEs (Colaninno,
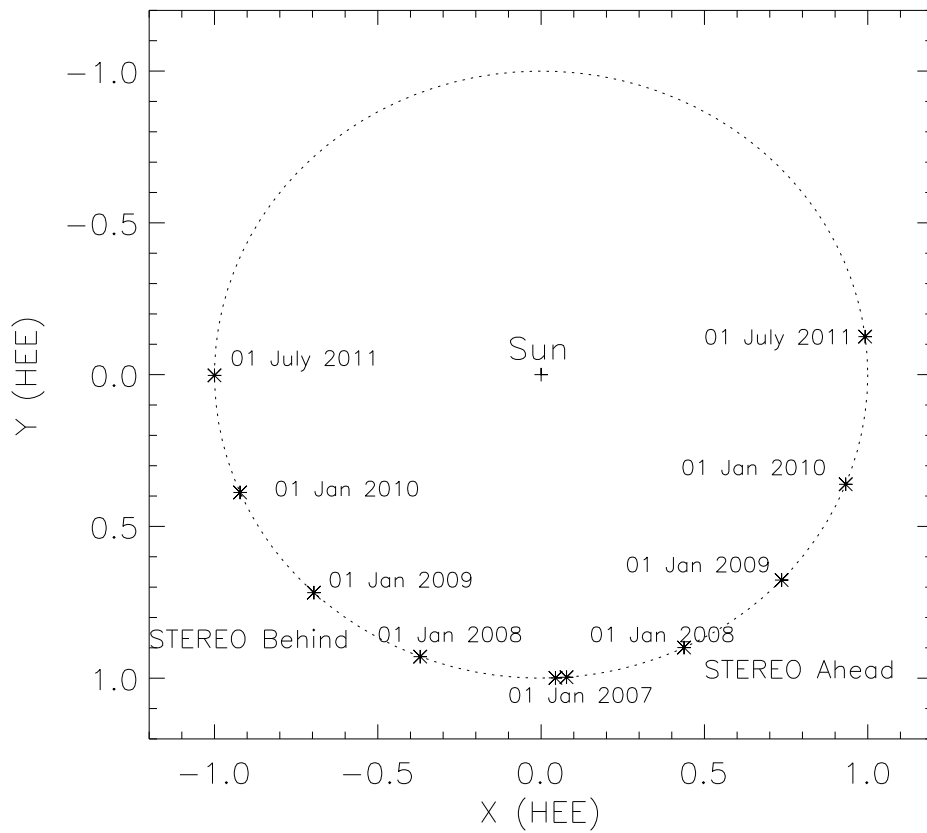
Figure 2.1: The respective position of the *STEREO*-A and *STEREO*-B spacecraft with respect to the Sun and Earth is illustrated. Both the spacecrafts separate from the Earth at an angle of 22° per year.
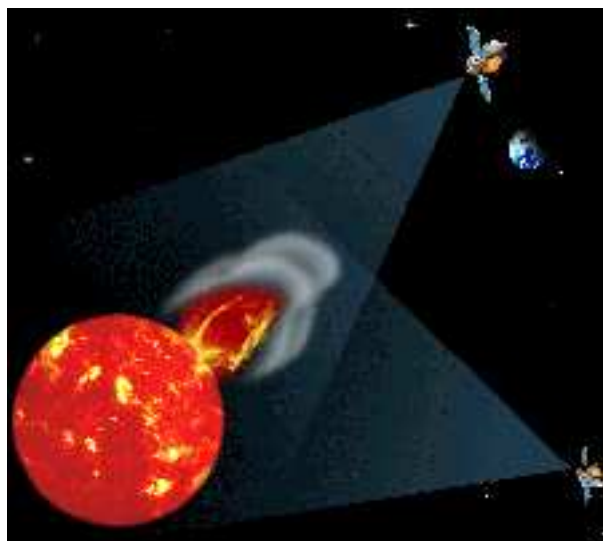
Figure 2.2: An artist's image of the *STEREO* mission concept.

Vourlidas & Nieves-Chinchilla, 2011; Vourlidas, Syntelis & Tsinganos, 2011) and radio emission processes (Gopalswamy et al., 2009) so that we can improve the accuracy and range in predicting the geoeffectiveness of a CME, i.e., the impact and consequences of a CME impact on Earth's atmosphere and to life activities on the planet. There are four instrument packages which are employed to study the nature of CME:

(a) STEREO Waves (SWAVES) : This set of instruments will track the CMEs in radio through the interplanetary medium. SWAVES probes the CME from lift-off to Earth by detecting the coronal and interplanetary (IP) shock of the CME's. It has also determined the radial velocity of these shock waves through spectral imaging. It is also capable of measuring in-situ the properties of the IP shock, magnetic cloud and density variations in the fast Solar wind stream[1].

(b) In-Situ Measurements of Particles and CME Transients (IMPACT) : This instrument is used to sample the plasma characteristics of the Solar energetic particles and the local magnetic vector component. Its objective is to understand the causes and mechanism of CME initiation, mechanism and sites of energetic particle acceleration and to develop a 3D time-dependent model of the magnetic field. It consists of seven instruments to enable an overall investigative approach[2].

(c) Plasma and Suprathermal Ion Composition (PLASTIC) : This is used in the study of the characteristics of the individual plasma particles like protons, alpha particles and heavy ions from the Solar wind which impacts on the spacecraft. It will provide key measurements, like the mass and charge state of the heavy ions, in order to characterise the CME plasma from normal coronal plasma (Galvin et al., 2008).

(d) Sun Earth Connection Coronal and Heliospheric Investigation (SECCHI) : This package contains imaging instruments and coronographs to study the evolution of a CME from the Sun surface till it impacts the Earth and makes stereoscopic images of these CMEs. It is comprised of four instruments: An extreme ultraviolet imager, two white-light corono-graphs and a heliospheric imager. The suite has three main parts - SCIP (Sun Centered Imaging Package) made up of three telescopes, the HI (Heliospheric Imager) comprised

---

[1]http://swaves.gsfc.nasa.gov/swaves_science.html
[2]http://sprg.ssl.berkeley.edu/impact/about_impact.html

Figure 2.3: A schematic representation of the *STEREO* spacecraft with its instrument package.

of two telescopes and the SEB (Secchi Electronics Box). It also has a guide telescope which also acts a fine Sun sensor. The SECCHI instruments will be described in detail in Section 2.2.

2.2   SECCHI Instruments

SECCHI package consists of two white light coronographs - COR1 and COR2, an Extreme Ultraviolet Imager (ENVI), collectively called as the Sun Centred Imaging Package (SCIP), a Guide Telescope and the Heliospheric Imager (HI). Refer to Fig. 2.3 for the location of the various instruments on the *STEREO* spacecraft. The whole system with its nested baffles weighs around 16.3kg at launch and measures $800 \times 550 \times 260$ mm in dimension.

(a) Extreme Ultraviolet Imager (EUVI) : This instrument focuses on the initiation and early evolution of the CME in three dimensions. It covers the innermost portion of the FOV,
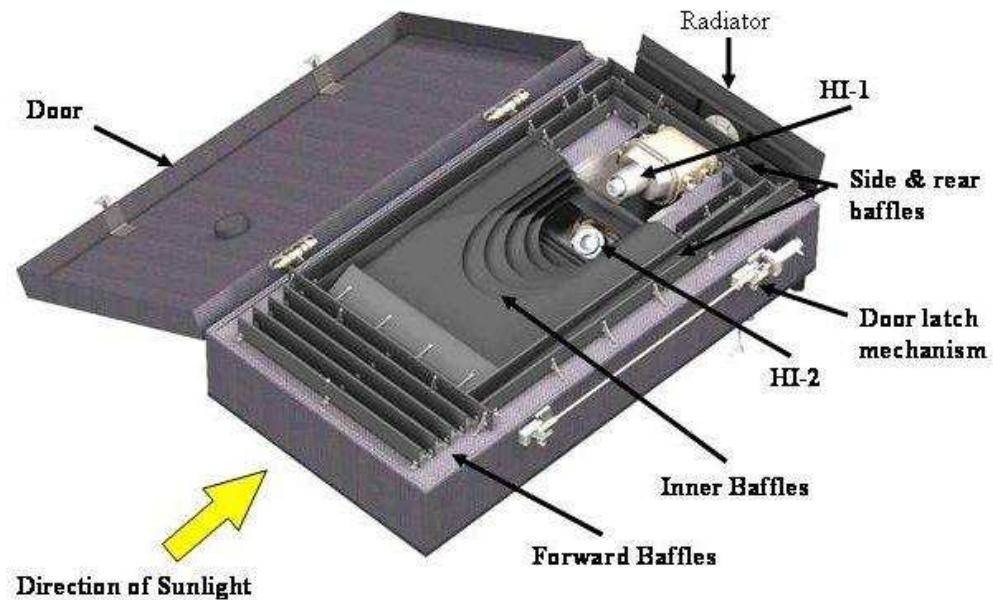
Figure 2.4: The HI camera system with its nested baffle system, measuring $800 \times 550 \times 260$ mm.

from the Solar chromosphere to the inner corona at 1.7 Solar radii ($R_\odot$; Wuelser et al., 2004). It images this region in four different emission lines characteristic of coronal temperatures at 17.1nm (Fe IX), 19.5nm (Fe XII), 28.4nm (Fe XV) and 30.4nm (He II) (Thompson et al., 2003). It provides twice the spatial resolution compared to the EIT instrument aboard SOHO spacecraft.

(b) Inner Coronograph (COR1) : This is an internally-occulted reflective coronograph with a FOV ranging from 1.3 to 4 $R_\odot$. It uses a linear polariser to filter out the light scatter and to extract the polarised brightness signal from the Solar corona (Thompson et al., 2003).

(c) Outer Coronograph (COR2) : This is also a set of reflective coronographs observing the region from 2 to 15 $R_\odot$. This studies the corona with five times the spatial resolution and three times the temporal resolution of LASCO/C3 on SOHO[3].

---

[3]www.secchi.nrl.navy.mil
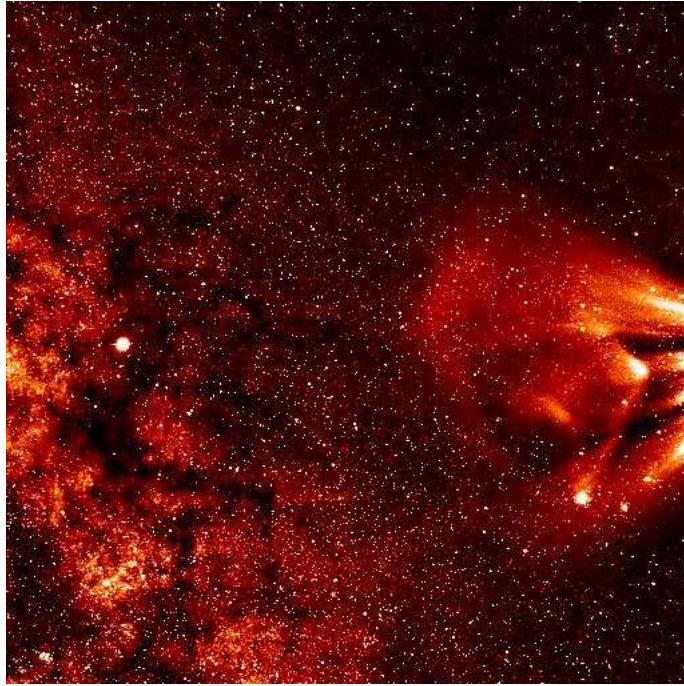
Figure 2.5: Image of the Galactic Plane, Jupiter and a CME taken by HI-1A instrument. Background stars up to a visual magnitude of 11, which are present in the image, will be used as data in the present project.

(d) Heliospheric Imager (HI) : This is described in detail in the forthcoming section.

2.3 Heliospheric Imager

The Heliospheric Imager is a wide angle imaging system designed to study the evolution of CMEs in the interplanetary medium and to enable a stereo-graphic reconstruction of CMEs propagating through the inner heliosphere. It consists of two specialised camera systems in a nested baffle configuration in each spacecraft: HI-1A and HI-1B centred at $13.28°$ from the Solar disk ($\sim 50 R_{\odot}$) with a $20° \times 20°$ field of view and resolution of $35.15''$/pixel. The HI-2A and HI-2B cameras are centred at $53.36°$ from the Solar disk ($\sim 200 R_{\odot}$) with a field of view of $70° \times 70°$ and resolution of $2.05'$/pixel.

Since the cameras are wide angled, they will be facing a large portion of the sky, experiencing a huge variation in the background light levels (the natural sky background B, from near the Sun till Earth varies from 10^{-7} - 10^{-14} times the brightness of the Solar disk -B_0). With these huge variations, it is important to avoid stray light reflection in the CCDs. This is achieved

by using a system of baffles - a linear forward baffle, a perimeter baffle and the internal baffle (Fig. 2.4). They together avoid the stray light from the Solar disk, the photosphere, from the stars, planets, Earth, zodiacal light and also the internally reflected light from the spacecraft. The low instrumental background as a ratio of B/B_0, described in Table 2.1 is achieved by this baffle configuration. Reduction of this background noise also results in the high Signal to Noise Ratio (SNR) achieved by these instruments, enabling the study of extremely faint and potentially geoeffective CME's as well. Thus, these baffle systems play a very important role in the high sensitivity configuration achieved for this HI instrument. A review of the performance of these baffles on-board can be found in Halain et al. (2007). The HI-1 instrument is sensitive to wavelengths ranging from 6500Å - 7500Å which lies close to the R-band of the electromagnetic spectrum, whereas for HI-2 instrument it is a large passband varying from 4000Å to 10000Å ,thus extending into the infrared region from near the blue end of the spectrum. Fig. 2.6 is a plot of the quantum efficiency as a function of wavelength (Eyles et al., 2009). The quantum efficiency reaches nearly 93% at around 550nm and drops down further on either side. The CCD has achieved its nominal operating temperature of $-80°C$ when its doors were opened. The dark current at this temperature is negligible and gives a high tolerance to radiation damage effects.

Fig. 2.7 is a graphical representation of the field of view covered by various SECCHI instruments (Socker et al., 2000). The smaller red circle represents the HI-1 coverage and the larger blue circle represents the field of view as covered by HI-2 cameras with the Sun and position of the Earth during various stages of the mission. It also shows the region viewed by the coronographs on board. We can see that when all the instrument data is covered, we can get a complete picture of the path of CME from when it is ejected from the Solar surface to when it passes the Earth's orbit. Fig. 2.7 also contains the detection threshold for the SECCHI instruments as a function of the elongation angle (angular representation of the Sun - Earth plane).

The HI-1 and HI-2 cameras have individual exposure times (equivalent to single shutter open/close cycle) of 40 and 50 seconds. Thirty such exposures are summed up together on board the spacecraft, after correcting for cosmic ray effects, finally providing an integrated cadence of 40 minutes for each HI-1 image transmitted to the ground. Similarly 99 exposures of
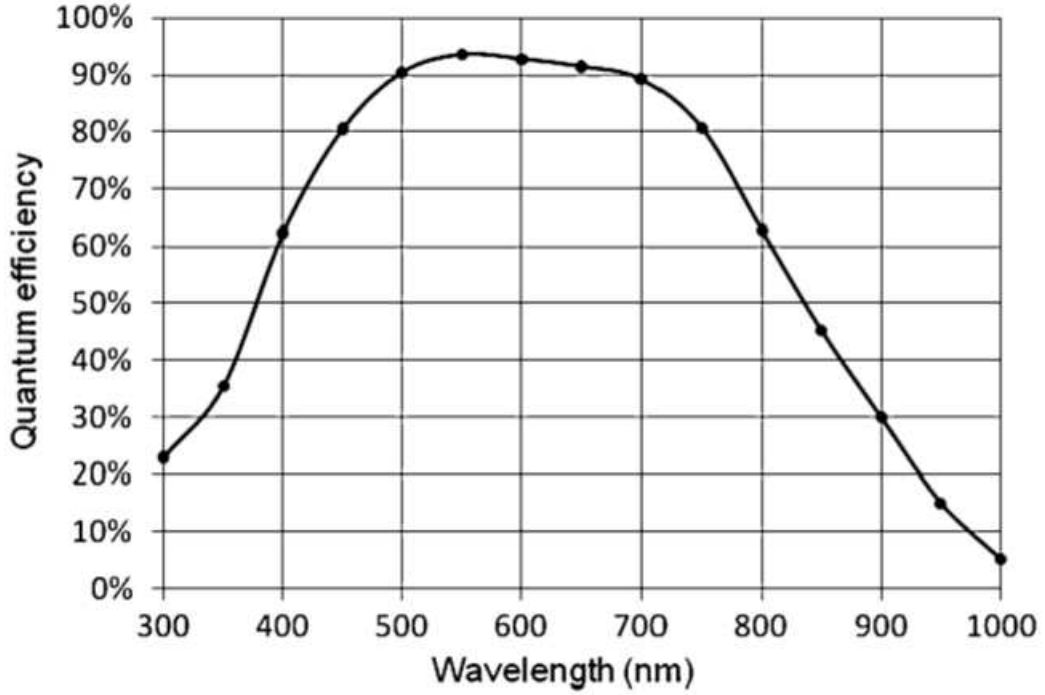
Figure 2.6: The quantum efficiency of the HI CCD measured as a function of wavelength (Eyles et al., 2009).

	HI-1	HI-2
Field of View		
Angular	20°	70°
Centre	$49.8R_{\odot}(13.28°)$	$200R_{\odot}(53.36°)$
Inner Cutoff (unvignetted)	$12.3R_{\odot}(3.28°)$	$72.8R_{\odot}(19.4°)$
Outer Cutoff (unvignetted)	$87.3R_{\odot}(23.28°)$	$332R_{\odot}(88.6°)$
CCD Format	$2048 \times 2048 \times 13.5\mu$m	$2048 \times 2048 \times 13.5\mu$m
Pixel sizes		
CCD	$35.15''$/pixel	$2.05'$/pixel
Image (binned onboard)	$70.3''$	$4.1'$
Telescope parameters		
Aperture	16.0mm	7mm
Focal Length	78.4mm	19.74mm
F-ratio	f/4.9	f/2.8
Passband	**6500Å - 7500Å**	**4000Å - 10000Å**
Instrumental Background	$< 3 \times 10^{-13}B/B_0$	$< 5 \times 10^{-13}B/B_0$
Cadence		
Nominal exposure time	40 sec	50 sec
Exposures per image	30	99
Integrated cadence	**40 min**	**2 hrs**

Table 2.1: HI Instrument Characteristics
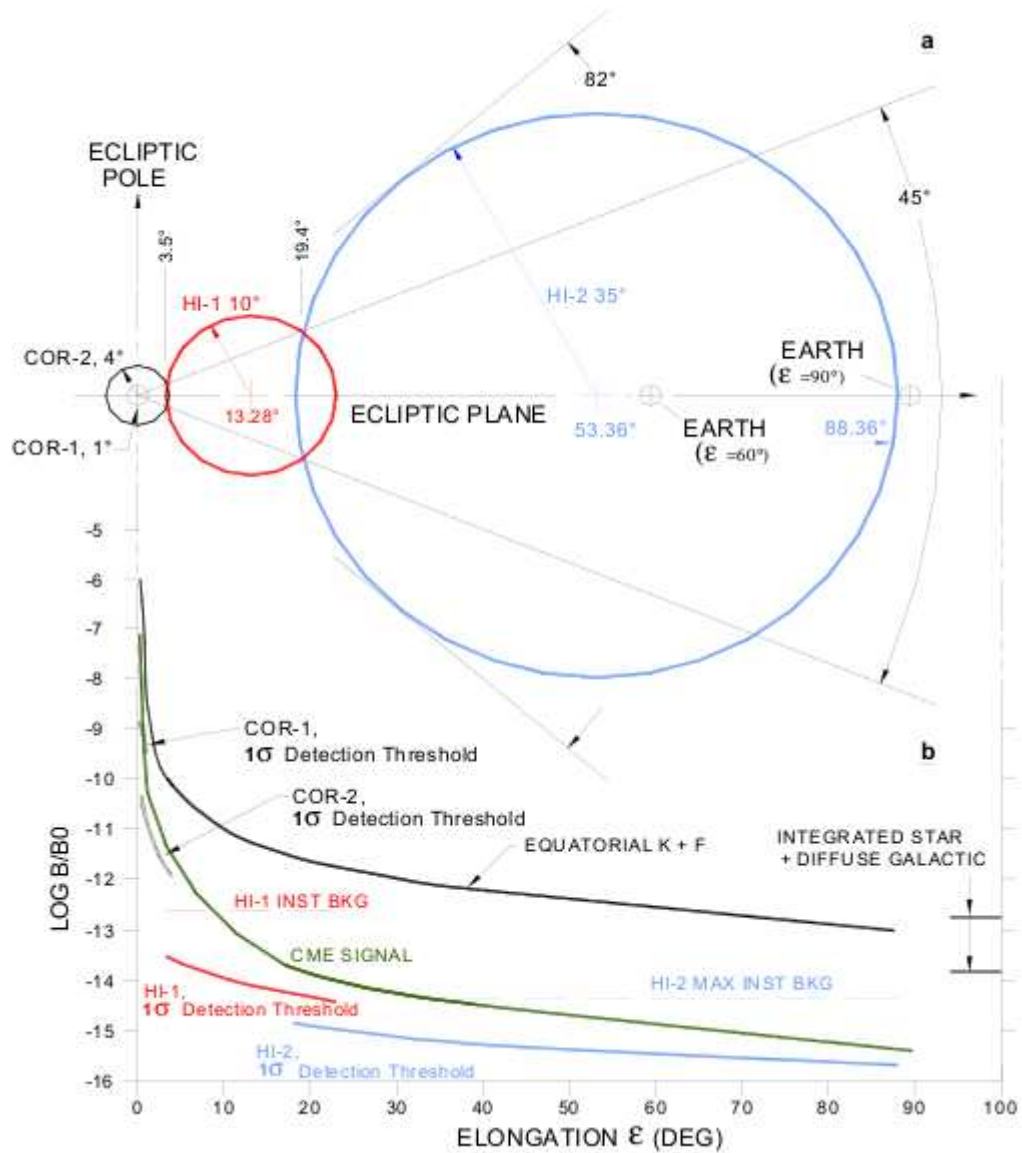
Figure 2.7: This diagram shows the FOV of the various SECCHI instruments and their signal detection conditions (Socker et al. (2000); see text for details). The Sun is at the centre of the black circle and the position of Earth at different points in its orbit is also shown. It also contains the background F-coronal intensity with respect to the detection threshold of the various instruments.

HI-2 are summed to provide a digitally integrated image cadence of 120 minutes (refer Eyles et al., 2009 for detailed explanation). The actual 2048×2048 images are also 2×2 binned, resulting in 1024×1024 images for all the instruments, thus giving a final resolution of $70.30''$/pixel and $4.10'$/pixel for HI-1 and HI-2 respectively.

The project detailing the identification and analysis of exoplanet transits and stellar variability using these four HI instrument is STRESS - STEREO TRansiting Exoplanet and Stellar Survey. In the forthcoming chapters, I will describe the type of images which are downloaded, the data analysis pipeline, discuss the characteristics of the derived light curves and variability analysis of the data with results. In all these chapters, the data used is only from the HI-1A instrument, even though the pipelines and algorithms developed can be modified for use with all HI instruments by changing some of parametric values used, e.g., FOV, aperture radius, etc. The HI-1A instrument has a well defined and stable PSF as well as devoid of many of the instrumental artifacts present in other HI cameras (see discussions in Chapter 3 and Chapter 4). For this reason, HI-1A is preferred for this initial development over the other instruments and the analysis and reduction pipeline using this instrument as an example is explained in the following chapters.

Chapter 3

Basic *STEREO* Data Reduction

3.1 Introduction

This chapter will describe the different types of *STEREO* data images available for download and the various methods which were used to reduce the data in order to do stellar photometry. Here we describe the first stages of the semi-automated photometry pipeline developed with the help of the Rutherford Appleton Laboratory (RAL) using IDL solar soft packages. The initial reduction pipeline was developed by RAL and further methods and analysis developed at Birmingham. Data from all the *STEREO* spacecraft instruments are readily available for download from the UK Solar System Data Centre (UKSSDC) website[1], in an interactive way, allowing the user to download images for various periods of interest from the different SECCHI instruments.

3.2 Pre-processing of data

There are three different levels of processed data available for download- i) raw summed images (L0), ii) semi-processed images (L1) and iii) pre-processed images (L2). As mentioned in Section 2.3, the cameras have no shutter and for HI-1, an image is taken every 30 seconds. Each of these 30 second images are scrubbed for cosmic rays, before 30 such images are summed on board the spacecraft, to give a final cadence of 40 minutes. These are the raw summed images and represented by the tag 's4' in the data file name and are generally of size 4MB. In between the regular science images, the raw, unsummed and unscrubbed images are also transmitted, which are about 8.5MB. These images are used for calibration purposes and tagged as 'n4' in

---

[1]http://www.ukssdc.rl.ac.uk

the image file. Some of these images are transmitted in beacon mode, when the images are used mainly as a warning of solar weather phenomenon. These images are cosmic ray scrubbed, summed and also highly compressed (1MB) which enables them to be transmitted quickly and often. The estimated rate of Cosmic rays (CR) impinging on each CCD is 45 pixels per second. In order to remove the CR, each exposure is compared on a pixel-by-pixel basis. If the value of the pixel is greater than five σ, with σ being the noise standard deviation predicted based on the number of photoelectrons, then the pixel value is replaced with the previous image pixel value. The number of such replaced pixels is represented by the last line in the final integrated image. Most of the information presented here regarding the characteristics of the HI instrument as well as the reduction pipeline developed by RAL are explained in detail in Eyles et al. (2009) and Brown, Bewsher & Eyles (2009).

The L0 images cannot be used directly for scientific purposes as they still need to undergo some basic reduction procedures, which are explained below. All these different corrections are contained in an IDL program *secchi_prep*. This is a common routine for the complete SECCHI suite of instruments.

The down-linked images have the cosmic ray scrubbing values and scanning information in the last row. The first part of the *secchi_prep* routine removes these final row values and replaces them with values from the adjacent row. The next section contains the various corrections and adjustments done to the data during the *secchi_prep* routine in detail.

Missing Data : The data are compressed using Rice loss-less compression method before transmitted to the ground. Due to the overflow integer problem, they are compressed as 'high' and 'low' words separately before transmission. Any loss of communication during telemetry, results in blocks of data being lost. These missing data blocks are removed and replaced by appropriate values if the loss is small and compensable without losing the image quality. In other cases, when a large amount of data blocks are lost, the image is tagged as bad file, explained below.

Bad Files : Certain images are tagged as bad due to various reasons like corrupted data, corrupted file, High/Low word problem when the 'high' and 'low' words, explained above do not match with each other. Another important problem, specially for stellar photometry are files with a split star field when the frame appears to have been shaken and hence it is impossible to identify individual point sources (refer to Fig. 3.1 for an example). Some other problems include integer overflow, a mathematical corruption which occurs during transmission of images as well as when the image is re-stitched back in ground, composite images from multiple cameras and also some of them have an unclassified problem. These are identified using a bad file list provided by the UKSSDC and removed from further analysis. Refer to Fig. 3.1 for an illustration of the different types of bad files tagged during this process. These images are from both HI-1A and HI-1B instrument. For HI-2, the bad files are similar to those ones shown here in Fig. 3.1.

Blooming - The CCD pixels saturate for bright objects and bleed vertically in that column and in some cases even across columns for extremely bright objects. In HI-1, the pixels saturate for objects with a magnitude brighter than 4 ($m < 4$) and HI-2 pixels saturate for magnitudes brighter than 1 ($m < 1$). Since these saturated columns overthrow the shutterless correction, such columns are identified and replaced with NaN values.

Shutterless Read-out - Due to mechanical constraints, the HI cameras do not have a shutter. A shutterless camera with a significant read-out time results in a differential exposure time for different parts of the image. The images are smeared vertically in the read out direction. A well understood mathematical correction of the form given below is applied to the image to remove this effect from the images and used for further processing (see Eyles et al., 2009 for a detailed explanation). In order to correct for this effect we need to accurately calculate the total time taken for the whole read-out procedure (refer Fig. 3.2). The total response R for a pixel (n, m) is given by

$$R(n, m) = [T_{exp} \times I(n, m)] + \Sigma_{(y=m+1, 2047)}[T_{clear} \times$$
$$\times I(n, y)] + \Sigma_{(y=0, m-1)}[T_{read} \times I(n, y)]$$

$$(3.1)$$
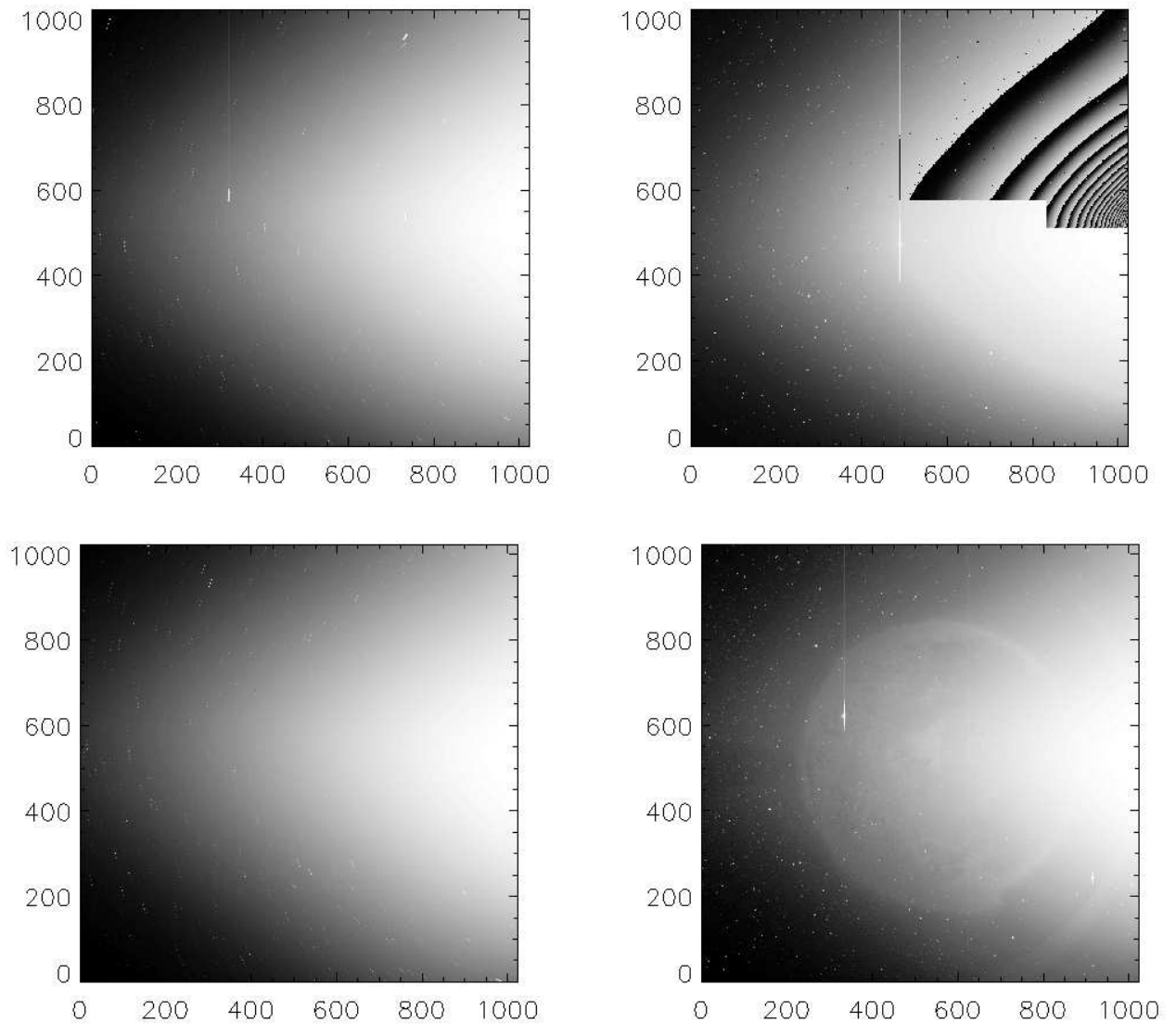
Figure 3.1: An illustration of different types of bad files which were removed during the initial data reduction process. A list of all these types of files are available in the UKSSDC website. Top left: Example of an image corrupted by the roll of the spacecraft, Top right: The image was missing many blocks of data which are probably lost during transmission to ground, Bottom left: This image has been corrupted by a bad calibration file, Bottom right: Images which are the composite of frames from different cameras.(*Continued over...*)

where the count rate for the pixel (x, y) is given by $I(x, y)$, T_{exp} is the exposure time, T_{clear} and T_{read} are the line transfer time during the previous exposure clearing and current exposure reading. This procedure also changes the pixel values from DN (Data Number) to DN/s. This process can be understood with Fig. 3.2, which indicates the various time taken for the different procedures involved in the final calculation of the CCD response.
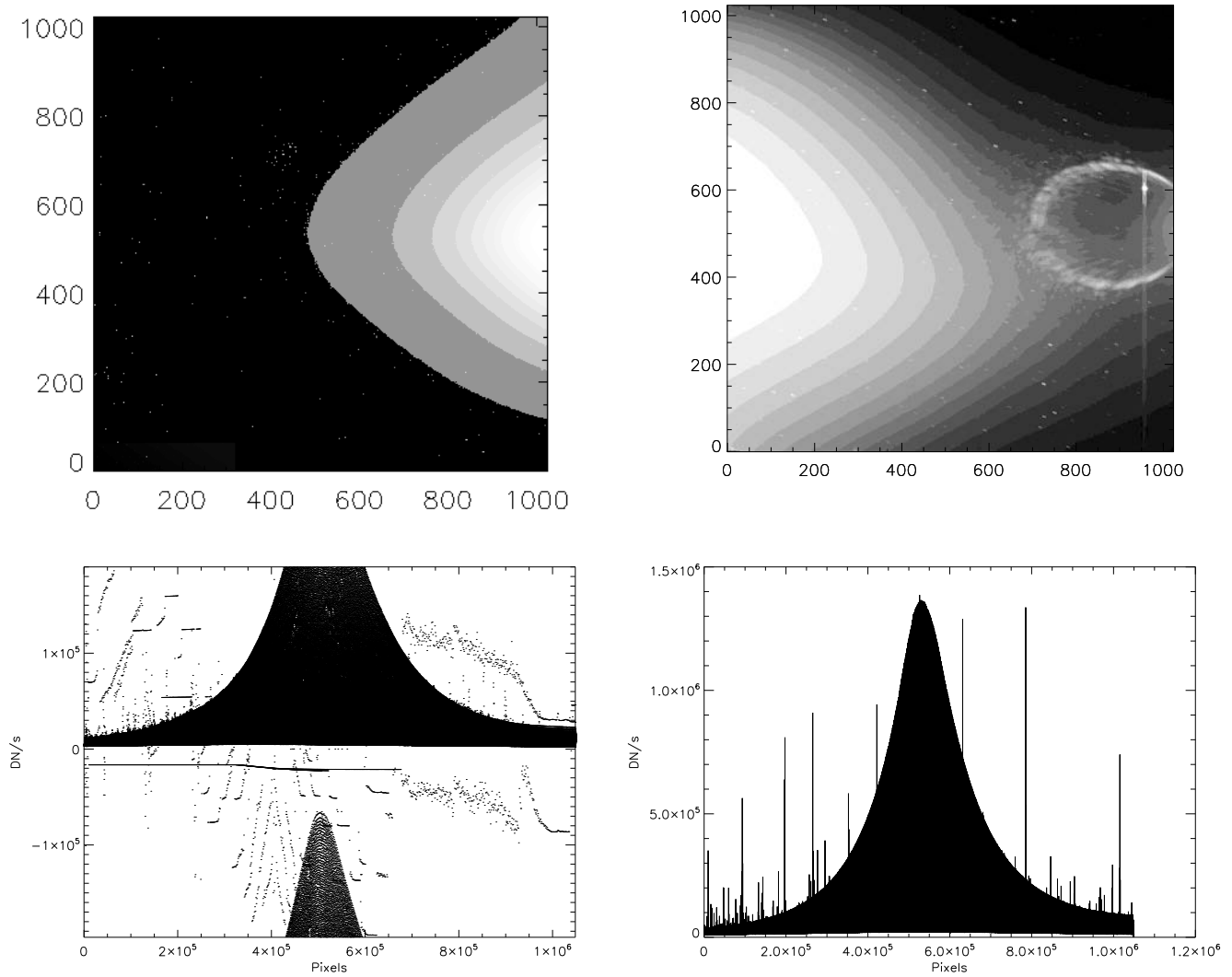
Figure 3.1: (*Continued..*) Top left: Example of a high/low word problem in recomposing 32 bit image from 16 bit data set , Top right: An example of a frame shaken during the exposures which results in a split star field, thus rendering it non-usable for photometry, Bottom: The left image is produced when there is an integer overflow when the images are stitched back after receiving and the right is a good image for comparison.
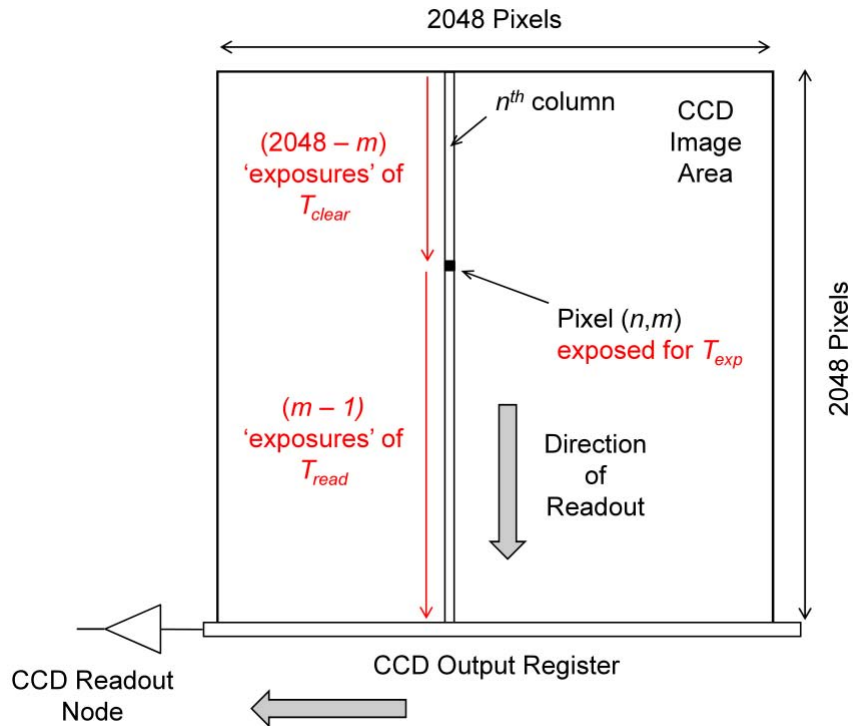
Figure 3.2: CCD read-out procedure for the HI instruments.

Flat-field - In any CCD, the quantum efficiency or the response of all the pixels are not same. Hence it is routine in a CCD reduction to correct for the effect of this differential pixel gain. This correction factor is called the flat-field. In the case of the HI instruments, a flat field map is produced before launch during the ground calibration of instruments. Due to measurement constraints, a flat-field could only be measured in the centre line and not across the whole CCD. Thus, an initial flat-field was derived by extrapolation of the centre line values (Eyles et al., 2009). This led to an observable effect in the final light curves. Hence a more robust flat-field was later derived from the images itself using the star-field and was applied during this part of the reduction (Brown, Bewsher & Eyles, 2009).

Updated Pointing Information and Optics Parameters - During the initial part of the analysis, until August 2008, the raw data contains only nominal instrument pointing information and camera optics as calculated from pre-launch parameters. These values were found to be inaccurate since a drifting star-field was observed, i.e., the calculated centroid based on the spacecraft position and the visually observed centroid of the stars when over plotted, were offset

from each other. The offsets observed were much larger than those predicted from pre-launch
calculations of the spacecraft pointing errors. This led to a recalculation of the spacecraft point-
ing parameters and to the inclusion of a new keyword *Ravg* in the header, which is the average
radial error in pixels between the position of stars in image and the catalogue used. Thus, this
keyword gives the accuracy of the pointing calibration of the instrument, on an image by image
basis. An *Ravg* value of < 1 pixel is good, a value less than five pixels is acceptable (Brown,
Bewsher & Eyles, 2009). By default, those frames with a *Ravg* greater than five pixels are
deemed unacceptable and removed from further analysis. The keyword also has some default
values representing different error codes as shown below :

-999 : No pointing calibration has been applied. This value is found for all L0 data.

-993 : When none of the catalogue stars are matched to the tracked stars, this code is given to
the keyword.

-992 : When no stars are found in the data, i.e., no pixel has value greater than the threshold
value used for identifying the stars.

-991 : When no catalogue stars are found in the field-of-view.

-883 : When the offset errors are greater than 5 (default) or any user defined value.

-882 : When there is no pointing calibration file for the particular frame.

-881 : When there is no pointing calibration file for the whole day of interest.

Fig. 3.3 is a plot of *Ravg* values for the month of February 2008. We find the error residuals
have a median of only 0.157 pixels with a scatter of 0.012 pixels. This indicates the accuracy of
tracking of the stellar objects in terms of position in *STEREO* light curves. Refer to Figure 16
in Brown, Bewsher & Eyles (2009) for a plot of this error value for more than a full year and we
see that the values are less than one pixel for nearly 99.29% of the observed stars, acceptable
limits for nearly 0.32% and unacceptable for only around 0.38%, thus rendering this method
very efficient (cf. Table 4 of Brown, Bewsher & Eyles, 2009).

The raw CCD images, which are 2048×2048, are binned on board, to give a final image
size of 1024×1024. When the images are calibrated during this *secchi_prep* process, in order
to compensate for the summations on board, the pixel values are divided by four to give a more
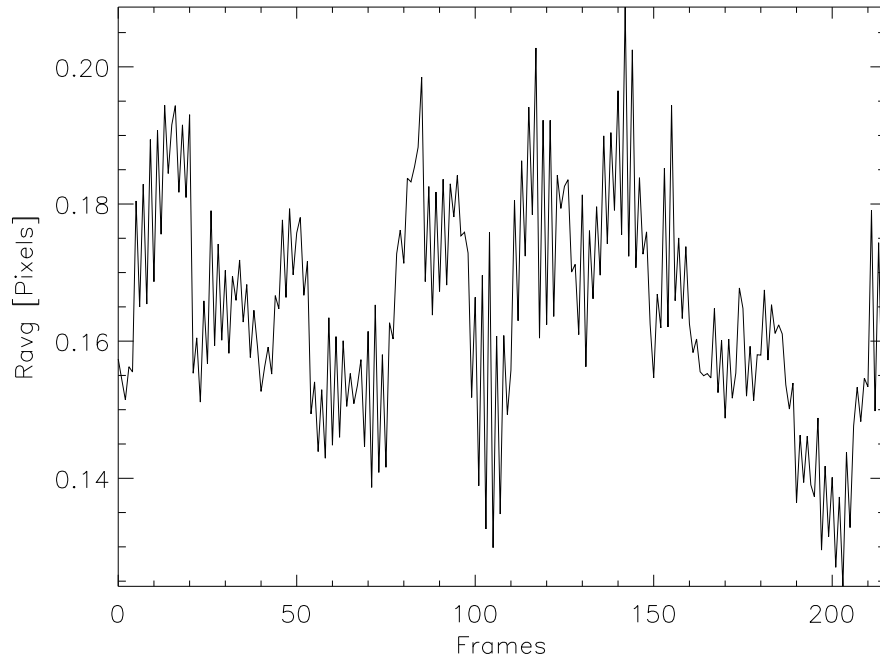
Figure 3.3: This plot shows the accuracy of the pointing calibration. This is a plot of residuals between the star positions in the images and star positions found in catalogues, measured by a parameter *Ravg*, in Pixels. This keyword is updated to the header during the *secchi_prep* analysis.

accurate estimate of the flux value. However, for stellar photometry, we need the exact input flux and hence these values are changed back to original values by multiplying by four in later stages of our stellar photometry pipeline. Apart from these instrument specific calibration, the CCD bias current is also subtracted at this stage. Since the instrument has a fairly steady dark current at its operating temperature of $-80°C$, the dark current subtraction is not needed.

Even after these processes, we found that a few frames were still corrupted due to various additional reasons. These frames, though very small in number, have to be removed in order to get an accurate F-corona map. The most common cause among them was found to be due to certain spikes and black outs in the data which are not tagged by the previous routines. These bad frames were removed using a five sigma clipping of the image pixel values in our pipeline. The general data for all the instruments which are produced after these processes in *secchi_prep* is called the L1 data. Nowadays, this data can be downloaded straight from the UKSSDC website as well.

3.3   F-corona subtraction

The data at this stage (L1) is dominated by the presence of the solar F-corona (Fraunhofer corona). The F-corona, which is the scattered sunlight from the dust particles in the inner Heliosphere, is already studied in detail in literature (e.g. Kimura & Mann, 1998). This light extends to high elongation angles from the Sun up till Earth's orbit where it merges with the zodiacal light. The brightness of this corona varies from 10^{-12} to 10^{-14} times the brightness of the solar disk, B_0 and hence in the case of *HI* instruments dominates its CCD frame. The intensity of CME's in these images are in the region of 10^{-13} to $10^{-14}B_0$ across the frame of HI-1A. Hence, removal of this corona is essential to study the evolution of CME's as well. Fig. 3.4 illustrates the need for the removal of the F-corona in order to do stellar photometry. The dark line running through the column is due to blooming of objects brighter than $m = 3$, which can be seen prominently in the bottom figure. Fig. 3.5 contains a plot of the image through the middle row with the F-corona present and when it is removed. It is clear from this plot, the dominance of F-corona in the data. The stars which are almost invisible in the top plot becomes clearly visible in the bottom plot. The brightest of the objects at around 400 and 600 pixels are 6th magnitude stars and the fainter ones which are barely visible in the bottom plot are the 11th and 12th magnitude stars.

In order to remove this corona, during the initial stages of the project, a minimum background map is obtained by computing the minimum value of each pixel for a stack of images. After experimenting with different number of images, we found a days worth of images, i.e., 36 images, to be a best fit for our need. The background map is then subtracted from each of the images which removes most of the F-corona. This is a crude method which still leaves a residual corona where the variations of the F-corona itself are quite distinct and prominent. Moreover, for photometry, the CME variation in the field also affects the resulting light curves. Hence, it became absolutely essential that we remove both these variations in order to do a effective stellar photometry.

The light curves computed after this method of F-coronal removal has a dominant signal of around one day corresponding to the minimum map used for subtraction and this made
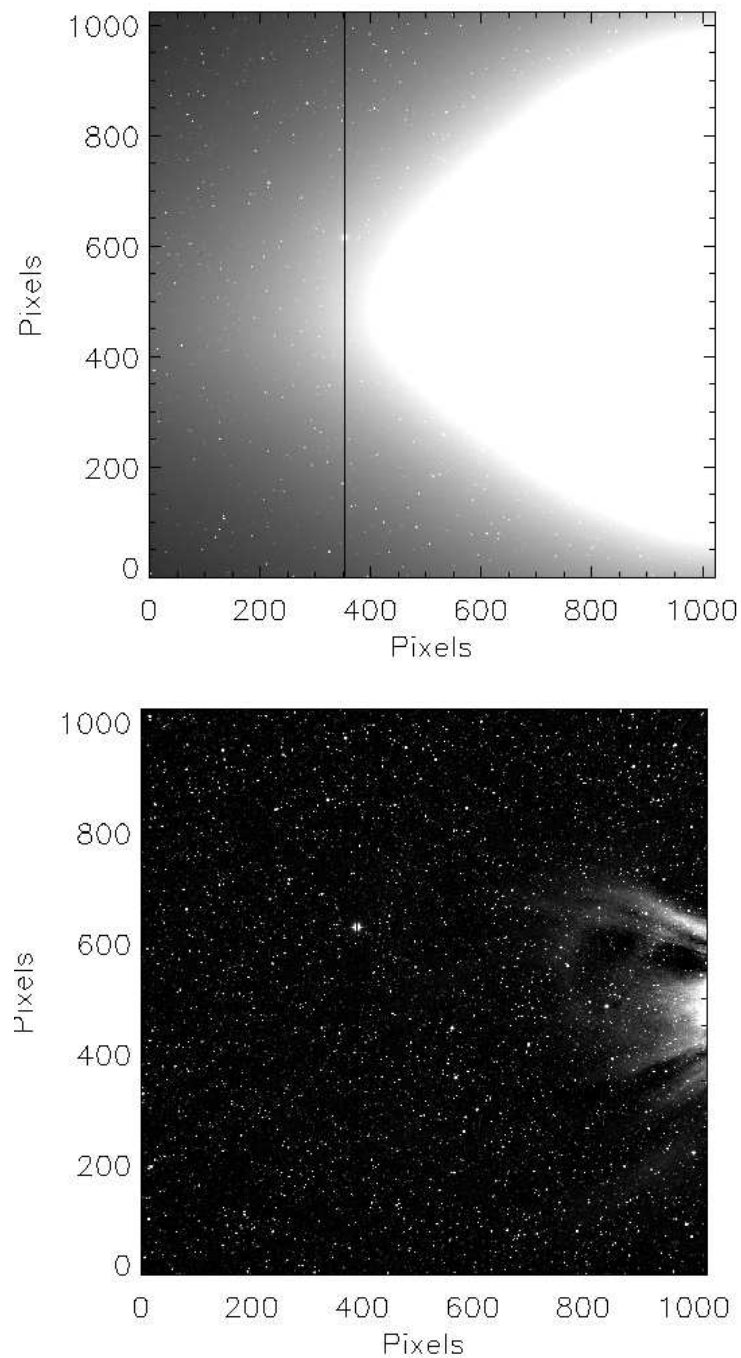
Figure 3.4: Top - Sample image with the solar F-corona present. The line running across the image is the blooming effect caused by a planet which can be seen prominently in the bottom image. Bottom - The same image after the subtraction of a running window map of one day which removes most of the contribution from the F-corona, thus enabling us to do stellar photometry. We can also identify the planet as well the solar streamers in this image, which were impossible in the top image. The bottom image is the data image that we use in further analysis. We can also understand from this figure the influence of the solar phenomenon in further photometric reduction and analysis.
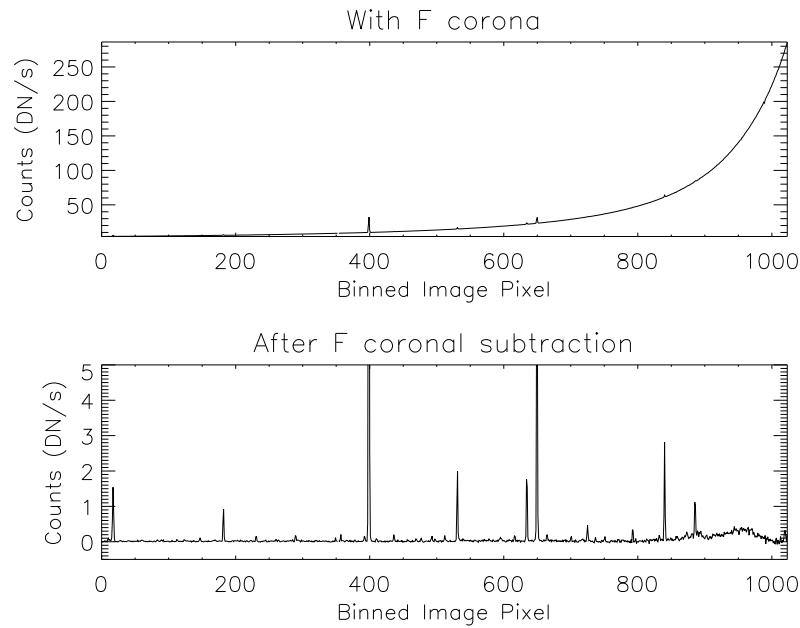
Figure 3.5: This is a slice of the image through the centre line before and after the F-corona subtraction. The spikes at around 400 and 600 pixels are the 6th magnitude stars in the field which are barely visible in the top image and the faint stars which are barely visible in the bottom image are the 11th and 12th magnitude stars which are completely invisible in the top plot.

studying periods around one day impossible. Early part of this project involved using different methods for removing this corona including modelling the F-corona using our data. However, modelling the corona is beyond the scope of this project and the various methods attempted were quite unsuccessful in improving the photometric efficiency. In late 2009, RAL formulated a new method of F-corona subtraction using the running window method[2]. This method was developed in consultation with Andy Buffington, who used a similar method in the analysis of data from the *SMEI* spacecraft. This method involves using a running window of either one day or eleven days. The lowest 25 per cent of the data is selected in this window to form an average on a pixel by pixel basis. This minimum map is then subtracted from the images. Two different window lengths are used in order to study different parts of the CME evolution.

Fig. 3.6 shows the difference and the need for two window lengths in order to study different solar phenomenon. The top image is the one day running window subtracted and it removes the CME streamers, thus is better suited for stellar photometry. The bottom image is produced by subtracting the eleven day running window minimum map. The solar streamers are clearly

---

[2]http://www.ukssdc.rl.ac.uk/solar/stereo/documentation/HI_processing_L2_data.html
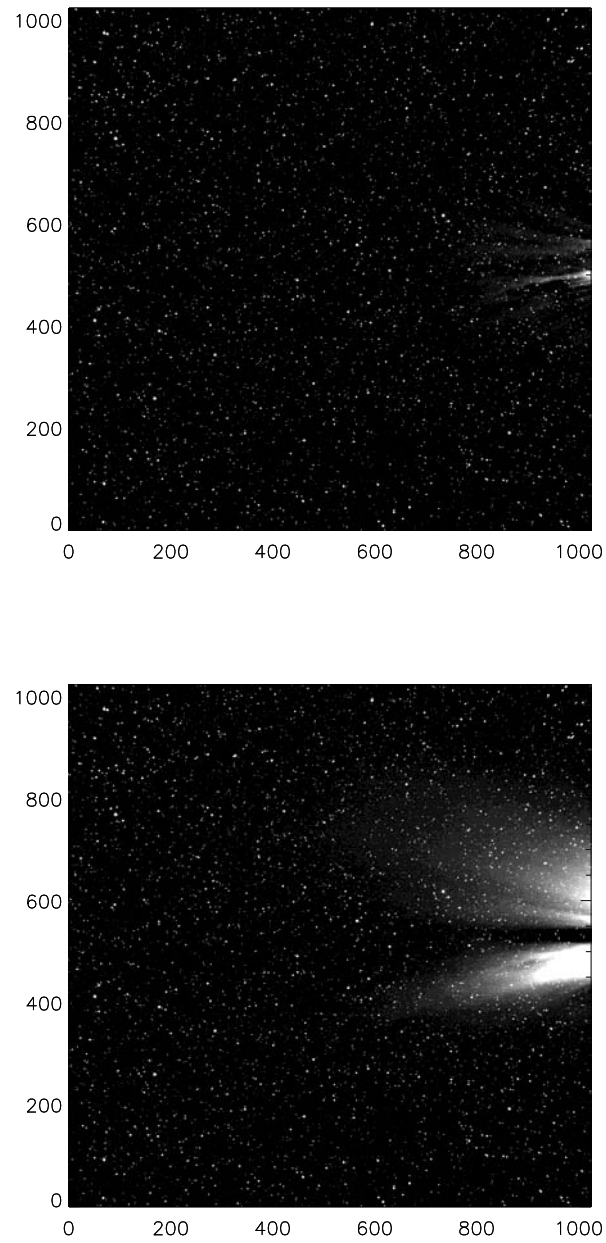
Figure 3.6: This plot is an illustration of the two different types of running window method formulated for the F-coronal subtraction. The top plot is made using the one day running window background map. It clearly removes most of the F-corona and stars are clearly visible for photometry. The bottom plot is made using the eleven day running window. Here, the CME is clearly seen with its various streamers but the background stars are dominated by the solar ejections and hence renders this method inefficient for photometry.

present and thus are mainly suited for the study of CME variation than stellar photometry. As in Fig. 3.6, for the case of stellar photometry, the one day window seems to improve the photometry of the data much better than the eleven day window. To further understand the effect of this new background subtraction in our data, we computed a parameter, the Fill factor (F), which essentially gives the percentage of data points available for an individual object or how well the star is tracked through a sequence of images. This is computed by dividing the number of data points present in the final light curve for the star by the total number of frames present from the moment when the star enters the field of view to when it leaves the field. In Fig. 3.7, we plot this Fill factor, F against the standard deviation, σ of the light curves.

The data used in this plot is from around 7700 stars which are tracked when the instrument faces the Galactic centre and hence involves a very crowded field, in order to consider an extreme case. The σ values are computed from raw light curves which contain the long term trends mainly caused by the variation of F-corona in time scales of the data. A Fill factor of one indicates that the star is completely tracked through all the images and $F < 1$ indicates a loss of data during its presence in the field of view. We find that in the running window method, there are many stars which are completely tracked whereas in the old background subtraction method none of the stars are tracked completely. And also from the data, we can infer that the 'raw' scatter of the data is close to 10 per cent photometric precision. There are a few objects below that scatter close to the zero value of the Fill factor which are the result of very few data points available for computation and are not indicative of better photometry. However, it has to be pointed out that this precision is obtained with all the systematic effects present in the data and hence there is a huge scope for improvement after treatment with various trend-removing methods. We do reach a σ of less than 1 per cent for stars brighter than 8 magnitude, when the systematics are removed, as discussed in Chapter 5.

Fig. 3.8[3] shows the complete cycle of processes that happens during this pre-processing stage of the data. The images at this stage are ready for extraction of stellar photometric information. We have developed the photometry pipeline used for all the HI instruments and is explained in

---

[3]Courtesy:http://www.sstd.rl.ac.uk/stereo/documents.html
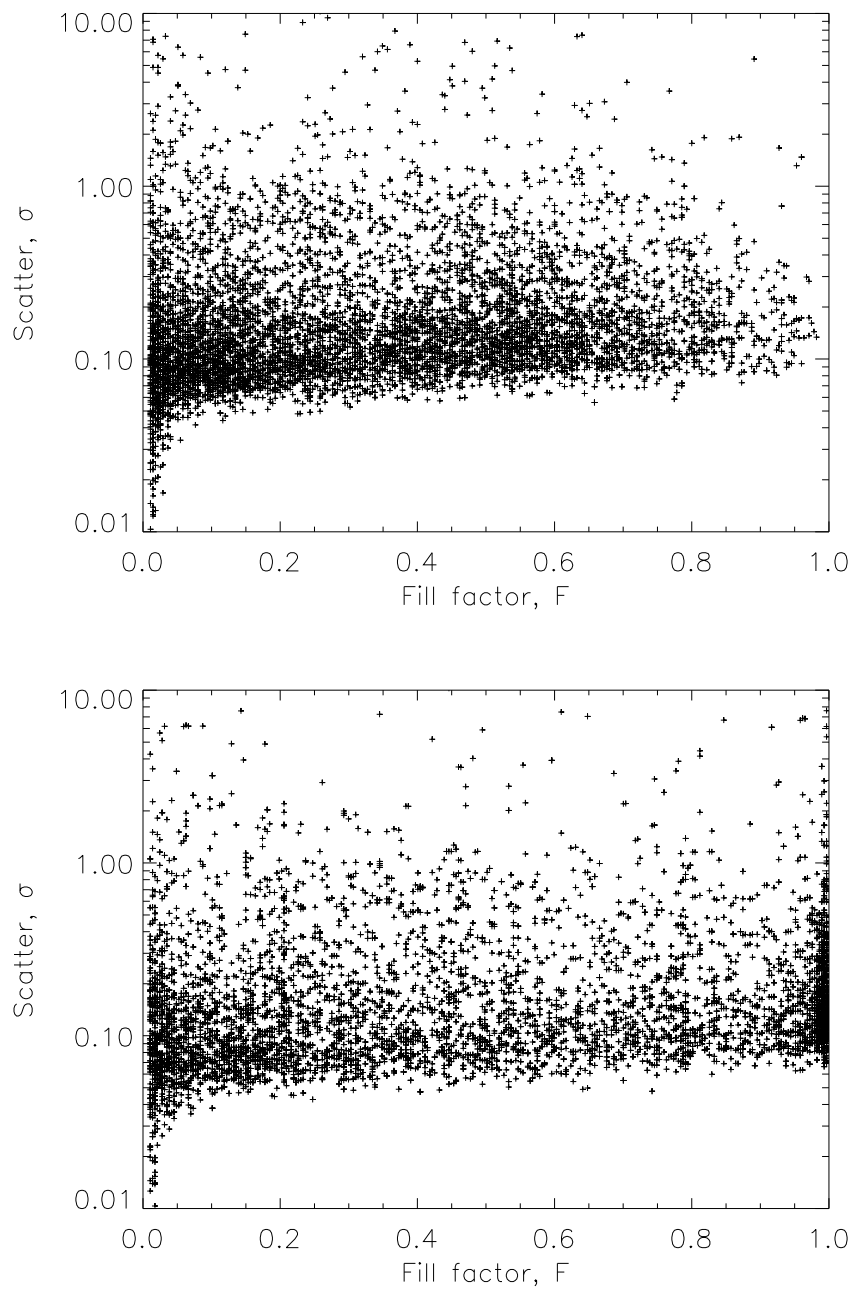
Figure 3.7: Plots of scatter vs Fill factor, F for a sample data of around 7700 stars tracked for 10 days in the month of October 2007. The top plot is using the one day minimum map as a background and the bottom plot is using the new running window as the background for F-corona subtraction (see text for further explanation).

detail in Chapter 4.
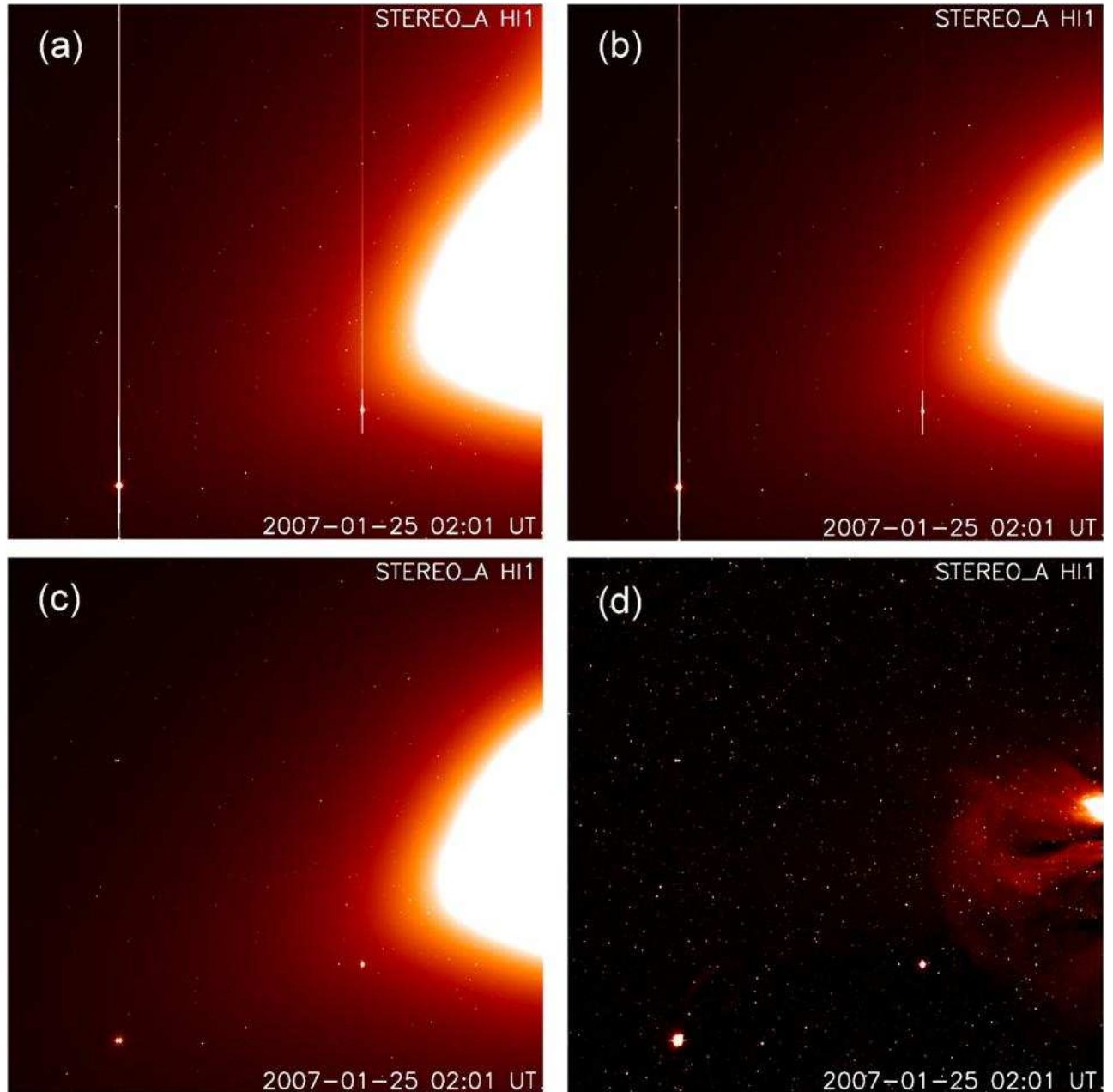
Figure 3.8: This set of images explains the various stages that are involved in the pre-processing pipeline explained in this chapter. From top to bottom, left to right : The first plot is the original data which is down linked from the spacecraft. The next one is produced after application of the shutterless correction, flat fielding, bias correction and pointing and optical calibration, but still contains the bloomed pixels. The next image is produced after the blooming and the vertical overflow pixels are modified. And the last one is after the removal of F-corona revealing the background stars.

Chapter 4

Stellar photometry

4.1 Introduction

The next stage in the STRESS project is the photometric reduction of the HI data and understand the general data characteristics. This chapter will briefly discuss the various catalogues used for the compilation of the STRESS input catalogue, some of the statistics of the target stars and the stellar photometry pipeline developed for the analysis of this data.

4.2 Stellar Catalogue

The data which is downloaded from the spacecraft undergoes various initial reduction processes described in Chapter 3. Photometric information is now extracted from these images using the NOMAD (Zacharias et al., 2004) as input catalogue for stars brighter than $R = 12$ by our semi-automated pipeline which is explained below. NOMAD - Naval Observatory Merged Astrometric Dataset has information for nearly 1.1 billion stars, derived from Hipparcos, Tycho-2, UCAC-2 and USNO-B1 catalogues for astrometric and optical data and 2MASS catalogue for infra-red photometric information. Some of the catalogues are complete down to a V magnitude of 21 (Monet et al., 2003), which is the limiting magnitude for the USNO-B1. Since it is a merged catalogue, it has the advantage of cross referencing between different source catalogues, thus reducing the error in astrometric positions and it also contains the various photometric bands information. We specifically need the R band data, so that it can be used with the HI-1A instrument, which is the first of four instruments that is analysed in this project (see Section 2.3). We download the J2000 position as well as the B,V,R,I magnitudes from their

interactive webpage[1] as a base for our input catalogue. The final version of our compiled input catalogue is shown in Fig. 4.1. The rectangular features seen in the catalogue is due to the fact that we have downloaded the catalogue region-by-region. Also, the gaps are when the spacecraft FOV passes through the Galactic Centre and the region is too crowded for proper photometric analysis with our instrument, as explained in forthcoming sections. Hence we ignore these regions during the catalogue compilation. Thus, the final catalogue comprises of 874000 stars from magnitude 2 till 12. Even-though we can track objects brighter than $R = 2$, we choose to ignore them due to blooming/saturated pixels as explained in Section 3.2. Fig. 4.2 shows the magnitude distribution of our input catalogue. In this plot, we can see that the number of stars peak at around $R = 11.5$ and then declines as we go to fainter magnitudes. In an ideal case, the star density at fainter magnitudes are higher than at brighter magnitudes (example at $V = 9$, on average there are around 3 stars $\deg^{-2}$ and at $V = 12$, there are around 60 stars $\deg^{-2}$), hence we should expect a lot more stars in the region towards the right end of the plot. Since, we apply a filter to reduce contamination (will be explained in forthcoming sections), we find a decline in the number of stars in our final input catalogue towards the fainter end.

4.3   Photometry

In general, there are two main types of aperture photometry techniques used by the community - PSF (Point-Spread Function) fitting and Aperture Fitting (AF). PSF fitting involves generating a function to model the PSF of the target, example, a 2-d elliptical Gaussian and hence extracting the flux from the target. This method performs much better than aperture photometry in crowded regions and generally used in photometry of clusters and supernova analysis. In *STEREO*, there is only a small region in the sky - Galactic Centre and anti-centre, where this method is absolutely essential to do photometry. Even though the PSF of HI-1A is well defined and circular almost throughout the CCD frame, the PSF of the other HI instruments are not so. This is specially the case for HI-2B where the PSF at the corners of the frame are much more elongated than at the centre. Moreover, the HI FOV is scanning and hence the PSF of an object also changes depending on its position in the CCD. Hence, developing a general pipeline

---

[1]http://www.nofs.navy.mil/data/FchPix/
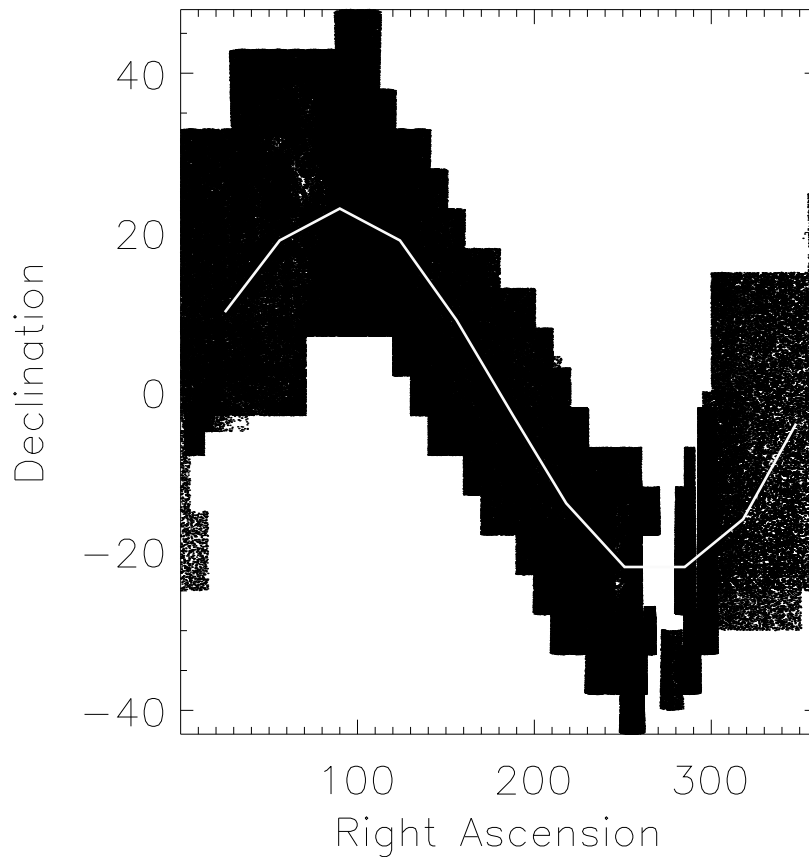
Figure 4.1: The input catalogue derived from the NOMAD catalogue up to $R = 12$ is plotted in this graph. The red line shows the spacecraft path in the sky and our attempt to cover a little more than 10 degrees on either side of the spacecraft position in order to maximise the use of CCD images.

for all HI instruments by PSF fitting is ruled out and aperture photometry is used instead.

4.3.1 Aperture Photometry

Aperture photometry means defining an aperture around the target in the image and calculating the integrated flux in that region. The aperture is generally circular but can also be elliptical depending on the target in question and the radius is generally few times the Full-Width-Half-Maximum (FWHM) of the PSF of the target. An annulus is defined around the target where the background flux is assumed to be constant and unaffected by the target. This background flux is then subtracted from the target mean flux value and then the pixels values inside the aperture are summed to provide the final flux value. In order to determine the size
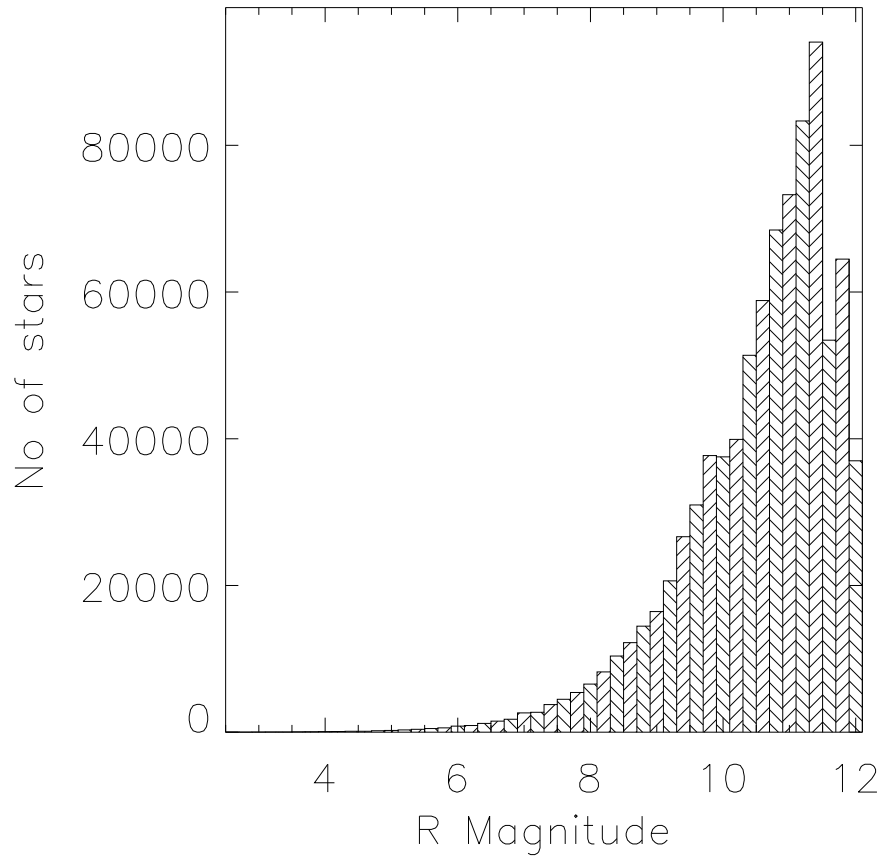
Figure 4.2: This plot describes the magnitude distribution of our input catalogue. It contains around 874000 stars and peaks at $R = 11.5$ and declines at fainter magnitudes due to a filter which we apply to remove crowded field contamination.

and shape of the aperture to be fitted around the target to extract the information contained, it is essential to understand the PSF of the instrument and various noise sources that can be present inside that aperture. We need to find the optimum aperture size that can maximise the SNR, i.e., in a random sized aperture, the SNR increases and then falls off, as we move from one edge to the other, with the maximum set at the target. We can identify the optimum aperture by either modelling the noise sources along with the PSF of the target or by a trial and error method of fitting various sized apertures, analysing the end result, thus finding the most suitable size and shape of aperture. In the case of HI, the PSF is approximately circular and hence a circular aperture is best for this purpose. A general rule of thumb used in literature is to use $2*FWHM$ of PSF as the aperture size (Bewsher et al., 2010). Since the SNR varies across the aperture, fitting different aperture is also used as way of identifying the blending of

two objects which are very close to each other. Pollacco et al. (2006) uses this method in WASP project to identify such contaminants. Another downside for using this method in our project, is that the background flux is not constant but changing, depending on the solar conditions and this offers another concern in our flux calculation. In spite of these two major issues, we find that the raw light curves derived by this method, have a standard deviation close to ten percent for stars brighter than $R = 8$ and this scatter decreases drastically close to 0.1 per cent for the brighter stars and close to a few per cent for fainter objects, when the systematics are removed as explained in Section 5.4

The image header contains information about the spacecraft positions - pitch, roll and yaw angles as well as the distortion parameters and X and Y offset values from the centre of HI optics on CCD. We use this information to convert from the spacecraft position to the corresponding sky co-ordinates in RA and DEC. In the initial stages of the project, we were using the nominal pointing information obtained from the guide telescope and spacecraft star trackers along with the preflight CCD offsets. We observed that, the final calculated position of the target in the image is systematically offset from the catalogue positions. In Fig. 4.3, the diamonds should contain a star from the image within them, but we find that the diamonds were systematically offset from the actual stars (dots) [2]. Brown, Bewsher & Eyles (2009) correct for this effect by iteratively fitting the position of stars and reducing the error between catalogue position and image position. These error values in the position, are then fed back into the image header as *Ravg* values described in Section 3.2. Using these new position information, we transform the spacecraft pointing to the RA and Dec positions in the sky and hence, used in identifying targets in the images. The first part of this transformation involves converting the RA-DEC to an instrument frame of reference and then transforming them to the HI camera image plane (see Section 2.2 of Brown, Bewsher & Eyles, 2009 for a detailed discussion of this transformation).

The bright spots in Fig. 4.3 are the stellar sources. In order to identify them, we look for a local maximum in a surrounding area. In the process, we define a sub-square of certain value, centered on the pixel and when the difference between the pixel value and the mean value of

---

[2]Courtesy: Brown, Bewsher & Eyles (2009)
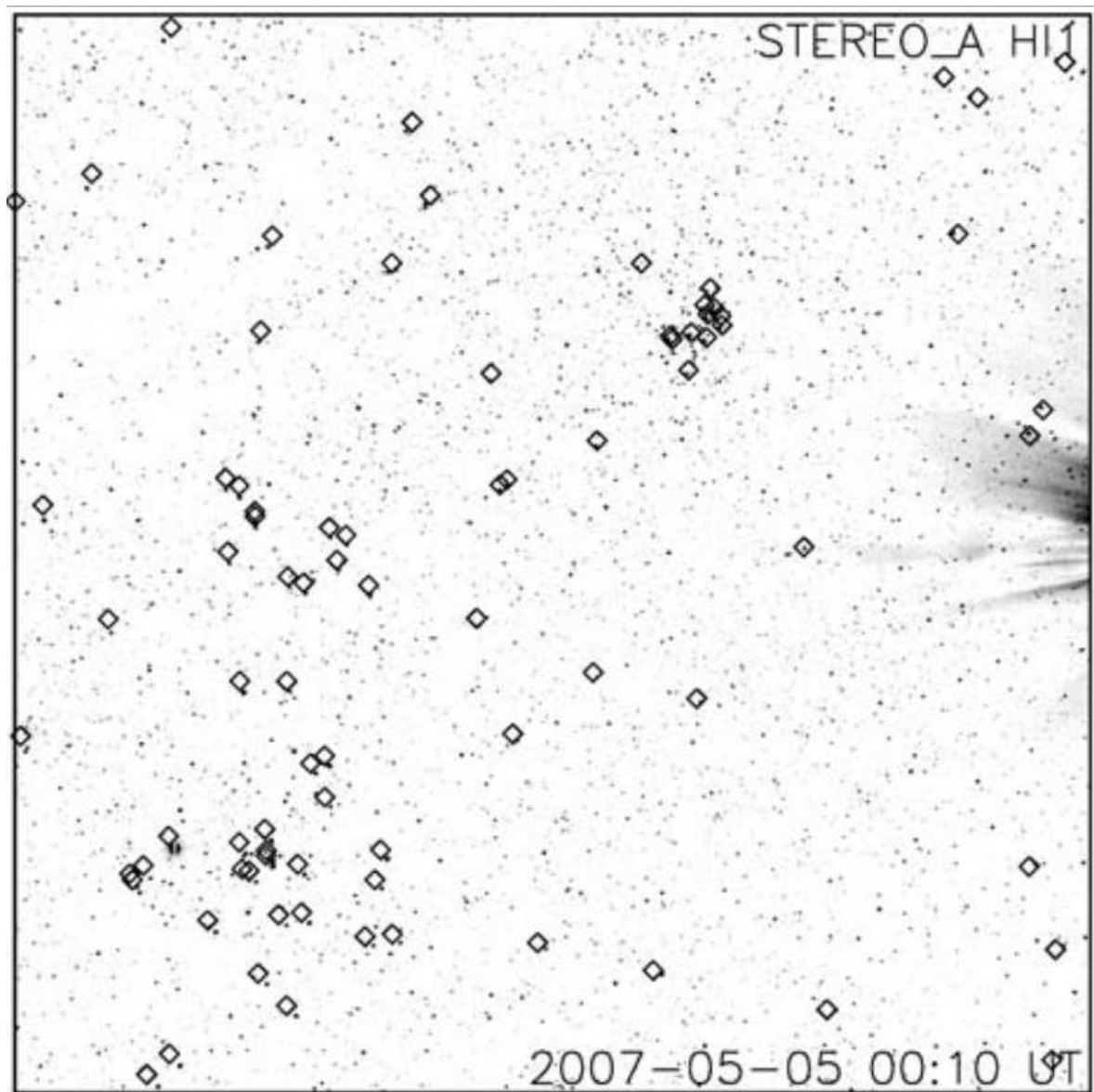
Figure 4.3: HI-1A image with the catalogue stars over plotted with diamonds. Ideally, each diamond should contain a star within, but we can see that there is a systematic offset between the image position and catalogue position of the objects (Brown, Bewsher & Eyles, 2009).

the sub-square exceeds a threshold, we identify the pixel to contain a stellar source. In the case

of HI-1A, we use a surrounding 4×4 area and a threshold of 1.5 times the flux value in this

sub-square. Since the calculated PSF of HI-1A is 1.6 pixels (Bewsher et al., 2010), this region

is adequate to find the peak image pixel[3] but for other HI cameras, this sub-square needs to be

increased in order to find the correct target. Also, in the HI-2 cameras the size of the pixels

are larger (cf. Table 2.1) such that contamination is an important issue to deal with, before

---

[3]The term pixel in all further discussions refer to the image pixel rather than the CCD pixel unless specifically mentioned.

processing further. When cross-identifying image sources with catalogue objects, for HI-1 cameras, an error margin of 2 pixels is used, i.e., if the difference in position between the image and catalogue is within 2 pixels, the pipeline still matches the image with the catalogue target and for HI-2 cameras, this value has to be smaller in order to compensate for the larger pixels.

The large pixels ($70.3''$ and $4.1'$) of the HI images induces PSF contamination between close targets of comparable magnitudes. We apply a basic magnitude-distance criterion to identify and remove a large portion of this effect and to ensure that the object is the brightest in the local neighbourhood. Thus, a pair of objects which are closer than 10 pixels, corresponding to $11'$ and have a magnitude difference of less than 0.5 are removed from the target list. We also remove the fainter star in these pairs within the same radius but with a magnitude difference of 2.5 or lesser, i.e., in a region of 10 pixel square, if R_a and R_b are the magnitude of two stars, then,

(a) $|R_a - R_b| \leq 0.5$, remove both R_a and R_b.

(b) $|R_a - R_b| \leq 2.5$, remove the fainter object in the pair.

This is explained in Fig. 4.4, where in the left panel, there are two stars which are less than 10 pixels away and with a magnitude difference less than 2.5, that both the PSF's contaminate each other and it is not possible to correctly identify the fainter source. In this case, the algorithm, will remove the fainter star from the catalogue source and hence we will be tracking only the brighter object. In the right panel, there are two stars of comparable magnitude (magnitude difference less than 0.5), which are clearly distinguishable when separated by more than 10 pixels. In the case of bright stars, this just resolves the two PSF. We have to mention that the algorithm only removes the object from our final tracking targets and the contamination from the nearby fainter objects are still present in the final light curves of bright objects. In order to remove this contamination, we use three different aperture sizes for calculating the flux. By using the biggest aperture for the brighter star and the smallest for the fainter star, we try to rule out further confusion from nearby objects. In spite of all these measures, we still find that targets are either missed out or mistaken for nearby closer and brighter ones. We discuss these issues in detail in Section 5.4.
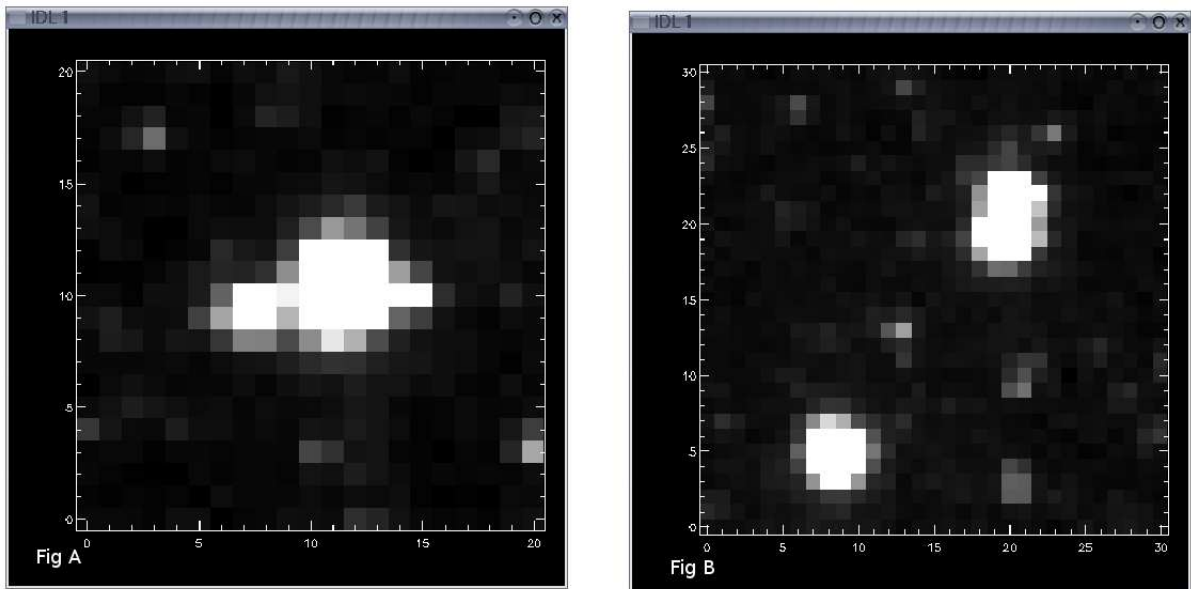
Figure 4.4: *STEREO* HI-1A images of individual stellar PSF. Left panel: An example of two stars which have a magnitude difference less than 2.5 and whose centers are separated by a distance of less than 10 pixels contaminating each other. Right panel: An example of two bright stars of magnitude difference less than 0.5, which are well separated, such that their PSF does not overlap and are well distinguished at a separation of more than 10 pixels.

Even-though the PSF of HI-1 is well behaved in most regions of the CCD, it deviates a little towards the edges of the frame, due to the wide-angle optics of the camera. The PSF in the centre of the field is highly circular whereas it starts to become elliptical at the edges. This variation from circularity in PSF, is given by a parameter, μ, the distortion parameter. This parameter is important specially at the edge of the circular fields, where the variation is visible. Hence this parameter is also used in the conversion of catalogue stars to HI field. In the case of HI-1, it is around 2% and for HI-2, around 10% at the corners of the square field (cf. Brown, Bewsher & Eyles, 2009). Hence, a more robust method of identifying the centroid of the targets is essential. Therefore, we use the method developed by Steve Spreckley for use with *SMEI*, where the PSF is of a very odd shape (Spreckley, 2010). This method determines the centroid of the target at a sub-pixel level.

Aperture photometry is done on these selected targets with the IDL routine, *aper*, which is adapted using DAOPHOT algorithm . An ideal aperture radius to be used for the PSF of HI-1A is 2*FWHM (Bewsher et al., 2010), but we use three different aperture radii, 2.5, 3.0 and 3.5 pixels, as discussed before in order to decouple the contamination of nearby objects. This

IDL routine, calculates the total flux inside these aperture, with the target centred at sub-pixel level, as determined by the previous routine. We then define an annulus, between pixel radii 6.0 and 10.0, around the target to calculate the mean sky background, which is then subtracted from the total flux. The variance of the sky background calculated from the flux in the sky annulus, random photon noise calculated from the measured total flux and the uncertainty in the mean sky brightness contribute to the measured error values of the flux. The flux values so far described in this analysis and hence used are in DN/s. In order to convert these values to actual flux counts and then to magnitude, we have to multiply them with the photon-electron conversion gain and then by the CCD quantum efficiency factor. For the HI cameras, the CCD is assumed to have a quantum efficiency of 90 per cent, which is the maximum response in the wavelength range 300 - 1000 nm (cf. Fig. 5 of Bewsher et al., 2010 and also Fig. 2.6) and the photon-electron conversion gain to be used is 15 units (i.e., 1 DN = 15 photoelectrons; Bewsher et al., 2010).

The image header contains various time values, such as date and time of file generation, time at the start of observation and the mid observation time. We use the time stamp at the start of observation, which is in UTC. We found that due to the complicated integration of the data on board the spacecraft, the start of observation is found to be the best parametric representation of time. We convert this to Julian date and thence to convert the observations to Sun-centered time, we use another header keyword which gives the time travel between the Sun centre and the spacecraft. Thus, the time stamp in the final data, is the time at the start of the observation in Geocentric Julian dates with correction for the Sun-spacecraft distance applied and hereafter, referred to as the Modified Julian date (MJD). We would like to note that this time of observation will vary across the CCD from the region closer to the sun to the farthest due to the large sky area covered. However, this difference is very small compared to the actual cadence, which is 40 minutes for HI-1 and 2 hrs for HI-2 and hence ignored in this analysis. Fig. 4.5 gives a schematic of the analysis pipeline discussed so far.

In order to process this data, we have to overcome a major obstacle in the form of computational requirements. The final pipeline is based on the initial algorithm developed at RAL.
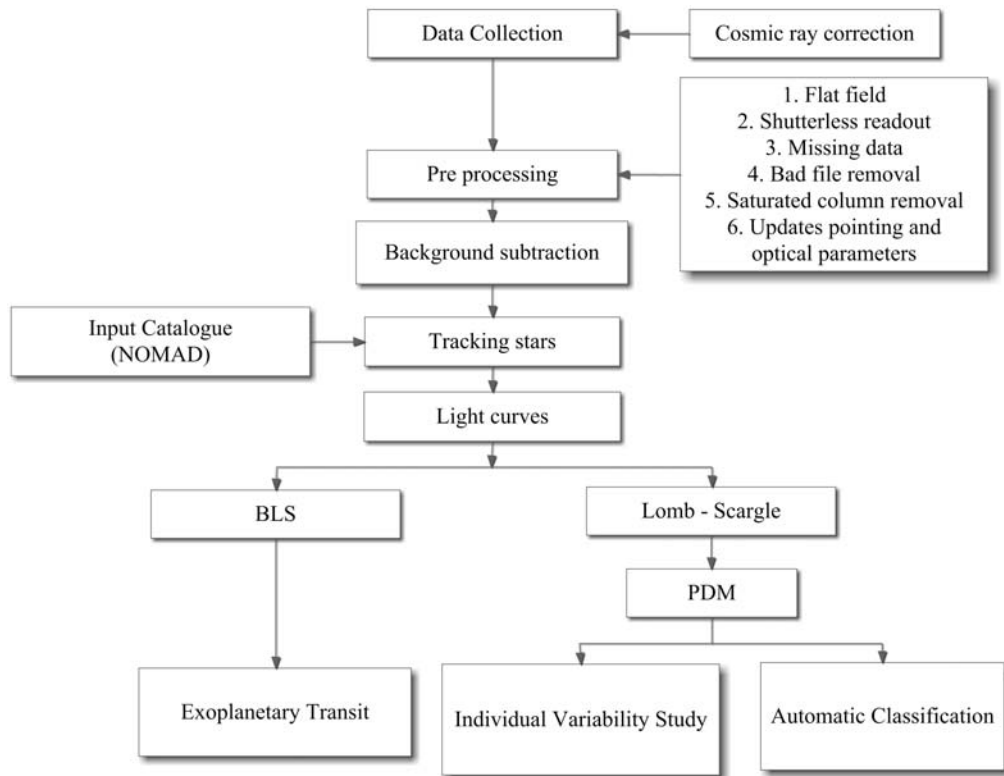
Figure 4.5: A schematic of the *STEREO* HI data analysis pipeline, showing the various stages of data preparation involved in producing light curves from raw images (see text for explanation).

Thus, the pipeline is based on a stacking algorithm. Information is first gathered for all stars in one frame. When the next image is processed, we just stack the information from the current frame to the one from the previous image. As such, we build up an individual IDL data arrays for each of the following information for all stars together - positional information, R magnitude, calculated flux for 3 apertures, error values for the flux information, sky background value for each star, its error values, the keyword *sun_time* containing the delay time between the Sun centre and the spacecraft used in the calculation of the final time array, two time arrays, one as the original UTC time from the header and another as the corrected and modified Julian date. The stacking of the various information for all the stars during the processing, results in a need for enormous Random Accesses Memory (RAM) space (typically more than 5 GB), which is difficult with a normal desktop computer. Hence, we made use of the cluster at the University of Birmingham, BlueBEAR [4]. It is a cluster of 1,500 processing cores and approximately 150 TB

---

[4]BlueBEAR is part of BEAR, which is Birmingham Environment for Academic Research. http://www.bear.bham.ac.uk/bluebear/

of user disk space, with a wide range of compilers. We used the IDL-VM version in all our processing. The total disk space required to store the processed light curve data for the 37 months of data, which has been processed is around 34GB. This value gives an idea about the data logistics involved in this project. In future, we are planning to process the data for the other HI instruments as well and combine them, to get a longer coverage of the light curve resulting a precise measurement of the periods involved. This part of the work has a larger computational requirement, which has to be planned in advance.

This chapter dealt with the various aspects of the basic photometry pipeline developed for the *STEREO* HI instruments, in view of using it for stellar variability and Exoplanetary transits. The next chapter contains some of the general characteristics of this processed data - light curves and a brief introduction and analysis of some of the systematic trend removal methods experimented using this data.

Chapter 5

STEREO HI-1A Data Characteristics and Trend Removal Methods

5.1   Introduction

The previous chapter describes the general photometric pipeline developed for the reduction of data from the *STEREO* HI instruments in general as well as HI-1A specifically. This chapter contains the general characteristics of the reduced data and some of the methods used in understanding and removing the various systematics in the data. This includes an overview of the some of the common variations like the effects of CME's, Poisson Noise limits, etc as well as methods like Inter-quartile filter, box-car smoothing, TFA and PDM.

The typical time taken by an object to transit the HI-1 field of view is about 20 days and for the HI-2 instrument it is about 70 days, corresponding to 700 and 2200 data points respectively. An integration of data from all the four instruments will give us around 180 days worth complete data set for each orbit of the spacecraft around the Sun. Since the cadence of the HI-1 instrument is 40 minutes, the Nyquist frequency for this set of data is around 18 cycles per day $(\mathrm{c}d^{-1})$ and for HI-2 is 6 $\mathrm{c}d^{-1}$. So far we have processed data for nearly a million stars for 3.5 years for the HI-1A instrument alone.

5.2   General Trends

The *STEREO* HI-1A data in general exhibits systematic trends which are the result of the basic instrumental design as well as trends which are common to all the stars in the neighbourhood, either due to some part of the processing pipeline or due to a systematic change in the basic conditions of the images.

5.2.1   CME Effects

One of the most common phenomena is the increase or decrease in the median intensity of the flux due to the presence of either a Solar-wind transient event or a CME in the image. Since the flux value is related to both the actual pixel intensity as well as the different backgrounds which are being subtracted in various stages. Initially, we subtract the F-corona and during photometry, in order to calculate the flux, we also subtract a sky background in an annulus around the target (refer to Section 3.3 and 4.3.1). Hence, the increase in intensity can be attributed to an increase in the flux of the whole region and a decrease in intensity to a higher value of coronal background being subtracted and consequently, we find this effect being manifested for the data point from the particular frame for all light curves. In order to illustrate this effect, we have plotted the median intensity of all the light curves in a particular region for a period in which we know the camera encountered a Solar transient event[1] in Fig. 5.1. The data used in the top plot of Fig. 5.1 is from a very small strip-like region of CCD comprising of 30 pixels by 1024 pixels and we can clearly identify the time when the Solar wind cloud passed through this region imaged by the CCD from the high scatter of all the light curves in that portion. We calculate the median by first normalising the light curves with its median value before adding all the intensities and dividing by the total number of light curves, i.e.,

$$Value = \frac{\sum(flux/median(flux))}{N}.\tag{5.1}$$

The bottom plot of Fig. 5.1 is one of the light curves, where we find that the flux remains constant till it encounters the CME and the intensity stars fluctuating. Fig. 5.2 is an image taken when a CME is passing through the frame of CCD. The image contrast is adjusted to clearly

---

[1]The event list is provided at the STEREO site http://www.sstd.rl.ac.uk/stereo/HIEventList.html
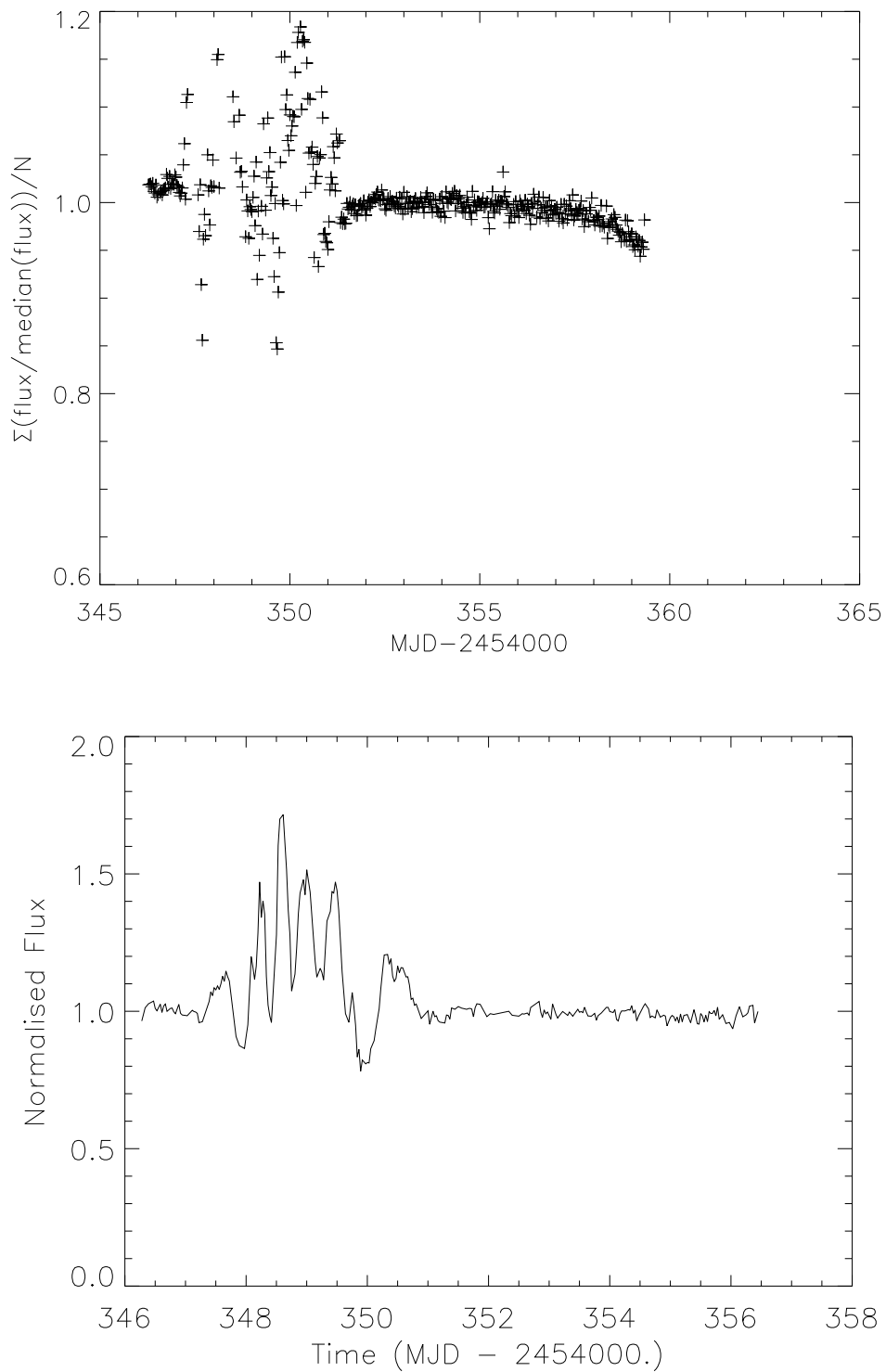
Figure 5.1: Top: A plot of the median value of all the light curves (12) against time in a small strip of 30 pixels by 1024, which encounters a Solar wind transient event. Bottom: A single light curve in the same small region as used for the above plot, whose flux remains steady until it encounters the CME, when the intensity starts fluctuating. This clearly illustrates the effect of such events in all the objects in the region.
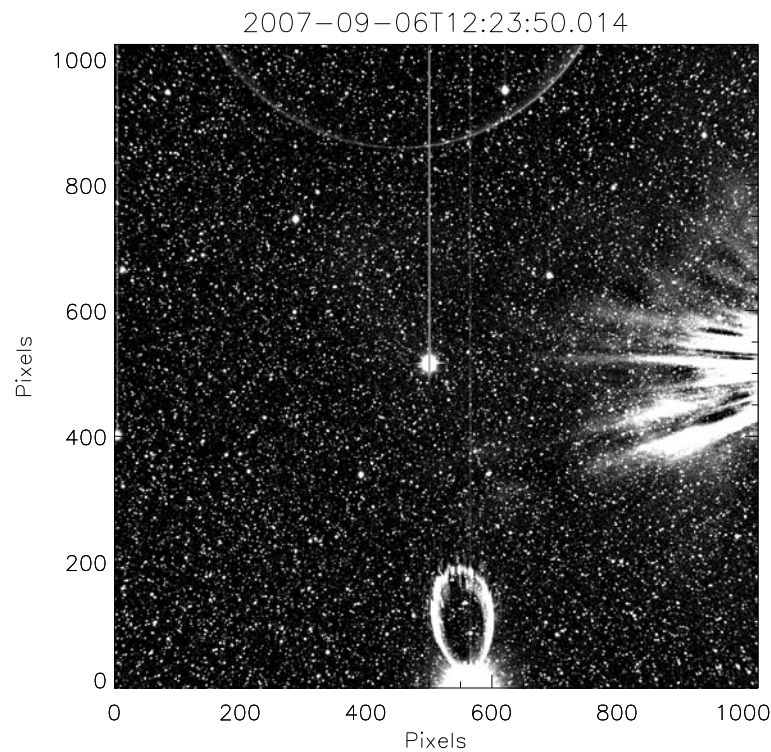
Figure 5.2: Image taken during the passage of a CME event, on 6th September 2007. The image contrast has been adjusted in order to identify the details of CME and its effect on stellar photometry. The Sun is in the right of the image, with the CME event moving from right to left. The bright streamers in the right are from the CME event. We can also observe two blooming events - the bright object in the centre of the frame with a vertical CCD overflow line and the petal-like blooming effect of a bright object at the bottom of the frame. The actual object in the latter case, is not exactly in our field of view and hence only the effect is observed.

identify the effect of the CME.

5.2.2 End-of-frame Effect

Another common trend is illustrated in the top plot of Fig. 5.1, which is the loss of flux at the end of the frame. Sometimes such a loss is also found at the start of a light curve. For the HI-1A instrument, the Sun is present at the right end of the image frame, and hence gives rise to a higher F-corona being subtracted at this end. This results in a higher noise when stars near this edge of the frame. This gives rise to another end-of-frame effect. Fig. 5.3 illustrates this effect in both cases. In most cases, we can clearly see that the light curves are fairly constant throughout until they reach the edges of the CCD. The right portion of CCD also contains the Solar streamers and other CME effects, which induces random fluctuations in the light curve at the edges.
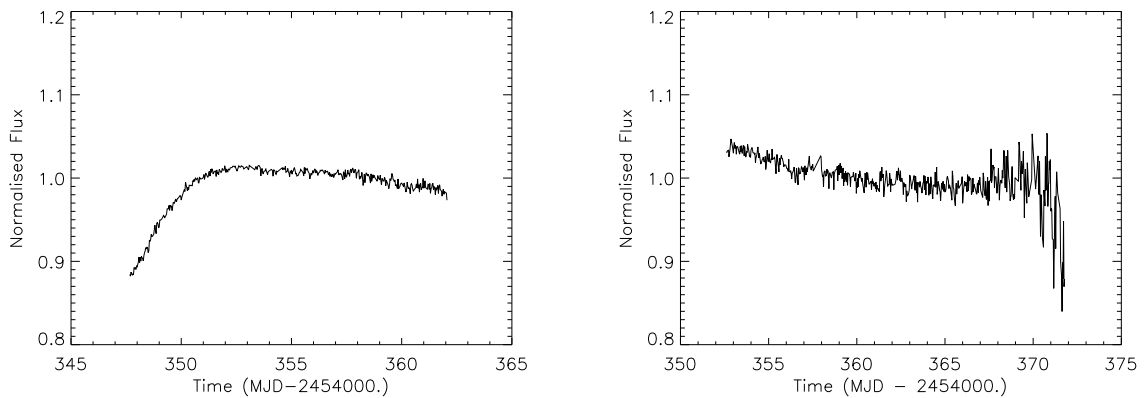
Figure 5.3: These two plots illustrate the two types of end-of-frame effect. The left plot shows a slow increase in flux as the star moves completely into the frame and we believe this is due to the PSF variation at the edges. The light curve on the right is consistent until it reaches the edge of the CCD where a noise is induced due to the presence of Solar streamers and the higher F-corona subtraction.
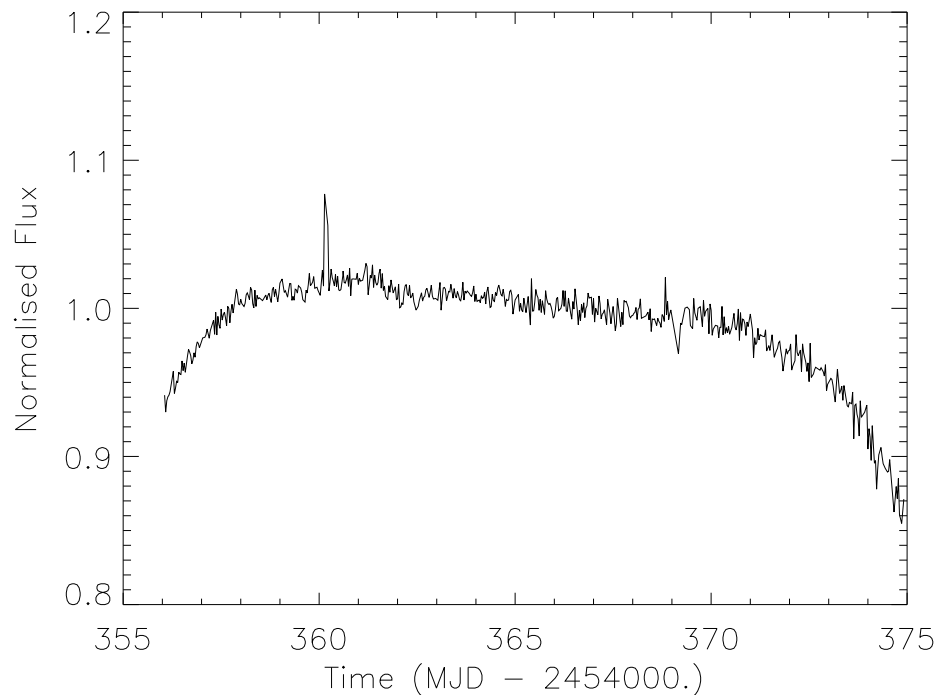
Figure 5.4: This light curve is an illustration of the large scale flat field effect, which was observed in the early stages of the project. This effect in large was removed in the latest update to the secchi_prep pipeline. However, there is still a residual of this flat field in the light curves in small scales.
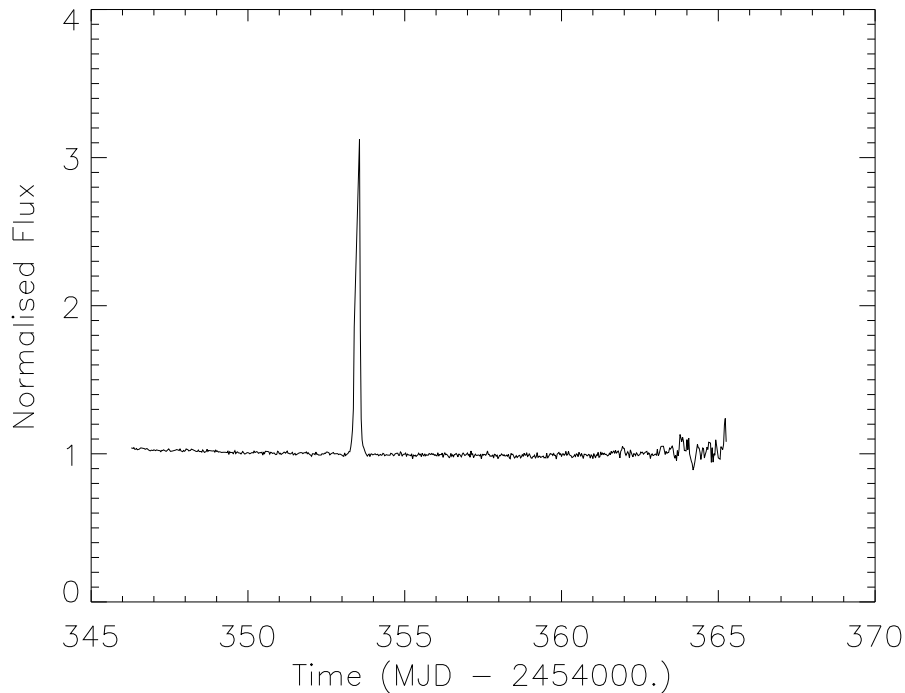
Figure 5.5: The above light curve is an example for the flare-like event observed in our data. This is the light curve of HD 114679, an F2 star with a $V = 9.7$. These events are highly transient and hence it is essential that we remove these events from the light curve before searching for short-period variabilities like extrasolar planets.

5.2.3   Flare-like Events

Another feature which we find in many light curves are sudden flux increases. We find that the flux has almost doubled or tripled in a space of a few frames. These can be caused by a variety of reasons, mostly when the star passes through a Solar streamer or when the object passes close to a bright object or through a blooming caused by a bright object. These are highly transient and temperamental. These events are removed from the actual light curve by the use of sigma clipping as explained below (sigma refers to the standard deviation of the light curve). Since these events are very short lived, they do not affect the summary statistics of the individual light curves, if robust methods are used for the calculation, for example, using a median value instead of median for further calculations. Fig. 5.5 is a light curve of star HD 114679, showing a sudden increase in intensity in a fairly constant light curve. this is an F2 star with a $V = 9.7$, which is not classified as a flare star. In order to remove the above two noise sources - end-of-frame and Flare-like effect, we calculate the difference in standard deviation for each half of the light curve and then compare it to the trend observed with all light curves

from that month. We remove all those light curves whose difference lies more than 5 times the standard deviation of the whole sample from that month.

5.2.4   Flat-field Effect

We also find another trend that is a systematic variation across the whole light curve for all the light curves. Since we can remove this trend by fitting a 4th order polynomial, which is the same as the CCD response function, we think this effect is a manifestation of the response function (Bewsher et al., 2010). The CCD response function, as measured before launch, in the centre line of the CCD is given as,

$$I = I_0(1 + a\theta^2 + b\theta^4) \tag{5.2}$$

where θ, is the angle relative to the optical axis and I_0 is the response on the axis, defined by Eqn. 3.1. $\theta = 0$ in this case, corresponds to the optical axis at the CCD centre. Fig. 5.4 shows the manifestation of this effect on the light curve, which is similar to Fig. 2 of Bewsher et al. (2010). The flat fielding effect has been largely removed in the current data after application of the correction mentioned in Section 3.2, still we find in most light curves a residual CCD response. See Fig. 15 and Fig. 16 of Bewsher et al. (2010) for the initial and corrected CCD response curves and the flat field for both HI-1A and HI-1B. The removal of this effect is discussed in Section 5.4

5.3   Poisson Noise

All these above trends, though systematic, are highly time dependent and specific to this project. However, there are some basic noise sources present in any observation. Even in an ideal case, the light curves have an inherent statistical noise - photon-shot noise. This noise is associated with Poisson statistics. When there is a very small number of particles under observation, the uncertainties in the distribution become significant and the random fluctuations in this case, gives rise to shot noise or photon-shot noise. For a CCD, this shot noise σ_{ph} is

defined as,

$$\sigma_{ph} = \sqrt{S_{ph}},$$ (5.3)

where, S_{ph} is the signal expressed in electrons[2].

In stellar photometry, the number of electrons received is governed by the magnitude of the object under observation and hence, this photon-shot noise varies as the magnitude of the object. This noise also sets a lower limit to the observable precision of stellar variability. Hence, in any project, systemic noise removal techniques are applied to reach this lower limit. One of the advantages of a space based mission is that, it is not encumbered by some of the uncertainties with a ground based observation like clouds, atmospheric variability, etc. However, in this project, the basic instrumental aim is studying Solar phenomena and its interaction with the inter-planetary medium, and this governed the characteristics of the instrumental design. This gives rise to *STEREO* specific issues like pixel size, CME's, planets, etc. Thus, we have some unexplained noise sources, which sets the scatter of our data, much higher than that of the photon-shot noise limit for stars fainter than $R = 10$. In spite of this and all the above mentioned noise sources, we can reach a precision of less than one per cent for stars brighter than $R = 8$ and a few per cent for fainter objects.

Bewsher et al. (2010) did a similar analysis when calibrating the HI instrument but on a smaller scale. They used very similar input catalogues derived from the NOMAD catalogue but used a smaller region of the CCD field of view and applied some additional constrains to arrive at their sample. A summary of the parametric difference between their methodology and ours, can be understood from Table. 5.1. They also use some additional filters, in order to arrive at a more universal result. They are

(a) Consider only light curves with more than 18 data points.

(b) Subtract a 70 frame box-car average smoothed function from all light curves.

(c) Ignore objects that have a scatter within the top 25%.

---

[2]http://astro.corlan.nnet/gcx/html/node10.html

	Bewsher (2010)	STRESS
Aperture Radius	3.2 pixels	2.5, 3.0 &3.5 pixels
Inner sky annulus	4.7 pixels	6.0 pixels
Outer sky annulus	7.9 pixels	10.0 pixels
FOV	100 pixel radius	Entire CCD
Confusion criteria		
Magnitude Difference	1.5 units	2.5 units
Neighbourhood	8 pixels	10 pixels

Table 5.1: This table summarises the difference in parameters between STRESS and that used by Bewsher et al. (2010), for their calibration of the HI instrument. The differences though substantial gives rise to very similar results, as explained in text.
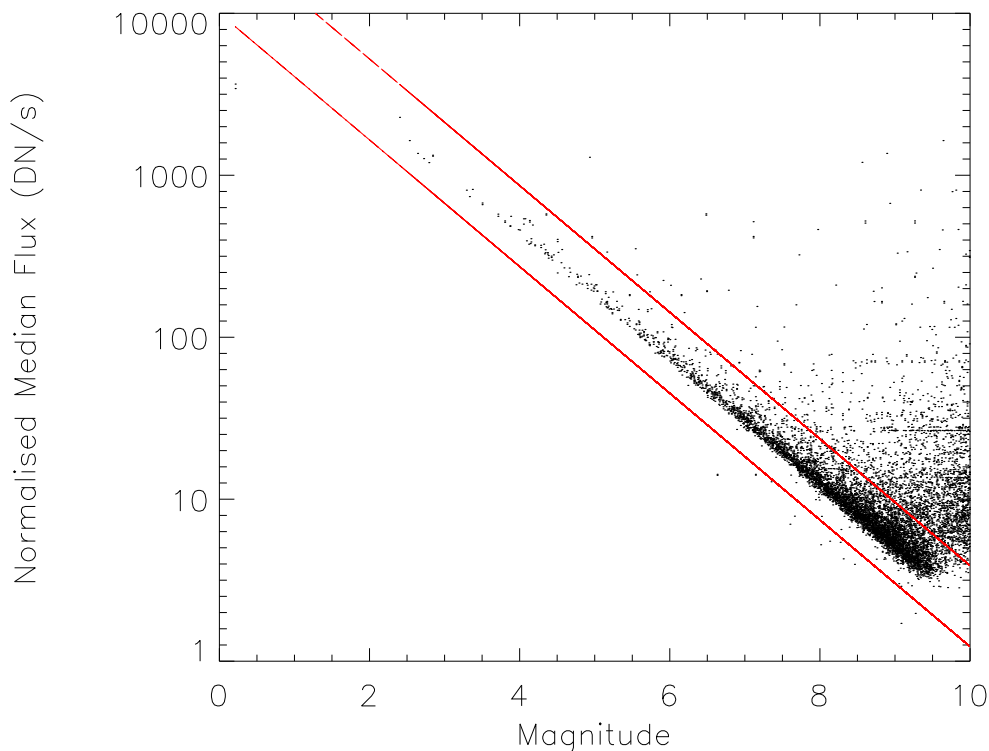
Figure 5.6: This is a plot of the median intensity values for the same set of stars as Fig. 5.7. One of the points to note is the apparent floor in the data, which is an unexplained feature as of now, but is a work in progress.

Fig. 5.6 is a plot of the median of observed intensity against their magnitudes for stars tracked during September and October of 2008. The sample is representative of general STRESS characteristics since the later part of the sample is through the highly crowded galactic plane region and the former is through a relatively sparsely distributed region. This plot is the equivalent of Fig. 13 (a) and Fig. 14 (a) of Bewsher et al. (2010). In our case, we can clearly observe a floor-like lower threshold to the observable flux value. This is also observed in their plot, by the
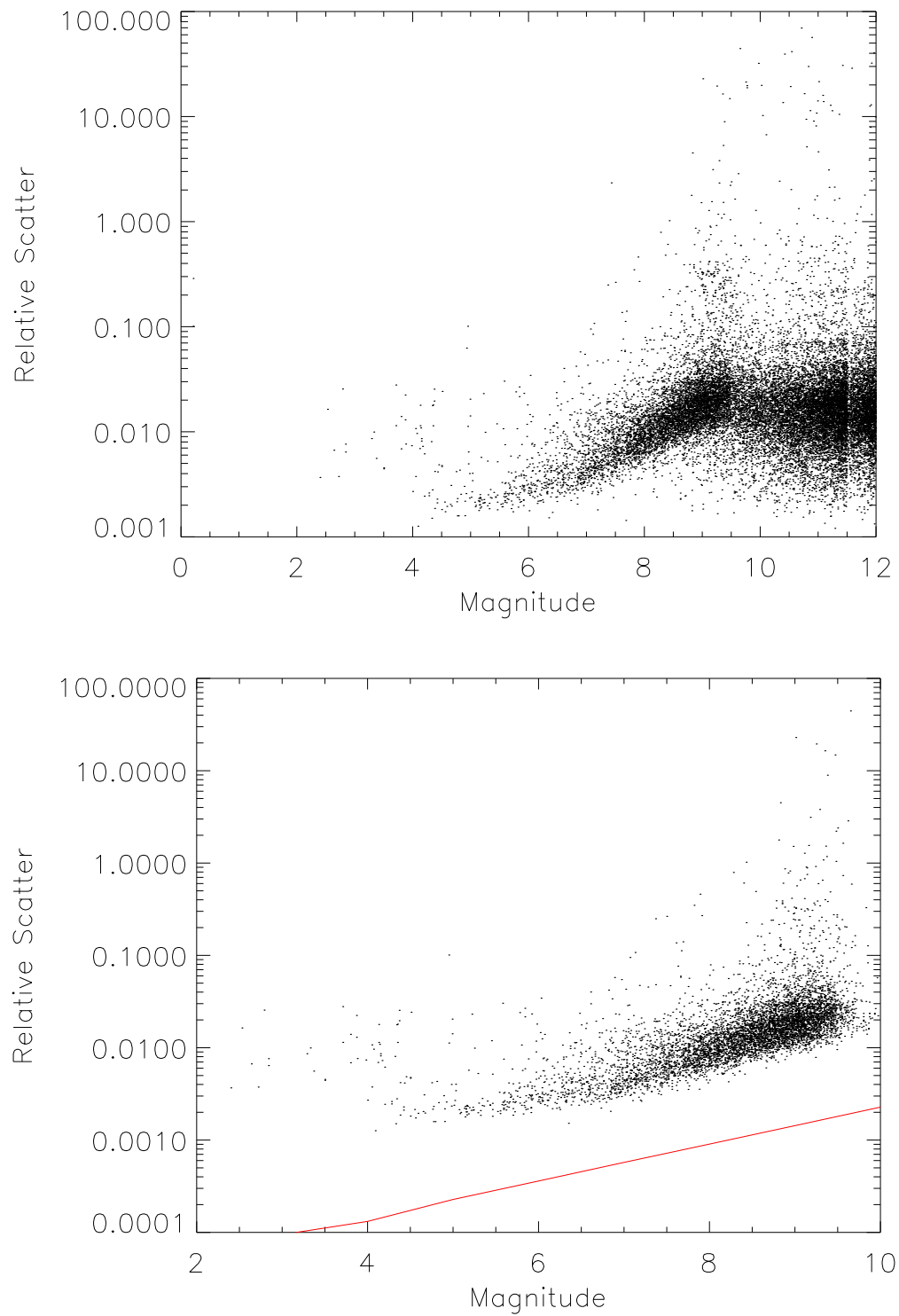
Figure 5.7: A plot of relative scatter of intensity against magnitude of the stars tracked during September and October of 2008. This is a representative sample of the whole data set. The top plot is for the complete data which is tracked during this period (22100) and the bottom plot is using data within the two lines in Fig. 5.6.

abrupt cut-off of data points at around 9th magnitude. This is, at the time of writing, an un-explained phenomena under investigation. Due to this deviation from normal behaviour in our data, caution is followed when analysing data for fainter stars. One thing to note, is that, both the plots, follow the expected linear trend until it reaches the floor. This trend also validates the calibration of the data. We also believe that this apparent floor in the data is also responsible for the observed scatter of data for fainter stars below the photon-shot noise limit as observed in the bottom plot of Fig. 5.7. Fig. 5.7 is a plot of the relative scatter (by relative scatter we mean the variance found after normalising the light curves with their median values) against R magnitude for the same set of stars. When the complete data for this period is plotted, we see the deviation from the expected trend, in fainter magnitudes in the top plot. But when the data is selected to represent the linear trend from Fig. 5.6 (data used is in between the two lines), we find that there is a clear observance of the expected trend. The line in the bottom plot represents the photon-shot noise limit for the HI instrument as calculated from the CCD configurations (RAL Technical Note, 8/8/2003). The light curves are smoothed using a running box-car filter with a length of one day before calculating the standard deviation. The effect of this box-car smoothing will be discussed in Section 5.4.2. This is to remove the residual effects of the F-coronal subtrac-tion and the small scale flat field effects. These features in Fig. 5.6 and 5.7 are being investigated with the help of the team at RAL. One of the reasons suggested for the behaviour of the data at around $V = 9.5$, is the effect of different sources in the NOMAD catalogue. Since NOMAD is a merged catalogue, the completeness varies with magnitude as per the individual catalogue. The data for most object with $V < 9.5$ comes from the Tycho catalogue and the difference in median value of our input catalogue at this point could be attributed to inaccuracies in the input magnitude for fainter objects, the data for which is obtained from other source catalogues.

5.4   Post processing of Light curves

From the above section, we find that there are many systematic variations present in the light curves. In this section, some of the methods attempted and finally used in removing or reducing these effects so as to do periodogram analysis on these data are presented. Some of the well

established methods in use for this type of analysis are the filtering methods like median smooth-
ing, box-car average smoothing and polynomial fitting. There are also various new specialised
algorithms like Sys-Rem (Tamuz, Mazeh & Zucker, 2005), Trend Filtering Algorithm (TFA;
Kovács, Bakos & Noyes, 2005), Photometric Detrending algorithm (PDT; Kim et al., 2009) for
this same purpose and which are used in many recent and on-going variability projects. These
are generic algorithms which have to be adapted to our very specific need. One of the main
differences between many of these projects and STRESS, is that the STRESS has a scanning
field rather a staring field. Hence, this gives rise to two simultaneously dependent variables, in
the form of the systematics across the CCD as well as the variation with respect to time. All
these have to be tackled and decoupled before we can clearly identify and understand the actual
variabilities present in our data. Below some of the basic methods which were adopted in this
project are discussed in brief.

5.4.1   Inter-quartile Range (IQR) Filter

Before the application of any of these trend removing methods, we have to tackle the issue of
blending due to the crowded field. Even though the confusion criteria mentioned in Section 4.3.1
removes close neighbours from the catalogue, specifically in a bright-faint pair, we find that we
are tracking the two objects alternatively depending on the component closest to the centre of
our aperture. This inadvertently raises the scatter of the light curves (Fig. 5.8).

Even though these light curves make up less than 1 per cent of the total, these 'alternating'
light curves generally create a very periodic signal, since the tracking algorithm shifts the cen-
troid very systematically depending on which star has the maximum flux inside that photometric
aperture at that point and hence often these light curves get flagged up as a periodic signal dur-
ing the period analysis. Even a robust sigma clipping does not identify these objects since the
actual variance of the light curve itself is very high and thus is an inaccurate measurement of
the scatter in the system. In order to flag these objects, we use the IQR measurements. After
calculating the IQR of all the objects in the same month, we normalise the value with their
respective median, find the standard deviation of these normalised IQR values and then remove
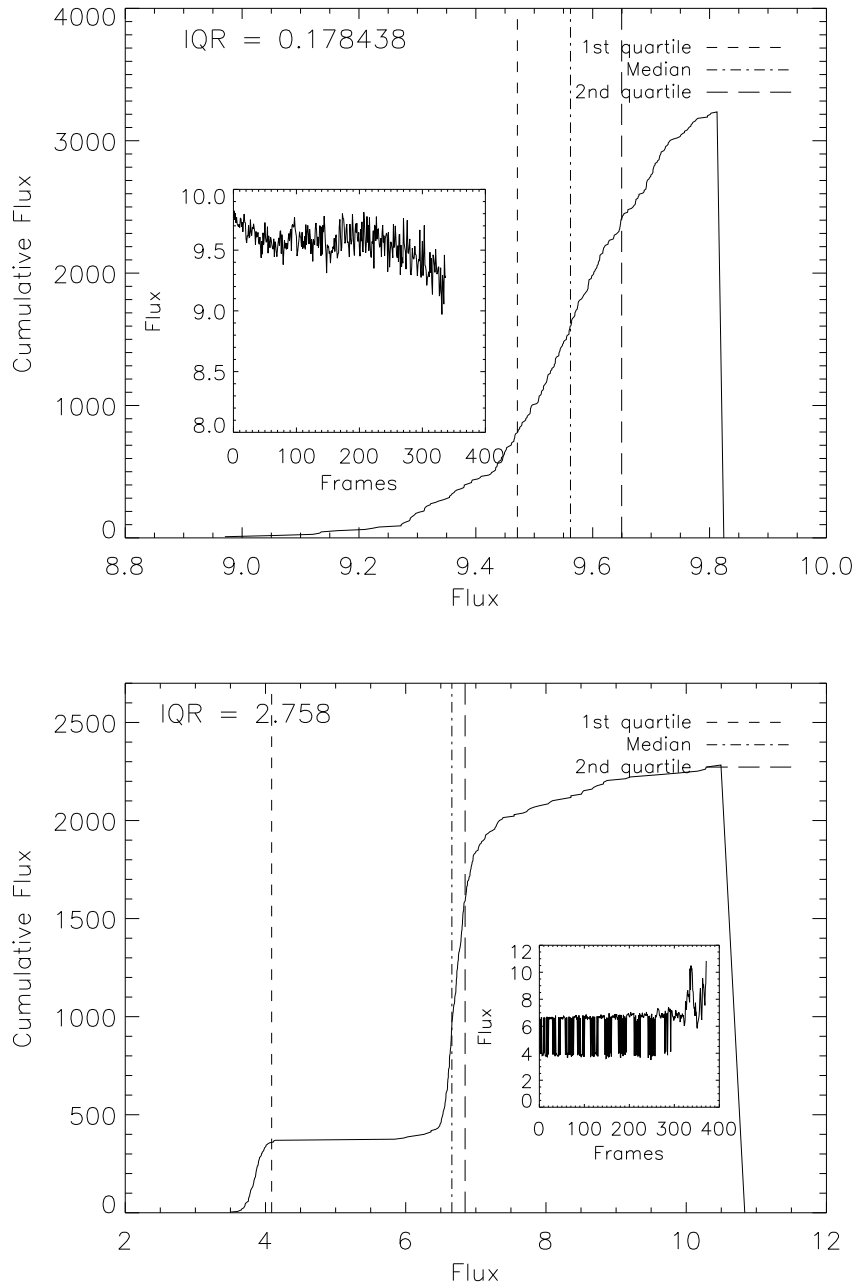
Figure 5.8: Quality control in light curves: Above is a plot of the cumulative frequency distribution of a light curve against actual intensity. The top plot is the distribution for a 'good' light curve, i.e, where only a single star is tracked throughout the CCD. The actual light curve is given as an insert in the plot. The bottom plot is the same distribution when two stars are tracked alternatively such as to give rise to a periodic signal. We can easily identify the 'bad' light curve from the inter-quartile range (IQR) which is given. The higher this value, the larger the spread of the light curve. This is a fairly robust method of identifying wrongly tracked objects in our data.

those objects which fall outside the 5 sigma value for this set of data. Fig. 5.8 illustrates this method in detail. The top plot is the cumulative flux against flux for a good light curve (see insert) and the bottom is the same for a highly scattered light curve, which is the tracking of two objects almost alternatively. Another method which was used initially as a cross-check, is the difference in intensity for the various apertures. The fainter stars have the same flux value irrespective of the aperture size and the brighter stars just increase their flux values as we go to larger apertures since most of the star PSF is captured in larger apertures. But when there is a confusion between two stars, there is a random change in flux values irrespective of the increase in aperture size. Hence, they can also be used to identify wrongly tracked objects but we use IQR for the rest of the analysis.

5.4.2   Box-car Smoothing

The light curves after passing through this IQR filter, still have a few other systematic variations which need to be removed, before analysing them for periodicity. The next step in the process is to remove the end noise in the power spectrum, namely, the noise at the high frequency and low frequency end of the spectrum. We use a simple box-car average smoothing to remove both these noise levels. It is understood that these moving averages or fitting of polynomial methods, introduces spurious peaks in the power spectrum (cf. Owens, 1977) but these peaks are assumed to be within the error limits of our analysis and hence this method is used.

High Frequency Noise Removal

We first remove the high frequency noise by using a box-car filter corresponding to a box width of around 3 hrs. Since this is also the time scale of any planetary transiting event, we implement this filter only for stellar variability analysis, which is explained in Chapter 6. The methodology used for planetary transit identification is described in Chapter 7. We use the resultant smoothed function for further analysis and the residual in the calculation of our signal-noise ratio when a stellar signal is confirmed. Fig. 5.9 illustrates how the use of smoothed function removes the high frequency noise in the data for a randomly selected standard star
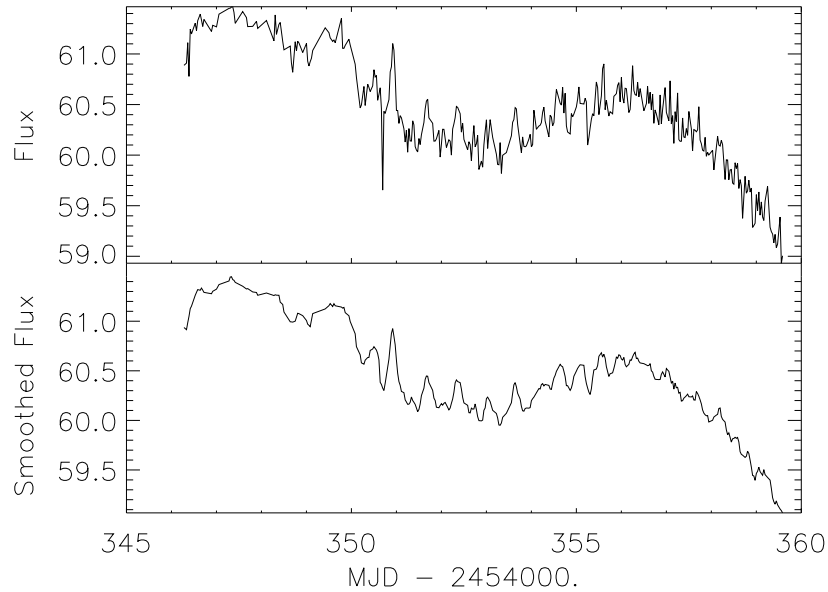
Figure 5.9: The top plot is the original light curve of a $V = 6.7$ star, HD 112281, before applying any correction and the bottom plot is the 3 hr box-car smoothed function which is used as the light curve for further analysis. This ensures the removal of the high frequency noise in the data. The box-car filter is applied only for stellar variability analysis and not for planetary transit analysis.

(HD 112281), which has no known variability. This has a V mag of 6.7[3] and a fairly non-interesting star, which will be used in all further processing examples. However, this is a K0 star at a distance of 172 parsecs with $B = 7.72$. Since it is a K0 star, it can be a giant and the residual variations observed could be attributed to its oscillations. However, the variations are below significance limit to derive a conclusion. The usage of the smoothed function as the light curve for further analysis does not significantly affect the standard deviation of the light curve.

Low Frequency Noise Removal

In order to remove the long term trends, two different methods were experimented with - a box-car average smoothing and polynomial fitting. After trying different box widths, a one day smooth was found as a best fit to many light curves. Removing a 4th order polynomial also tend to remove most of the long term trend in the data. From Bewsher et al. (2010) we know that the CCD response varies exactly as a fourth order polynomial across the CCD central

---

[3]All the magnitude mentioned hereafter refer to the data from the Simbad database for ease in cross checking.
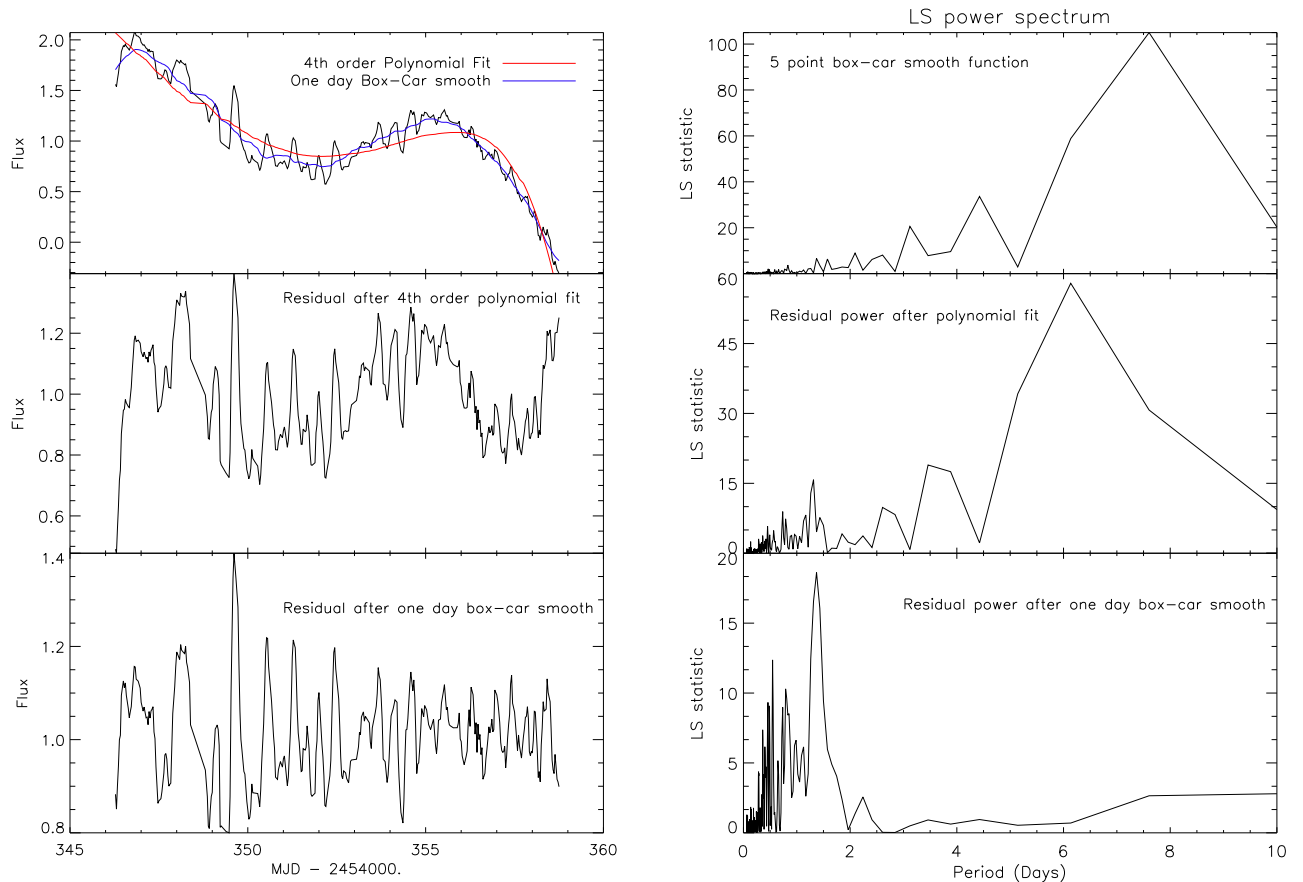
Figure 5.10: The figures in the the left are the light curve and the various smoothing functions operated on it and the plots on the light are the corresponding power spectrum. The top plot illustrates how both a 4th order polynomial as well as a one day box-car average smoothing function fits to the light curve of HD 112281. The bottom two plots give the residuals after removing the polynomial and the box-car filter respectively and their corresponding LS spectrum. It can be clearly seen that the box-car smooth function completely removes the low frequency power.

axis, and hence could be causing this systematic trend across all the light curves. One of the disadvantages of the polynomial fitting is that, when there is a larger stellar variability signal, this fitting becomes inaccurate and thus increases the variance of the final light curve. Even though, the polynomial is a theoretical fit to the trend, when the pipeline is automated, box-car smoothing is more reliable in removing the systematics only. Fig. 5.10 shows how a box-car average of one day and a 4th order polynomial fits the long term trend in this sample light curve. We also find that the polynomial fit does not remove all the long term trend, as seen from their corresponding power spectrum on the right side plot, but only that specified by the CCD function whereas the smooth function removes all types of long term variation. This is

certainly helpful when looking for variability periods relative to the length of the light curve.

STRESS data is unique in many ways and ironically, one of them is its systematic noise sources. In order to remove these systematics in our data, we turned to the help of some of the established trend removing algorithms used in the community. The basic method in all these algorithms is to find a template light curve which is a representative of the various trends found in the data and then reduce the residuals between the trend and the other light curves. The following section describe the two different methods, which were used in STRESS data in order to identify the systematic trends and remove them. However, there is also another method which is widely used by the community (Cameron et al., 2007; Karoff et al., 2007) developed by Tamuz, Mazeh & Zucker (2005). This method, known as Systemic Removal (Sys-Rem) was initially developed to correct the effects of atmospheric extinction but later found can be applied to any systematic effect found in many light curves. Thus, this method evolved to be used in general removal of systematic effect in a large number of light curves without any a priori information. This algorithm fits linear effects simultaneously in a large number of light curves and tries to reduce their residuals to achieve the best fit and then the process is repeated iteratively until all the linear effects are identified and removed. At present, this method has not been yet used in STRESS data for the removal of systematic effects. The following are the two different methods, which were eventually used in STRESS data in order to identify and remove the systematics.

5.4.3 Trend-Filtering Algorithm

One of the widely used algorithm for removal of systematic trends is the Trend Filtering Algorithm by Kovács, Bakos & Noyes (2005). This algorithm has been tested time and again in many transit search programs, but specifically with the HAT project (Bakos et al., 2004 and references therein). The basic methodology as the name suggests, is to identify the various trends in the light curve and then remove them by using a filter. This particular method can be used for two different purposes - one is to identify and remove the trends and the other is to reconstruct the shape of the present signal.

The algorithm selects a set of template time series randomly sampled throughout the CCD, in order to ensure uniform representation of the variations in the CCD position. The only cut-off employed at this stage is that stars with low data points, low intensity and high standard deviation are not included in the template. One of the assumptions in the next stage of processing is that all the light curve are sampled at the same time and have the same number of data points. Hence, they define a uniform template time base, which is the time base of light curve containing the largest number of data points, to be used in the filter. Once a template set has been defined, a weighted linear combination of the template stars provide the actual trend. The filter function $\{F(i); i = 1, 2, ..., N\}$ is built up from the following linear combination of the template time series $\{X_j(i); i = 1, 2, ..., N; j = 1, 2, ..., M\}$

$$F(i) \quad = \quad \sum_{j=1}^{M} c_j X_j(i) \ . \tag{5.4}$$

The template time series are assumed to have zero averages. The coefficients $\{c_j\}$ are determined through the minimisation of the following expression:

$$\mathcal{D} \quad = \quad \sum_{i=1}^{N} [Y(i) - A(i) - F(i)]^2 \ . \tag{5.5}$$

where $\{Y(i); i = 1, 2, ..., N\}$ is the target time series which is to be filtered, and $\{A(i); i = 1, 2, ..., N\}$ denotes the current best estimate of the *detrended* light curve. Fig. 5.11 shows the comparison in the power spectrum with and without TFA filtering (Fig. 6 of Kovács, Bakos & Noyes, 2005). It is clearly evident that detrending using TFA removes most of the aliases in the power spectrum and leaves only the noise behind.

One of the major assumptions in this code is that, "all photometered stars have the same number of data points distributed in the same moments of time" (Kovács, Bakos & Noyes, 2005) and this is not valid in the case of STRESS light curves. Still an attempt was made to modify this code for our purpose. Initial work is in adapting the code to read STRESS data and setting up bits and pieces which were specific to our data set. Due to the above mentioned assumption, i.e., all stars have their data points at the same moment in time, one of the major assumptions used in engaging the code is to write all the data sets to the same time base, irrespective of their
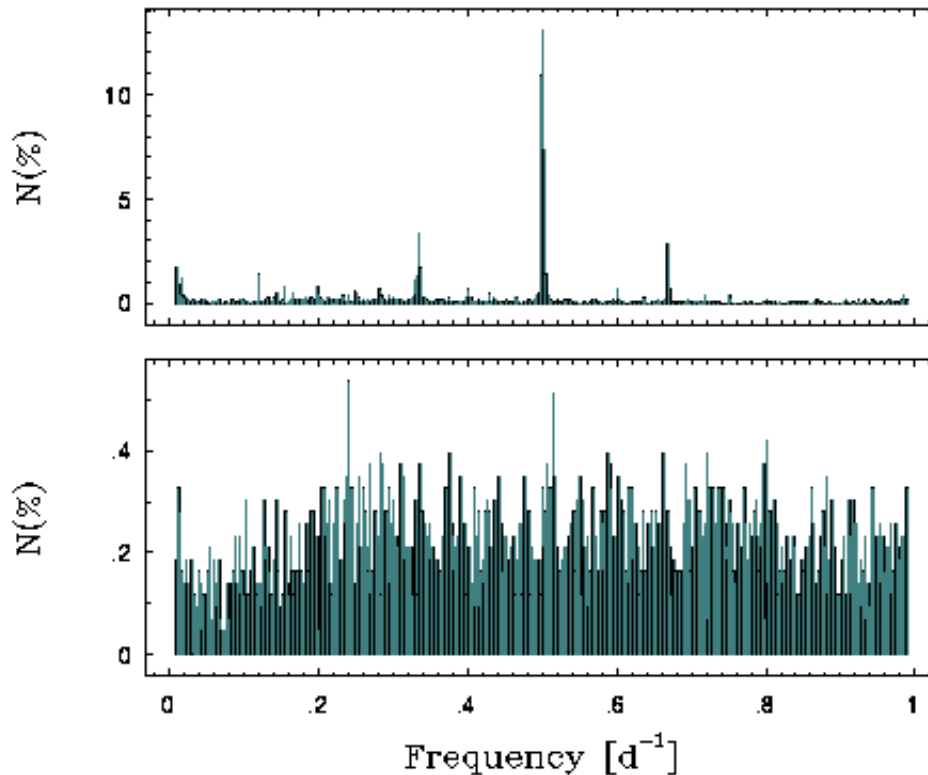
Figure 5.11: The plot is a comparison of power spectrum when the light curve is not detrended and when TFA detrending is applied (Fig.6 of Kovács, Bakos & Noyes, 2005). The top plot shows the one day alias in the data, whereas after TFA detrending, the power spectrum is dominated by the noise only.

temporal or spatial distribution. Until we made this assumption in our version of the code, the results of the analysis were mostly unnecessary, unwanted and totally irrelevant changes to the light curve. Hence this assumption was used. However, we do not force all light curves have the same number of data points and this does not seem to affect the results, as long as all of the light curves have a minimum number of data points - in our case, 50. Fig. 5.12 shows the effect of TFA filtering in our light curve scatter. Clearly, many of the objects have a decreased scatter when subjected to TFA filtering, but there are also objects for which the deviation has increased. This is a work in progress and needs more analysis to confirm.

Fig. 5.13 is an example of how the filtering algorithm helps us to identify true variability. As seen from the right plot from the figure, the long term trend in the light curve completely masks the actual binary signal and when TFA is applied to the algorithm, we can clearly see
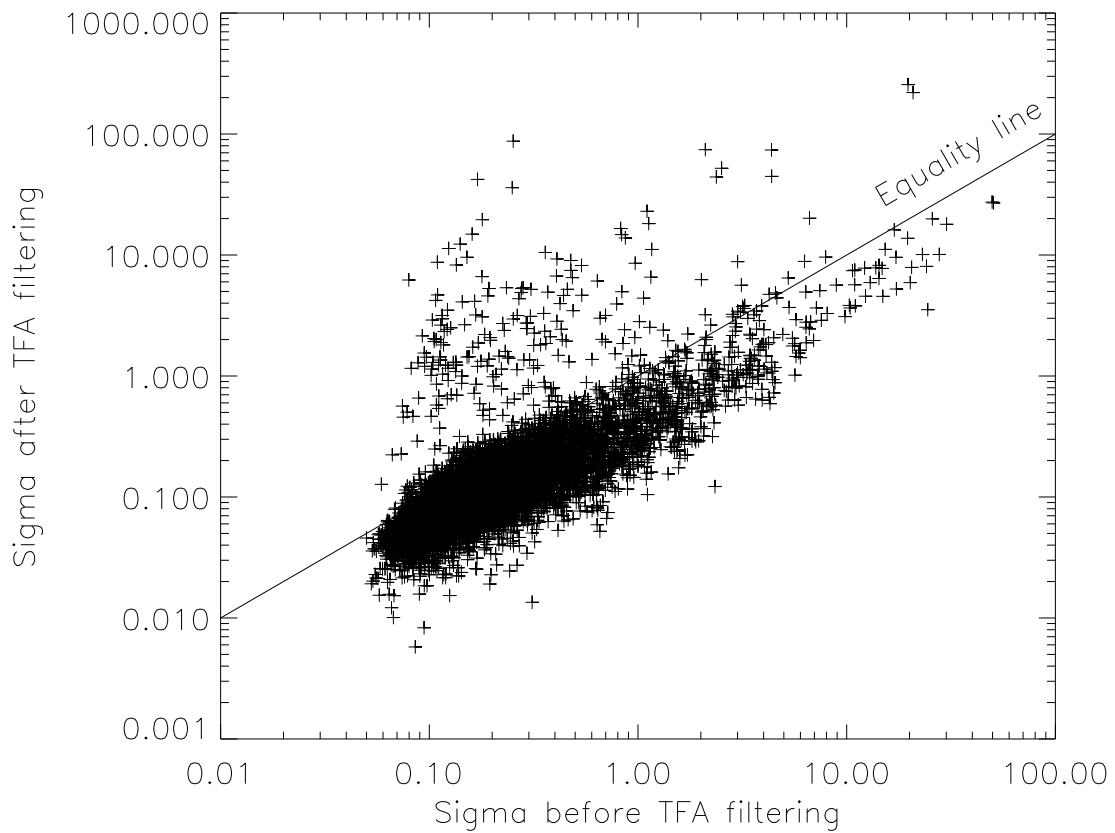
Figure 5.12: The above is a plot of the standard deviation of the light curves before and after processing with TFA. The dashed line represents the equality between the two sets of data. We can clearly observe that many objects have moved below the line, indicating that there is a reduction of scatter in many objects. However, there are also objects for which the scatter has actually increased when the TFA is applied. Hence caution and more tests are needed to confirm these two observations before TFA can be implemented into our pipeline.
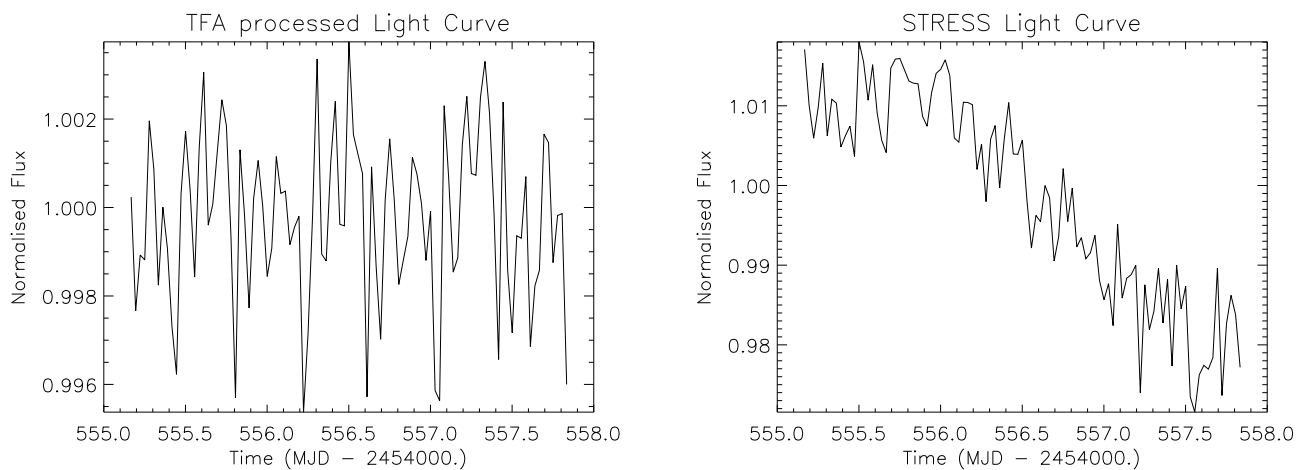
Figure 5.13: The right plot is the actual normalised light curve obtained with STRESS pipeline. It clearly shows the long term trend which is common to most light curves. When the TFA algorithm is applied to this light curve, it reveals well defined eclipses in the left plot, which were not seen in the right plot.
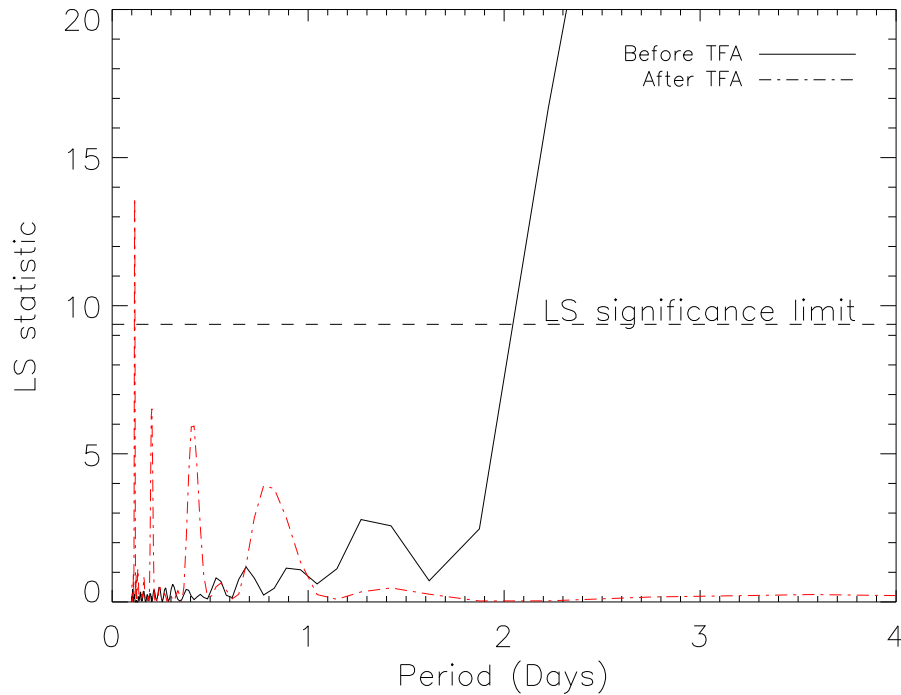
Figure 5.14: This shows the obtained Lomb-Scargle spectrum for the original STRESS light curve and the TFA processed light curve. The continuous line represents the spectrum before processing with TFA and the dashed line represents the TFA processed light curve spectrum. In the former case, the long term trend is masking the true variability which is picked up after filtering with TFA.

the variability signal. This is an F5 star with a $V = 7.9$, which Simbad categorises as a double star. Fig. 5.14 is a Lomb-Scargle power spectrum of the same star. The red dashed line is the power spectrum after TFA filtering and the continuous line is the actual light curve spectrum. We can clearly see the removal of the long term signal from the light curve. Fig. 5.15 is another comparison, where TFA actually modifies the light curve so much that any true variability signal, if present, will be impossible to decouple. Since we observe both these cases, utmost caution and more tests are needed to fit the TFA well to our data and use it. In spite of this inconsistencies when TFA is applied to the whole sample set, as from Fig. 5.16, we can understand that it does remove some of the long term signal from our data, as seen by the gap in the upper left corner of the plot. At the same time, it also does have some of the long term trends in the data. Investigation is going on as to how much of these long term trends are true signal and to determine the rate of false positive. At the time of writing, we only know that we can potentially use TFA for our data and TFA has not been integrated into our analysis pipeline.
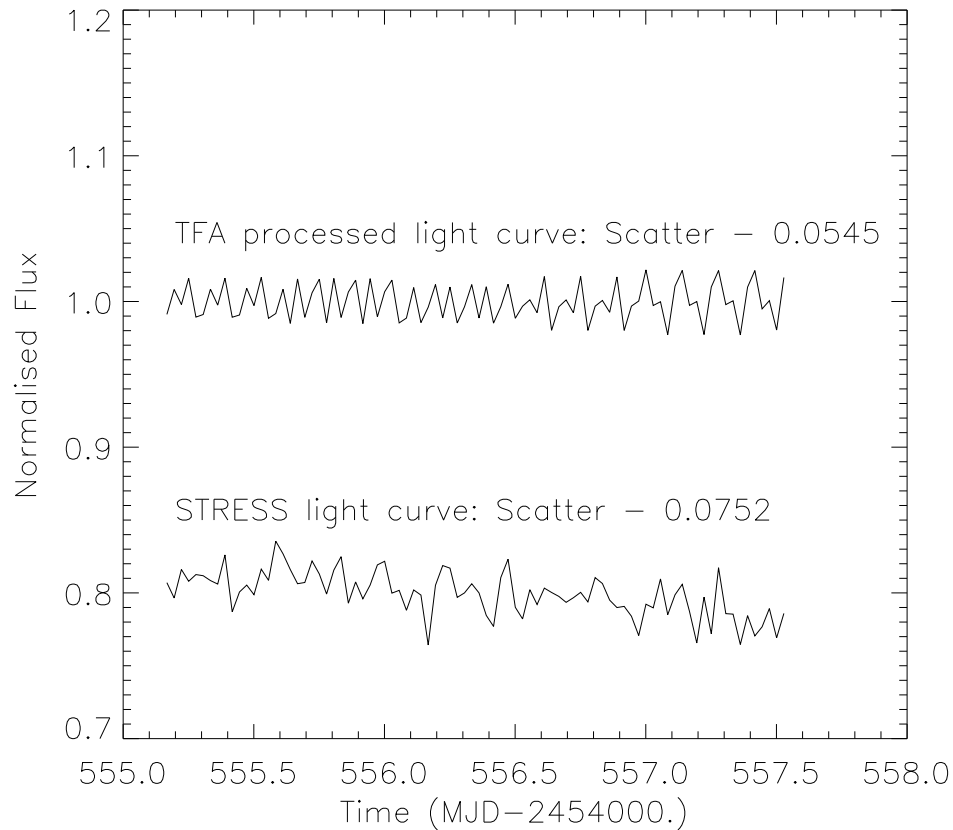
Figure 5.15: A comparison plot between TFA filtered and actual STRESS light curve, in the case when the filtering actually distorts the light curve even though it appears to decrease the scatter in the signal. The top plot is the TFA filtered light curve and the bottom is before filtering, offset from one for easy comparison.

5.4.4   Photometric Detrending Algorithm

Another algorithms which has been investigated in order to remove systematics from the STRESS light curves is PDT, developed by Kim et al. (2009). It follows the general assumption that systematic variations are present in clusters of stars and hence, efficient identification of these clusters is paramount to the removal of these trends. The first step is to select a template star light curve, which is highly correlated with other light curves. This is done using a similarity matrix formed using the Pearson correlation values between the considered template and other stars. One of the main conditions is that the number of measurements for every light curve has to be the same. Clusters, which are group of strongly correlated light curves are formed using the Hierarchical Tree Clustering algorithm (HTC; Jain, Murty & Flynn, 1999) and an agglomerative merging (bottom-up merging) algorithm. The tree clustering algorithm produces
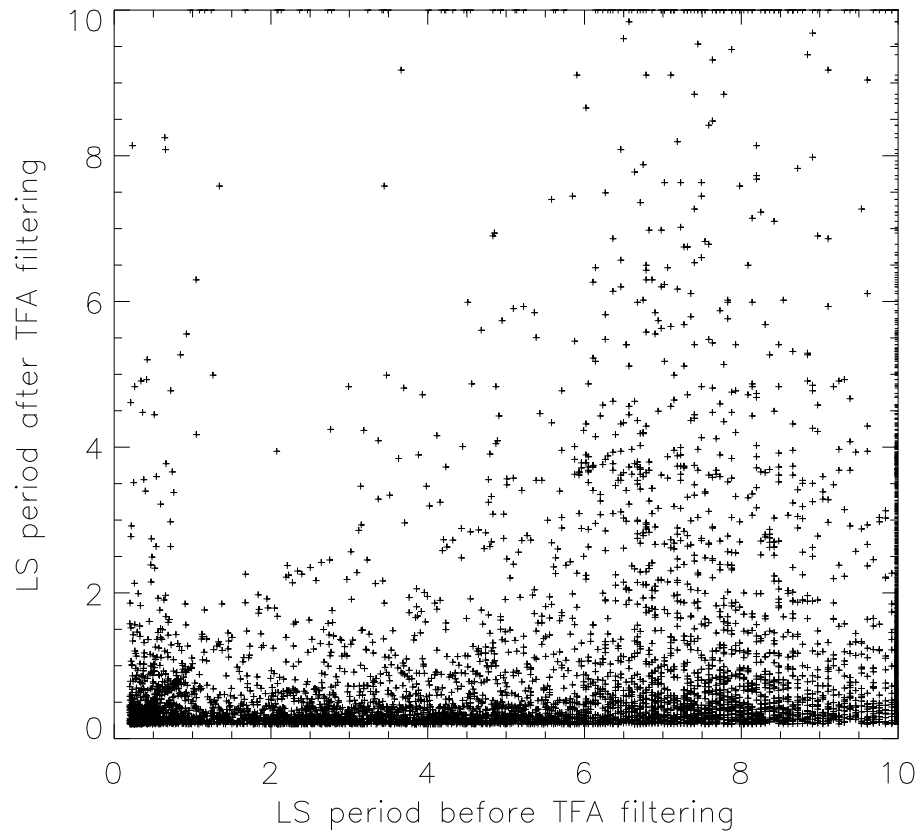
Figure 5.16: This is a plot explaining the LS periods found in light curves before and after applying the TFA filter. We can clearly observe that most of the long term signal is removed from the data when TFA is applied, from the gap in the top left of the plot, corresponding to the long term signal observed after detrending with TFA.

pairs of objects progressively based on certain threshold criteria, until finally all the objects are connected in some way. As we go higher up in the tree, the objects become more dissimilar or away from the initial threshold criteria. Then the bottom-up merging algorithm is used to arrive at the final cluster. This algorithm is used in order to identify outliers of a particular trend and to decouple clusters with more than one trend. This is achieved by the assumption, that when the clusters are converted to a normal distribution by Fisher's transformation (Fisher, 1915), the outliers and multiple trend clusters will fail the normality test. Clusters which have a single trend - 'master trend', are identified by an iterative method. A representative trend light curve is then produced by the weighted sum of all light curves from each cluster. The trends are removed by minimising the residuals, r_i between each light curve and the trend light curve,
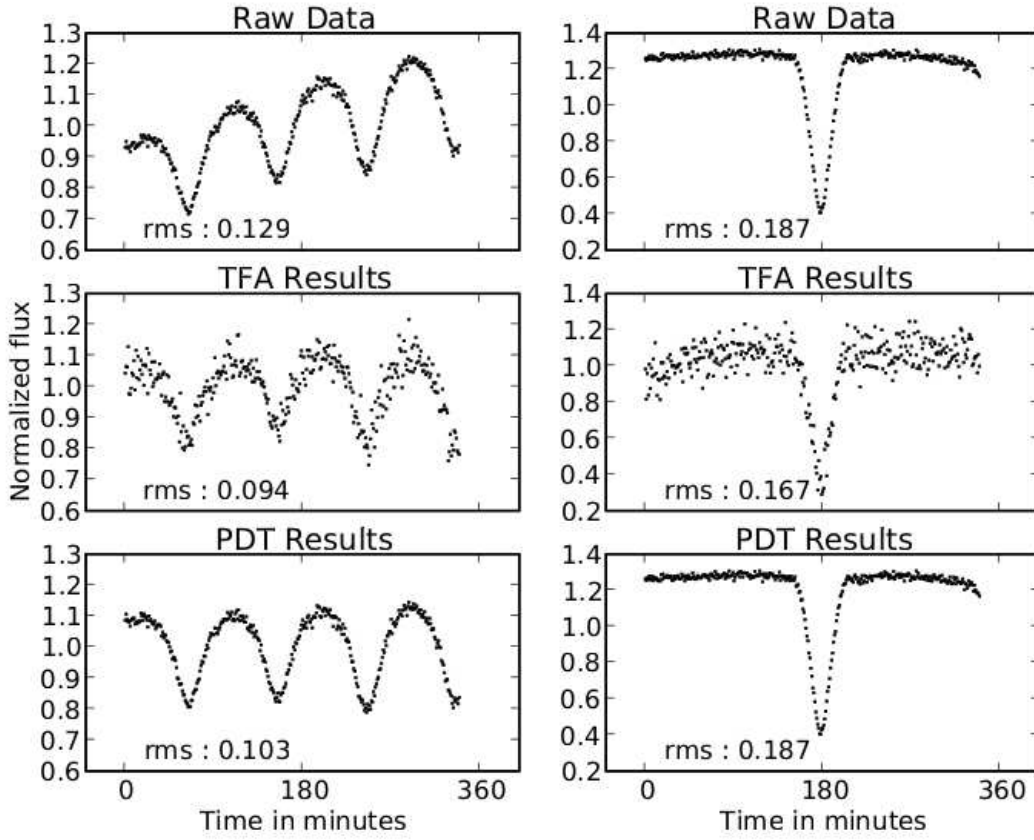
Figure 5.17: Comparison plots between PDT and TFA using simulated eclipsing binary (Kim et al., 2009). The rms value of the light curves are also indicated in the plot.

where r_i is defined as,

$$r_i = \sqrt{\frac{1}{n}\sum_t \left[L_i(t) - \lambda_i - \sum_k^m \beta_{ik}T_k(t)\right]^2} \tag{5.6}$$

where, r_i is the rms of the residuals, $L_i(t)$ is the flux of the stars i at time t, n is the total number of measurements, $T_k(t)$ are the trends for cluster k, m is the total number of clusters, β_{ik} and λ_i are free parameters to be calculated for each light curve. Fig. 5.17 is a plot of the comparison between the PDT and the TFA for a set of simulated eclipsing binaries (Fig. 6 of Kim et al., 2009).

One of the most important assumptions in this algorithm is that, all the light curves consist of equal number of data points. However, in STRESS light curves, due to various reasons, many of the light curves does not have continuous coverage and has data drop-outs (Fig. 3.7). The

bottom plot of Fig. 3.7 is a representative of the actual data and even though, some light curves do have complete coverage, there are lot of stars which have data losses and these are an important factor, if PDT has to be applied to these light curves. We attempted various methods, such as padding with zeros, interpolating the light curve, padding with median values of the surrounding region to make sure that all light curves have the same number of data points. The identified trend in all these above methods were highly influenced by the additional points, such that the real trend was lost and the rms value of the light curves actually increases. Therefore, until we find a way to completely track all stars, it is not possible to detrend the data using PDT.

5.5   Final Comments

Now that we understand the various trends in the data and have also investigated methods to remove these trends, we can analyse this data for variability using algorithms like Lomb-Scargle, Phase Dispersion Minimisation and Box-fitting Least Squares method. In the next chapter, we will discuss these algorithms in brief, describe the variability analysis pipeline developed for identifying eclipsing binaries and present the results of this analysis on HI-1A data.

Chapter 6

Stellar Variability Analysis

6.1 Introduction

The next step in this process, is to identify the different interesting objects. Owing to the huge data set, it is essential to have this process automated. One of important characteristics of the data is that it is unevenly sampled and hence, we have to resort to specialised algorithms like Lomb-Scargle (LS; Scargle, 1982), Phase Dispersion Minimisation (PDM; Stellingwerf, 1978), etc, rather than the traditional simple Fourier analysis. We use LS to flag up potential variables in our data and then confirm the periodicity with PDM. LS seem to be sensitive to the period used in smoothing whereas PDM was uniformly sensitive to all periods. Thus, this two step process using both the algorithms increased the probability of finding a true variable and eliminating false positives. The first half of this chapter will deal with a brief description of LS, PDM and the pipeline developed to identify different variables using these algorithms and the later part, will deal with the results of this analysis. We also used Box-fitting Least Squares method (BLS; Kovács, Zucker & Mazeh, 2002) to look for exoplanetary transits, which will be discussed in the next chapter.

6.2 Phase Dispersion Minimisation (PDM)

PDM is well suited for the use of identifying non-sinusoidal variation in irregularly spaced data. The methodology of this algorithm is to reduce the variance of the data with respect to the mean light curve. The method of reducing the variance in order to identify the period has been in use even before the usage of Discreet Fourier Transform (DFT; Kovács, Zucker & Mazeh, 2002). This is achieved by computing the phase of each data point for various periods and finding

the best fit phase by least-squares fitting. Since the technique minimises the dispersion of the data at constant phase, it is known as 'Phase Dispersion Minimisation'. The data is binned in phase-space and a minimum value is obtained for each bin by simultaneously obtaining the least-squares light curve and fitting for the best period. Thus, this technique can also be described as a 'Fouriergram' of infinite order. Hence the statistic,

$$\Theta = s^2/\sigma^2, \tag{6.1}$$

where,

$$s^2 = \frac{\sum (n_j - 1)s_j^2}{\sum n_j - M} \tag{6.2}$$

and

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1}. \tag{6.3}$$

In the above equations, we consider a light curve with N points of flux value $x_i, i = 1, 2, ...N$ with a mean value of $\bar{x}$. s_j^2 represents the variance of a subset of this light curve, and when M such subsets are considered, then s^2 represents the overall variance of the data. For a trial period Π, Θ varies between 0 and 1, with more smaller the value, the higher the probability of the presence of a signal. The signal-to-noise ratio in this case is given by,

$$\epsilon = \sigma_0/\sigma_N \tag{6.4}$$

where, σ_0 represents the maximum variance produced by a signal of amplitude A, i.e., for a sine wave it will be $\sigma_0^2 = A^2/2$ and σ_N is the variance of the observational noise in the data and sum of these two will give the overall variance of the system. The PDM technique is efficient in finding all the periodic components including the sub-harmonics in the system. In order to identify only the fundamental frequency, it is advised to use a coarse bin size when folding the light curve. In the case of a fewer number of observation points, N, the signal-to-noise ratio can be obtained as,

$$\epsilon = \left[\frac{F - 1}{1 - (\alpha/N_b)^2 F} \right]^{1/2} \tag{6.5}$$

where F represents the F distribution and consequently the probability density of the Θ statistic, N_b is the number of phase-bins and α represents the curve shape of the Θ statistic. From Fig. 2 of Stellingwerf (1978), we find that the algorithm can effectively identify a signal even when there are as few data points as 10 for $\epsilon > 1$. This is an added advantage in the case of

STRESS since we can look for periodicities even when the object is not tracked very well and this aids in adding statistical significance to the identified objects. This method has been used as a confirmation in other space data analysis like the *COROT*. For example, Chadid et al. (2011), uses this method along with Period04, etc to confirm the presence of Blazhko effect in RR Lyrae stars and this method has been used in studies of eclipsing binaries. Schwarzenberg-Czerny (1997), did an analysis of the PDM method, to determine the correct probability distribution. He came to conclusion that PDM follows the beta distribution rather than F distribution assumed by Stellingwerf (1978). Schwarzenberg-Czerny (1997) also concluded that with the regular usage of the Fourier transform methods, the motivation for the use of this method has weakened.

6.3   Lomb-Scargle (LS)

If a data contains a periodic signal, the common method of resolving this signal is by a Fourier transform or periodogram analysis. Scargle (1982) modified the well known periodogram method to suit for unevenly spaced data. This method removes the statistical difficulties associated with a classical periodogram for an unevenly spaced data. The new definition, renders the new method an equivalent of the least-squares fitting of sine waves for the data similar to that derived by Lomb (1976). This method is hence known as the Lomb-Scargle periodogram.

Consider a time series of a physical variable, for example, flux, as $\{X(t_i), i = 1, 2, ..., N_0\}$, the classical periodogram, $P_X(\omega)$ is defined as,

$$
\begin{aligned}
P_X(\omega) &= \frac{1}{N_0}|FT_X(\omega)|^2 \\
&= \frac{1}{N_0}\left|\sum_{j=1}^{N_0} X(t_j)\exp(-i\omega t_j)\right|^2 \\
&= \frac{1}{N_0}\left[\left(\sum_j X_j \cos\omega t_j\right)^2 + \left(\sum_j X_j \sin\omega t_j\right)^2\right].
\end{aligned}
$$

(6.6)

When using the above function, when X contains a true sinusoidal period of frequency, ω_0, then at and near $\omega = \omega_0$, all the factors are in phase contributing to a maximum power value

and when it is not a true period, both the values in second equation have random phases, thus cancelling each other. Scargle (1982), redefined this periodogram as,

$$P_X(\omega) \;=\; \tfrac{1}{2} \left\{ \frac{\left[\sum_j X_j \cos \omega \left(t_j - \tau\right) \right]^2}{\sum_j \cos^2 \omega \left(t_j - \tau\right)} + \frac{\left[\sum_j X_j \sin \omega \left(t_j - \tau\right) \right]^2}{\sum_j \cos^2 \omega \left(t_j - \tau\right)} \right\}, \qquad (6.7)$$

where τ is defined by

$$\tan\left(2\omega\tau\right) = \frac{\left(\sum_j \sin 2\omega t_j \right)}{\left(\sum_j \cos 2\omega t_j \right)}. \qquad (6.8)$$

The main advantage of this definition is that this method is the equivalent to the reduction of the sum of squares in least-squares fitting of sine waves to the data. This definition, in the case of evenly spaced data, reduces to the definition of the classical periodogram and also has the property of time-translation invariance, i.e, if all the data is shifted in time simultaneously, the power spectrum still remains the same, much like the classical periodogram. Scargle (1982) derives the probability distribution using the cumulative distribution function. Two important statistical functions, which are essential to the usage of this periodogram is defined as follows:

(a) **False Alarm Probability, P_0:** The probability that the detected signal is false within a given confidence limit is called as the False alarm probability of a function. The detection threshold power for this periodogram, z_0 can be defined as,

$$z_0 = -\ln\left[1 - (1 - P_o)^{1/N} \right] \qquad (6.9)$$

where N is the number of frequencies searched and P_0, the False alarm probability. For a small P_0, for example, $P_0 = 0.01$, the threshold reduces to

$$z_0 = 4.6 + \ln\left(N\right). \qquad (6.10)$$

(b) **Detection Efficiency, DE:** The minimum probability of a method finding the peak signal in the data is known as the Detection Efficiency of the method. In this case, for a peak
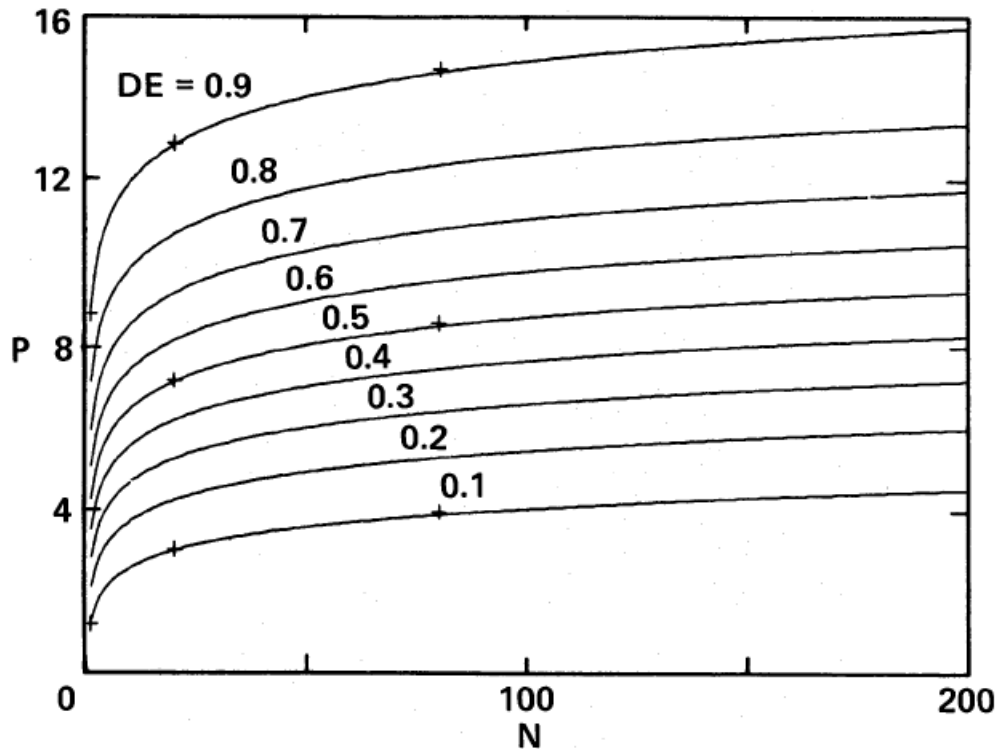
Figure 6.1: This is Fig. 2 from Scargle (1982), shows contours of constant detection efficiency (DE) as a function of the number of frequencies (N) searched and the corresponding signal strength. A higher detection efficiency is achieved even with a smaller sample of frequencies when the signal strength is high, which is as expected.

power P, this probability is given by,

$$\text{DE} = 1 - p^* (N, P) \tag{6.11}$$

where $p^*(N, P)$ is the CDF evaluated at the detection threshold for N sampled frequencies. The detection efficiency depends considerably on the number of frequencies searched as shown by Fig. 2 of Scargle, 1982 (Fig. 6.1 reproduced here).

Scargle (1982) defines this periodogram by making an assumption about the noise which is present in the signal - that is, the noise is independent of each other, normally distributed with zero mean value and the variance of noise is unity. The significance limit calculated with these assumptions is greatly influenced by the number of frequencies the signal is searched for. Hence, it is essential to determine the number of searched frequencies in order to identify a true signal present within the noise at a certain significance. In our data, we normalise the light curve with the median value in the initial stages and the *scargle.pro* normalises the light curve with the variance. Thus, we comply with two of the assumptions made by Scargle (1982). However,

we also make the assumption that the noise in our data is independent to the first order, even though we do find some correlation in our light curves. This dictates that we confirm the results from LS with another method which can identify the periodic signal in the presence of correlated noise. For this purpose we use the PDM method as a confirmation of the periodicity found by LS method.

The LS method is the commonly preferred method today for analysis of time series in all fields (example, Thong et al., 2004 used this method in bio sciences to characterise heart rate data) and is found to be efficient at finding and testing the significance of weak signals or signals in a noisy data. Horne & Baliunas (1986), clarified the usage of the total variance of the data as a method for normalising this periodogram. They also gave an empirical definition of the correct number of frequencies searched in order to estimate the false alarm probability for a particular signal detection. Thus, the number of searched independent frequencies, N_i is given as,

$$N_i = -6.362 + 1.193N_0 + 0.00098N_0{}^2 \qquad (6.12)$$

where, N_0 is the number of data points. For example, in the case of HI-1A instrument, the maximum number of data points, $N_0 = 700$ and the number of frequencies searched, N_i is approximately 1300. They also comment on the usage of the false alarm probability for second prominent frequency, in the case of the presence of more than one signal period as well as the calculation of the uncertainty of the frequency. Even though, Horne & Baliunas (1986), insisted on the use of only the total variance of the data as a proper way to normalise the periodogram, Schwarzenberg-Czerny (1998) disputes this in part, by claiming that it is valid to normalise the periodogram by either the total variance of the data or by the residual variance calculated after the subtraction of a sinusoid. Schwarzenberg-Czerny (1998) also comments that LS is biased in the sense that the amplitude estimates do not converge to the true value of amplitude.

The calculation of this periodogram has a computational price of the order of N^2 operations to analyse N data points. Press & Rybicki (1989), made this method computationally attractive by the usage of Fast Fourier Transform (FFT), to compute the individual bits of the equation, thus speeding up the process to an order of $NlogN$ operations. This method came to be known as *FASPER*. We use the IDL function *scargle.pro* from the IAAT library, in our computation of

power spectrum. This is a normalised LS periodogram (with total variance of the data) which, uses the Horne & Baliunas (1986) equation for calculating the number of searched frequencies and uses an in-built $fap = 0.01$, unless otherwise specified. Another inbuilt IDL function for computing LS spectrum is the *LNP_TEST* function, which has some additional functionality over the basic *scargle.pro*, like an oversampling factor parameter and also, it tests the hypothesis that the found signal is a true period against the hypothesis that they represent random noise in the data. The results of both these methods are fairly similar and the preference of usage is biased on the user. For the general flagging of interesting objects, we search in a broad range of periods from 0.06 days to 10.0 days. As mentioned in Section 6.1, the Nyquist period for the *STEREO* HI-1A data is 0.05 days and the maximum span of our data is 20 days. For a detailed analysis using LS, refer to Section 6.4.

6.4   Variability Analysis Pipeline

The previous sections gave a brief overview of the two main algorithms used in the variability analysis of STRESS data. This section will deal with the description of the semi-automated pipeline devised for the identification of the variable stars in the data. The data is now available for more than three epochs for most of the targets. This enables the search in two different period spaces - the 20 day chunks of data and the combined data from either three or four epochs. For the initial identification pipeline, the smaller 20 day data sets are used as a starting point. Once a variable is identified, then data from all epochs is combined and the period is confirmed by individual analysis, mainly using Period04 software (Lenz & Breger, 2004). Even though the pipeline is basic, we use various regions of the sigma-period space (a plot of the standard deviation of the light curve against the period found by LS), in order to identify different types of variables. This is a very basic method but seems to yield substantial results as seen from Section 6.5. The various process given below follow the methods explained in Section 5.4, which is also summarised in Fig. 6.2.

(a) The input files are the light curve information stored in the IDL format for each month, which contains the various flux and its error values, RA and DEC position with magnitude, CCD X and Y positions, sky background and its error values as well as the time array.
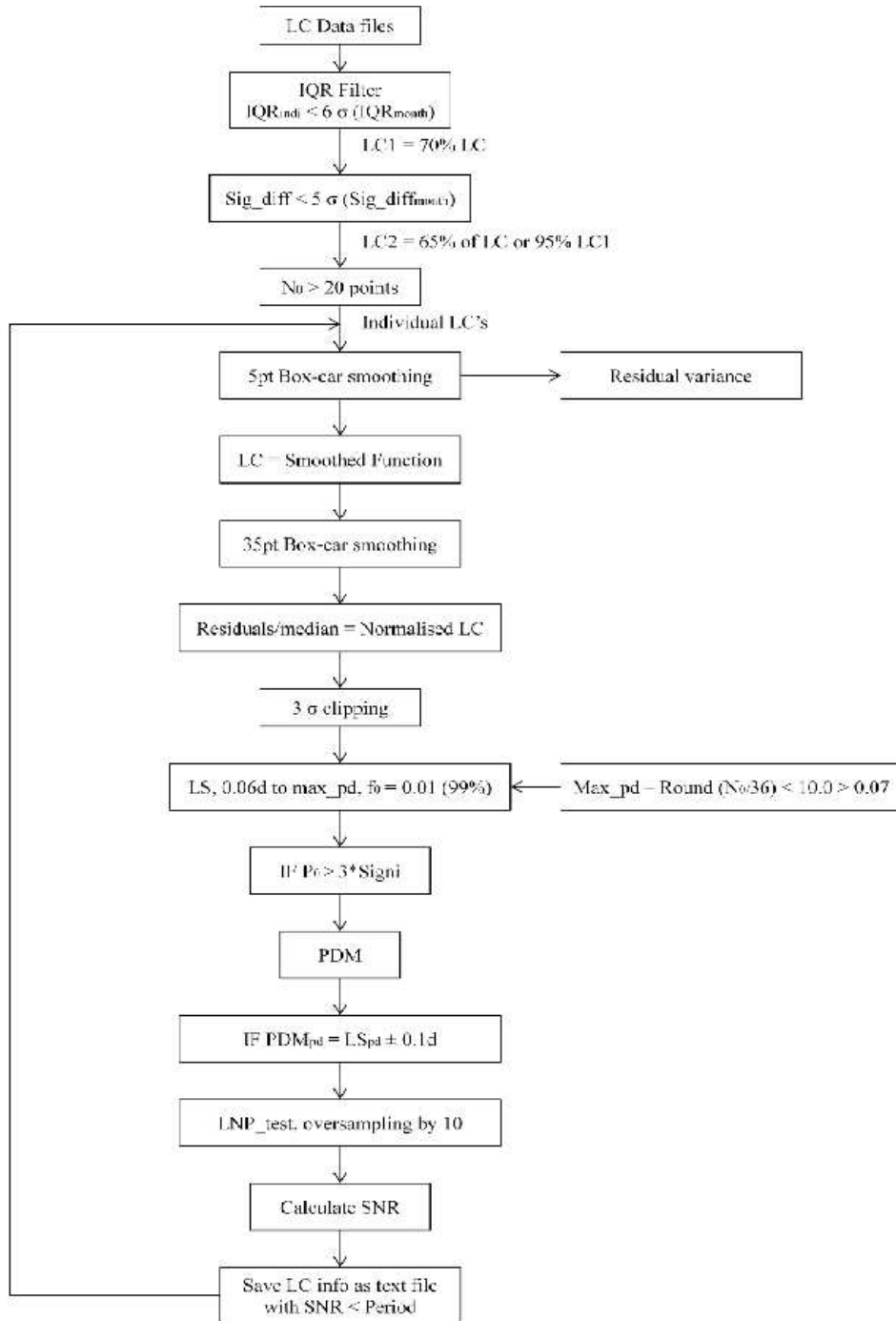
Figure 6.2: Schematic of the pipeline used in the stellar variability analysis

Only the flux from the smallest aperture is considered in this basic analysis.

(b) We use the IQR filter described in Section 5.4.1, in order to identify and remove the confusion resulting from alternative tracking between the target and nearby stars. We select only those stars, whose IQR values are within 6σ (standard deviation of the light curves for this particular month) range and discard the rest from further analysis. On an average, this filter selects around 70% of the total light curves to proceed to the next step.

(c) The targets selected by this method undergo another filter in order to ensure that these are not heavily contaminated by the Solar effects. It is in effect a sigma clipping of a sort. To this effect, we calculate the standard deviation of the light for the first and second halves separately and calculate the difference in them, sig_diff. If this value, sig_diff is outside a range of 5 standard deviations of the group, then we reject these light curves as well. The number of those targets selected after this step is column 3 of Table 6.1. As seen, this step filters out more than 35% of the total light curve or selects about 95% of the total light curve for further analysis.

(d) For each individual light curve considered, those with fewer than 20 data points are discarded from further analysis. Then we perform a 5 point box-car average smoothing, corresponding to 3.3 hrs, as mentioned in Section 5.4.2, to remove the high frequency noise in the data. The smoothed function is used as a light curve in further analysis and the variance of the residuals of this smoothing are used in the noise calculation in later stages.

(e) The long term trend in the light curve is now removed by using a box-car average smooth of 35 points, corresponding to the number of images collected in a day (see Section 5.4.2). The residuals after smoothing are now used in further analysis. The light curve is also normalised at this time by dividing it by the median value of the flux [1]. In order to make sure to remove all outliers from the data, we again perform a sigma clipping on the light curves using a 3 σ range limit.

(f) The light curves are now analysed for periodicity using the Lomb-Scargle periodogram described in Section 6.3. We fix the period range to be searched for based on the individual

---

[1]The term 'flux' henceforth refers to this normalised flux value unless mentioned

light curve itself. The minimum period searched for is 0.06 days corresponding to the
Nyquist frequency of the data. The maximum period depends on the length of the data.
We find the length of data to the nearest day and subtract half a day, to get the maximum
period for the LS. For the general analysis, we use the inbuilt false alarm probability of
$P_0 = 0.01$ (99% significance).

(g) The result of the previous analysis raises a flag, if the maximum power in the spectrum
exceeds three times the theoretically calculated power ('signi' keyword). Those light curves
not meeting this criteria are ignored from further analysis.

(h) The light curve is now analysed for periodicity using the Phase Dispersion Minimisation
(PDM; see Section 6.2) algorithm. If the period is confirmed to be almost equal (± 0.1
days) of the LS period, the SNR is calculated based on Eqn 23 of Horne & Baliunas (1986),
which is given below:

$$\xi = -1 \left\{ 1 + \frac{N_0}{2 \ln \left[1 - (1 - F)^{1/N_i} \right]} \right\}^{-1} \tag{6.13}$$

where, ξ is the SNR, N_0, the number of data points, N_i is the number of independent
frequencies searched and F is the false alarm probability. For confirmation, we again do
the LS search but by oversampling the frequency range by a factor of 10, in order to gain
the most accurate estimate of the period. We then save the period information from all
the algorithms in a text file before repeating the cycle from step (d) for the next light
curve.

The result of the analysis from the above methodology is presented in the next section. As
can be seen from the rest of this chapter, the method is efficient to identify some very interesting
known and also new variables.

6.5   Results

The results presented in this section are just to give the reader a flavour of the complete set of
results. These results in its entirety will be published in the future as a catalogue. Those stellar

objects which vary in their intensity are known as 'Variable stars'. The *STEREO* instrument has the capability to observe many of the different types of variability mentioned in Table. 1.2. It is well suited to observe δ Scuti's, γ Doradus, Eclipsing binaries, RR Lyrae's, β Cepheids, a few classical Cepheids as well as a variety of rotating variables like α^2 CVn and RS CVn stars. Previous studies of Eyer & Mowlavi (2008) has shown that nearly 10% of the Hipparcos observed stars and around 4% of the observed stars from ASAS are identified to be varying at some level. In STRESS, we find a very similar contribution from the variable stars, which is around 4% as can be found from Table. 6.1. This table lists the total number of stars observed in each month, the total number of selected objects for variability after the application of the various filters explained in Section 5.4 and the final column contains the number of stars confirmed to be a variable using the pipeline explained in Section 6.4. From a rough estimation of these values, we can infer that nearly 4% of the total objects observed in each month are periodic, varying in periods from 0.05 days to 8 days. There are exceptions to this statistic for a few months, which can be attributed to the fact that, as we move closer to the Galactic Centre, the stellar density increases, resulting in more stars rejected due to wrong tracking or confusion. Thus the number of variables identified from STRESS is in step with the other big surveys in the past.

Figures 6.3, 6.4 and 6.5 represent the period distribution of the stars at various points in the pipeline. After the various post-processing methods, the light curves are searched for periodicity using LS. Fig. 6.3 represents the period distribution of this selection of stars. This contains around 319,853 non-unique data points, since the star will be observed in more than three epochs in most cases, and this reduces the unique data points to nearly one-third of this value - 100,000. Fig. 6.4 is a distribution of the period of those object which are confirmed to have similar periods (± 0.1 day) by both LS and PDM. Fig. 6.5 contains the periods of those stars which are presented in the appendix as a representative sample. The sample presented is not exhaustive but to give a flavour of the different types of variable stars which can be studied using this data. To illustrate this, below a subset of the sample is presented.

Month	Total LC	Selected LC	LS variables
2007-04-01	10401	7324	379
2007-05-04	20646	16632	670
2007-06-02	28319	23063	1050
2007-07-03	11242	8908	455
2007-08-03	9554	6771	392
2007-09-02	11041	7539	427
2007-10-01	13460	9932	756
2007-11-01	20362	14549	1102
2007-11-30	159667	11661	870
2007-12-21	51525	6045	507
2008-01-18	7191	5253	482
2008-02-20	7964	5602	418
2008-03-18	5378	3646	238
2008-03-29	12629	8437	561
2008-04-26	20459	13296	573
2008-05-12	32468	21855	1141
2008-06-12	13498	8879	476
2008-07-09	9602	6232	336
2008-08-06	11281	6995	312
2008-09-05	16374	10665	491
2008-10-06	16331	10448	226
2008-11-02	26038	12885	532
2008-11-14	142865	13005	964
2008-11-28	62941	6811	649
2008-12-28	7392	5409	411
2009-01-30	8066	5646	405
2009-02-25	10400	7351	544
2009-03-30	25409	17529	1152
2009-04-29	23650	15530	752
2009-05-26	12424	8426	459
2009-07-01	8930	6146	394
2009-07-31	12928	8628	360
2009-09-01	13249	9578	414
2009-10-30	90671	4034	140
2009-11-20	6101	4458	369
2009-12-18	7647	5449	372
2010-01-20	8207	5895	376
2010-02-18	7107	5038	307
2010-03-07	7020	4548	233
2010-03-16	13577	9015	396

Table 6.1: This table contains the total number of light curves in each month, the number of light curves which are selected for further variability analysis by the different post-processing methods and finally the number of confirmed variables. The total number of all the months is not unique, since the same star will be observed more than once during this three year project and can attribute to the final number of identified variables more than once. On average 4% of the total light curve in each month is found to be a variable by this pipeline, which is similar to other surveys in the past.
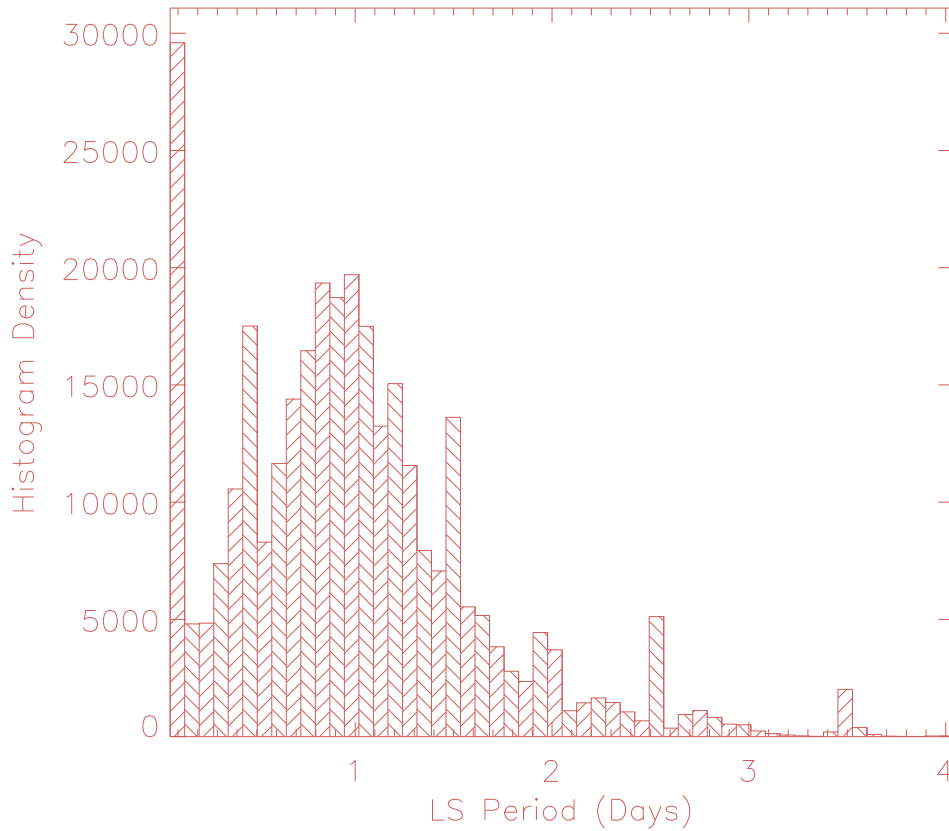
Figure 6.3: The is the distribution of period found by LS before it is confirmed by PDM. There are a few periods longer that 4 days given in this graph but for ease of understanding that has been omitted. The distribution is single humped around one day. At this time, we are not sure of the reason for this preference, but it is being investigated.

6.5.1 Individual Stars

The results presented here are a subset of the sample presented in Appendix. This subset contains some of the most interesting objects analysed during this thesis, though it is not exhaustive. This is a random representation of the whole data in terms of magnitude, different variabilities observed as well as quality of the light curve in terms of scatter. For example, the left plot in Fig. 6.6 shows the whole data set for the δ Scuti variable 44 Tau, for the three epochs, each made up of 20 days and the right plot is a blow-up of a few days of data where the variability is clearly evident.

HD 26322: (= 44 Tau) A F2 star, δ Scuti variable with 49 known pulsational frequencies (V mag = 5.39) in 15 independent modes (Lenz et al., 2010 and references therein). The
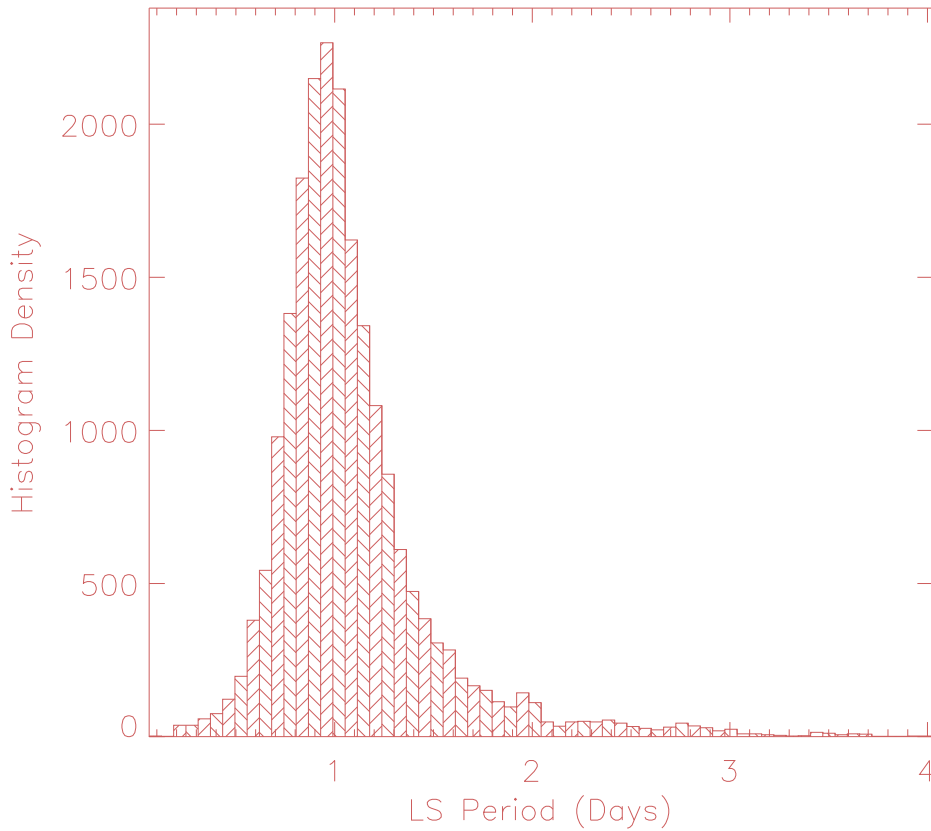
Figure 6.4: The above plot represents those periods that are initially identified by LS but also confirmed by PDM as described in Section 6.4. In this case as well, we see the preference for periods around one day as in Fig. 6.3, indicating that these might be true periods as well.

evolutionary phase of this well studied object is still unclear and under investigation. STEREO has data for three epochs (Left plot in Fig. 6.6) and we were able to identify at least 15 pulsational frequencies using the Period04 package (Lenz & Breger, 2004). We find a dominant period of 0.144 days. The next four most prominent periods with STRESS data as given by Period04 are 9.83, 0.1045, 0.1116 and 0.109 days see Table 6.2 for a comparison of the frequencies identified from STRESS data with that of those identified by Lenz et al. (2010)). These periods were obtained by using data from all three epochs. When 20 days data from only one year is used, even though the dominant period is the same, the other prominent periods shift to 9.707, 0.1115, 0.1046 and 2.85 days. This is to give an example of using data from all epochs and even when the data has large gaps, it can help in refining the period to a greater precision. Fig. 6.7 explains this property in detail. The left plot is the power spectrum using data from all three epochs and the right is constructed using only 20 days of data. The plot on the right is clearly more noisier
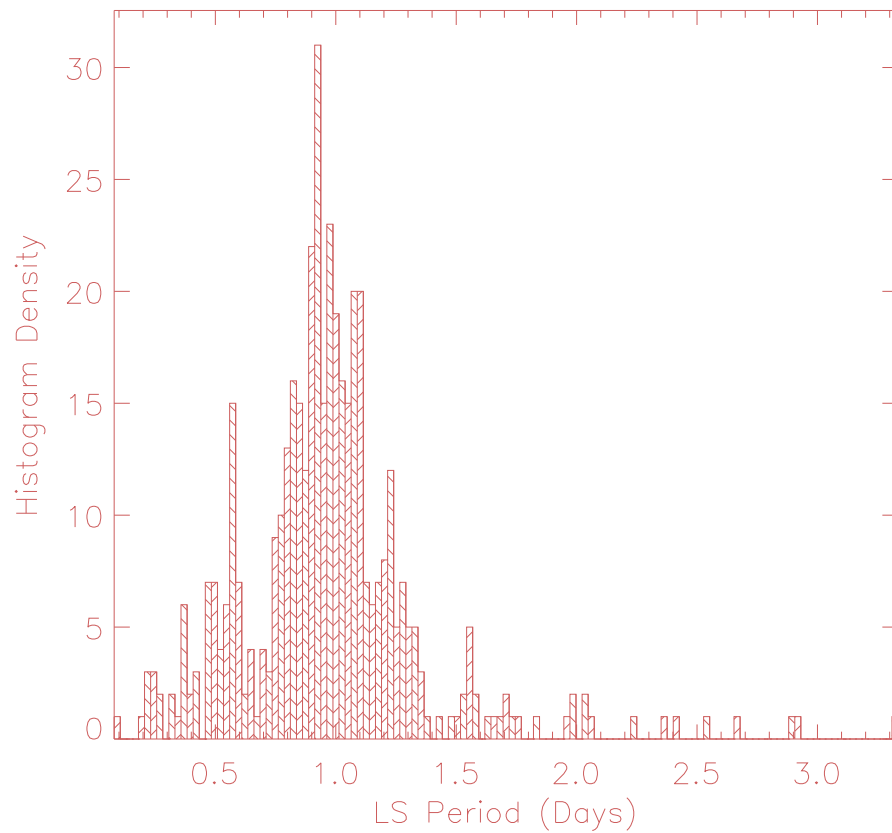
Figure 6.5: The period distribution of the sample set of light curves presented in appendix section. This sample is selected based on a visual inspection and no other criteria, and hence the drop in the period between 0.5 days and one day is not significant.
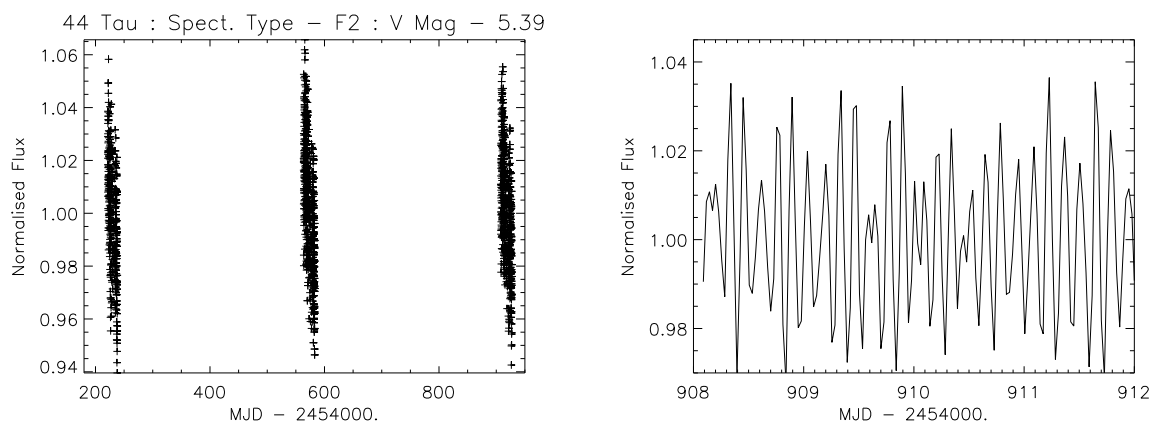
Figure 6.6: The above is the STRESS light curve of the 44 Tau system. The plot on the left is an illustration of the data availability for more than three epochs as of now. This aids in more precise period determination as well as characterising period changes in an object. The right plot is a zoomed in version displaying clearly its symphony of frequencies.
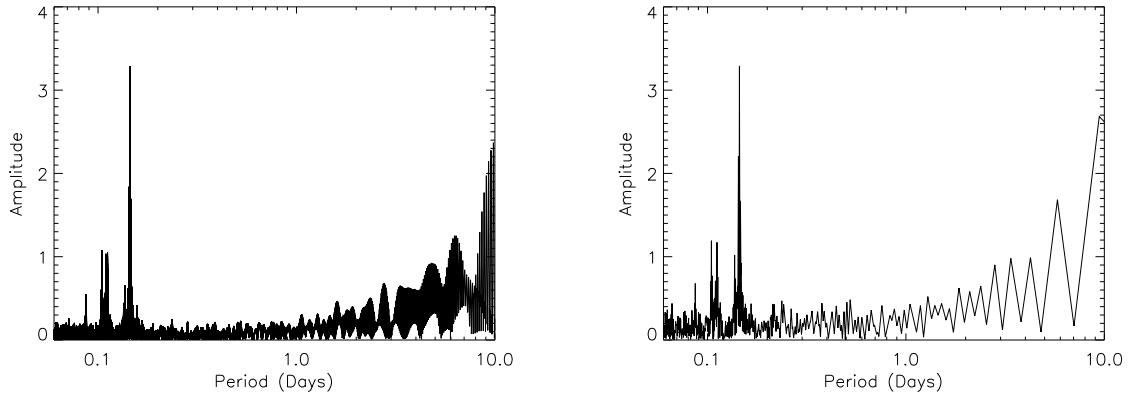
Figure 6.7: Left plot is the power spectrum for the whole data set spanning three epochs. This spectrum exhibits a dominant period at 0.144 days and four other prominent periods at 9.83, 0.1044,0.1116 and 0.109 days. The plot in the right is the power spectrum of only one epoch, containing an individual 20 day data. Even though, the dominant period remains the same, the other periods shift slightly to 9.707, 0.1115, 0.1046 and 2.85 days. Thus, the precision in the identified frequencies increase as we include more data, as expected in any observation.
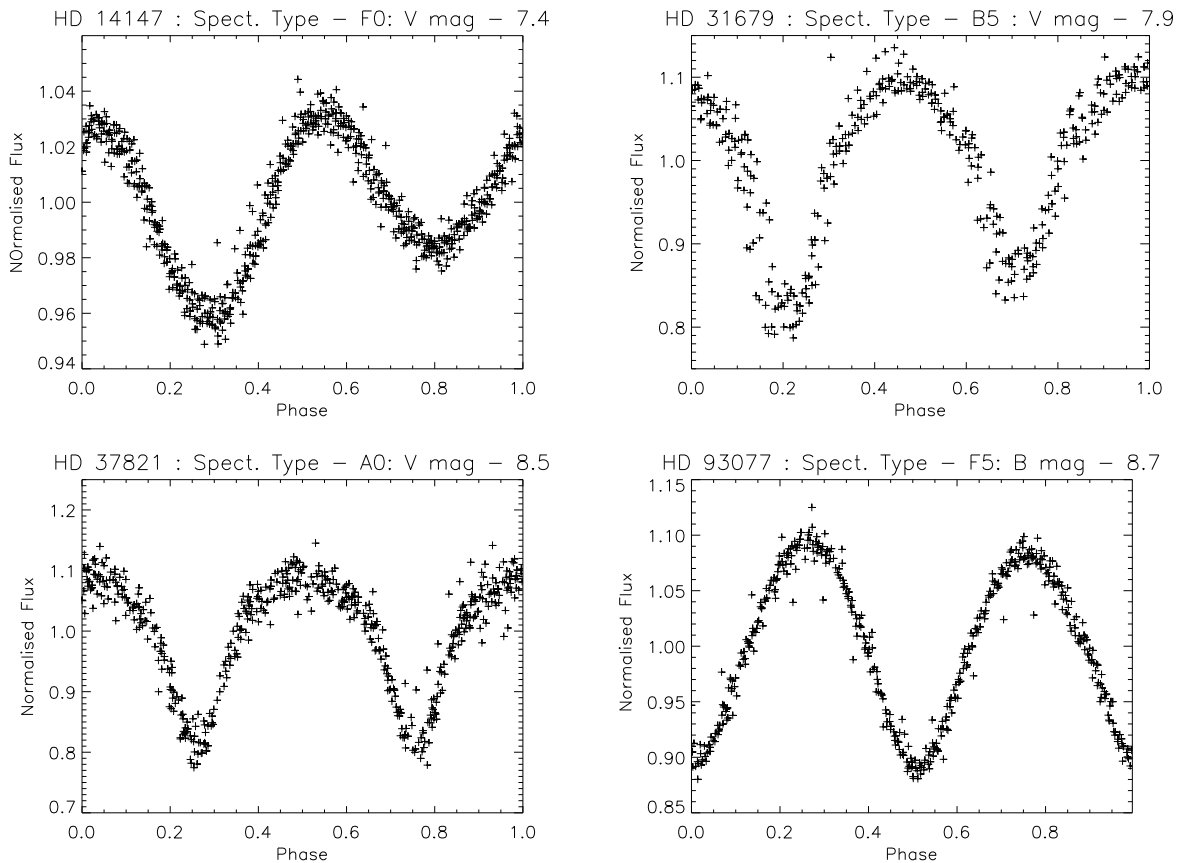
Figure 6.8: STRESS sample light curves : The above are plots of known variable stars of different types of variability folded on the period found from the STRESS light curve. Again we are stressing that this is only a sample of our data and even these data needs to be further analysed to realise the full potential of the science that can be derived from them.

Lenz et al. Freq [cd^{-1}]	STRESS Freq.[cd^{-1}]
6.8980 (*f1*)	6.9009
9.5611 (*f6*)	9.5614
8.9606 (*f5*)	8.9577
9.1174 (*f3*)	9.1175
7.3031 (*f7*)	7.3000
11.5196 (*f4*)	11.5168
6.7955 (*f8*)	6.7922
7.0060 (*f2*)	7.0057
9.5828 (*f9*)	9.5745
6.3390 (*f10*)	6.3476

Table 6.2: Comparison of the observed frequencies in 44 Tau between Lenz et al. and STRESS. However, we also identify a few other strong periods in the data not given by Lenz et al. (2010) outside their range of analysis.

than the left plot. However, when dealing with long data with big gaps, the problem of aliasing and spectral dissipation has to be taken in to account. In the left plot, we find that, specially the peak at 0.111 days is split whereas, on the right, it is a single clear peak. These issues are both an advantage as well a disadvantage and have to be dealt with when we have two sets of data of different lengths.

HD 14147: (= AD Ari) A F0 δ Scuti variable of V mag of 7.43 - Top left light curve of Fig. 6.8. Even-though Zhou (2002), identifies two different periods, Rodríguez, López-Gonzalez & López de Coca (2000) find only one period at 0.2699d. Our analysis, using Period04, results in two distinct but harmonic peaks at 0.269d and 0.54d. The phase folded plot (top right of Fig. 6.7) at 0.54d period made us suspect that it might be an eclipsing binary star. However, a photometric study by Handler & Shobbrook (2002) concludes that this is an ellipsoidal variable rather a δ Scuti or a classical binary system. They also find inconclusive evidence that there might a δ Scuti period of about 0.06d, which however cannot be confirmed using STRESS data due to our time sampling.

HD 31679: (= V1061 Tau) A recently discovered (Terrell, Williams & Kaiser, 1995) B5, eclipsing binary of the β Lyrae type with a $V = 7.95$ (Fig. 6.8 - top right plot). They found it to be a system with a mass ratio of 2.407 and also speculate on the probable presence of a third light in the light curve, which is not confirmed yet. STRESS light curve exhibits a period
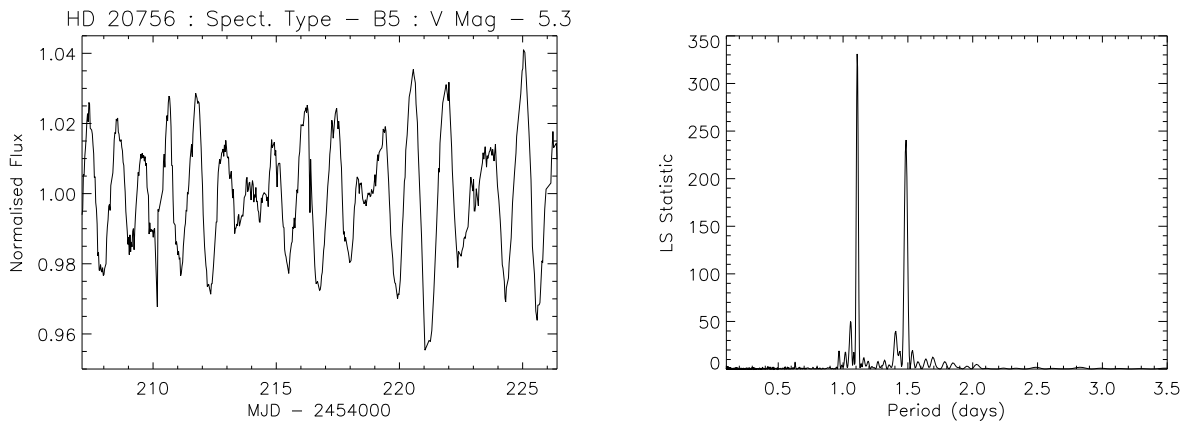
Figure 6.9: HD 20756 exhibits two prominent period in the STRESS light curve. The left plot is a light curve for one epoch of our data and the right is the LS periodogram. The data exhibits two clear periods at 1.105d and 1.492d.

of 1.379d and resembles an Algol type light curve.

HD 37821: (= V1238 Tau) An A0 eclipsing binary of the W Uma type of a $V = 8.5$. The All Sky Automated Survey Catalogue of Variable Stars (ACVS) defines a period of 1.12173d. From the STRESS light curve (bottom left of Fig. 6.8), we find a period of 1.11 days. However, the difference in period between GCVS and STRESS is not significant given the fact that the time resolution of our data is much larger than the difference in period observed.

HD 93077: (= EX Leo) A F6 eclipsing binary star of the W Uma Type with a known period of 0.40860d with a B mag of 8.71 at the bottom right of Fig. 6.8. Eker et al. (2009), used this star as a sample in their re-calibration of the Period - Luminosity Colour relationship of the W UMa type Binaries. It was also studied by the N2K consortium using the KPNO low resolution spectra in order to identify targets for the Keck Planet search program (Robinson et al., 2007).

Below, we present a sample of new variables (Fig. 6.9 and 6.10) discovered in our data using our variability analysis pipeline. This analysis algorithm and complete catalogue will be discussed in a forthcoming paper in this series. Hence we would like to stress the fact that these results are presented to just give a flavour to the reader and needs further investigation.
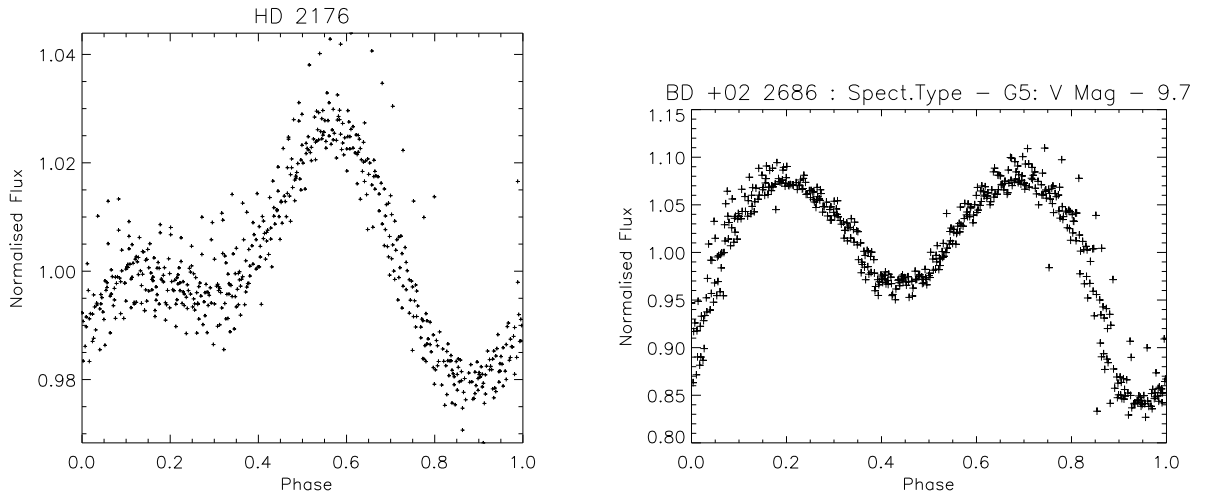
Figure 6.10: Two new variables with no previous literature on its variability. The left is a plot of HD 2176, A K0 star with $V = 9.2$. We can observe non-symmetrical pulsations which are similar to those observed in *KEPLER* ID 8505670 (Reinhold et al., 2011), indicating could be caused by star spots. The right is a plot of BD+02 2686, a faint eclipsing binary with no known literature, having a period of 0.77d.

HD 20756: (= 61 Tau) A B5 star of V mag 5.3 in a multiple system as per Simbad's classification and as a three body system by Eggleton & Tokovinin (2008). GCVS gives a binary period of 2.203 days. This object was studied by Hoffleit & Jaschek (1991) and suspected of being a double or triple system. From our light curve, Fig. 6.9, we find two prominent periods at 1.105d and 1.492d. The period at 1.105 days could be an alias of the GCVS period of 2.203 days. However, the period at 1.492 bears no such resemblance. Further spectroscopic analysis can identify blends within the system and will help identify the components of the system as well as verify the period observed. Also a high resolution photometry can also help in refining the period of the system.

HD 2176: A K0 star of $V = 9.2$ and a high proper motion star according to Simbad. The STRESS data finds a period of 1.17 days, with an unusual light curve shape (Fig. 6.10 - left). This light curve looks similar to *KEPLER* object - 8505670, with a KEP-mag of 15.6 (Reinhold et al., 2011). Fig. 6.11 is the light curve of *KEPLER* ID:8505670, showing very similar shape. The light curve shape in this object is known to have caused by star spots and hence we believe that HD 2176 and other similar light curves given in appendix (HD 4649, BD+02 4751, etc.) could be due to the presence of star spots. This star was observed by the N2K consortium in
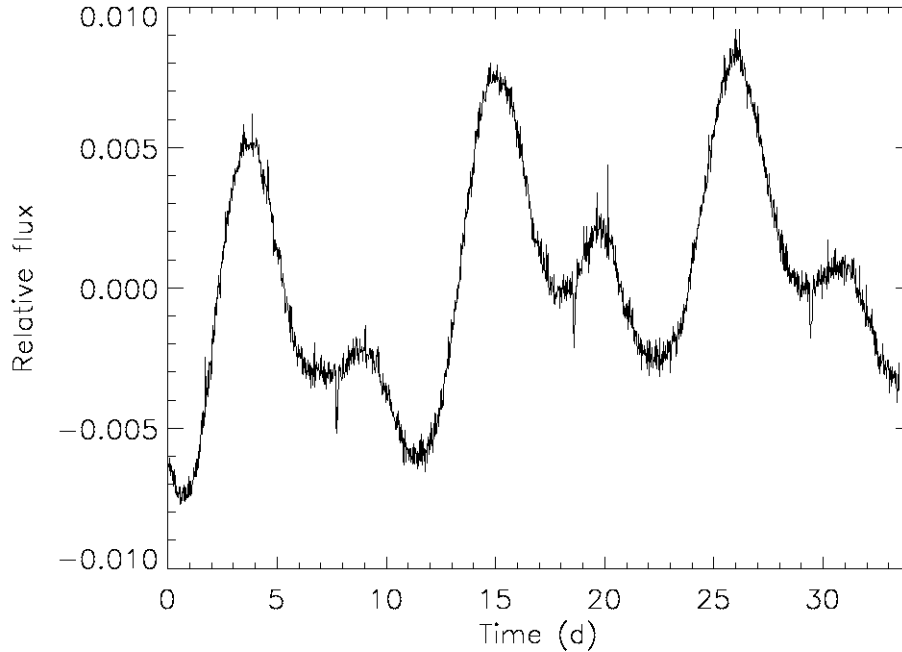
Figure 6.11: This the *KEPLER* light curve of 8505670, showing features similar to HD 2176, STRESS light curve. Since 8505670 is understood to contain star spots (Reinhold et al., 2011), implied by its light curve shape, we assume that HD 2176 also contains stellar spots.

their search for planetary hosts based in Keck observatory but did not make it to their observation campaign, which would be in agreement to our inference, of this being an active star. This hypothesis can be confirmed by studying the line-bisector variations in a spectroscopic observation. Also the presence of Ca II H and K emission lines will indicate a high level of chromospheric activity indicative of spots.

BD +02 2686: Known as a G5 star of $V = 9.75$ with no other literature. The right plot of Fig. 6.10 clearly shows the presence of both primary and secondary eclipses with a primary period of 0.774d.

Given the continuous availability of data, an attempt was made to study Solar-like oscillations in giants. The first step is to identify the giants in our data. Given the fact that the basic parameters known about the targets are extremely limited, in many cases, only their R magnitude and position, this was an almost impossible analysis to do en masse. However, a basic analysis was done to probe whether it is possible to observe solar-like oscillations given
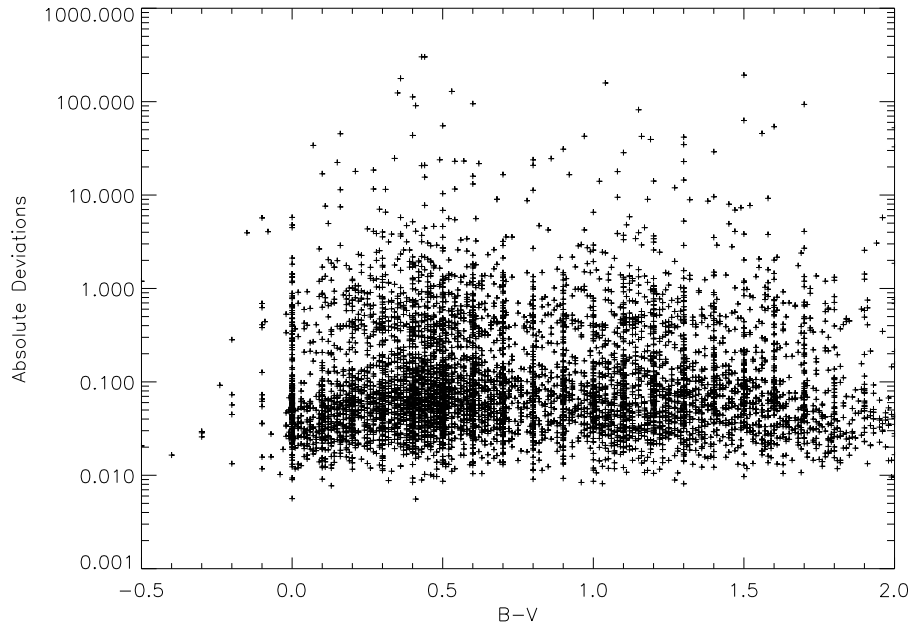
Figure 6.12: Variation of absolute deviations of the light curves from the STRESS data as a function of their $B - V$ values. This corresponds to the analysis by Basri et al. (2010) for the observation of Solar-like oscillation in our data. However, we find that our level of variability if far below than the variability observed in Sun-like stars.

the time sampling of STRESS data. We used 12633 objects from a months' data, for which the B, V magnitudes were available from Simbad database. We calculated the absolute deviations of the light curves for these objects using a method corresponding to the one described by Basri et al. (2010), so that we can compare our level of variations with their plot of T_{eff} against Range (Fig. 3). STRESS plot of the same is given in Fig. 6.12 with $B - V$ in the X axis and Absolute deviations in Y axis. The absolute deviations can be compared to the range observed by Basri et al. (2010) and the $B - V$ values can be corresponded to the effective temperatures of stars. We can understand from the comparison that the data obtained by *STEREO* are unable to reach the level of variability needed to detect Solar-like oscillations in the stars in general. However, if we are to isolate the giants in the data and study them individually, we might be able to re-move the systematics to reach that level of variations and detect the presence of these oscillations.

6.6 Final Comments

This chapter represents a very important part of this thesis, in the sense that it illustrates the enormous potential of this data. It is to be stressed again that the sample presented here or in appendix is by now means exhaustive but just a representative. According to current statistics, we expect the data to yield at least a few hundred new variable objects of different types. Moreover, the analysis at this stage is also very basic and crude and can be refined in many ways to yield better results. The refinement can include the use of automated algorithms for systematic noise reduction as well as identification of the different variables along with follow-up of interesting candidates. Future work will also involve detailed studies of different classes of objects as well as of individual stars. For this purpose, the data in its current form, will be published as a catalogue in future for the use of scientific community and to extract the maximum information from them.

Chapter 7

Exoplanetary Transits with *STEREO*: Analysis and Results

7.1 Introduction

One of the major goals of this project is to identify and analyse transiting exoplanetary candidates. As discussed in Chapter 6, the data is a prime sample consisting of very interesting objects of many different genre of variability. From Chapter 2, we can also understand that the spacecraft is positioned looking at a very interesting part of the sky not generally covered by the other transit search projects, both spatially and temporally. Fig. 1.4 shows that the portion of sky covered by *STEREO* is fairly empty of transiting planets. This gives STRESS an added advantage to identify potential candidates, generally missed by other programmes. The data specification and the noise characteristics render the possibility of identifying Jupiter-sized planets around a Solar-like star in a two different period regime - with a maximum period of 6 days with data from only the HI-1A instrument and maximum period of around 50 days, when the data from all the HI instruments are combined. As of 1 August 2011, 28% of all exoplanet discovered so far have an orbital period less than 6 days and 42% of them fall within an orbital period of 50 days. In the case of transiting planets alone, this is much higher, since transit searches are more sensitive to shorter period planets. The statistics in this case is that, nearly 83% of them fall under a 6 day period and almost 97% of them have under a 50 day period.

At present, there are only 10 stars brighter than $V = 9$, which hosts a transiting planet. Presently STRESS has the capabilities to identify transit signals of depth above and around 0.7%, corresponding to a planetary radius of $0.8 R_{Jup}$ around a star of Solar radius. Among

111

	All Planets	Transiting Planets
P< 6 days	28%	83%
P < 50 days	42%	97%

Table 7.1: Orbital period statistics of the currently confirmed planets, that fall within the 6 day period and within 50 day period regime. Courtesy: http://exoplane.eu

these 10 stars, there are only four of them, that have depths greater than 0.7%, that is, only these four objects would have been identified in our data using our current pipeline, if they were present in our data. Ironically, these are four most studied transiting planets amassing information about the physical properties of the planet, its atmospheric profile, its interaction with the host star, etc. In order to estimate the probable number of bright transiting planets yet to be discovered, we identify the total number of bright individual stellar objects from Simbad database as 100,917, which are classified as stars not belonging to any type of multiplicity and fainter than $V = 9.0$. According to von Braun et al. (2005), 0.7% of isolated stars have a planet within an orbital radius of $a < 0.05AU$. Hence, we can estimate that there must be atleast 706 transiting exoplanets in this magnitude range. To further illustrate this point with respect to STRESS data, the following calculation was done to arrive at the probable number of transiting planets in our data as a first estimate. This estimate is only of the first order and contains assumptions at every step of the calculation. However, I believe it is correct to the first order and also re-iterates the point that this data contains transiting exoplanetary candidates.

We consider the 82,863 light curves which were left over in step 2 of the pipeline described in section 7.4. We use the Simbad database to gather information about the B and V magnitudes, their spectral types, parallax and proper motion information, by defining a circular area of radius 1 minute around the STRESS target object. Since our input catalogue is NOMAD, SIMBAD may not contain information for many of the objects studied by STRESS. This script returned 39100 objects from Simbad database, out of which 19136 were catalogued to be just stars and spectral data was available for only 16359 stars. There were 13067 of these objects which were classified as F, G and K spectral types. Kaltnegger et al. (2010) identifies 57% of F, G and K stars to be binary, which gives 7448 stars in our list as individual systems. Cumming et al. (2008) estimates that 12 - 19% of stars to contain a gas giant planet within 20 AU. They also give "the fraction of FGK stars with a planet in the mass range $0.3 - 10M_{Jup}$ and period

range 2 - 2000 days is 10.5%". Using these two values, we find that the number of planetary systems in our data to be $0.105 * 7448 = 782$ stars may contain a planetary system out of which $0.17 * 782 = 132$ systems may contain a gas giant within 20 AU. We can also use the probability values given by Cumming et al. (2008) and find that the probable number of Jupiter-like systems around a Sun-like star as 6.9%*7448 = 513 systems. Using this final value along with the Geometric tansit probability (Borucki & Summers, 1984), for STRESS, the probable number of observable Jupiter-like transits around a Sun-like star as 0.077%*513 = 0.39 planets. However, there are two other probabilities given in literature by Marcy et al. (2005); von Braun et al. (2005). The former states that 1.2% of FGK stars have a Hot Jupiter within the orbital radius of less than 0.1AU and this estimates gives, in the case of STRESS, 0.0126*7448 = 89 planetary systems. Based on the study by von Braun et al. (2005), who states that 0.7% of isolated stars have a planet within $0.05AU = 0.007*7448 = 52$ planets and out of which between 10% to 20% of these Hot Jupiters have favorable orientation, which translates in our case to around 8 Hot Jupiters in STRESS data. Hence, we can conclude from the available statistics, considering the various assumptions that went into this back-of-the-envelope calculation, we can conclude that up to 10 planets can be detected in STRESS data.

The work presented here is done in collaboration with Gemma Whittaker, of the University of Birmingham, UK. The radial velocity follow-up described later in the chapter, is done with the help of Dr.Chip Kobulnicky, University of Wyoming, US.

7.2 Box-fitting Least Squares (BLS) Algorithm

When a planet crosses in front of its parent star, it produces a corresponding distinct box-like signal in the light curve. The LS periodogram and other methods, which searches for a signal by fitting a sinusoidal, fails in this case. Hence the need for specialised methods like the matched-filter approach (Jenkins, Doyle & Culler, 1996), Deeg's approach (Doyle et al., 2000), Bayesian approach (Defaÿ, Deleuil & Barge, 2001) and Box-fitting Least Squares (BLS; Kovács, Zucker & Mazeh, 2002). Tingley (2003a), did a rigorous comparison of these different types of methodology using a set of simulated light curves. Tingley (2003a) also compared two other methods, namely, correlation method and Aigrain's approach (Aigrain & Favata, 2002). Tingley (2003a)

initially concluded that the matched filtered gave the best performance in terms of identification of candidates but cautioned that BLS searched for not only the best period but also tries to fit the best phase simultaneously. However, on further analysis (Tingley, 2003b), they reported that all the methods performed similarly. He arrived at this conclusion, based on the fact that the ability of any algorithm depends on the 'transit signal energy' $E_S = N_{in} \times d^2$, where N_{in} is the number of observations within the transit and d is the depth of the transit compared to the scatter in the light curve. The *CoRoT* consortium uses different types of methods individually as well as a combination of these methods for their detection. However, Barge et al. (2006) reported that BLS seem to perform much better than other algorithms, but the clear demarcation of identifying the false positives depend on the detrending method used. In contrast, *KEPLER* uses a wavelet-based, adaptive matched filter to identify its candidates (Jenkins et al., 2010b). BLS has also been widely used in successful identification of planets (HAT - Bakos et al., 2007; SuperWASP - Lister & Street, 2007; OGLE - Snellen et al., 2007. In the analysis of STRESS light curves, we use the readily available BLS method for identification of planetary candidates. This algorithm is summarised below.

When a planet starts to transit its host star, the light intensity from the star drops suddenly and when the planet is completely in front of the star, the light signal remains constant till the planet reaches the outer rim of the star, when the intensity rises back to the actual stellar value. The phenomenon described above, in the ideal case, describes a box shaped signal in the light curve, repeated according to the period of the planet around its host. The Box-fitting Least Squares algorithm (BLS; Kovács, Zucker & Mazeh, 2002), utilises this special form of the signal and the fact that the duration of the transit is relatively small compared to the orbital period and also the shallowness of the transit. It assumes that the signal has two discrete values - a high and a low corresponding to when the planet is out of transit and in transit respectively. It also ignores the effect of ingress/egress and limb darkening effects. The algorithm folds the light curves at the various trial periods and fits a step function to it. It then calculates the residuals between the fitted function and the data and then performs a least squares minimisation of these residuals for the different periods. Kovács, Zucker & Mazeh (2002) then define their frequency

spectrum as a function of Signal Residue, SR, for the time series at any given period, given by,

$$SR = MAX \left\{ \left[\frac{S^2 (i_1, i_2)}{r (i_1, i_2) [1 - r (i_1, i_2)]} \right]^{1/2} \right\} \tag{7.1}$$

where, $S^2(i_1, i_2)$ is the contribution from the signal and $r(i_1, i_2)$ represents the percentage contribution from the transit level with respect to the whole light curve. The difference between the indices i_1 and i_2 conforms to the suspected fractional transit length. For this algorithm to be computationally feasible for large data sets (as in the case of STRESS), the folded time series is divided into bins before evaluating SR. An exact solution for the above equation can be arrived at, as long as the bin size is lesser than the transit duration, thus compromising between the efficiency of the algorithm against the computational constraints.

Kovács, Zucker & Mazeh (2002) define a quantity, the Signal Detection Efficiency, SDE, which is comparable to the false alarm probability or the significance limit in other algorithms.

$$SDE = \frac{SR_{peak} - \langle SR \rangle}{sd (SR)} \tag{7.2}$$

where SR_{peak} is the maximum SR, $\langle SR \rangle$ is the average of the spectrum and $sd(SR)$ is the standard deviation of SR over the frequencies tested. This value of SDE depends on the total data length as well as the length and position of the frequencies searched. Intuitively, it can be inferred that a large number of bins results in a highly stable SDE even in the presence of noise, i.e., if the transit is to cover at least 2-3 bins, there is a higher probability of recovering the signal and hence a higher SDE value. It is very essential to use the correct number of bins in the computation, as this decides the efficiency of the algorithm to detect box-like signals. The size of the bins are important in the estimation of the transit duration. In simple terms, the transit duration is the product of the time size of each bin and the number of bins in which the signal is in the 'L' state or the 'Low state', i.e., when the planet is transiting the host. Hence, the accuracy of this estimate largely depends on the size of our bins and hence has to be kept lesser than the transit duration. Thus, the number of bins along with a minimum number of data points is essential for BLS to perform efficiently and this is taken into account when using this method in STRESS data.

7.3 False Positive Detection

One of the most important issues to be dealt with in any exoplanet transit search method is the false positives present in the data. These are light curves which mimic a planet-like signal so effectively that they are flagged as planetary events. There are three main identifiable causes of these false detections. One of them is the systematic artifacts present in the data, like the instrumental defects or as in the case of STRESS, it can be due to the presence of solar phenomenon, which can cause inconsistency in a light curve, produce outliers, thus resulting in a planet-like signature. Another cause is the effect of statistical noise, specially in the case of low amplitude signals. Finally, it will be the presence of physical systems, nearby or in line-of-sight of the target, which gives rise to a false signal. Brown (2003) describes in detail the third case for a ground based search, identifies 12 false-positive probabilities and explains in detail five of the major cases, which affects the transit detection rate (cf. Table 1 of Brown, 2003). The initial and the most important of these classes is the true planetary signal caused by the presence of a giant planet around the main-sequence star. To summarise the major false positives:

(a) Grazing eclipses - a binary system of two main-sequence stars in an inclined orbital plane to our line-of-sight. These represent the highest false detection rate according to this study.

(b) Eclipses caused by a binary system with a giant star component.

(c) Binary systems whose light is diluted by the presence of a foreground (or background) star, which is not physically related to the system but present in the line-of-sight of the observer.

(d) Triple systems, which include a binary and a bright star, physically related to the system, can cause in dilution of stellar light mimicking the presence of a planet.

One of the more generally used method to eliminate some of these false positives is to pre-select the target from the input catalogue, that is, removing the dwarfs and giant population. Ironicaly, the recent developments with regard to planets around giants (Howell et al., 2012; Santerne et al., 2011) induce us to believe that it is essential that we search around giant population as well, as against ruling out the giant and the dwarf population completely. However, it is imperative to understand that the probability of finding such a planet in a bright sample using transit is very slim, since we need to have a highly sensitive instrument and also that

such a sample might be biased towards more evolved stars. In the case of STRESS, there are two conflicting parameters which influence the transit search. One is the continuous coverage of large number of bright targets, which increases the probability of identifying a transit and the other is the large integration time and the lower sensitivity of the data in the presence of excessive noise. The compromise between these two state restricts our ability to search for Jupiter and smaller sized planets around any population of stars. However, for the sake of completion, we can use the method described by Gould & Morgan (2003), using reduced proper motion of stars to differentiate between the population of giants, dwarfs and main-sequence stars in our catalogue. Since our catalogue is derived from the NOMAD catalogue which is a compilation of various catalogues such as Tycho, 2MASS, etc (described in section 4.2), we can apply the statistics derived for its population by Gould & Morgan (2003), as a first estimate. Accordingly, Gould & Morgan (2003) rejects nearly 70% of the transit targets in a magnitude limited sample with $V < 11$. Using equation 4 of Gould & Morgan (2003), with a limiting magnitude of $V = 11$ for STRESS, it would be possible to distinguish between giants and dwarfs for $V - I > 0.05$, which is at the very edge of the distribution for the Tycho-2/2Mass catalogue.

For wide-field transit surveys like $COROT$, HAT and SuperWASP, the number of false detections are substantial corresponding to the total number of observed stars (Almenara et al., 2009; Bakos et al., 2004; Faedi et al., 2011). In the case of $COROT$, which has a fairly large aperture ($320'' \times 320''$; Deeg et al., 2009), 83% of the initial candidates were attributed to false detections and only 12% of these were so far confirmed as planets. However, in the case of $KEPLER$, which has a much smaller effective aperture ($2''$) Morton & Johnson (2011), gives a false positive probability of only 2%. This probability seem to have a strong relation with the depth of the survey and the Galactic latitude of the target. Hence, in the case of STRESS, with the smallest aperture at $351.5''$ and a PSF of $222''$, can be compared to have a false positive rate similar to that of $COROT$, i.e., there is a 27% probability that the identified candidate can be false positive, using the results from Morton & Johnson (2011).

In order to identify these false positives in our analysis, a number of "false alarm" tests are used, some of which are presented below:

7.3.1   The Exoplanet Diagnostic Tool

Tingley & Sacket (2005) gave a diagnostic tool to weed out the false positives in a large sample of probable planetary candidates. This statistic, η^* is equivalent to the ratio of the observed transit duration, D_{obs} and the expected transit duration, D:

$$\eta \equiv \frac{D_{obs}}{D} \tag{7.3}$$

The expected transit duration, D, is derived based on the depth, δ and period of the transit P, the orbital inclination of the system and the stellar radius. Hence, this statistics gives the probability that for the parameters obtained from BLS, the data can contain a true planetary signal as opposed to it being a false positive. The observed transit duration is calculated by the BLS. The calculation of D, differs for the space-based and ground-based projects and this results in two different Exoplanet diagnostic formulations. There are two unknown parameters in the equation, namely, the stellar radius R_* and the planetary radius R_P and during the calculation one of them is taken to be a constant based on the sensitivity of the instrument. For a ground-based survey, the photometric precision is limited and hence, the radius of the planet can vary by a factor of two and the stellar radius will vayry by a factor of three. Thus, it is best to take the planetary radius to be a constant in this formulation. Whereas in the case of a space-based transit search, the instruments potentially have a higher photometric precision leading to it being sensitive to a large range of planetary radii. This results in the stellar radius being assumed to be a constant, as in the calculation below:

$$\eta^* = \frac{D_{obs}}{2Z\left(1 + \sqrt{\delta/1.3}\right)} \left(\frac{2\pi GM_\odot}{P}\right)^{1/3} R_*^{-7/12} R_\odot^{-5/12}. \tag{7.4}$$

In the above equation, BLS calculates the transit duration D_{obs}, the transit depth, δ and P the orbital period of the candidate. Z represents the orbital inclination of the system and the separation between the planet and star at the time of transit. The maximum value of Z is equal to one for a central transit, i.e., a planet transiting the star at its equator to our line-of-sight, corresponding to an orbital inclination of $90°$. Since this value is not known at this point in analysis, this is taken to be $Z = 1$, as in an ideal case for a transiting planet and the radius of the star is assumed to have a Solar radius for simplicity. As a result in the highly probable case of the transit caused by the presence of a planet, this results in the statistic being less than or equal

to one and for a false detection, a much higher statistic. The value of statistic is modulated by the two values, the transit duration and the transit depth. In the case of a binary system, both the theory and the observation agrees that the transits are usually longer and always deeper than for a planetary scenario. Thus, the calculated η^*, in this case will be much larger. The same is true for the case of a blended eclipse, where even though δ is reduced, D_{obs} remains large and results in $\eta^* > 1$. When the binary system has two eclipses of almost equal depths, the period found by BLS should be half of the true period, resulting in an increased statistic. Alternatively, visual inspection can rule out false positives in the form of binary systems with unequal eclipse depths. The η^* statistic has a probability of deviation in the case of a grazing eclipsing binary, where both the depth and duration of the transit can exactly mimic that of a planetary system, thus leaving behind some false positives in the candidate sample. This tool has been successfully used in many exoplanet search programs like $COROT$ (Almenara et al., 2007), SuperWASP (Wilson, 2007) and LCOGT follow up of SuperWASP candidates (Lister et al., 2007).

7.3.2 Stellar Classification and Spectral Type

As seen from Section 7.3.1, this statistic is also dependent on the stellar radius. This can be easily calculated, if stellar distance and temperature are known. Public astronomical databases like Simbad, ASAS, GCVS contain these information for large sample of stars. The database can also give information as to the spectral type of the target and this also aids in ruling out false positives. For e.g., at 1% planet depth level, it is impossible to detect planetary signal around massive stars like O, B and early-A type as they are too luminous. However, a signal around a late-A star might fall at the very edge of the detection limit in this case. Another constraint is that, they are short lived and hence theoretically unable to have a disk life time that allows the growth of a planetary embryo. The earliest known planetary host is an A5 star, HR8799, which harbours three known planets (Fig. 1.9). Thus, most of the current planet search programs are centred around F, G, K and M stars similar to the Sun in age and composition. Since these databases are formed from many other detailed and survey studies, they also contain information about probable variability within and nearby a target system, thus aiding in the

reduction of false positives.

7.3.3   Visual Inspection Tests

Last but not the least, all the light curves selected by any or all of the above methods are given a visual inspection. This rules out the cheaters, which mimic the planetary signal, in spite of all the tests mentioned previously. Also, a visual inspects helps in confirming that the light curves display plausible transit signals with a characteristic box-shaped signals as well. Based on the model devised for SuperWASP light curves by Lister et al. (2007), the following methodology is adopted for STRESS light curves:

1. Transit period > 0.5 days and transit depth < 4.5 %. This value of $\delta < 4.5\%$ was chosen from Fig. 3 of Brown (2003), which gives the expected distribution of δ for transits by planets in within a magnitude limit of $R = 12$. This can also be inferred from the distribution of known transiting planets.

2. Equal depth in all transits.

3. Observe at least two transits.

4. Realistic transit duration, ranging from a few hours up to half a day.

5. Box-shaped or 'U' shaped transits rather than 'V' shaped, indicating a grazing eclipse. This is very relevant to STRESS, due to its large integration time, which can smooth out box-shaped signals to give a more 'V' shaped signal. Hence, precaution is taken when ruling out candidates just with their transit signal.

6. Fairly stable out-of-transit noise level.

7. Evenly distributed data points in the phase-folded light curves, i.e., no significant gaps or concentration in the data. The continuous observation of *STEREO* helps in this case, as opposed to ground-based projects, which can have problems if the orbital period is around one day.

7.4   Exoplanetary Candidate Identification Pipeline

In this section, the basic pipeline used for identifying planetary candidates, developed by Gemma Whittaker, taking into account the various methods described in Section 7.3, is explained in detail. It also uses a similar methodology described in Section 6.4 but with modifications to suit the search for short-period and shallow events. The same STRESS data files saved in the IDL format are used in this analysis as well. A summary of this pipeline is given in Fig. 7.2. An important characteristics of the data, which we have to consider when searching for transits, is its time sampling. Even though, the data is available continuously for almost 20 days, the sampling rate is quite low. The time integration which results in the final image is also an important factor to consider when searching for exoplanetary transits, in that it can mask the actual transit event in the case of short transits. This 40 minute integration can actually result in the non-identification of an actual transit or the transits having 'V' shapes resulting in them being classified as a variable star. Thus, it is extremely important to account for this effect at least partially if not completely when identifying candidates. In STRESS, we account for this finite integration time by the method described in Pont, Zucker & Queloz (2006) during our analysis.

(a) As seen from Section 7.2, a planetary signal attains a higher significance in the presence of more transit events. Hence, the first step is to concatenate data for a single star from all three years together. In many cases, this results in three epochs of data and in some even four epochs. This resulted in data for a total of 124,565 objects.

(b) Similar to the LS pipeline, only the flux from the smallest aperture is considered for this analysis. The average value is subtracted from the data and we also remove the NAN values which might be present. Use only those light curves which has more than 30% coverage. Thus, even if a star is observed only once in the three years, this will give more than 200 data points. This resulted in only 82,863 light curves for further analysis.

(c) The next step is the detrending procedure in order to remove the long term variation in the data. This is achieved in three steps explained in Fig. 7.1:

(i) The first step is a sigma clipping using the individual light curve standard deviation.

> We use a lower cut-off of 4σ and a upper cut-off of 2σ.

(ii) The next step is the subtraction of a box-car smooth function with a box width corresponding to one day, i.e., 36 data points. As explained in Section 5.4.2, this removes the long term variations of the order of one day effectively, most probably caused by the F-coronal background subtraction, thus still enabling us to search for variabilities around one day.

(iii) The last step is another outlier removing procedure. In this case, we use a upper cut-off of 1σ and a lower cut-off of 4σ. The two sigma-clipping procedures ensures that all the deep eclipses are removed from the signal and hence results in a better identification of planet-like transit events.

(d) We next use the BLS algorithm to search for shallow, box-like transit events in the data. We use the Fortran program from Kovács, Zucker & Mazeh (2002). We ran the algorithm with 300 bins, in the frequency range from 0.1 cd^{-1} to a maximum frequency of 2 cd^{-1}. The frequency range was then divided into 8 ranges with 6800 frequencies in each range, so that the frequency resolution was $3.49 \times 10^{-5} cd^{-1}$. BLS returns the results which has the highest SDE value.

(e) In order to rule out false positives, only those objects with an $SDE > 9$ and which have at least one data point inside the transit (as calculated from the transit duration), are saved and used in further analysis. The cut-off was determined after visual inspection of preliminary light curves. Some statistics at this stage : among the 82,863 light curves which were input into the algorithm, 908 were found to have a candidate transit with an SDE value greater than 9.

(f) The Exoplanet diagnostic tool η^* is then calculated in order to rule out false positives. As explained in Section 7.3.1, any object with a value of $\eta^* << 1$ and a depth, $\delta < 4.5\%$, has a higher probability of being a planetary candidate. The transit duration is also modified based on Kipping (2010), to account for finite integration time of the data. Fig. 7.3 is a plot of the η^* values against the δ values and the lines represent the division into the four categories, as explained below:

(i) Primary Planet Candidates : $\eta^* < 1.0$ & $\delta < 4.5\%$. 222 objects fall into this category.
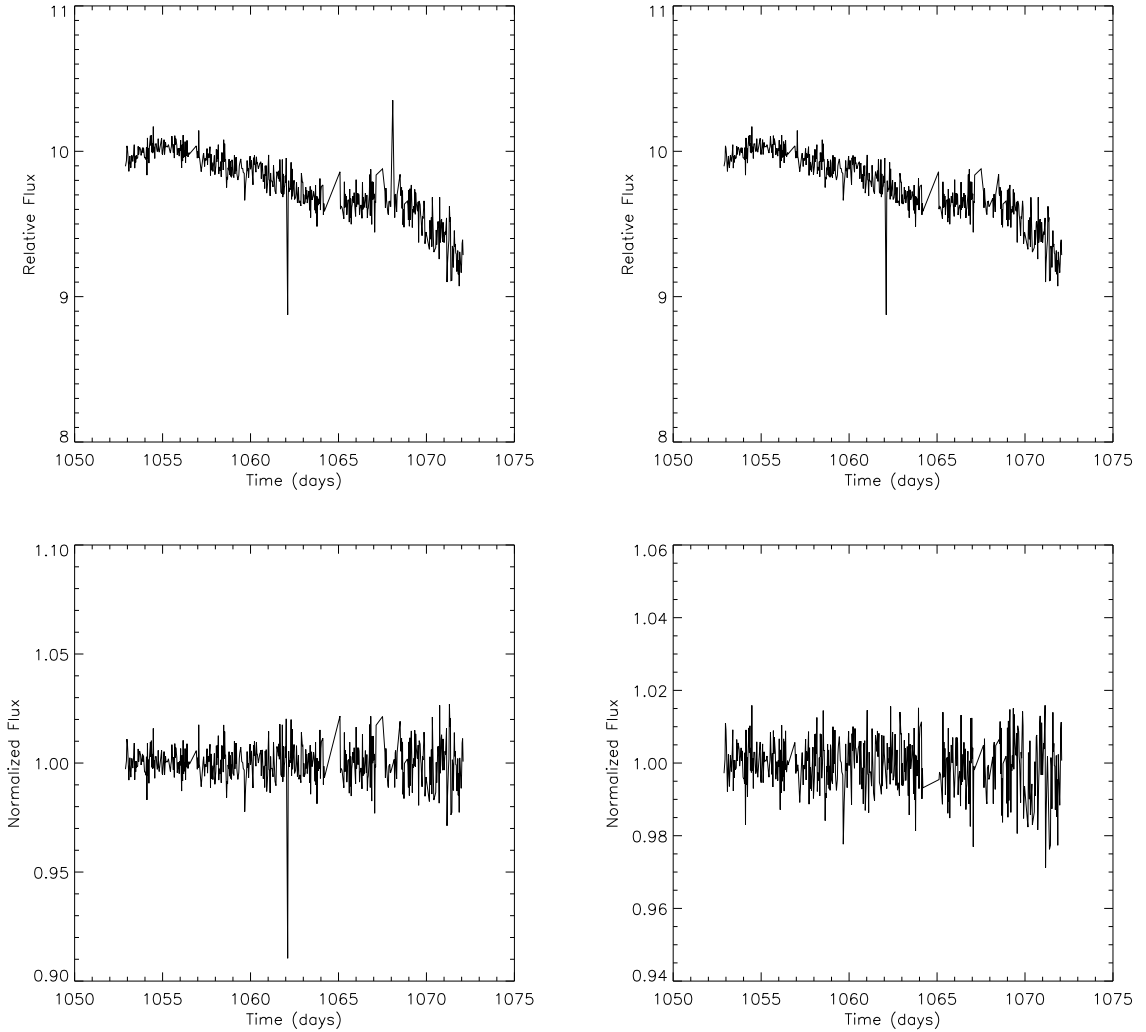
Figure 7.1: The various steps followed in the detrending procedure for the identification of shallow transit-like events using BLS. Top: Left plot is the raw STRESS data while is the right is a plot after the first sigma clipping procedure. Bottom: Left plot represents the data after the subtraction of a one day box-car smoothing function and the right plot is the resultant of the second sigma clipping procedure. The bottom right plot is the data used in BLS searching for transits.

(ii) Secondary Planet Candidates : $\eta^* > 1.0$ & $\delta < 4.5\%$. There are 421 light curves in this group.

(iii) Primary Eclipsing Binary : $\eta^* < 1.0$ & $\delta > 4.5\%$. 209 objects are in this class.

(iv) Secondary Eclipsing Binary : $\eta^* > 1.0$ & $\delta > 4.5\%$. This group of 56 are unlikely to contain a planetary transiting event in the data and most likely to be eclipsing binaries. For completion of the analysis, these are investigated further as well.

We have started a visual inspection of all the light curves based on the groups in the above
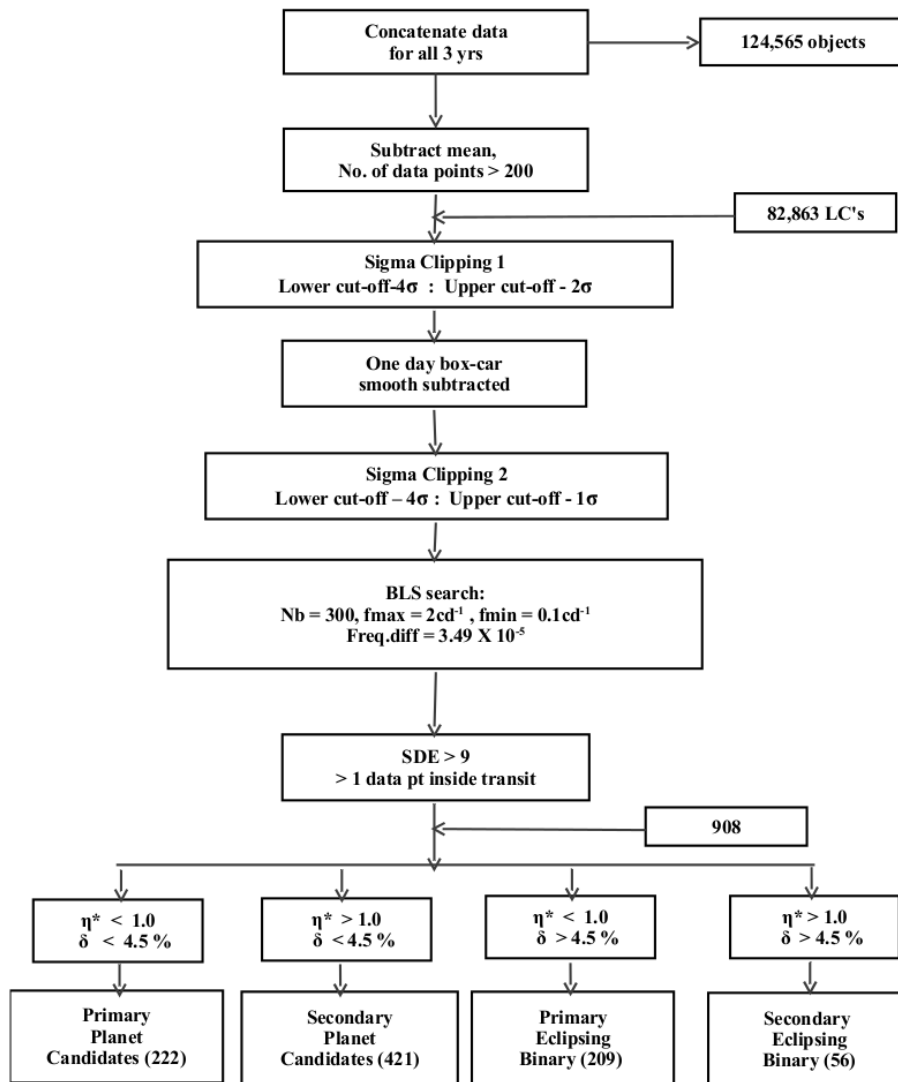
Figure 7.2: This is a schematic of the analysis pipeline developed for the search of transit signal as explained in Section 7.4.

order. We use the Simbad database to cross reference the spectral type of the candidates

and to select objects for follow-up.

Type	η^*	δ	% of total 82,863
Primary Planet Candi.	< 1.0	< 4.5%	0.2
Secondary Planet Candi.	> 1.0	< 4.5%	0.5
Primary EB Candi.	< 1.0	> 4.5%	0.2
Secondary EB Candi	> 1.0	> 4.5%	0.06

Table 7.2: This table summarises the results from the BLS pipeline and our classification into four classes based on the two parameters.

ID	Vmag	Spectral Type	Period (days)	SDE	η^*	δ	T_{dur} (days)
HD 128143	7.74	F5	0.729	15.8	0.11	0.004	0.008
HD 222891	8.07	F8	1.594	30.973	0.045	0.017	0.004
HD 213597	7.81	F0	2.423	30.1	1.03	0.024	0.11
HD 27268[1]	8.70	A0	4.126	40.5	0.96	0.088	0.11
HD 110366[1]	10.16	A0	3.144	15.4	0.88	0.087	0.097

Table 7.3: A sample of stars which are flagged up by the BLS pipeline as potential planetary candidates as well as new and interesting eclipsing binary. The first two objects are identified using our pipeline, the third one is identified as a potential candidate by Wraight et al. (2011), the final two are new eclipsing binaries.
1 - The periods given here are twice the period identified by BLS.

7.5 Preliminary Results

In this section, a flavour of the results obtained from the above BLS pipeline is presented. As mentioned, the final data is classified into four classes, which is summarised in Table 7.2. This is a work in progress and hence holds a huge potential in terms of adding invaluable data to the existing science. In order to illustrate the uniqueness of the data, Fig. 7.4, gives a comparision between the spectral types of the known transiting planet hosts and our candidates. Spectral information is available from Simbad database for 31 known planet hosts and only 137 of STRESS candidates. We present a sample of five objects selected from this pipeline, the various parameters of which are summarised in Table 7.3.

HD 128143: This 7.74 mag star, at an approximate distance of 77.4 parsec is an F5 dwarf, with a known X-ray source (Haakonsen & Rutledge, 2009). However, there is no previous reported optical or X-ray variability. The STRESS BLS pipeline, however, identifies this to be a planetary candidate with a period of $P = 0.72$ days with an SDE value of around 15. A BLS calculated depth of $\delta = 0.0049$ gives a planetary radius of $0.81 R_{Jup}$, for a stellar radius of $1.2 R_{\odot}$. Due to the presence of the X-ray source, two difference scenarios can be considered - the presence of X-ray companion star could be the origin of these eclipses or
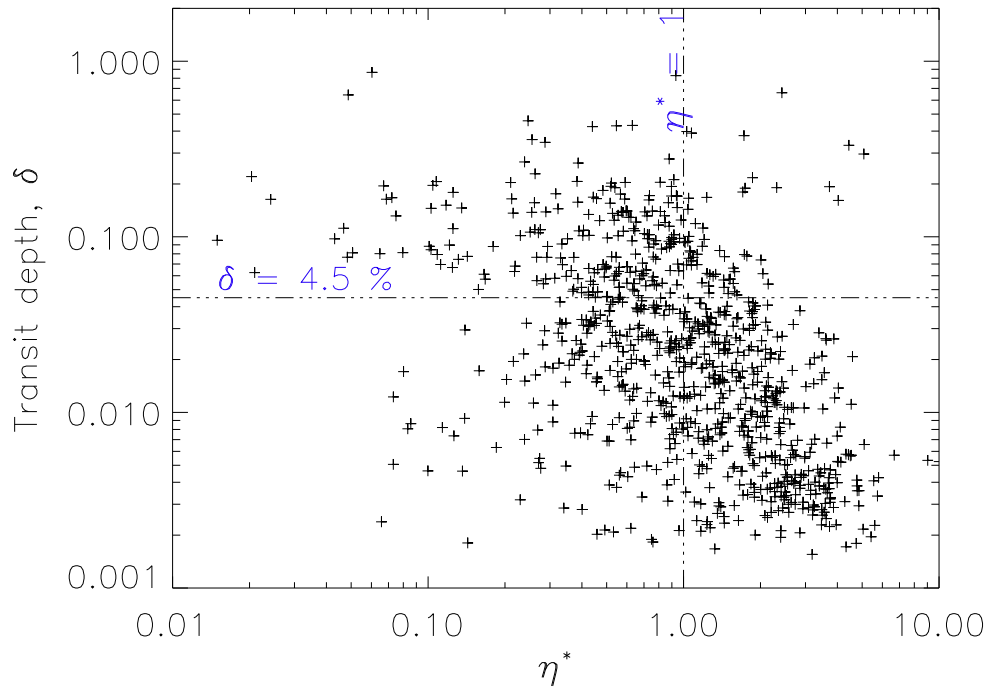
Figure 7.3: The relationship between η^* and δ for all the 908 candidates selected from the transit pipeline.
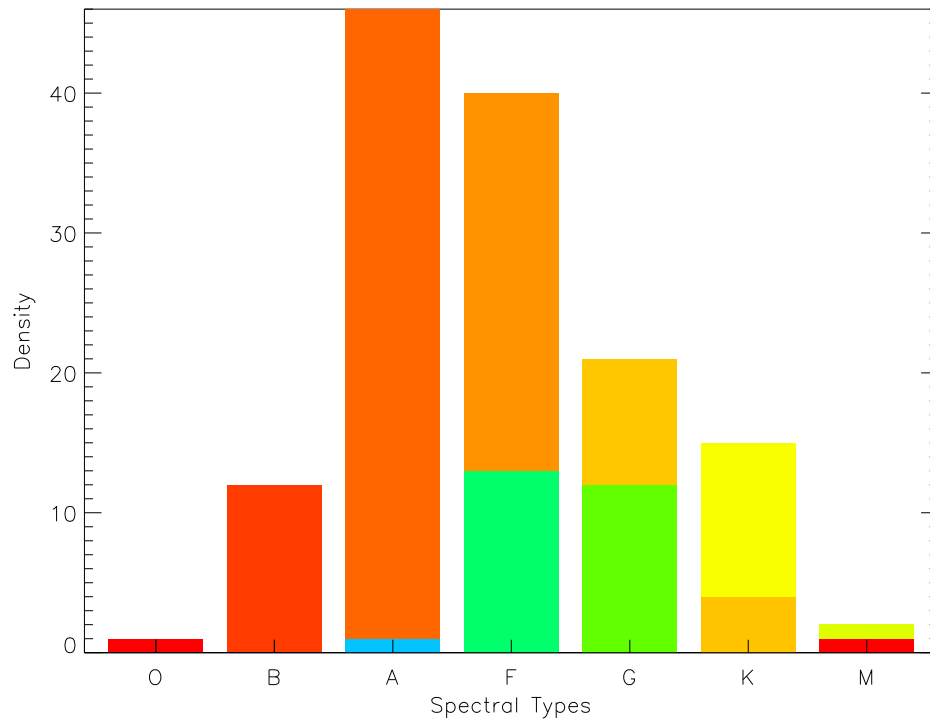
Figure 7.4: The plot illustrates the different spectral types of the *STEREO* candidates and the known exoplanets. The lower portion of the vertical bar represent the the spectral type of known transiting planets and the upper portion, the STRESS candidates.

this is a very young star, with an age < 50 Myr, with a planetary companion (see Güdel & Nazé, 2009 for a review).

HD 222891: A metal rich F8 star at a distance of 102 parsec, studied by N2K consortium as a potential exoplanet host star (Robinson et al., 2007). N2K consortium rejected this after their initial analysis from the potential targets for high precision radial velocity using the Keck telescope based on its higher temperature (Dr. Sarah Robinson, personal communication). No other literature of variability is found for this object. STRESS light curves gives a period of 1.59 days with an incredibly high SDE value of 30.9. From a BLS depth of 0.017, a planetary radius of $1.58R_{Jup}$ can be calculated, considering the radius of the star to be $1.25R_{\odot}$. The V shape of the eclipse can be attributed to either a systematic of the STRESS data or to the morphology of the actual system itself. The systematic is caused by the integration of the data, which smooths out a transit if the duration is close to integration time, 40 minutes in the case of STRESS. Another scenario is the case when one of the binary component is comparatively smaller than the other, for example, a binary system with G and M star companions (Fleming et al., 2011). Hence, as suggested by Don Kurtz (private communication), this light curve could be explained as a binary system.

HD 213597: This star is of interest, considering that Wraight et al. (2011) proposed it as an exoplanet candidate object based on an analysis using the same *STEREO* HI-1A data. It is an F0 star at a distance of approximately 114 parsecs and a V mag of 7.81. If we use a temperature of 7240K and a radius of $1.86R_{\odot}$ for the F0 star, for the period from Wraight et al. (2011) and a BLS depth of 2.4% found from our analysis, it can host a planet of $2.7R_{Jup}$. Wraight et al. (2011) give a transit depth of 25 mmag, which is close to the depth used in this calculation. This is actually very large compared to any known exoplanet and can probably be a K dwarf with a radius of $0.27R_{\odot}$.

HD 27268: An A0 star with no previous literature for variability cited in Simbad, presents an interesting object with a binary period of 4.126 days with a transit depth of 8.8%. This is a new variable with on-going analysis. The companion star can be an object of radius $0.57R_{\odot}$, if we assume the primary stellar radius as $2.0R_{\odot}$.

HD 110366: An A0 eclipsing binary with a binary period of 3.14 days. There is no previous study reported on this object. Even though secondary eclipses are clearly visible, the primary eclipse also exhibits a flat-bottom. A flat-bottom eclipse indicates that the eclipse is caused by a body which is much smaller than the primary body, like an M star around a B or A star or a planetary companion around a stellar body. The flat-bottom of these eclipses can also be the consequences of data integration, i.e., when the eclipse duration is long enough, our sampling smoothes the real eclipses, resulting in flattened transits rather than curved transits.

7.6 Radial Velocity Follow-up

The next step in the project, is to confirm the identified candidates using follow-up methods like high cadence photometry and /or high precision spectroscopy. Photometric follow-up can give precision transit duration and depth and thus, helps in refining the companion characteristics, while, a spectroscopic follow-up, measuring the radial velocity can confirm the companion as either a planetary or a stellar body and can also weed out blends from foreground stars. In the case of STRESS, we have attempted to get some spectroscopic data from the Wyoming Infrared Observatory (WIRO), Wyoming, US. This is a 2.3m Cassegrain telescope, with an integral-field, holographic spectrograph, within the optical limits of 400 - 1000nm [1]. WIRO would be able to confirm the presence of eclipsing binaries with its radial velocity variations but would be out-of-limits to be able to confirm the presence of transiting planets. Hence, for a true transiting planet, the negative result would be step further in confirming their presence. To that extent, currently, some test data has been taken for a sample of 3 stars, which looks promising from our light curve. This is at an infant stage, where we are attempting to a basic reduction of the spectra, in order to get radial velocities which suit our need and also to understand the spectroscopic data reduction process in general. We use the IRAF in-built procedures in this process. So far, we have reached a point where we can reduce the data and attain a radial velocity of a few km/s for bright stars in our sample. Future work including, identifying a promising set of candidates for further observation and analysing their data.

---

[1]http://www.physics.uwyo.edu/~chip/wiro/wiro.html
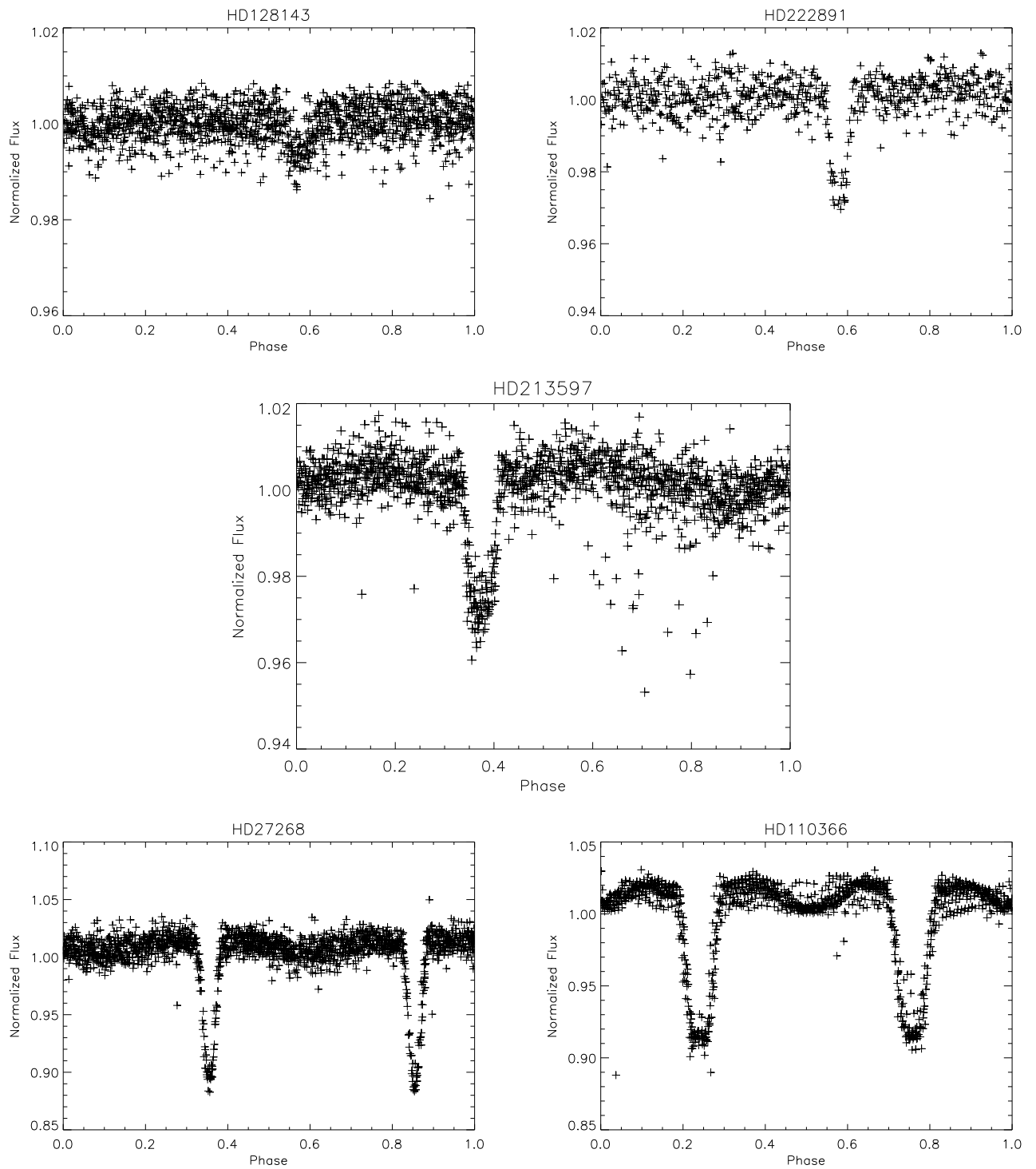
Figure 7.5: A flavour of the various objects flagged up by the Exoplanet transit analysis pipeline using the STRESS data. The top two objects are transit candidates, the middle plot belongs to a candidate suggested by Wraight et al. (2011) using the same data and the bottom two plots represent new eclipsing binary stars identified in this pipeline.
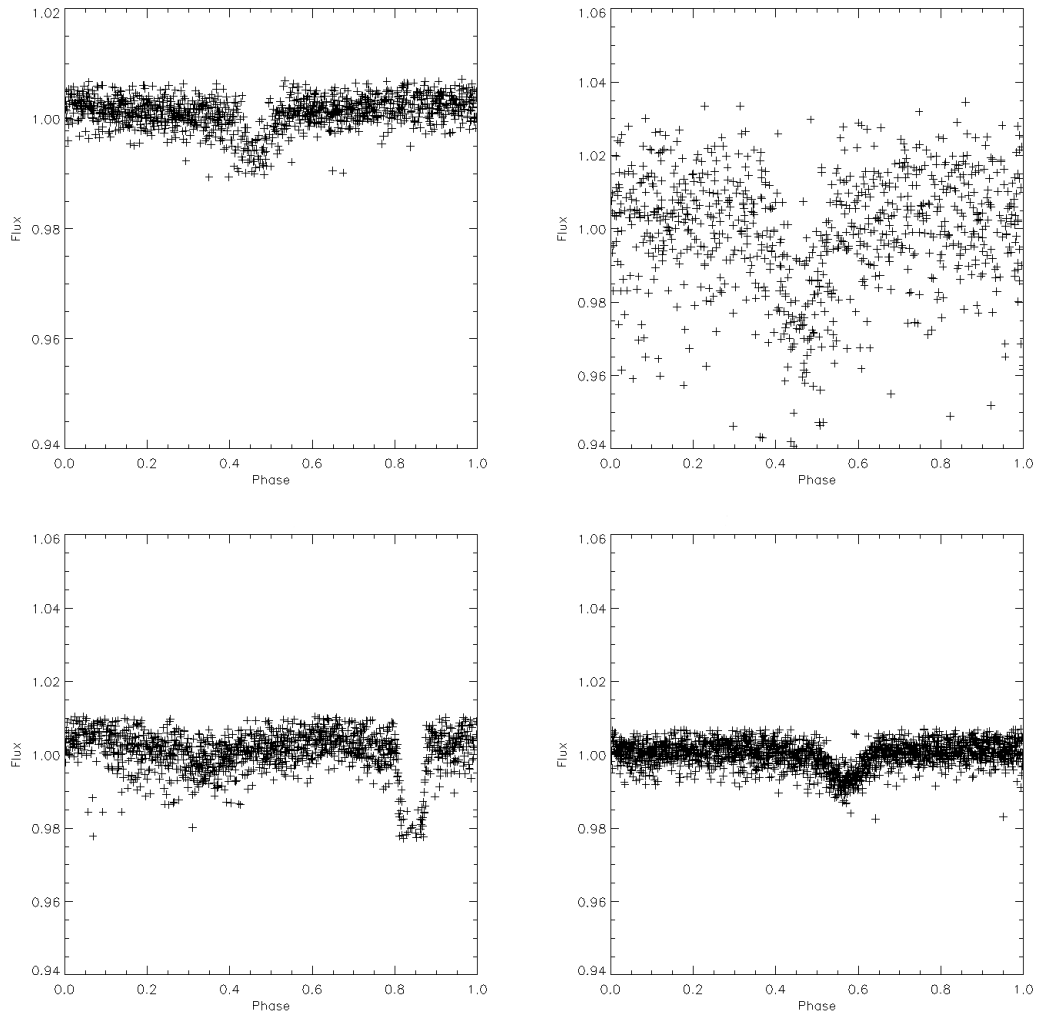
Figure 7.6: The following are some more examples of the interesting objects identified by our pipeline. These plots also stresses the level of variability that we can observe using *STEREO*. They are not identified to preserve the anonymity of the candidates.

7.7   Remarks and Comments

In summary, one of the main aims of this project is to identify planetary candidates in our data. The more stars looked at, the more likely we are to find a planet. Owing to the volume of data available, its coverage and depth in the sky and our preliminary analysis so far, this statement is reinforced again. As mentioned earlier, STRESS can have two lengths of data depending on the instruments used. We can calculate the theoritcal approximate transit duration in both these cases as follows: For a Solar type star with $M_* = 1 M_\odot, R_* = 1 R_\odot$ and a planet with $R_p = 1 R_{Jup}$, with a period of $P = 1$ days, $T_{dur} = 2.02$ hrs and with a period of $P = 6$ days. To compare it with the work by Wraight et al. (2011), who describe the eclipsing binaries

that are found in the data and our analysis from Chapter 6 give very similar results. In fact, our semi-automated pipeline has identified many of the same eclipsing binaries. In the case of planetary transits, the data in its current form present some major difficulties in order to identify candidates. However, we are currently attempting to reduce the systematics by the use of TFA and also enhance our understanding of the noise characteristics of the data by various simulations and methods used in literature (Pont, Zucker & Queloz, 2006). The results so far in both these regimes are promising to yield a better quality data and thus, we believe that the data indeed has the potential to identify planetary candidates in the future, specifically when combined with the data from the other three HI instruments. This preliminary investigation has shown a large potential for future work, in terms of follow-up analysis as well as more candidates when data from the other HI instruments are combined. The complete list of candidates arrived at from this pipeline will be published in a forthcoming paper by Whittaker et al. Currently, work is also ongoing to combine the data from other HI instruments for this list candidates at least, so that their parameters can be confirmed and false positives ruled out. This is a work in progress with much analysis and results to be announced to the scientific community in the future.

Chapter 8

Summary and Future Work

In this era of the search for extrasolar Earth-like planets, *STEREO* HI instruments offers a very promising capability for this specific purpose. This instrument is part of the SECCHI suite of instruments developed for the study of the evolution and structure of CME's in the inter-planetary region between the Sun and Earth in three dimensions. The potential use of this instrument as a stellar photometer was recognised during the early mission phase and was considered to be a continuation of similar work done using SMEI spacecraft.

This thesis in general has described the use of the HI instruments for the search of extrasolar planets and stellar variability in detail. The twin *STEREO* spacecrafts, as such are observing a portion of the sky in a unique time domain. The wide FOV of the instruments allow us to observe a substantial number of stars, thus increasing our probability of discovery. The HI-1 cadence of 40min is acceptable for the search of planets in an orbit of few days, whereas the HI-2 cadence falls right at the edge of the limit of detection if used alone. However, for candidates selected by HI-1 instruments, the HI-2 instrument can act as a confirmation ground by the availability of its extended data set. The data from HI-2 is excellent for studying many pulsating stars, both in terms of identifying them as well as studying them. Both the instruments together also cover nearly 60% of the sky and can probe objects as faint as $R = 12$, in a wavelength close to the R band. The data undergoes a basic reduction process, which is common to all type of science goals. The instrument faces a huge contamination due to the presence of the Solar F-corona. This has to be removed for both the CME analysis as well as stellar photometry. The next stage of the thesis describes the methodologies adopted for this purpose in detail, since it is a crucial part of the project. Specifically, for stellar photometry, it is absolutely essential that we remove as much of the F-corona as possible, since dynamic variations in this corona affect the photometric signals observed in the light curves.

After this process, the data is ready for photometric analysis. In order to simplify the problem in the development of the photometric pipeline, only data from the HI-1A instrument is used in this thesis. The input catalogue is crucial to this process as this determines the target, which will be eventually tracked for light curves. Then, the various catalogues available and the NOMAD catalogue which was finally used are discussed. Aperture photometry is performed in all these target stars using three different apertures and the final data is available for nearly a million stars for a period of 37 months from April 2007 till March 2010. Thus, accumulating data for individual stars for at least three epochs. The next part of the thesis dealt with a discussion of the various general characteristics of this data - the systematics and transient events observed due to the instrument itself or the presence and variation of the Solar component in the data or the effects of the analysis procedure. There is also a discourse of the Poisson noise limits achievable by the present analysis procedures and future prospects for the data, if that limit can be closely matched. The light curve at this time, can achieve a relative scatter, less than 1% for objects brighter than $R = 8$ and a few percent for fainter objects. This confirms the initial expectation that the data can be used successfully for the identification of exoplanet transit candidates which are Jupiter sized around a Sun-like star.

Various methodologies which were used for the removal of the systematic noise sources in data, is described in the next section along with some of the automated trend removal methods commonly used in the community, like the TFA and PDT. The prospect of using those methods with STRESS is also reviewed in this section. The next chapter described one of the two main branches of this project - Stellar variability analysis. The two different specialised algorithms used in this part of the project, namely Lomb-Scargle (LS) and Phase Dispersion Minimisation (PDM) are debriefed before describing the semi-automated pipeline devised for the data. This chapter also dealt with the result of this analysis with the identification of many new variables.

The next chapter described the exoplanetary transit analysis pipeline. This work describes the commonly used algorithm, Box-fitting Least Squares (BLS) and the various false positive detection tests employed during this analysis. These tests involve the computation of

the η^* statistic, validating the probability of the result being caused by a planetary signal as opposed to stellar signal along with Wald Test statistics, the host stellar classification information as well as a final visual inspection of the candidates for confirmation. Again the semi-automated pipeline used in this process in discussed in detail with some of the very early results. This is an active work in progress with more detailed analysis to be presented in a forthcoming publication. The radial velocity follow-up which was initiated in order to confirm these candidates are also described in this chapter in brief.

The semi-automated pipeline described for stellar variability is mostly complete and the project is in a highly productive stage, when the candidates are being visually inspected and confirmed. The necessary follow-up work needed for the confirmation of these objects are also underway at the time of writing. Currently there are candidates for different variability like δ Scuti, γ Doradus, pulsating B-stars, MCPs, Cepheids, Semi-regular variables and many more. The classification at this stage is not automated and hence it is a time consuming process to rule out false positives by individual analysis of each object and to make a detailed study of all these probable variable stars.

At the time of writing there are many active projects both in space and ground undergoing stellar variability studies. The information that will be collected from the STRESS data can only add to these ongoing studies, since the data is unique in many ways, such as its capability to study one day periods, the bright star regime, the continuous tracking of objects in the ecliptic plane at that particular time when the space based mission point away from the Sun and the ground based missions can obtain a time limited sample only. To stress this point, lets consider the Chemically Peculiar (CP) stars. The set of stars are generally observed in the upper main sequence of the HR diagram. The chemical anomalies found in the atmosphere of these objects are generally attributed to the diffusion process (See Kurtz, 2000 for a review of this type of objects with the various nomenclature and explaining mechanisms). There are two types of CP stars, namely Magnetic CP (MCP) and the non-magnetic type. The origin and the pulsation mechanisms of the CP stars in general are still unclear, especially the MCP stars which show more complicated phenomena (Gopka et al., 2010). One of the working theories explaining the

origin of MCP stars is that it is a binary system, where one of the component is a supernova remnant and hence its chemical peculiarity can be explained (Gopka et al., 2008). The other school debates that the anomalies develop as a result of atomic diffusion during the regular chemical evolution of the star. Photometric variability has been observed in many CP stars. These variations are thought to be either caused by pulsation modes or due to the presence of star spots caused by the concentration of elements, mostly near the magnetic poles. STRESS data, with its cadence, continuous observing capability as well as the pass band, presents an excellent opportunity to study and understand the physical mechanisms of these objects through their variability analysis. Recently, a detailed analysis of B-stars have been started by Dogus Ozuyar, University of Birmingham, UK. At present, there are nearly 50 MCP stars in the HI-1A FOV, 106 Be stars and about 44 early B-stars (B0 to B4). Fig. 8.1 is an MCP star found in STRESS which is currently under investigation by Dogus Ozuyar. We expect this work to follow in the lines of similar work done using the Solar Mass Ejection Instrument (SMEI) in the bright regime up to $V = 6$. For example, Goss et al. (2011) have done a detailed analysis of the Be star, Achernar using SMEI data. Fig. 8.2 shows the change in amplitude of two frequencies over a period of five years. Space-based monitoring of these objects also presents opportunities to study the changes in pulsational modes owing to the long period continuous observation available. For example, the study of the pulsation frequency changes in the Cepheid variable Polaris, by Spreckley & Stevens (2008) and also the work by Berdnikov & Stevens (2010), searching for random fluctuations in short period Cepheids using the SMEI data.

Still, the amount of data is extensive to be analysed by a few people alone. A catalogue of all these variables is in preparation to be published to the community for effective use of this huge data set. Another prospect which was initiated is to use the automatic classifier algorithm developed by Debosscher et al. (2007). This collaboration was initiated at the early stages of the project and there are plans to re-initiate this methodology now for STRESS data.

The exoplanetary transit analysis pipeline is also in its infant stage, where an initial list of candidates are available. The current stage involves a visual inspection and confirmation of these candidates. The analysis will be followed by various ground base and space based follow
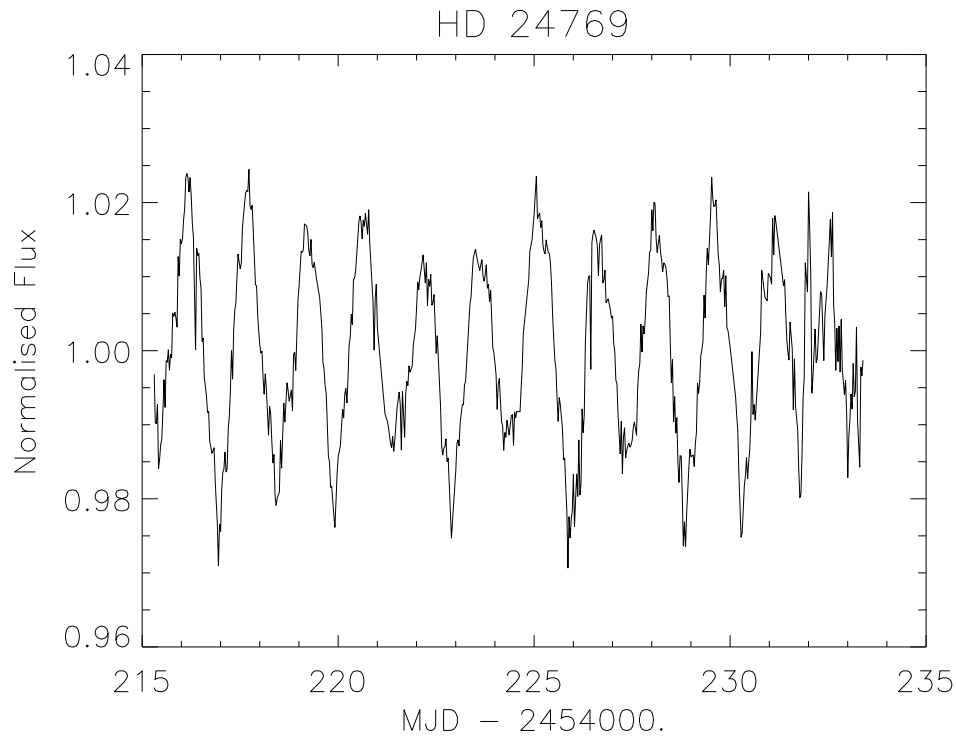
Figure 8.1: An example MCP star of $V = 6.0$ and B9.5 observed by STRESS. It is currently being studied in detail by Dogus.

up of promising candidates in order to confirm the presence of planets. We have received interest from the Micro-variability and Oscillation of Stars (MOST) project, to be involved in follow-up studies with promising candidates.

This project as a whole is never-ending, that is, the extraction of useful information from the data obtained by these instruments is not time limited. Once this information is published as a catalogue to the public, information can be obtained, referenced, inferred and extracted for the next ten years at the least. The data put together in this project is enormous and covers a large part of the parameter space. Target catalogue for this project contains around 840,000 which are present in the HI-1A FOV alone. Since, HI-2 FOV is three times bigger, we expect at least twice the size of this catalogue in the final run with all the instruments. For HI-1, an object is tracked for 20 days/yr, resulting it 700 data points for each HI-1 instrument or 1400 data points together and combined with the mission life of around five years gives an whopping $5,880,000,000$ photometric data points for the HI-1 instruments alone and at least twice this amount if HI-2 instruments are also covered. The potential of such a huge database is unlimited
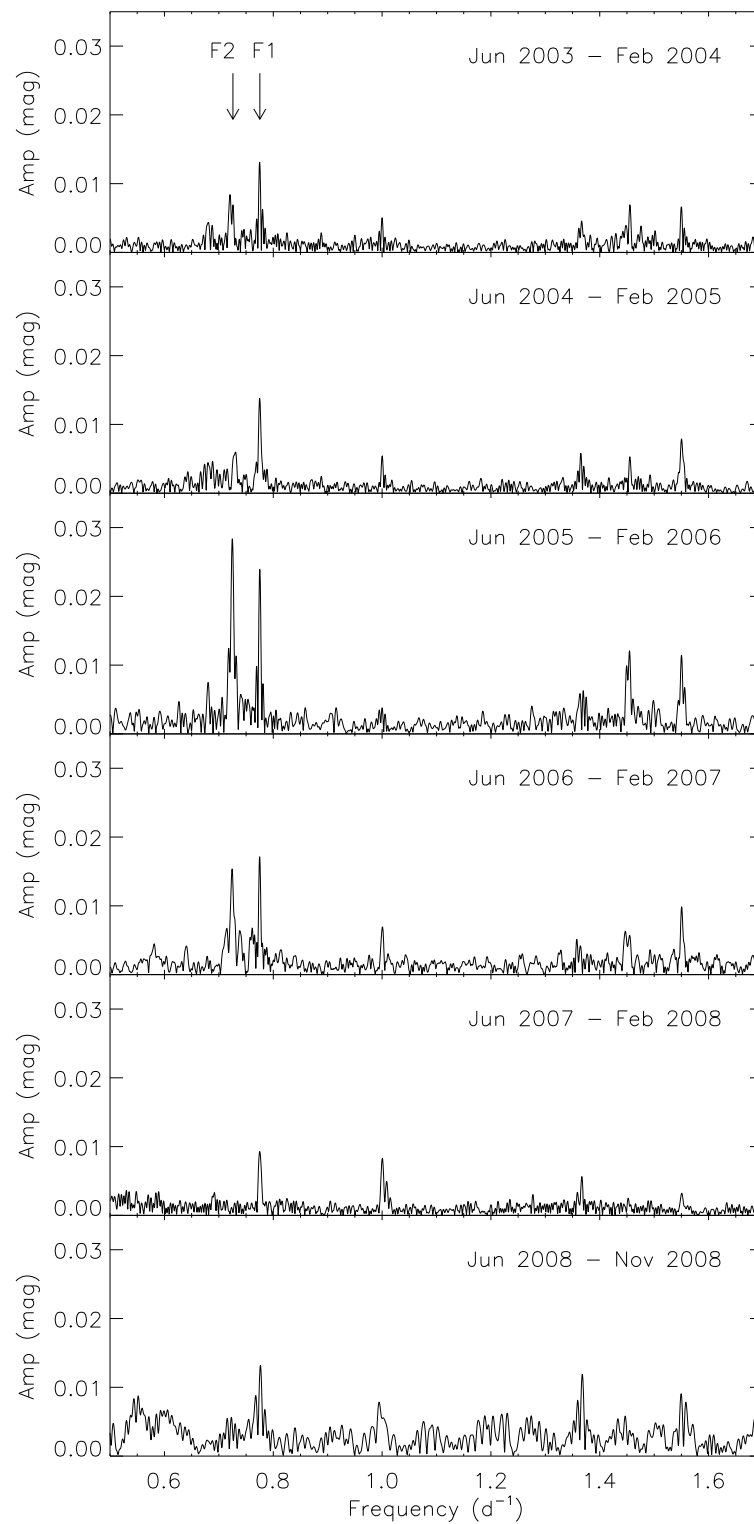
Figure 8.2: The figure shows the changing amplitude of two prominent period in Achernar over a period of five years. The data used is from SMEI spacecraft. Courtesy: Goss et al. (2011).

for different types of study ranging from individual study to statistical derivatives. There have
been many new bright variables discovered with it but the list is not exhaustive. We have exo-
planetary candidates, but hopefully new exoplanets as well in the near future. We believe that
this data will attain its scientific potential maybe a decade from when the spacecraft stops taking
data. Thus, this project STRESS - *STEREO* Transiting Exoplanet and Stellar Survey, holds a
huge potential for the advancement of our knowledge in the field of stellar and planetary studies.

References

Aigrain S., Favata F., 2002, Astronomy and Astrophysics, 395, 625.

Aigrain S. et al., 2009, Astronomy and Astrophysics, 506, 425.

Agol E., Steffen J., Sari R., Clarkson W., 2005, MNRAS, 359, 567.

Alibert Y., Mordasini C., Benz W., Winisdoerffer C., 2005, Astronomy and Astrophysics, 434, 343.

Alonso R. et al., 2004, ApJ, 613, L153.

Almenara J. M., Deeg H. J., Régulo C., Alonso R., 2007, Transiting Extrasolar Planets Workshop ASP Conference Series, 366, 183.

Almenara J. M. et al., 2009, Astronomy and Astrophysics, 506, 337.

Baglin A., Vauclair G. & The COROT Team, 2000, Journal of Astrophysics and Astronomy, 21, 319.

Bakos G.Á., Lázár J., Papp I., Sári P. & Green E. M., 2002, PASP, 114, 799, 974.

Bakos G. Á., Noyes R. W., Kovács G., Stanez K. Z., Sasselov D. D., Domsa I., 2004, PASP, 116, 817, 266.

Bakos G. Á. et al., 2007, ApJ, 656, 552.

Bakos G. Á., Hartman J. D., Torres G., Kovács G., Noyes R. W., Latham D. W., Sasselov D. D., Béky B., 2011, EPJ Web of Conferences, 11.

Balona L. A., 2011, MNRAS, 415, 1691.

Balona L. A et al., 2011, MNRAS, 413, 2403.

Barge P. et al., 2006, Proceedings of "The CoRoT Mission Pre-launch Status - Stellar Seismology and Planet Finding", ISBN 92-9092-465-9, 419.

Baruteau C., Masset F., 2012, arXiv:1203.3294.

Basri G. et al., 2010, The Astrophysical Journal Letters, 713, L155.

Baumann P., Ramirez I., Meléndez J., Asplund M., Lind K., 2010, Astronomy and Astrophysics, 519, A87.

Beck P. G. et al., 2011, Science, 332, 205.

Bedding T. R., Kjeldsen H., 2003, PASA, 20, 203.

Bedding T. R., Timothy R., 2011, arXiv:1107.1723.

Bedding T. R. et al., 2010, ApJ, 713, L176.

Benedict G. F. et al., 2002, ApJ, 581, 115.

Bennett D. et al., 2008, ApJ, 684, 663.

Berdnikov L. N., Stevens I. R., 2010, Variable Stars, the Galactic halo and Galaxy Formation, Proceedings of an international conference held in Zvenigorod, Russia, 12-16 October 2009.

Bewsher D., Brown D. S., Eyles C. J., Kellett B. J., White G. J., Swinyard B. M., 2010, Solar Physics, 264, 433.

Black D. C., 1980, Space Science Reviews, 25, 35.

Blomme J. et al., 2010, ApJ, 713, L204.

Bond I. et al., 2004, ApJ, 606, L155.

Bonfils X. et al., 2011, Astronomy & Astrophysics (submitted), arXiv:1111.5019.

Borucki W. J. & Summers A. L., 1984, Icarus, 58, 121.

Borucki W. J., Dunham E. W., Koch D. G., Cochran W. D., Rose J. D., Cullers D. K., Granados A., Jenkins J. M., 1996, Astrophysics and Space Science, 241, 111.

Borucki W. J. et al., 2010, Science, 327, 977.

Borucki W. J. et al., 2011, ApJ, 736, 19.

Bouchy F. et al., 2005, Astronomy and Astrophysics, 444, L15.

Brown D. S., Bewsher D., Eyles C. J., 2009, Solar Physics, 254, 185.

Brown T. M., 2003, ApJ, 593, L125.

Bryson S. T. et al., 2010, ApJL, 713, L97.

Budding E., 1986, Ap&S, 118, 241.

Buffington A., David L. B., Bernard V. J., Paul H., Aaron C. S., 2005, ApJ, 637, 880.

Cameron C. et al., 2007, MNRAS, 380, 1230.

Charbonneau D., Brown T. M., Latham D. W., Mayor M., 2000, ApJ, 529, L45.

Chadid M., Perini C., Bono G., Auvergne M., Baglin A., Weiss W. W., Deboscher J., 2011, Astronomy and Astrophysics, 527, A146.

Cherepashchuk A. M., 1992, Proceedings of the 151st. Symposium of the International Astronomical Union, 123.

Christensen-Dalsgaard J., 1998, Astrophy. Space Science, 261, 1.

Christensen-Dalsgaard J., 2011, in Pere L. Pallé, eds, Chapter "Asteroseismology", Canary Islands Winter School of Astrophysics, Vol. XXII, Cambridge University Press.

Colaninno R. C., Vourlidas A., Nieves-Chinchilla T., 2011, American Geophysical Union, Fall Meeting, SH23A-1946.

Cumming A., Butler R. P., Marcy G. W., Vogt S. S., Wright J. T., 2008, PASP, 120, 531.

Debosscher J. et al., 2007, Astronomy and Astrophysics, 475, 1159.

Debosscher J. et al., 2009, Astronomy and Astrophysics, 506, 519D.

Deeg H. J. et al., 2009, Astronomy and Astrophysics, 506, 343.

Defaÿ C., Deleuil M., Barge P., 2001, Astronomy and Astrophysics, 365, 330.

di Mauro M. P. et al., 2011, MNRAS, 415, 3783.Publication Date: 08/2011

Doyle L. et al., 2000, ApJ, 535, 338.

Driesman A., Hynes S., Cancro G., 2008, Space Science Reviews, 136, 17.

Dunham E.W., Mandushev G. I., Taylor B. W., Oetiker B., 2004, PASP, 116, 1072.

Eggleton P.P., Tokovinin A.A., 2008, MNRAS, 389, 869.

Eker Z., Bilir S., Yaz E., Demircan O., Helvaci M., Astronomische Nachrichten, 330, 68.

Eyer L. & Mowlavi N., 2008, Journal of Physics: Conference Series, 118, 012010.

Eyles C. J. et al., 2009, Solar Physics, 254, 387.

Faedi F., West R. G., Burleigh M. R., Goad M. R., Hebb, L., 2011, MNRAS, 410, 899.

Fischer D. et al., 2004, ApJ, 620, 481.

Fisher R.A., 1915, Biometrika, 10, 507.

Fleming S. W. et al., 2011, The Astronomical Journal, 142, 50.

Ford E.B., Rasio, F.A., 2008, ApJ, 686, 621.

Galvin A. B. et al., 2008, Space Science Reviews, 136, 437.

Gaudi B.S., 2010, Refereed chapter in Exoplanets, Space Science Series, University of Arizona
    Press, arXiv:1002.0332.

Gilliland R. L. et al., 2000, ApJ, 545, L47.

Gilliland R. L. et al., 2010, PASP, 122, 131.

Gopalswamy N. et al., 2009, Solar Physics, 259, 227.

Gopka V. F., Ulyanov O. M., Andrievsky S. M., 2008, AIP Conference Proceeding, in Origin
    of matter and Evolution of Galaxies: The 10th International Symposium on Origin of Matter
    and Evolution of Galaxies: From the Dawn of Universe to the Formation of Solar System,
    1016, 460.

Gopka V. F., Ulyanov O. M., Andrievsky S. M., Shavrina A. V., Yushchenko V. A., 2010, Odessa Astronomical Publications, 23, 41.

Goss K. J. F., Karoff C., Chaplin W. J., Elsworth Y., Stevens I. R., 2011, MNRAS, 411, 162.

Gould A. & Morgan C. W., 2003, The Astrophysical Journal, 585, 1056.

Güdel M., Nazé Y., 2009, Astronomy and Astrophysics Review, 17, 309.

Haakonsen C. B., Rutledge R. E., 2009, ApJS, 184, 138.

Halain J. P., Mazy E., Defise J. M., Moses J. D., Newmark J. S., Korendyke C. M., Eyles C. J., Harrison R. A., Davis C. J., 2007, in Fineschi S., Viereck R. A., eds., Proc. SPIE Vol 6689, Solar Physics and Space Weather Instrumentation II.

Handler G., Shobbrook R.R., 2002, MNRAS, 333, 251.

Harrison R. A. et al., 2009, Solar Physics, 256, 219.

Hekker S., Broomhall A. M., Chaplin W. J., Elsworth Y. P., Fletcher S. T., New R., Arentoft T., Quirion P. O., Kjeldsen H., 2010, 402, 2049.

Hekker S., 2010, Astronomische Nachrichten, 331, 1004.

Hekker S. et al., 2011, Astronomy and Astrophysics, 525, 131.

Hernanz M., José J., 2008, New Astronomy Reviews, 52, 386.

Hess P., Zhang J., 2011, Bulletin of the Americal Astronomical Society, 43.

Hoffleit D., Jaschek C., 1991, New Haven, Conn: Yale University Observatory, 5th revised edition, edited by Hoffleit D., Jaschek C.

Holman M. J., Murray N. W., 2005, Science, 307, 1288.

Horne J. H., Baliunas S. L., 1986, ApJ, 302, 757.

Howard R. A. et al., 2008, Space Science Review, 136, 67.

Howard A. W. et al., 2009, ApJ, 696, 75.

Howell S. B. et al., 2005, ApJ, 746, 123.

Israelian G. et al., 2009, Nature, 462, 189.

Jain A. K., Murty M. N., Flynn P. J., 1999, ACM Comput. Surv., 31, 264.

Jenkins J., Doyle L., Culler D., 1996, Icarus, 119, 244.

Jenkins J. M. et al., 2010a, Proceedings of the SPIE, 7740, 77400.

Jenkins J. M. et al., 2010b, ApJL, 713, L87.

Kaiser M. L., Kucera T. A., Davila J. M., St. Cyr O. C., Guhathakurta M., Christian E., 2008, Space Science Reviews, 136, 5.

Kalas J. R. et al., 2008, Science, 322, 1345.

Kaltenegger L., Eiroa C., Fridlund C. V. M., 2010, Astrophysics and Space Science, 326, 233.

Karoff C. et al., 2007, The Astronomical Journal, 134, 766.

Kim Dae-Won., Protopapas P., Alcock C., Byun Y., Bianco F., 2009, MNRAS, 397, 558.

Kimura H., Mann I., 1998, Earth, Planets and Space, 50, 493.

Kipping D. M., 2010, MNRAS, 408, 1758.

Kjeldsen H., Bedding T. R., 2011, Astronomy and Astrophysics, 529, 8.

Kovács G., Zucker S., Mazeh T., 2002, Astronomy and Astrophysics, 391, 369.

Kovács G., Bakos G., Noyes R.W., 2005, MNRAS, 356, 557.

Kurtz D. W., 2000, in Breger M., Montgomery M., eds., Delta Scuti and Related Stars, Reference Handbook and Proceedings of the 6th Vienna Workshop in Astrophysics, Austria, ASP Conference Series, 210, 287.

Latham D. W., Stefanik R. P., Mazeh T., Mayor M., Burki G., 1989, Nature, 339, 38.

Lebzelter T., 2011, Astronomy and Astrophysics, 530, A35.

Lenz P., Breger M., 2004, in Zverko J., Ziznovsky J., Adelman S. J., Weiss W. W., eds, IAU Symposium 224: The A-Star Puzzle, Cambridge University Press, Cambridge, UK, 786.

Lenz P., Pamyatnykh A A., Zdravkov T., Breger M, 2010, Astronomy and Astrophysics, 509, A90.

Lister T., Street R., 2007, Bulletin of the American Astronomical Society, 39, 970.

Lister T. A. et al., 2007, MNRAS, 379, 647.

Lomb N. R., 1976, Applied Space Sciences, 39, 447.

Maciejewski G. et al., 2010, MNRAS, 407, 2625.

Mattei J. A. et al., 1987, AAVSO Monograph, Cambridge University Press, Cambridge.

Matthews J. M. et al., 2000, ASPC, 203, 74.

Marcy G., Butler R. P., Fischer D., Vogt S., Wright J. T., Tinney C. G., Jones H. R. A., 2005, Progress of Theoretical Physics Supplement, 158, 24.

Marois at al., 2008, 322, 1348.

Maxted P. F. L. et al., 2011, PASP, 123, 547.

Mayor M., Queloz D. , 1995, Nature , 378 , 355.

Mayor, M. et al., 2003, Messenger, 114, 20.

Mayor M. et al., 2009, Astronomy and Astrophysics, 507, 487.

McCullough P. R., Stys J. E., Valenti J. A., Fleming S. W., Janes K. A., Heasley J. N., 2005, PASP, 117, 783.

Monet D. G. et al., 2003, AJ, 125, 984.

Mordasini C., Alibert Y., Benz W., Naef D., 2008, Extreme Solar Systems, ASP Conference Series, 398, 235.

Mordasini C., Alibert Y., Benz W., Klahr H., 2011, in Bouchy F., Diaz R. F., Moutou C., eds, Proceedings Haute Provence Observatory Colloquim: Detection and Dynamics of Transiting Exoplanets, arXiv:1102.4146.

Morton T. D., Johnson J. A., 2011, ApJ, 738, 130.

Naef D. et al., 2001, Astronomy and Astrophysics, 375, L27.

Norton A. J. et al., 2007, Astronomy and Astrophysics, 467, 785.

Nataf D. M., Stanek K. Z., Bakos G. Á., 2010, Acta Astronomica, 60, 261.

Owens A. J., 1977, ICRC, 9, 222.

Payne M.J., Veras D., Ford E. B., 2010, Bulletin of the American Astronomical Society, 41, 928.

Perryman M. et al., 2005, ESA-ESO Working Group Report, http://www.stecf.org/coordination/eso-esa/extrasolar/report.pdf.

Pojmański G ., 2000, Acta Astronomica, 50, 177.

Pollacco D. L. et al., 2006, PASP, 118, 1407.

Pollack J. B., Hubickyj O., Bodenheimer P., Lissauer J. J., Podolak M., Greenzweig Y., 1996, Icarus, 124, 62.

Pont F., Zucker S., Queloz D., 2006, MNRAS, 373, 231.

Press W. H., Rybicki G. B., 1989, ApJ, 338, 277.

Queloz D., Mayor M., 2001, Messenger, 105, 1.

Quirrenbach A. et al., 2011, The Astrophysics of Planetary Systems: Formation, Structure, and Dynamical Evolution, Proceedings of the International Astronomical Union, IAU Symposium, 276, 545.

Reiners A., Bean J. L., Huber K. F., Dreizler S., Seifahrt A., Czesla S., 2010, ApJ, 710, 432.

Reinhold T., Reiners A., Basri G., Walkowicz L. M., 2011, ASP Conference Series (in press), arXiv:1012.2217.

Robinson S. E. et al, 2007, ApJs, 169, 430.

Rodríguez E., López-Gonzalez M.J., López de Coca P., 2000, Astronomy and Astrophysics Suppl. Series, 144, 469.

Sangaralingam V. & Stevens I. R., 2011, MNRAS, 418, 1325.

Santerne A. et al., 2011, Astronomy and Astrophysics, 528, A63.

Scargle J. D., 1982, ApJ, 263, 835.

Schwarzenberg-Czerny A., 1997, ApJ, 489, 941.

Schwarzenberg-Czerny A., 1998, MNRAS, 301, 831.

Seager S., 2008, Space Science Review, 135, 345.

Slawson R. W. et al., 2011, The Astronomical Journal, 142, 160.

Snellen I. A. G., van der Burg R. F. J., de Hoon M. D. J., Vuijsje F. N., 2007, Astronomy and Astrophysics, 476, 3, 1357.

Socker D. G., Howard R. A., Korendyke C. M., Simnett G. M., Webb D. F., 2000, SPIE, 4139, 284.

Spreckley S. A., Stevens I. R., 2008, MNRAS, 388, 1239.

Spreckley S., 2010, PhD Thesis, University of Birmingham, UK.

Stellingwerf R. F., 1978, ApJ, 224, 953.

Stello D. et al., 2010, ApJ, 713, L182.

Struve O., 1952, The Observatory, 72, 199.

Swain M. R. et al., 2008, Nature, 452, 329.

Tamuz O., Mazeh T., Zucker S., 2005, MNRAS, 356, 1466.

Terrell D., Williams D. B., Kaiser D. H., 1995, PASP, 107, 653.

Tinetti G. et al., 2007, Nature, 448, 169.

Tingley B., 2003a, Astronomy and Astrophysics, 403, 329.

Tingley B., 2003b, Astronomy and Astrophysics, 408, L5.

Tingley B., Sackett P. D., 2005, ApJ, 627, 2, 1011.

Thong T., McNames J., Aboy M., Oken B, 2004, Biosignal 2004, Biomedical Signal Processing (BSP) Laboratory at Portland State University, BIOS1GNAL2004 (http://bsp.pdx.edu/Publications/2004/BIOSIGNAL_Thong.pdf).

Thompson W. T., et al., 2003, SPIE, 4853, 1.

Twicken J. D. et al., 2010a, SPIE, 7740, 774023.

Twicken J. D. et al., 2010b, SPIE, 7740, 77401U.

Udalski A., Szymanski M., Kaluzny J., Kubiak M., Mateo M., 1992, Acta Astronomica, 42, 253.

Udalski A. et al., 2005, ApJL, 628, L109.

Vidal-Madjar A. et al., 2003, Nature, 442, 143.

Vidal-Madjar A. et al., 2004, ApJ, 604, 69.

von Braun K., Lee B. L., Seager S., Yee H. K. C., Mallén-Ornelas G., Gladders M. D., 2005, PASP, 117, 141.

von Braun K. et al., 2011, ApJ, 740, 49.

Vourlidas A., Syntelis P., Tsinganos K., 2011, Solar Physics Topical Issus, "Advances of European Solar Physics, arXiv:1201.0162.

Walker G., Matthews J., Rainer K., Johnson R., 2003, PASP, 115, 1023.

Warner B., 1995, Cataclysmic Variable Stars, Cambridge University Press, Cambridge.

Wilson R. E., 1979, ApJ, 234, 1054.

Wilson D. M., 2007, Transiting Extrasolar Planets Workshop ASP Conference Series, 366, 187.

Winn J. N., 2009, Transiting Planets, Proceedings of the International Astronomical Union, IAU Symposium, 253, 99.

Wolszczan A., Frail D. A., 1992, Nature, 355, 145.

Wraight K. T., White G. J., Bewsher D., Norton A. J., 2011, MNRAS, 416, 2477 .

Wuelser J.P. et al., 2004, SPIE, 5171, 111.

Zacharias N., Monet D. G., Levine S. E., Urban S. E., Gaume R., Wycoff G. L., 2004, Bulletin
    of American Astronomical Society, 36, 1418.

Zhou, A., 2002, Chinese Journal of Astronomy and Astrophysics, 2, 43.

Appendix: Sample Variables

Table 8.1: A random sample of variable stars as found from the analysis of STRESS data. The table contains the Right Ascension (RA), Declination (Dec) and R magnitude values as from our input catalogue using NOMAD. The ID and spectral information are as found from Simbad database. 'n/a' in first and fifth column indicate that no corresponding object is found from Simbad and the rest of the symbols are default from Simbad.

ID	RA	Dec	R Mag	Spectral Type	Period Found (days)
HD 224908	0.426796	-0.217777	8.28	G5	0.905
HD 224986	0.571702	-2.13914	9.41	F8	0.912
HD 225129	0.924630	-1.55556	8.71	K0	0.984
BD+09 5315	0.959310	10.4511	8.66	M1	0.583
TYC 4663-208-1	1.25229	-1.54534	9.92		0.915
HD 225294	1.26467	1.03771	8.37	G5	0.721
BD+02 4751	1.31341	3.78163	9.67	K0	1.163
HD 45	1.35354	-1.44022	9.06	F8	1.162
HD 199	1.71815	1.86103	8.69	F8	1.063
** GRV 7	1.90365	4.35685	9.84		0.878
BD+02 94	10.5230	2.84396	9.17	G	0.564
HD 3952	10.5329	-1.20391	9.03	K0	0.655
BD+10 77	10.6011	11.1742	9.39	F8	1.619
HD 4046	10.7442	6.81160	7.72	K0	0.852
TYC 610-682-1	10.8239	12.6257	10.09		1.283

150

TYC 1188-1597-1	10.8945	15.1767	9.23		1.021
HD 54294	107.323	23.6397	8.15	F	0.233
BD+12 1471	108.803	12.7318	9.75	A0	1.771
TYC 1921-1909-1	109.089	29.3575	11.57		0.553
TYC 1346-1064-1	109.546	15.7053	10.71		1.074
BD-00 108	11.1120	0.525110	8.44	K0	0.805
TYC 17-976-1	11.3725	5.63139	10.27		0.881
BD+11 93	11.6212	12.3492	9.15	K0	0.219
BD+13 105	11.6537	14.4306	11.26	K5	0.965
HD 4423	11.6601	14.0031	8.46	B8	1.024
HD 4503	11.8076	8.64400	8.63	G0	0.855
HD 4570	11.9114	-4.49765	9.11	A	0.797
HD 4552	11.9215	12.8793	8.58	F8	1.049
TYC 11-825-1	11.9380	0.742124	11.17		0.932
TYC 1351-10-1	110.581	17.3257	9.51		0.929
TYC 1355-1000-1	110.747	20.2219	10.38		1.149
TYC 1922-1463-1	111.655	28.7801	11.12		0.955
N/A	112.520	15.1935	12.01	n/a	1.255
N/A	112.960	15.9392	11.87	n/a	1.075
TYC 777-795-1	113.685	13.2325	9.21		0.989
TYC 777-1070-1	114.416	14.3114	11.46		1.985
HD 4649	12.1384	0.382053	7.75	G0	0.936
N/A	12.4008	-3.40597	11.45	n/a	0.875
TYC 607-117-1	12.5190	11.8512	11.35		0.915
HD 68934	123.789	15.8276	8.86	F	1.093
TYC 1382-1593-1	124.343	18.3685	11.17		0.952
N/A	128.837	12.9997	11.18	n/a	0.956
HD 72982	129.178	25.9373	7.48	A0	0.953
TYC 797-709-1	129.188	9.28803	11.22		1.422
TYC 809-1187-1	129.953	13.1653	11.23		1.244

TYC 805-1043-1	129.959	12.1522	11.56		1.077
TYC 608-938-2	13.0317	10.6009	9.15		0.867
TYC 611-999-1	13.1035	14.2592	10.43		0.971
TYC 15-257-1	13.9004	2.53296	10.89		0.913
BD+24 1976B	130.491	23.7971	9.44		1.228
BD+15 1880	130.935	14.7272	8.34	M0	1.523
TYC 810-948-1	130.984	9.81155	10.42		1.238
BD+15 1890	131.307	15.3881	9.33	A	0.422
BD+16 1815	131.582	16.0367	8.55	F	0.372
TYC 1393-2099-1	131.631	16.3040	10.76		0.372
HD 74815	131.635	8.47794	7.06	B9	1.344
N/A	133.201	18.3975	11.15	n/a	1.253
HD 77444	135.813	27.2113	7.01	K4III	1.226
BD+07 2069	135.960	7.29592	9.41	F	2.921
TYC 818-1298-1	136.277	12.6908	10.18		1.351
HD 79426	138.637	18.6439	7.81	K	0.572
V* GN Cnc	139.018	20.0732	8.56	F	0.573
HD 5558	14.3700	9.51972	8.35	F8	0.219
TYC 4674-520-1	14.4772	-2.42331	10.65		0.924
TYC 831-723-1	147.007	11.7254	11.22		0.894
HD 5859	15.0235	2.63296	6.73	K	2.369
N/A	151.086	15.5293	11.56	n/a	0.222
BD+03 2335	153.309	2.41919	9.35	K	1.025
HD 89773	155.474	19.7932	8.63	G	0.875
TYC 256-1402-1	157.601	3.63457	9.67		0.577
HD 91686	158.830	5.80693	8.85	F8	1.554
TYC 259-504-1	158.848	5.85323	10.62		1.585
TYC 259-121-1	159.318	6.95656	11.97		0.685
HD 92111	159.522	-2.23552	8.96	G0	1.099
N/A	159.609	1.75643	11.99	n/a	1.305

HD 6422	16.2881	11.6781	7.64	K5	0.432
BD+07 170	16.7278	7.88154	9.85	F8	0.261
HD 92390	160.006	-0.146969	8.26	A	0.817
HD 92706	160.591	0.862139	6.85	K	1.294
HD 93832	162.470	-2.69728	8.02	K0	1.291
HD 94539	163.705	-2.84711	8.47	F0	0.885
V* HR Leo	166.176	-2.97222	8.71	F	0.539
2MASS J11044901-0253196	166.204	-2.88885	11.89		0.541
V* FK Leo	168.023	14.3062	8.25	F5III	0.875
* 69 Leo	168.440	-0.0695020	5.42	A0V	0.651
BD+10 2269	169.113	9.34086	10.05	G5	0.546
HD 98156	169.396	4.32296	7.53	G5	0.637
BD+06 2429	169.733	5.29397	9.27	K0	3.309
TYC 855-320-1	170.995	8.29711	11.17		1.267
HD 99287	171.407	8.03567	8.41	K0	1.044
BD+07 2447	171.555	6.78426	8.93	K0	0.771
HD 100285	173.073	-5.62927	11.67	K	1.285
[ZEH2003] RX J1132.8+1210 3	173.206	12.1736	8.92		0.788
TYC 274-22-1	173.484	2.60150	10.54		1.021
N/A	173.602	-5.58108	11.31	n/a	1.268
TYC 4936-232-1	173.684	-5.35413	11.12		1.305
HD 101209	174.706	-7.60004	8.05	F	0.685
N/A	176.889	5.63388	11.65	n/a	0.566
N/A	18.3158	3.01238	11.07	n/a	0.875
HD 7321	18.3465	-2.55083	6.55	K0	0.465
N/A	180.356	10.6379	11.54	n/a	0.687
HD 109158	188.187	-12.7847	8.46	K1III	0.816

BD+04 2632	189.543	4.18645	8.42	M0	1.189
N/A	190.270	-9.56377	11.15	n/a	0.568
BD-04 3380	194.095	-5.40844	8.82	K5	0.567
TYC 4960-112-1	195.638	-2.90154	11.08		0.564
BD-11 3434	196.369	-11.9333	10.52		0.789
HD 113718	196.428	-12.3700	8.78	F8	0.784
BD+00 3022	196.842	0.103155	11.33	A0	0.565
HD 113970	196.851	-12.6684	9.01	G3V	0.839
N/A	196.921	-14.1150	11.68	n/a	1.201
HD 114302	197.428	-3.38954	8.27	F	0.716
HD 114430	197.632	-10.8146	7.24	K0	1.333
TYC 302-95-1	198.162	2.72341	9.85		0.576
TYC 302-1012-1	198.222	2.93260	11.45		0.549
V* UW Vir	198.836	-17.4714	9.01	A8IV/V	0.915
BD+04 4	2.12344	5.62687	8.62	K0	1.093
HD 518	2.43205	-3.67842	9.28	F5	0.323
BD+07 9	2.46373	8.45331	9.09	G7V	0.947
TYC 597-1216-1	2.47648	10.4145	10.71		0.814
TYC 4666-822-1	2.48293	-4.51417	11.97		0.855
N/A	2.54135	-4.71067	8.85	n/a	1.074
TYC 4667-1-1	2.56490	-2.50316	9.39		0.855
BD+04 9	2.57145	4.89238	9.25	F8	0.838
HR 29	2.57862	-5.24859	5.25	K1III	1.067
BD+04 11	2.60437	5.36299	10.13	F8	0.971
2MASS J01220532+0339117	20.5222	3.65326	11.61		0.842
HD 116045	200.301	-9.99911	6.67	G5	1.161
TYC 4962-503-1	200.864	-4.75379	9.86		1.289
N/A	201.107	-4.61651	8.56	n/a	1.307
BD+02 2676	201.641	1.73338	8.65	K	0.797

BD-03 3469	201.669	-4.60374	9.89	F8	0.848
HD 119095	205.345	-10.4525	6.91	M...	1.985
N/A	209.038	-5.12824	11.45	n/a	0.764
N/A	209.642	-21.2695	10.61	n/a	0.811
TYC 6149-241-1	211.469	-22.2357	11.25		1.237
BD-12 3971	212.230	-13.2689	9.18		0.755
TYC 6142-822-1	212.301	-18.5194	10.87		0.796
TYC 4978-686-1	212.324	-4.04155	11.35		0.538
BD-09 3872	212.854	-10.2288	9.67		0.754
HD 125002	214.329	-20.4713	8.83	F8V	0.246
2MASS J14243782-1057134	216.158	-10.9538	11.32		0.246
BD-13 3925	218.403	-13.9009	10.22		1.228
N/A	218.952	-18.0365	8.53	n/a	0.488
BD-10 3953	221.266	-10.6710	10.42		0.492
TYC 6155-608-1	222.370	-16.4512	10.65		0.761
TYC 6756-480-1	224.767	-27.3069	10.49		1.064
N/A	226.986	-25.0589	11.78	n/a	0.657
HD 9947	24.3139	2.58245	8.04	K	0.868
BD+14 248	24.7314	14.7447	9.35	F8	0.995
N/A	24.8376	8.17022	10.05	n/a	0.897
TYC 621-695-1	25.3075	8.47885	9.45		0.846
HD 10570	25.7775	3.47240	9.25	F8	1.285
V* EU Psc	26.2229	19.8568	8.17	F0	0.846
HD 10966	27.0213	21.6588	8.53	F0	0.918
HD 11349	27.9839	17.0796	7.54	G5	0.785
TYC 1212-1328-1	28.5094	21.9913	11.27		1.155
BD+17 282	28.5959	17.8955	11.44	F	0.738
* bet Ari	28.6600	20.8080	2.56	A5V	0.787

TYC 6867-910-1	282.402	-27.5465	10.65		1.065
HD 174091	282.456	-23.3623	8.78	K1III	0.994
CD-26 13532	282.629	-26.8629	9.41		1.007
HD 174312	282.719	-23.5846	11.02	G	0.968
TYC 6860-392-1	282.726	-22.5175	10.81		1.006
HD 174312	282.735	-23.6000	8.98	G	0.954
HD 174331	282.737	-22.5824	9.26	F5V	1.053
BD-22 4883	282.747	-22.6325	9.53	G5	1.015
N/A	282.748	-18.0430	10.85	n/a	1.105
TYC 6281-1361-1	282.794	-15.8625	11.35		0.904
HD 174597	283.034	-22.4243	9.04	K0III	1.058
CD-23 14792	283.315	-23.2507	10.71		1.098
CD-23 14800	283.363	-22.8727	8.93	K5	1.005
TYC 6860-1608-1	283.381	-23.9466	10.85		1.023
N/A	283.541	-23.4011	11.45	n/a	0.957
HD 175067	283.599	-21.3052	8.32	K0III	1.127
HD 175160	283.801	-26.6693	8.35	M	1.035
HD 175832	284.571	-24.7617	8.68	K5	1.052
TYC 6873-287-1	284.692	-22.8506	9.45		1.096
HD 176472	285.323	-22.1175	8.65	K0III	1.093
TYC 6295-2469-1	286.057	-20.8244	10.93		1.214
TYC 6295-120-1	286.065	-22.3667	10.69		2.026
TYC 6295-839-1	286.088	-20.7680	10.61		1.225
TYC 6873-42-1	286.095	-22.5523	11.69		2.025
TYC 6291-1861-1	286.166	-20.4896	10.35		1.219
N/A	286.822	-18.6879	11.98	n/a	1.177
HD 177891	286.849	-21.6287	11.49	F3IV/V	1.113
TYC 6295-2233-1	287.383	-21.4155	11.61		1.071
TYC 6295-2173-1	287.420	-21.3803	11.79		1.138
N/A	287.635	-21.5772	11.84	n/a	1.121

HD 178727	287.643	-21.6725	8.52	F3V	1.065
HD 178753	287.653	-21.3734	8.65	K1III	1.177
HD 178786	287.745	-25.6721	10.25	Ap...	0.754
HD 178813	287.764	-25.6673	8.26	K	0.771
TYC 6295-2118-1	287.786	-21.8783	10.92		0.925
TYC 6878-920-1	287.791	-25.3624	10.35		0.774
N/A	288.088	-17.5656	10.65	n/a	0.929
TYC 6874-251-1	288.162	-23.0543	11.72		0.932
HD 179233	288.193	-24.9062	8.69	A9V	0.821
HD 179266	288.194	-19.2425	11.84	G0V	1.375
HD 179294	288.207	-20.5287	9.12	G	1.195
TYC 5720-445-1	288.217	-14.1145	10.24		0.982
TYC 6308-719-1	288.490	-20.9758	11.71		0.948
HD 179610	288.529	-23.0668	8.27	G8III	0.912
TYC 6874-605-1	288.535	-23.0089	10.45		0.958
TYC 6308-1008-1	288.538	-21.1658	11.83		1.132
TYC 6304-1445-1	288.539	-20.3823	10.58		1.207
BD-20 5457	288.559	-20.5142	11.67		1.229
TYC 6874-196-1	288.564	-23.0624	10.89		0.967
HD 179612	288.574	-26.5719	8.57	M4/M5III	0.693
HD 179663	288.587	-22.7056	9.21	K1III	0.932
TYC 6878-582-1	288.598	-25.3377	10.64		0.762
TYC 6304-783-1	288.607	-20.5260	10.84		1.221
HD 179665	288.626	-25.3423	8.45	K	0.767
TYC 6304-256-1	288.669	-18.9260	11.28		1.546
CD-24 15169	288.976	-24.1269	9.64		0.837
TYC 6308-358-1	288.979	-22.0477	10.67		1.026
CD-24 15170	288.982	-24.0843	8.02	K	0.815
CD-24 15171	288.997	-24.1020	11.43		0.867
CD-24 15171	289.005	-24.1104	9.81		0.815

CD-24 15173	289.016	-24.5292	8.91	K0	0.808
TYC 6879-1367-1	289.027	-24.5796	10.18		0.791
TYC 6879-1367-1	289.036	-24.5771	11.98		0.793
TYC 6879-1367-1	289.047	-24.5729	10.09		0.817
TYC 6308-1691-1	289.203	-21.4545	8.62		1.105
TYC 5721-987-1	289.220	-14.3657	10.56		0.925
TYC 6308-928-1	289.517	-21.4907	12.01		0.961
TYC 6296-1864-1	289.520	-16.5857	11.85		0.601
TYC 6296-1864-1	289.525	-16.5923	10.71		0.601
TYC 6296-1912-1	289.582	-16.6510	11.93		0.608
HD 180672	289.591	-23.0129	8.59	M1III	0.925
TYC 6875-1147-1	289.608	-22.9447	10.79		0.927
TYC 6300-719-1	289.612	-17.0137	10.73		0.932
HD 180955	289.832	-21.9431	8.03	K	1.066
N/A	289.833	-28.8424	11.25	n/a	0.593
TYC 6308-1813-1	289.838	-20.8666	10.96		1.224
N/A	289.900	-22.2067	11.99	n/a	1.021
HD 181264	290.173	-23.0554	7.79	K0III	0.865
HD 181340	290.222	-21.5357	8.04	B9V	1.095
TYC 6309-2040-1	290.235	-20.7117	10.97		1.207
HD 181512	290.350	-19.8844	8.46	K1/K	1.323
CD-25 13980	290.617	-25.7143	8.57	K5	0.705
HD 182980	292.032	-22.1535	9.07	F3V	1.014
TYC 6306-2342-1	292.309	-19.5829	11.55		0.931
TYC 5730-584-1	292.790	-12.3978	11.87		0.935
HD 183797	292.986	-18.6559	7.25	M1III	1.026
TYC 6310-549-1	293.347	-22.3506	11.53		0.924
BD-22 5146	293.369	-22.3588	8.69	K0	0.914
TYC 6310-1303-1	293.686	-21.0966	10.82		0.922
HD 184269	293.686	-28.8228	8.98	F3V	0.895

TYC 6307-1972-1	294.301	-20.3107	10.99		0.993
HD 184864	294.306	-20.3758	11.55	A1V	1.078
TYC 5732-2357-1	294.918	-11.2745	11.85		0.977
HD 185591	295.141	-21.2531	7.82	G5V	1.003
N/A	295.383	-28.3093	11.85	n/a	1.064
TYC 6898-1433-1	295.441	-28.1033	11.76		1.058
[LDB2002] 23	296.038	-14.5920	9.39		1.208
CD-28 16117	296.113	-28.2782	10.19	F5	0.925
HD 186734	296.684	-15.0509	8.92	F1V	1.076
HD 187740	298.092	-22.7663	8.64	K3/K4III	1.024
HD 188375	298.882	-21.5086	8.11	K0V	1.001
CD-28 16302	299.579	-28.1609	9.35		0.901
LP 870-61	299.786	-21.9159	10.01		0.888
N/A	3.01893	-6.61421	11.84	n/a	1.355
HD 796	3.06983	-5.93434	9.59	F5	0.911
BD+07 16	3.10428	7.83217	8.03	M0	0.471
HD 931	3.44044	8.19350	7.16	F	1.049
BD+11 18	3.48391	12.1523	9.15	K0	0.826
N/A	3.78342	-7.10706	9.51	n/a	0.964
TYC 1207-745-1	30.2639	16.7665	10.47		0.507
TYC 1210-160-1	30.4125	17.9749	10.75		0.515
CCDM J02025+1924A	30.6313	19.4007	10.53		0.519
HD 191299	302.478	-21.5217	8.87	F3V	0.976
BD-15 5588	303.511	-15.2517	8.85		0.767
HD 192613	304.158	-25.7390	9.33	G3V	0.856
HD 192670	304.198	-20.9334	8.85	F0V	0.965
BD-09 5436	305.166	-9.39717	9.17		0.752
HD 193455	305.255	-18.3239	6.38	K3III	0.808
HD 193544	305.321	-18.5547	8.38	K1III	0.998

CPD-19 7763	306.612	-19.1343	9.36		0.977
HD 194995	307.386	-26.7753	8.84	F8V	0.615
HD 195424	308.029	-26.8698	8.76	M	0.515
HD 196246	309.112	-9.33266	8.25	A0	1.169
HD 197819	311.660	-19.4381	7.42	F0V	0.875
HD 200183	315.618	-17.6019	9.15	F0V	1.306
HD 200901	316.707	-19.7061	8.95	F5V	1.247
HD 201058	316.983	-22.3353	7.66	A9IV/V	0.935
* chi Cap	317.140	-21.1937	5.33	A0V	1.165
N/A	322.120	-18.8733	9.15	n/a	1.952
HD 204609	322.608	-25.1933	6.45	K7III	0.576
N/A	323.237	-17.5640	8.52	n/a	0.743
BD-08 5699	324.399	-7.74090	9.53	K0	0.963
HD 205829	324.558	-14.9100	6.55	K0III	1.489
HD 206648	326.015	-19.9378	8.93	A0V	1.008
HD 207877	328.141	-16.2588	8.92	F	1.567
HD 208068	328.516	-17.1615	8.32	G1/G	0.079
HD 208461	329.118	-7.56636	9.26	F8	1.101
BD+08 344	33.2144	9.45662	9.13	K0	1.321
HD 13802	33.7252	21.0059	7.91	K	0.479
HD 13859	33.7825	3.85125	7.94	F8	0.935
HD 208943	330.006	-16.9113	9.18	K1/K	0.564
HD 210278	332.377	-8.80055	8.46	F	0.969
BD-10 5850	332.544	-9.57415	9.28		0.974
HD 212022	335.350	-6.24313	7.41	A	0.394
HD 212478	336.220	-17.0798	8.95	G6IV	0.815
HD 212994	337.154	-15.6666	8.75	K1/K	1.216
2MASS J22294487-0056318	337.437	-0.942142	9.46		0.948

NLTT 54003	337.506	-12.9151	6.22	F	0.392
HD 214656	339.992	-17.9592	8.93	K0/K1III	1.573
TYC 1217-184-1	34.3931	18.4833	10.31		0.269
N/A	34.4392	18.4550	7.25	n/a	0.983
HD 214863	340.325	-6.24603	9.34	G0	0.975
HD 215708	341.803	-2.70971	7.45	F5	0.842
BD+02 4571	342.883	3.70679	8.57		0.821
HD 216344	343.089	-12.5521	8.94	G8III/IV	0.995
TYC 569-430-1	343.427	2.00958	9.52		0.931
HD 216677	343.748	-14.0430	8.99	G8IV	0.504
HD 217409	345.156	-2.47150	9.21	G0	0.336
HD 220079	350.256	-8.50160	9.26	K0	0.903
BD-00 4565	356.577	0.565467	9.93	F5	0.325
TYC 1174-815-1	359.985	11.2559	9.53	G5	0.929
V* DT Cet	36.4073	8.72162	8.81	A3	0.495
BD+03 338	36.4576	4.06401	8.78	K0	0.891
BD+16 298	37.7000	17.1910	9.15	K	0.895
TYC 52-174-1	38.2259	7.47964	9.93		1.006
HD 15952	38.6613	24.8928	8.21	F0	0.975
HD 15952	38.6613	24.8928	8.21	F0	0.481
HD 16113	38.9271	13.8173	7.25	K5	0.905
BD+20 436	39.4004	21.4157	8.97	G0	0.563
BD+20 436	39.4004	21.4157	8.97	G0	0.492
HD 16398	39.5859	20.2594	8.54	A	0.366
HD 16398	39.5859	20.2594	8.54	A	0.744
HD 16451	39.7538	24.0516	8.38	G5	0.929
HD 17024	41.1288	20.5875	7.92	K0	0.607
HD 17160	41.4493	20.6927	7.56	A0	0.614
BD+13 454	42.2402	13.9329	8.45	K	0.366
HD 17613	42.4682	8.93962	9.17	F8	0.189

TYC 1785-487-1	42.8007	25.4203	9.07			0.906
BD+10 385B	43.2205	10.6688	9.05	G5		0.368
BD+20 476	43.9247	21.0141	9.44	G0		0.905
BD+19 437	44.5266	20.5908	9.13	G5		0.984
TYC 1228-1399-1	45.4024	18.0891	11.25			0.951
TYC 651-922-1	47.2320	12.1195	10.83			0.965
LTT 11052	48.5200	18.3038	7.75			0.909
BD-03 39	5.27503	-2.81506	10.72			0.595
HD 1823	5.66971	1.82350	8.94			0.479
BD+08 37	5.72818	8.82062	9.73			0.975
HD 1935	5.92282	2.74839	10.49			0.594
N/A	5.95808	-3.06512	9.42			0.925
HD 20671	50.1652	28.8265	8.11			1.696
HD 20845	50.4569	9.23211	8.49			0.904
BD+21 446	50.7971	21.5680	8.87			0.924
BD+24 479	51.0316	25.1143	10.28			0.762
BD+18 475	51.1389	19.0634	8.78			0.925
TYC 656-523-1	52.7759	13.6766	9.39			2.066
TYC 1802-597-1	52.9110	25.7920	10.15			2.411
BD+19 556s	53.7437	19.8134	10.21			2.528
TYC 1810-729-1	54.1239	29.4850	10.45			0.518
* 14 Tau	55.9467	19.6650	5.52			0.932
[SCG94] X 38	56.0011	24.5570	7.95			0.966
TYC 1807-641-1	56.1703	28.0282	10.25			1.229
HD 281193	56.3097	29.7265	9.88			0.787
N/A	56.6451	25.8439	8.51			2.674
HD 23642B	56.8727	24.2883	6.78			1.245
N/A	56.9040	23.6091	7.04			0.909
BD+06 42	6.40146	7.73246	9.06			1.196
LTT 10146	6.47001	6.53100	8.66			1.175

HD 2331	6.80793	5.63744	9.25		1.174
TYC 1272-527-1	64.7974	20.3762	10.76		0.752
HD 27482	65.3136	27.3502	6.53		1.509
HD 283573	65.4798	28.2741	10.41		1.565
HD 284375	66.1308	23.0008	9.43		2.887
HD 286736	66.2581	14.1151	9.18		1.534
BD+30 665B	67.2194	30.3648	7.66		1.835
HD 28592	67.9007	29.9271	8.21		0.821
HD 283677	68.0456	29.0156	8.95		0.826
BD+18 660	69.2191	19.2904	9.45		1.113
2MASS J04392369+1638372	69.8488	16.6436	9.62		0.815
CCDM J00288+0250C	7.20904	2.83879	8.13		0.725
N/A	7.82192	-2.93855	11.97		0.475
HD 2868	7.98855	-6.31650	7.37		0.815
HD 29726	70.3860	17.4129	8.18		1.153
N/A	70.4225	18.3154	11.48		1.725
HD 285977	70.9979	17.9641	8.37		1.096
N/A	74.6671	24.1946	11.08		1.047
TYC 1832-1667-1	74.8964	23.7436	10.92		0.855
V* V1361 Ori	75.7630	14.7657	8.79		1.067
HD 32388	75.8821	16.0946	8.78		1.075
TYC 1281-630-1	75.9279	16.2240	11.58		1.096
HD 282841	76.0328	31.5156	8.32		1.095
BD+30 774	76.1381	30.4041	8.72		1.095
V* V411 Aur	76.1437	30.3804	7.43		1.099
HD 284083	76.7102	26.2676	11.34		1.118
N/A	76.8111	23.0743	10.78		0.828
HD 240723	76.8645	31.1204	8.82		1.135

HD 34547	79.7205	13.5672	7.41		1.685
HD 34683	79.9944	16.7308	7.54		1.702
HD 2879	8.03496	3.46990	9.08		1.203
TYC 1300-1352-1	80.9167	18.4824	9.85		0.989
CCDM J05263+3020AB	81.5782	30.3366	8.68		0.963
TYC 1848-1352-1	82.5717	23.7086	10.46		0.955
TYC 1856-224-1	83.5656	26.8082	9.32		1.065
TYC 1865-1158-1	83.8115	25.0458	10.72		1.005
HD 36976	84.0445	19.1731	8.15		0.992
N/A	84.3989	31.0407	11.57		1.099
HD 247209	86.5505	31.3878	9.34		1.121
TYC 1870-1510-1	86.7822	27.9232	10.55		1.318
TYC 2409-1534-1	87.2415	33.2927	11.02		1.073
HD 39208	88.0659	29.5180	7.07		1.039
TYC 1320-2089-1	88.3830	19.5784	9.79		1.045
TYC 1324-1955-1	88.3960	22.2484	9.94		1.083
TYC 1320-183-1	88.8435	18.7860	11.02		0.985
TYC 2410-2525-1	89.1095	33.5875	10.62		1.053
TYC 1312-1933-1	89.3865	15.6507	11.58		1.105
TYC 1875-2323-1	89.6671	29.5472	9.12		1.665
HD 249827	89.7537	22.7304	9.31		1.051
TYC 1324-2723-1	89.8539	20.6671	9.26		1.105
TYC 10-862-1	9.41963	2.41630	11.05		0.902
TYC 606-425-1	9.57733	12.1818	10.62		0.899
HD 3556	9.64366	6.02176	8.38		0.568
HD 3706	9.94804	-5.26878	9.42		0.831
TYC 1876-242-1	90.0157	29.3332	11.05		0.425
TYC 729-191-1	90.6944	14.7520	10.42		1.087
HD 250990	91.1022	25.6363	10.25		0.365

TYC 1868-2813-1	91.2257	25.2825	10.71		2.245
TYC 2419-506-1	91.4364	31.0346	10.18		1.033
HD 251622	91.6796	24.1339	9.44		1.016
TYC 1321-637-1	91.8171	20.1773	11.85		1.546
TYC 1321-960-1	91.8384	20.2064	10.26		1.067
TYC 742-1812-1	92.5894	13.9466	11.18		0.995
TYC 1318-194-1	92.6965	18.6949	10.02		1.325
HD 253683	93.5317	18.9649	9.37		1.171
TYC 1891-17-1	96.3176	28.2622	11.25		0.969

221.266_−10.6710 — 0.492 days

222.370_−16.4512 — 0.761 days

224.767_−27.3069 — 1.064 days

226.986_−25.0589 — 0.657 days

24.3139_2.58245 — 0.868 days

24.7314_14.7447 — 0.995 days

24.8376_8.17022 — 0.897 days

25.3075_8.47885 — 0.846 days

25.7775_3.47240 — 1.285 days

26.2229_19.8568 — 0.846 days

27.0213_21.6588 — 0.918 days

27.9839_17.0796 — 0.785 days

295.141_−21.2531 — 1.003 days

295.383_−28.3093 — 1.064 days

295.441_−28.1033 — 1.058 days

296.038_−14.5920 — 1.208 days

296.113_−28.2782 — 0.925 days

296.684_−15.0509 — 1.076 days

298.092_−22.7663 — 1.024 days

298.882_−21.5086 — 1.001 days

299.579_−28.1609 — 0.899 days

299.786_−21.9159 — 0.888 days

3.01893_−6.61421 — 1.355 days

3.06983_−5.93434 — 0.911 days