Photometric transit search for planets around cool stars from the Western Italian Alps: the APACHE survey

Ph.D. program Director: Prof. Paolo Camerini

Ph.D. student: Paolo Giacobbe

Thesis Supervisors:
Prof. Mario G. Lattanzi
Prof.essa Francesca Matteucci
Dr. Alessandro Sozzetti

ANNO ACCADEMICO 2012/2013
Photometric transit search for planets around cool stars from the Western Italian Alps: the APACHE survey

Ph.D. program Director:
Prof. Paolo Camerini

Ph.D. student:
Paolo Giacobbe

Thesis Supervisors:
Prof. Mario G. Lattanzi
Professa Francesca Matteucci
Dr. Alessandro Sozzetti

ANNO ACCADEMICO 2012/2013
You find out that life is just a game of inches.

Al Pacino’s *Inch By Inch* speech in ‘Any Given Sunday’
Abstract

Photometric transit search for planets around cool stars from the Western Italian Alps: the APACHE survey

by Paolo GIacobbe

Small-size ground-based telescopes can effectively be used to look for transiting rocky planets around nearby low-mass M stars using the photometric transit method. Since 2008, a consortium of the Astrophysical Observatory of Torino (OATo-INAF) and the Astronomical Observatory of the Autonomous Region of Aosta Valley (OAVdA) have been preparing for the long-term photometric survey APACHE (A PAthway toward the Characterization of Habitable Earths), aimed at finding transiting small-size planets around thousands of nearby early and mid-M dwarfs. APACHE uses an array of five dedicated and identical 40-cm Ritchey-Chretien telescopes and its routine science operations started at the beginning of summer 2012.

Here I present the results of the ‘pilot study’, a year-long photometric monitoring campaign of a sample of 23 nearby dM stars, and of the APACHE survey first year data. In these studies, I set out to (i) demonstrate the sensitivity to $> 2R_\oplus$ transiting planets with periods of up to a few days around our programme stars, through a two-fold approach that combines a characterization of the statistical noise properties of our photometry with the determination of transit detection probabilities via simulations; and (ii), where possible, improves our knowledge of some astrophysical properties (e.g. activity, rotation) of our targets by combining our differential photometric measurements with spectroscopic information from the long-term programme GAPS with the HARPS-N spectrograph on the Telescopio Nazionale Galileo.

Furthermore, cool M dwarfs within a few tens of parsecs from the Sun are becoming the focus of dedicated observational programs in the realm of exoplanet astrophysics that will make use of astrometric measurements. I present numerical simulations to gauge the Gaia potential for precision astrometry of exoplanets orbiting a sample of known dM stars within $\sim 30$ pc from the Sun. I then investigate some aspects of the synergy between the astrometric data expected from the Gaia mission on nearby M dwarfs and the APACHE program.
Contents

Abstract ii

List of Figures vi

List of Tables x

1 Introduction 1
   1.1 A personal prospect .................................................. 1
   1.2 Scientific motivations ................................................. 2
   1.3 Chapter Summaries ..................................................... 6

2 Detection and characterization of extrasolar planets 8
   2.1 The radial velocity method .......................................... 8
   2.2 The transit method .................................................... 10
      2.2.1 Exoplanet atmospheres ........................................... 11
      2.2.1.1 Transmission spectroscopy .................................. 11
      2.2.1.2 Infrared Emission ............................................. 12
      2.2.2 The Rossiter-McLaughlin effect ............................... 12
      2.2.3 Transit Timing Variation (TTV) .................................. 13
   2.3 The astrometric method .............................................. 13
   2.4 The direct imaging method .......................................... 15
   2.5 The gravitational microlensing method ............................ 16

3 The physical properties of extrasolar planets 18
   3.1 Introduction .......................................................... 18
   3.2 A brief overview of observations ................................... 19
      3.2.1 Lessons from our Solar System ................................ 19
      3.2.2 Observed properties of exoplanets ............................. 20
   3.3 Planet formation ..................................................... 21
      3.3.1 Core-accretion model .......................................... 21
      3.3.2 Gravitational instability ...................................... 23
      3.3.3 Core-accretion versus gravitational instability .......... 24
   3.4 Interior structure properties ...................................... 24
      3.4.1 Earth-like to super-Earth planets (<10M⊕) ................. 25
      3.4.2 Neptune-like to super-Jupiter planets (>10M⊕) ............ 25
   3.5 Atmospheric properties ............................................ 26
      3.5.1 Biosignatures .................................................... 26
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.4.3.1</td>
<td>Transit Detection Efficiencies from the raw light curves</td>
<td>106</td>
</tr>
<tr>
<td>6.4.4</td>
<td>APACHE’s Ensemble Sensitivity: results</td>
<td>107</td>
</tr>
<tr>
<td>6.5</td>
<td>Summary and Conclusion</td>
<td>109</td>
</tr>
<tr>
<td>7</td>
<td>Astrometric detection of giant planets around nearby M dwarfs: the</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td>Gaia potential</td>
<td></td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>111</td>
</tr>
<tr>
<td>7.2</td>
<td>Simulation Scheme</td>
<td>113</td>
</tr>
<tr>
<td>7.3</td>
<td>Statistical and Numerical Analysis Tools</td>
<td>115</td>
</tr>
<tr>
<td>7.4</td>
<td>Results</td>
<td>116</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Detection Probabilities</td>
<td>116</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Orbit Determination</td>
<td>118</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Expected Planet Yield</td>
<td>122</td>
</tr>
<tr>
<td>7.4.4</td>
<td>Measuring Transiting Systems Configurations</td>
<td>124</td>
</tr>
<tr>
<td>7.4.5</td>
<td>Predicting Giant Planets’ Location and Brightness</td>
<td>127</td>
</tr>
<tr>
<td>7.5</td>
<td>Summary and Conclusions</td>
<td>130</td>
</tr>
<tr>
<td>8</td>
<td>On the synergies between GAIA and transits survey</td>
<td>134</td>
</tr>
<tr>
<td>8.1</td>
<td>Simulation Scheme</td>
<td>134</td>
</tr>
<tr>
<td>8.2</td>
<td>Statistical and Numerical Analysis Tools</td>
<td>136</td>
</tr>
<tr>
<td>8.3</td>
<td>Predicting the transit epochs</td>
<td>136</td>
</tr>
<tr>
<td>8.4</td>
<td>Summary and Conclusions</td>
<td>142</td>
</tr>
<tr>
<td>9</td>
<td>Summary and future steps</td>
<td>143</td>
</tr>
<tr>
<td>9.1</td>
<td>The APACHE survey</td>
<td>144</td>
</tr>
<tr>
<td>9.2</td>
<td>Astrometric detection of giant planets around nearby M dwarfs: the</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>Gaia potential</td>
<td></td>
</tr>
</tbody>
</table>

**Bibliography** 149
List of Figures

3.1 Mass VS Year of discovery diagram ........................................... 19
3.2 Mass of known extra-solar planets (in Mearth) as a function of orbital
distance (in AU) ................................................................. 21
3.3 Mass-radius diagram for transiting planets ................................. 22

4.1 The distribution of seeing measurements over the whole period of obser-
vation. The small black stars show 7864 individual data points (many of
them overlapping due to the seeing values having been approximated to
the first significant digit), each corresponding to an average seeing value
over a 1 min interval, rounded to 0.1. The large filled circles indicate the
median seeing value for each night ................................. 34
4.2 Photometric errors (RMS) VS instrumental mean magnitude in the field
of the planet-hosting star WASP-3 during a good observing night (and
for an exposure time of 35 s) .................................................. 35
4.3 Distribution of the mean RMS for each target. For each of the four
differential photometry methods, the panels show: the mean intra-night
RMS distribution IN1 for the sample (top), the mean intra-night RMS
distribution IN2 extracted from a full-period light curve for each target
(center), and the full-period RMS distribution FP for the sample (bottom). 42
4.4 Panel a: mean RMS of the unbinned light curves as a function of J mag.
Panel b: white noise term as a function of J mag. Panel c: red noise term
as a function of J mag. Panel d: the total error term for the binned light
curves as a function of J mag. Stars indicated by asterisks framed by a
square are objects showing clear signs of activity (see § 4.3.6 and § 4.3.5) 47
4.5 The behaviour of $\sigma_t$ and $\sigma_w$ vs number of points in time bins ....... 48
4.6 Left: Red noise (top), white noise (middle), and total RMS (bottom)
for the sample as a function of the difference ($V_r-K_r$-$V-K$) between
the colour index of the target and the average color index of the corresponding
comparison stars. Right: The same quantity is plotted as a function of
the average number of comparison stars for each red dwarf in the sample. 49
4.7 The search for periodicities in the full set of the photometric data for
the dM star LHS 3445 revealed the existence of a signal with period
$P=0.4752812$ days (corresponding to the absolute minimum in this peri-
odogram). ................................................................. 51
4.8 Photometric data for the dM star LHS 3445 folded according to its prob-
able rotation period $P=0.4752812$ days. Overplotted is the best-fit single-
spot model described in Section 4.3.5 ..................................... 52
4.9 Cross-correlation function (CCF) peaks as a function of the projected rotational velocity $v \sin(i)$ for two of our dM targets, LHS 3445 and GJ 1167A, and other M dwarfs used as a template (GJ 402) and check stars used to validate the procedure. The maximum values of the CCF peak functions correspond to the projected rotation velocities measured from the spectra taken with the Keck/HIRES echelle spectrograph. LHS 3445 and GJ 1167A appear to be rapid rotators. .......................... 53

4.10 The search for periodicities in the full set of the photometric data for the dM star GJ 1167A revealed the existence of a signal with period $P=0.2151997$ days (corresponding to the absolute minimum in this periodogram). ............................... 54

4.11 Photometric data for the dM star GJ 1167A folded according to its probable rotation period $P=0.2151997$ days. Overplotted is the best-fit single-spot model described in Section 4.3.5 .................................................. 55

4.12 Four flares seen in the LHS 3445 (upper and lower left panels) and LHS 2686 (lower right panel) light curves. In each case, the solid line shows the fit of Equation 4.4 to the light curve. For LHS 3445, two almost consecutives flares were recorded, with approximately equal decay times of $\sim 4.5$ min. .................................................. 60


4.15 Summary of simulation output for LHS 1976 (see text for details). Panels $a$: number of signals vs period; Panels $b$: detection probability vs period; Panels $c$: SDE distributions; Panels $d$: distributions of transit depths of the detected signals. .................................................. 65

4.16 Summary of simulation output for LHS 417 (see text for details). Panels $a$: number of signals vs period; Panels $b$: detection probability vs period; Panels $c$: SDE distributions; Panels $d$: distributions of transit depths of the detected signals. .................................................. 66

4.17 Summary of simulation output for LHS 3343 (see text for details). Panels $a$: number of signals vs period; Panels $b$: detection probability vs period; Panels $c$: SDE distributions; Panels $d$: distributions of transit depths of the detected signals. .................................................. 67

4.18 Average detection probability as a function of average phase coverage over the period range 0.5-5.0 days. Different symbols are used to show the trend of detection probability for the four regimes of transit depth simulated ($t_d = 2\%$: diamonds; $t_d = 1.5\%$: triangles; $t_d = 1.0\%$: squares; $t_d = 0.5\%$: crosses). .................................................. 69

4.19 Average detection probability as a function of average phase coverage for $P < 1.0$ days. All symbols are as described in Figure 4.18. ......................... 70

5.1 Detection probability VS Temporal Sampling. Different symbols are used to show the trend of detection probability for the four regimes of transit depth simulated ($t_d = 2.0\%$: diamonds; $t_d = 1.5\%$: triangles; $t_d = 1.0\%$: squares; $t_d = 0.5\%$: crosses). .................................................. 77
5.2 $T_{\text{eff}}C/T_{\text{eff}}L$ ratio distribution where $T_{\text{eff}}C$ is the effective temperature derived from [46] while $T_{\text{eff}}L$ is the effective temperature derived in [139].

5.3 Up: radius distribution of the AIC. Down: mass distribution of the AIC.

6.1 Top left: distribution of number of frames for each target. Top right: number of observing nights distribution for each target. Bottom left: time span distribution for each target. Bottom right: phase coverage distribution for each target.

6.2 Distribution of the RMS for each target. The continuous line shows the filtered full period RMS distribution, while the dotted line shows the photometric uncertainty distribution as predicted by our error model.


6.4 SDE distribution of a sub-sample of observed target with phase coverage $> 10%$.

6.5 folded1551

6.6 folded2961

6.7 folded3276

6.8 flares

6.9 Top: synthetic folded light curve with $t_d = 1.5\%$ and a period of 0.984718 days. The times of observation are those of the APACHE’s target 1820. Center: the actual phased light curve of the APACHE’s target 1820. Bottom: the combined light curve for the APACHE’s target 1820.

6.10 APACHE’s ensemble sensitivity as a function of planet radius and orbital period. These sensitivity curves include both the geometrical probability that a planet would transit and the probability that we would detect transits that do occur. The window function (solid line) shows how sensitive APACHE would be if all that was required to detect a planet were at least three in-transit points from at least three transit events.

6.11 Planet occurrence upper limits as a function of planet radius and orbital period.

7.1 Top left: distance distribution of the LSPM M dwarf sub-sample. Top right: The corresponding magnitude distribution in Gaia’s G band. Bottom left: the derived stellar mass distribution. Bottom right: number of individual Gaia field transits for the sample.

7.2 Gaia single-transit astrometric uncertainty as a function of $G$-band magnitude, including the effects of the adopted gate scheme.

7.3 Left: Gaia giant planet detection efficiency around a sample of 3150 M dwarfs from Lépine (2005) as a function of distance from the Sun (data are binned in 5 pc bins). Right: The same quantity expressed as a function of primary mass (data are binned in 0.1 $M_{\odot}$ bins).

7.4 The dependence of giant planet detection efficiency with Gaia astrometry as a function of distance and primary mass (lines of different styles) based on M dwarfs starcounts out to 100 pc from the Sun derived using the Besancon stellar population synthesis model.
7.5 Fractional error on the orbital period $P$ as a function of $P$ itself. Solid line: the full LSPM sub-sample. Dashed-dotted line: stars $d < 15$ pc. Dashed line: stars with $G < 12$ mag. ........................................ 119

7.6 Top: Fractional error on planetary mass as a function of orbital period. Bottom: the same, but as a function of distance from the Sun. In both cases, the large black dots represent the median value in a period or distance bin, respectively. ........................................ 120

7.7 Fractional error on planetary mass as a function of astrometric signal-to-noise ratio (all stars included). Black dots represent the median values in each $\varsigma/\sigma_{AL}$ bin. .................................................. 121

7.8 Fractional error on the inclination angle $i$ as a function of $i$ itself. Error bars correspond to the standard deviation in each inclination bin. ............ 125

7.9 Left: Degradation rate in the prediction for the value of orbital separation of the planet as a function of its orbital period. The results are averages over each 0.5 yr period bin. Right: The same, but for the position angle. In both panels, symbols are as in Figure 7.8. ........................................ 126

7.10 Left: Errors on the planetary phase function (assuming a Lambert sphere) as a function of the errors on the reconstructed planetary phase angle for the LSPM sub-sample. Each dot represents one target star, and the values correspond to differences taken between orbit-averaged derived and true quantities. Right: Absolute errors $|\Delta \Phi(\lambda)|$ as a function of $\lambda$. The values correspond to the median of all orbital configurations in 10-deg bins in $\lambda$. 128

7.11 Top: Predicted planetary emergent flux as a function of orbital semi-major axis from the M dwarf primary. Logarithmic error bars on $F_p/F_\star$ are phase-averaged values in each 0.05 AU bin obtained through error propagation of the uncertainties on $a_p$ and $\Phi(\lambda)$, assuming an error on the planetary radius ($R_p = 1 R_J$) of 5%. Theoretical values based on a Lambert sphere model are shown as a dashed-dotted line. Bottom: The ratio of $(F_p/F_\star)_O$, derived from the simulated data, to $(F_p/F_\star)_C$, computed from the model, as a function of orbital semi-major axis. .... 131

8.1 Top left: distance distribution of the dM stars sample. Top right: The corresponding magnitude distribution in Gaia’s $G$ band. Bottom left: the derived stellar mass distribution. Bottom right: number of individual Gaia field transits for the sample. ........................................ 135

8.2 Acceptance Rate vs Simulated Period $P$ in years. ........................................ 137

8.3 Fractional error on $P$ vs Simulated Period $P$ ........................................ 138

8.4 Fractional Error on $i$ VS Simulated Period $P$ ........................................ 139

8.5 Distribution of the differences (in days) between the transit epoch ($T_E$) as derived by the fitted orbital parameters and the simulated ("true") transit epoch $\Delta T_E = |T_{E_{\text{fitted}}} - T_{E_{\text{true}}}|$Acceptance Rate VS Simulated Period $P$ ........................................ 140

8.6 Top: $\Delta T_E$ VS $\tau_{\text{fitted}} - \tau_{\text{true}}$ Bottom: $\Delta T_E$ vs $(P_{\text{fitted}} - P_{\text{true}})/P_{\text{true}}$ .... 141
List of Tables

4.1 Summary of the main characteristics of the telescope and camera systems. 37
4.2 Log of observations for the target sample. 38
4.3 Characteristics of the M dwarf sample. 40
4.4 K-S test results (1). 43
4.5 K-S test results (2). 43
4.6 K-S test results (3). 43
4.7 Results of the star spot model applied to the fast rotators LHS 3445 and GJ 1167A. 57

5.1 Summary of the main characteristics of the telescope and camera systems. 74

7.1 Summary of relevant data on known Doppler-detected giant planets orbiting $M_* < 0.6 M_\odot$ dwarfs in the solar neighborhood. 124
Chapter 1

Introduction

1.1 A personal prospect

If one day one were to ask her/his friends what an astronomer does for a living, some will say: she/he observes the sky from remote places with expensive and big instruments. In the words of William Romanishin in his *An Introduction to Astronomical Photometry Using CCDs*: “They play with large telescopes: expensive toys for good boys and girls”. This is a fun definition but it does hit on target. Nowadays, the progress in scientific knowledge often takes advantage of large economic and technological investments. At the present time, the media, the public opinion and a big part of the scientific community consider this a golden rule. The costs and the dimensions of equipment are proportional to the results that are sought for. The story of the APACHE project tells us that things don’t always go this way. The ingredients of the APACHE project, the very core of my PhD thesis work, are cheap and simple: Take the technique of planetary transits, a simple and brilliant method of research, add a focused application for it, a bunch of little red stars in the solar neighbourhood, and finally choose off-the-shelf instrumentation and the dish is ready to be served.

One risks being dumbfounded when one discovers that the APACHE project could have a role in exoplanets science and ultimately it could help answering a question that sooner or later everyone ends up asking: “Are we alone in the universe or (better) in the Milky Way or (best) in the solar neighbourhood?” But the efficacy of this project does not only reside in its smartness. Its ‘human’ dimension (few people for a lot of work) cannot be underestimated. Here every person counts for 110%. Maybe this is the most fascinating aspect of this project, and this is certainly one of the characteristics that I have appreciated the most during the years spent as PhD student.
Working for APACHE has made me feel like a pioneer. I have seen through it all, from the simplest screws of the telescope mounts to the most abstract mathematical algorithms. But especially I have learnt how to make a choice at my own risk, and for this I have to thank all the collaborators and friends who have put me in a position to do so. I have understood that to reach a goal you have to proceed inch by inch and these inches must be gained at every step and in any direction, whether it’s learning how to deal with the screws of the instruments or with the physics of the problem. What makes the difference is earning each and every inch without prejudices, always with the curiosity of the pioneer. I know that I have to walk through many roads to become an astronomer, I have to study more math and physics and techniques. Through the APACHE ‘gymnasium’ I have learnt the fundamentals and I have started to build the man that I will become. I known not on the exoplanet nearest to us, but here on Earth, the most important thing is still what we are and what we will be.

1.2 Scientific motivations

M dwarf stars, with masses $M_\star \leq 0.6 \, M_\odot$, make up the vast majority of the reservoir of nearby stars within $\sim 25 - 30$ pc. These stars have not traditionally been included in large numbers in the target lists of radial-velocity (RV) searches for planets for two main reasons: 1) their intrinsic faintness, which prevented Doppler surveys in the optical from achieving very high radial-velocity precision ($< 5 - 10$ m/s) for large samples of M dwarfs (e.g., Eggenberger & Udry 2010 [1], and references therein), and 2) their being considered as providers of very inhospitable environments for potentially habitable planets (e.g., Tarter et al. 2007 [2]; Scalo et al. 2007 [3], and references therein). These two paradigms are now shifting. First, the application of the transit technique to M dwarfs presents several exciting opportunities, and the advantages are especially compelling for the detection of transiting habitable, rocky planets. These include, for example, improved observing windows due to the short periods of potential planets in the stellar habitable zone (the range of distances from a given star for which water could be found in liquid form on a planetary surface. E.g., Kasting et al. 1993 [4]), or the possibility to reach detection of rocky planets due to the small radii of M dwarfs, leading to deep transits ($\Delta \text{mag} \sim 0.005$ mag) easily detectable from the ground with modest-size telescopes ($30 - 50$ cm class), and readily confirmable with present-day precision RV measurements (owing to their moderately large RV amplitudes, on the order of 5-10 m/s). Second, while not all concerns about their habitability have been resolved yet, there has been a recent change in view for planets orbiting low-mass M stars, now often considered as potentially hospitable worlds for life and its remote detection (e.g., Seager & Deming 2010 [5]; Barnes et al. 2011 [6], and references therein).
Chapter 1. Introduction

Advancements in our knowledge of the complex processes of planet formation and evolution cannot be achieved without a detailed understanding of the role of the central star (through its properties such as mass and metal abundance) and its environment (the circumstellar disk within which the planetary population must form). For example, the theoretical expectations (within the framework of the standard core accretion model) that giant planet frequency and upper mass limits ought to be direct functions of stellar mass $M_*$ and metallicity $[\text{Fe/H}]$ (e.g., Laughlin et al. 2004 [7]; Ida & Lin 2004 [8], 2005 [9]; Kennedy & Kenyon 2008 [10]; Mordasini et al. 2009 [11]) have so far been confirmed on relatively firm statistical grounds only for stars (mid-F through mid-K type) with masses close to that of the Sun (Santos et al. 2004 [12]; Fischer & Valenti 2005 [13]; Johnson et al. 2007 [14]; Sozzetti et al. 2009 [15]), while results for stars with masses significantly different from that of the Sun still rely on small-number statistics (e.g., Endl et al. 2006 [16]; Johnson et al. 2012 [17]). Similarly, the statistical significance of the early evidence for a relatively high frequency of low-mass planets (Neptunes and super-Earths) around low-mass stars (e.g., Forveille et al. 2011[18], and references therein) is still hampered by the observational bias intrinsic to long-term RV surveys (only a few hundred objects monitored), and the recent, compelling evidence from Kepler photometry (e.g., Howard et al. 2011[19]) of increasing occurrence rates for small-radius, short-period planets around increasingly cooler stars still suffers from small-numbers statistics at the latest spectral types (only a few hundred of relatively bright M0-M1 dwarfs being included in the Kepler catalogue). Finally, the anticipated wild diversity of the structural and atmospheric properties of super-Earths (Seager & Deming 2010 [5], and references therein) can be most easily investigated using a sample of such planets observed as transiting companions to nearby M dwarf primaries, given that for low-mass stars the planet-to-star flux ratio is much larger than that for the Earth-Sun system 1, thus spectral characterization of the planet via, e.g., occultation spectroscopy is much more readily attainable.

These considerations have brought about renewed efforts to monitor photometrically as well as spectroscopically large samples of nearby cool dwarfs. The first spectacular success of the dedicated MEarth transit search for rocky planets around 2000 late M dwarfs was announced by Charbonneau et al. 2009 [20], with the detection of the low-density transiting super-Earth GJ 1214b ($M_p = 6.5 \, M_\oplus$, $R_p = 2.7 \, R_\oplus$) around a nearby M4.5 dwarf. The primary in this system is bright enough to enable the detailed spectroscopic characterization of the planet’s thick atmosphere over a broad wavelength range (Bean et al. 2011 [21]; Croll et al. 2011 [22]; Crossfield et al. 2011 [23]).

---

1For example, in the Rayleigh-Jeans limit, this flux ratio depends on the relative surface areas and brightness temperatures of the planet and star. For a 2-$M_\oplus$ super-Earth, this ratio is in the range 0.01%-011% for a mid- to late-M dwarf primary (M4V-M8V), compared to 0.00044% for the Earth-Sun system.
recent constraints on GJ 1214b’s atmospheric composition are not only essential for breaking the degeneracy between the mass, radius and composition of both the interior and a possible atmosphere in theoretical models of super-Earths (Adams et al. 2008 [24]; Rogers & Seager 2010 [25]; Miller-Ricci & Fortney 2010 [26]; Désert et al. 2011 [27]; Nettelmann et al. 2011 [28]), but they also constitute a remarkable test of planetary evolution models in a mass range (for both the primary and the planet!) not seen in our Solar System. Very recently, the M2K Doppler search for close-in planets around 1600 nearby M and K dwarfs has also started producing its first results (e.g., Apps et al. 2010 [29]). Decade-long Doppler monitoring has also allowed to detect the first Saturn-mass planet in the habitable zone of a nearby mid-M dwarf (Haghighipour et al. 2010 [30]). The early-M dwarf GJ 581, already hosting a system of four low-mass (Neptunes and super-Earths) planets, is currently the focus of a hot debate on the actual existence of a fifth planet with the mass of a super-Earth orbiting right in the middle of the habitable zone (Vogt et al. 2010 [31]; Tuomi 2011 [32]; Pepe et al. 2011 [33]; Gregory 2011 [34]).

There is a growing consensus among the astronomers’ community that the first habitable rocky planet will be discovered (and might have been discovered already!) around a red M dwarf in the backyard of our Solar System.

However, not all physical properties of low-mass stars are known precisely enough for the purpose of the detection and characterization of small-radius planets. Worse still, some of the characteristics intrinsic to late-type dwarfs can constitute a significant source of confusion in the interpretation in planet detection and characterization measurements across a range of techniques. First of all, there exist discrepancies between theory and observations in the determination of the sizes of M dwarfs, typically on the order of 10%-15% (Ribas 2006 [35]; Beatty et al. 2007 [36]; Charbonneau et al. 2009 [20], and references therein). It has been suggested that this problems might be stemming from the lack of a detailed treatment of the effects of non-zero magnetic fields on the properties of low-mass, fully convective stars (Ribas 2006 [35]; López-Morales 2007 [37]; Torres et al. 2010 [38], and references therein). As a result, the inferred composition of a transiting planet detected around an M dwarf might be subject to rather large uncertainties, particularly when it comes to super-Earths, for which, as mentioned above, degeneracies in the models of their physical structure indicate a wide range of possible compositions for similar masses and radii (Seager & Deming 2010 [5], and references therein). Indeed, for the two known transiting planets around M dwarfs, GJ 436b and GJ 1214b, uncertainties in the planetary parameters are dominated by the limits in the knowledge of the stellar parameters.

Second, there are at present difficulties in spectroscopically determining with a high
Chapter 1. Introduction

degree of precision M dwarf metallicities\(^2\) (Bean et al. 2006 [41]; Woolf & Wallerstein 2006 [42]; Woolf et al. 2009 [43]; Rojas-Ayala et al. 2010 [44]), which are only partially mitigated by recent attempts at deriving photometric calibrations (Bonfils et al. 2005 [45]; Casagrande et al. 2008 [46]; Johnson & Apps 2009 [47]; Schlaufman & Laughlin 2010 [48]). In addition, studies of the rotation-activity relation\(^3\) for M dwarfs using large stellar samples are limited to young and active stellar samples (e.g., Shkolnik et al. 2009 [52]; López-Santiago et al. 2010 [53], and references therein), often in young open clusters (e.g., Meibom et al. 2009 [54]; Hartman et al. 2009 [55], and references therein), while our understanding of the rotation-activity connection for M dwarfs with age greater than \(t \sim 0.5\) Gyr (e.g., Pizzolato et al. 2003 [56]; Reiners 2007 [57]; Jenkins et al. 2009 [58]; West & Basri 2009 [59]) is still subject to rather large uncertainties due to the sparseness of the data. All these issues hamper at present the possibility of determining precisely the ages of (particularly mid- and late-) M dwarfs in the field, and this in turn has a significant impact on the calibration of the fundamental evolutionary properties of the planets they might be hosting.

Third, as measurements of chromospheric activity indicators (H\(\alpha\) line) have shown how the fraction of active M dwarfs increases as a function of spectral sub-type (e.g., Bochanski et al. 2005 [60]; West et al. 2011 [40]), activity-related phenomena such as stellar spots, plages, and flares become increasingly a matter of concern for planet detection and characterization programs targeting late-type stars. Stellar surface inhomogeneities can hamper the detection, and sometimes even mimic the signal, of exoplanets (e.g., Queloz et al. 2001 [61]), and seriously complicate the characterization of their properties. This problem has already become acute in the case of active solar analogues hosting transiting planets. An illustrative example is provided by the ongoing debate on the actual mass of CoRoT-7b, varying (including 1-\(\sigma\) uncertainties as large as 20\%) between 1 \(M_\oplus\) and 9 \(M_\oplus\) (!), depending on how one decides to deal with the modelling of the planetary signal superposed to the much larger activity-induced stellar ‘jitter’ in both the photometric and the radial-velocity measurements (Queloz et al. 2009 [62]; Hatzes et al. 2010 [63]; Pont et al. 2011 [64]; Ferraz-Mello et al. 2011 [65]). Recently, the first serious studies attempting to gauge the limits to planet detection induced by stellar activity-related phenomena, and strategies aiming at minimizing such effects,

\(^2\)M-dwarf spectra are dominated by chemically complex molecular features. As a result, the identification of the continuum in an M dwarf spectrum is challenging, rendering line-based metallicity indicators unreliable. The poorly constrained molecular opacity data currently available make the determination of metallicity through spectral synthesis also difficult (e.g., Gustafsson 1989 [39]; West et al. 2011[40], and references therein).

\(^3\)The connection between stellar rotation and activity is usually investigated by means of 1) spectroscopic measurements of the rotational velocity \(v\sin i\), usually coupled to measurements of the H\(\alpha\) luminosity (e.g., Reiners & Basri 2010 [49], and references therein), 2) spectroscopic monitoring of temporal evolution of the \(R^*_HK\) activity index as determined from the Ca II H & K emission line cores (e.g., Wright et al. 2004 [50]), and 3) photometric determination of rotation periods for stars with significant spot coverage (e.g., Strassmeier et al. 2000 [51]).
have been undertaken. These have focused primarily on the impact of, and possibility of calibrating out, activity-induced jitter in high-precision radial-velocity and astrometric measurements (Makarov et al. 2009 [66]; Lagrange et al. 2010 [67]; Boisse et al. 2011 [68], and references therein; Dumusque et al. 2011 [69]; Sozzetti 2011 [70], and references therein). Very recently, the first analyses of the impact of starspots on radial-velocity searches for earth-mass planets in orbit about M dwarf stars have been carried out by Reiners et al. 2010 [49] and Barnes et al. 2011 [6], who also addressed the merit of moving from the optical to infrared wavelengths (where the starspots-induced RV noise might be significantly reduced).

All the above considerations clearly underline how achieving the goal of the detection and characterization of low-mass, potentially habitable, rocky planets around low-mass stars requires the construction of a large (all-sky) sample of nearby, relatively bright M dwarfs with well-characterized properties. This will necessitate the combined use of time-series of spectroscopic, astrometric, and photometric data of high quality. In particular, the jitter levels will have to be quantified in detail for each target individually, as the jitter properties may vary from star to star within the same spectral class, as suggested by recent findings based on high-precision Kepler photometry (e.g., Ciardi et al. 2011 [71]) and high-resolution, high-S/N spectroscopy (e.g., Zechmeister et al. 2009 [72]).

1.3 Chapter Summaries

In Chapter 2, I present a review of the detection and characterization techniques of extrasolar planets with particular attention to the transit method, radial velocity method and astrometry method.

In Chapter 3, I present a review of the status of extrasolar planet searches. I summarize the main statistical properties of extrasolar planets, and in particular correlations among orbital parameters and masses, and between planet properties and the characteristics of the host stars. Finally, I briefly review recent theoretical efforts to bring predictions from models of planet physical properties, planet formation, migration, and evolution in better agreement with observations.

In Chapter 4, I highlight the big opportunities related to the M dwarf targets and I summarize the milestones of four-year long efforts undertaken to build the final setup of the photometric transit survey APACHE, officially started in July 2012 at the site of the OAVdA.

In Chapter 5, I describe the architecture of the APACHE survey.
In Chapter 6, I present data and results from the first year of observations of APACHE.

In Chapter 7, I gauge the Gaia potential for precision astrometry of exoplanets orbiting an actual sample of thousands of known dM stars. This study highlights the capability of high-precision astrometry to reconstruct the underlying orbital element distributions and occurrence rates of the planetary companions and it helps in evaluating the expected Gaia recovery rate of actual planet populations around late-type stars. Finally, I investigate some elements of the synergy between the Gaia data on nearby M dwarfs and other ground-based and space-borne programs for planet detection and characterization.

In Chapter 8, I better quantify the relevance of the Gaia observations of the large sample of nearby M dwarfs in a synergistic effort to optimize the planning and interpretation of follow-up/characterization measurements of the discovered systems by means of transit photometry (e.g., APACHE) and simultaneous multi-wavelength spectroscopy (e.g., EChO, JWST).

In Chapter 9, I summarize the results of the thesis and I show some future prospects arising by the described work.
Chapter 2

Detection and characterization of extrasolar planets

Any planet is an extremely faint light source compared to its parent star. In addition to the intrinsic difficulty of detecting such a faint light source, the light from the parent star causes a glare that washes it out. For those reasons, fewer than 5% of the extrasolar planets known as of DECEMBER 2013 have been observed directly.

Instead, astronomers have generally had to resort to indirect methods to detect extrasolar planets. At the present time, more than 90% of the 1040 known planets were detected with indirect method. Many of these methods, often combined together, not only allow the detection but also the characterization of a planetary system. For example, in the best cases, we can interpret the measurements in terms of bulk or atmosphere composition.

In this Chapter, I describe these techniques of detection and characterization.

2.1 The radial velocity method

A star with a planet will move in its own small orbit in response to the planet’s gravity. This leads to variations in the speed with which the star moves toward or away from Earth, i.e. the variations are in the radial velocity of the star with respect to Earth. The radial velocity can be deduced from the displacement in the parent star’s spectral lines due to the Doppler effect. The radial-velocity method measures these variations in order to confirm the presence of the planet.
The velocity of the star around the system’s center of mass is much smaller than that of the planet, because the radius of its orbit around the center of mass is so small. However, velocity variations down to $1\text{m/s}$ or even somewhat less can be detected with modern spectrometers, such as the HARPS-N (High Accuracy Radial Velocity Planet Searcher, Cosentino et al., 2012 [73]) spectrometer at the TNG (Telescopio Nazionale Galileo) 3.58 meter optical telescope located in the Island of San Miguel de La Palma, Spain. The spectrograph HARPS-N is based on the design of its predecessor (HARPS-S) working at La Silla ESO’s 3.6m telescope, achieving unprecedented results on radial velocity measurements of extrasolar planetary systems. Thanks to an ultra-stable environment that utilizes a temperature-controlled vacuum chamber, these spectrographs allow measurements under $1\text{m/s}$ which will enable the characterization of rocky, Earth-like planets.

The method is distance independent, but requires high signal-to-noise ratios to achieve high precision, and so is generally only used for relatively nearby stars out to about 160 light-years from Earth to find lower-mass planets. Jovian mass planets can be detectable around stars several hundred light years away. It easily finds massive planets that are close to stars. Modern spectrographs can also easily detect Jupiter-mass planets orbiting 10 astronomical units away from the parent star but detection of those planets requires many years of observation.

With the measurement of radial velocity and considering the Star-Planet system like a two bodies problem, we are able to measure the minimum mass of the planet:

$$M_{\text{planet}} \sin i \simeq \left( \frac{P}{2\pi G} \right)^{1/3} K_\star M_\star^{2/3} (1 - e^2)^{1/2}$$

where $M_{\text{planet}}$ is the mass of the planet ($M_{\text{planet}} << M_\star$), $i$ the inclination of the orbit respect to the line of sight, $P$ the orbital Period, $G$ the gravitational constant, $K_\star$ the radial velocity’s amplitude of star, $M_\star$ the mass of the star and $e$ the eccentricity. From Equation 2.1 it’s clear that planets with orbits highly inclined to the line of sight from Earth produce smaller wobbles, and are thus more difficult to detect. Furthermore, one of the advantages of radial velocity method is that eccentricity of the planet’s orbit can be measured directly.

The radial-velocity method is fundamental to confirm detections made by using the transit method (see Section 2.2). Furthermore, when both methods are used in combination, then the planet’s true mass can be estimated.
2.2 The transit method

If a planet crosses (transits) in front of its parent star’s disk, then the observed visual brightness of the star drops a small amount. The amount the star dims depends on the relative sizes of the star and the planet, accordingly to:

$$\Delta F = \frac{F_{out} - F_{in}}{F_{out}} = \left( \frac{R_{\text{planet}}}{R_*} \right)$$

(2.2)

where $F_{out}$ is the star flux, $F_{in}$ is the star flux when the planet transit, $R_{\text{planet}}$ is the planet radius and $R_*$ is the star radius. Seeing the Equation 2.2, it is clearly that the first output of this technique is planet radius star radius ratio. Consequently, the planet radius $R_{\text{planet}}$ can be obtained providing a reasonable estimate of the stellar radius $R_*$. Further, if we consider the total transit duration for a circular orbit

$$d \simeq \frac{PR_*}{a} \sqrt{\left(1 + \frac{R_{\text{planet}}}{R_*}\right)^2 \left(\frac{a}{R_* \cos i}\right)^2}$$

(2.3)

where $d$ is the transit duration, $P$ is the orbital period, $a$ the semi-major axis of the orbit and $i$ the inclination of the orbit respect to the line of sight, inverting the equation it is possible to derive inclination of the orbit $i$. The accurate measure of the orbital inclination, $i$, allowing us to evaluate the planetary mass $M_{\text{planet}}$ directly from the minimum mass value $M_{\text{planet}} \sin i$ determined from radial-velocity observations (see Equation 2.1) and an estimate of the stellar mass, $M_*$. The present knowledge of the stellar physics allows an estimate of the stellar radius and stellar mass with an uncertainty of 10% (e.g. Ribas et al., 2006 [35]).

This method has two major disadvantages. First of all, planetary transits are only observable for planets whose orbits happen to be barely perfect aligned from the astronomers’ vantage point. The probability of a planetary orbital plane being directly on the line-of-sight to a star is, for circular orbit, the ratio of the diameter of the star to the diameter of the orbit $Probability \simeq \frac{R_*}{a}$. For example, if we consider a earth like planet orbiting a sun-sized star at 1 AU, the probability of a random alignment producing a transit is 0.47%. However, by scanning large areas of the sky containing thousands or even hundreds of thousands of stars at once, transit surveys can in principle find extrasolar planets at a rate that could potentially exceed that of the radial-velocity method. Several surveys have taken that approach, such as the ground-based SuperWASP Project ([74]) and HATNet ([75]) and the space-based COROT ([76]) and Kepler missions ([77]). Secondly, the method suffers from a high rate of false detections. A 2012 study found that the rate of false positives for transits observed by the Kepler
mission could be as high as 35% ([78]). For this reason, a transit detection requires additional confirmation, typically from the radial-velocity method (see Section 2.1) or transit timing (see Section 2.2.3) variation method. Radial velocity method is especially necessary for Jupiter-sized or larger planets as objects of that size encompass not only planets, but also brown dwarfs and even small stars.

The main advantage of the transit method is that the size of the planet can be determined from the light-curve. When combined with the radial-velocity method (which determines the minimum planet mass’) one can determine the density of the planet, and hence learn something about the planet’s physical structure. Furthermore, when a planet transiting its parent star become possible very detailed follow-up studies, described in the next sub-sections. The transiting planets are by far the best-characterized of all known exoplanets.

2.2.1 Exoplanet atmospheres

When a planet transit its parent star become it possible to study the atmosphere of the planet. The eclipsing geometry of transiting systems permits the spectrum of the planet and star to be disentangled through monitoring of the variation in the combined system light as a function of the known orbital phase. Detections and meaningful upper limits about the identity of their chemical constituents, the presence (or absence) of clouds, the fraction of incident radiation that is absorbed (and hence the energy budget of the atmosphere), and the ability of winds and weather patterns to redistribute heat from the dayside to the nightside, have been achieved using the following two techniques. A summary of the main results in the past decade is performed in the Chapter 3.

2.2.1.1 Transmission spectroscopy

The technique of transmission spectroscopy seeks to ratio stellar spectra gathered during transit with those taken just before or after this time, the latter providing a measurement of the spectrum of the isolated star. Wavelength-dependent sources of opacity in the upper portions of the planetary atmosphere, or in its exosphere, will impose absorption features that could be revealed in this ratio. This technique can be viewed as probing the wavelength-dependent variations in the inferred value of $R_{\text{planet}}$. 
2.2.1.2 Infrared Emission

At infrared wavelengths, the secondary eclipse (i.e. the decrement in the system flux due to the passage of the planet behind the star) permits a determination of the planet-to-star brightness ratio. Since the underlying stellar spectrum may be reliably assumed from stellar models (e.g. Kurucz, 1992), such estimates afford the first direct constraints on the emitted spectra of planets orbiting other Sun-like stars. In the Rayleigh-Jeans limit, the ratio of the planetary flux $F_{\text{planet}}$ to that of the star $F_\star$ is

$$\frac{F_{\text{planet}}}{F_\star} = \frac{T_{\text{eq,planet}}}{T_{\text{eff,\star}}} \left( \frac{R_{\text{planet}}}{R_\star} \right)^2,$$

(2.4)

where $T_{\text{eq,planet}}$ is the equilibrium temperature of the planet, $T_{\text{eff,\star}}$ is the effective temperature of the star and $(R_{\text{planet}}/R_\star)^2$ is simply the transit depth.

As a significant example, Charbonneau et al. (2005) and Deming et al. (2005a) employed the remarkable sensitivity and stability of the Spitzer Space Telescope to detect the thermal emission from TrES-1 and HD 209458b. These measurements provide estimates of the planetary brightness temperatures, which in turn can be used to estimate (under several assumptions) the value of $T_{\text{eq,planet}}$ of the planets.

2.2.2 The Rossiter-McLaughlin effect

High-resolution stellar spectra obtained during transits can be used to determine the degree of alignment of the planet’s orbital angular momentum vector with the stellar spin axis. As the main star rotates on its axis, one quadrant of its photosphere will be seen to be coming towards the viewer, and the other visible quadrant to be moving away. These motions produce blueshifts and redshifts, respectively, in the star’s spectrum, usually observed as a broadening of the spectral lines. When the secondary star or planet transits the primary, it blocks part of the latter’s disc, preventing some of the shifted light from reaching the observer. This causes the observed mean redshift of the primary star as a whole to vary from its normal value. As the transiting object moves across to the other side of the star’s disc, the redshift anomaly will switch from being negative to being positive, or vice versa. This phenomenon is known as the Rossiter-McLaughlin effect (Rossiter, 1924; McLaughlin, 1924), and has long been observed in the spectra of eclipsing binary stars. Queloz et al. (2000) and Bundy and Marcy (2000) detected this effect during transits of HD 209458. A full analytic treatment of the phenomenon in the context of transiting extrasolar planets has been given by Ohta et al. (2005) [79]. This effect has been used to show that as many as 25% of hot Jupiter
are orbiting in a retrograde direction with respect to their parent stars [80], strongly suggesting that dynamical interactions rather than planetary migration produce these objects (see Chapter 3).

### 2.2.3 Transit Timing Variation (TTV)

Timing variation’ asks whether the transit occurs with strict periodicity or if there’s a variation. When multiple transiting planets are detected, they can often be confirmed with transit-timing variation method. This is useful in planetary systems far away from the Sun where radial velocity methods cannot detect them due to low signal-to-noise ratio. If a planet has been detected by the transit method, then variations in the timing of the transit provide an extremely sensitive method which is capable of detecting additional non-transiting planets in the system with sizes potentially as massive as Earth. It is easier to detect transit-timing variations if planets have relatively close orbits and when at least one of the planets is more massive, causing the orbital period of a less massive planet to be more perturbed.

The main drawback of transit-timing method is that usually not much can be learned about the planet itself but the transit-timing variation can help to determine the maximum mass of the planet (e.g. the case of Kepler-36 [81]).

### 2.3 The astrometric method

The astrometric method consists of precisely measuring a star’s position in the sky and observing how that position changes over time. If the star has a planet, then the gravitational influence of the planet will cause the star itself to move in a tiny circular or elliptical orbit. The main observable (assuming circular orbits) is the ‘astrometric signature’, i.e. the apparent semi-major axis of the stellar orbit scaled by the distance to the observer and the planet-to-star ratio:

\[
\alpha = \left( \frac{M_{\text{planet}}}{M_{\text{Sun}}} \right) \left( \frac{M_{\text{planet}}}{M_\star} \right) \left( \frac{a_{\text{planet}}}{1\text{AU}} \right) \left( \frac{M_{\text{planet}}}{M_\star} \right) \left( \frac{pc}{\text{dist}} \right) \text{arcsec}
\]

(2.5)

However, by reconstructing the orbital motion in the plane of the sky, astrometry alone can determine the entire set of seven orbital elements (for the details see Chapter 7 and Chapter 8), thus breaking the $M_{\text{planet}} \sin i$ degeneracy intrinsic to Doppler measurements and allowing one to derive an actual mass estimate for the companion.
Chapter 2. Observational technique

The state-of-the-art astrometric precision is nowadays set to $\sim 1\,\text{mas}$ by Hipparcos and HST/FGS. By looking at Equation 2.5, one realizes how the magnitude of the perturbation induced by a 1 Jupiter-mass planet in orbit at 5 AU around a 1$-M_{\odot}$ star at 10 pc from the Sun is $\alpha \simeq 500$ micro-arcsec ($\mu$as). For the same distance and primary mass, a ‘hot Jupiter’ with $a_{\text{planet}} = 0.01\,\text{AU}$ induces $\alpha = 1\,\mu$as, and an Earth-like planet ($a_{\text{planet}} = 1\,\text{AU}$) causes a perturbation $\alpha = 0.33\,\mu$as. One then understands why, despite several decades of attempts (e.g., Strand 1943; Reuyl & Holmberg 1943; Lippincott 1960; van de Kamp 1963; Gatewood 1996; Han et al. 2001; Pravdo & Shaklan 2009), and a few recent successes primarily thanks to HST/FGS astrometry (Benedict et al. 2002, 2006, 2010; McArthur et al. 2004, 2010; Bean et al. 2007; Martioli et al. 2010) and recent analyses of the re-reduced Hipparcos intermediate astrometric data (Sozzetti & Desidera 2010; Sahlmann et al. 2011; Reffert & Quirrenbach 2011), astrometric measurements with mas precision have so far proved of limited utility when employed as either follow-up tool or to independently search for planetary mass companions orbiting nearby stars (for a review of the approach to planet detection with astrometry see, for example, Sozzetti 2005, 2010, and references therein). However, an improvement of 2-3 orders of magnitude in achievable measurement precision, down to the $\mu$as level, would allow this technique to achieve in perspective the same successes of the Doppler method, for which the improvement from the $\text{km}\,\text{s}^{-1}$ to the $\text{m}\,\text{s}^{-1}$ precision opened the doors for ground-breaking results in exoplanetary science. Indeed, $\mu$as astrometry is almost coming of age. Provided the demanding technological and calibration requirements to achieve the required level of measurement precision are met (e.g., Sozzetti 2005), future observatories at visible and near-infrared wavelengths, using both monolithic as well as diluted architectures from the ground (with VLTI/PRIMA, e.g., Launhardt et al. 2008) and in space (with Gaia (Casertano et al. 2008) or proposed ultra-high-precision observatories (Malbet et al. 2010) such as NEAT (Malbet et al. 2011)) hold promise for crucial contributions to many aspects of planetary systems astrophysics (formation theories, dynamical evolution, internal structure, detection of Earth-like planets), in combination with data collected with other indirect and direct techniques.

Some of the most important issues for which $\mu$as astrometry will play a key role in the next decade are summarized below. High-precision astrometry, particularly with Gaia, has the potential to significantly refine our understanding of the statistical properties of extrasolar planets, thus helping to crucially test theoretical models of gas giant planet formation and migration. For example, the Gaia unbiased and complete magnitude-limited census of $> 10^5$ M-F-G-K dwarfs screened for new planets out to 200 pc from the Sun will allow us to test the fine structure of giant planet parameters distributions and frequencies, and to investigate their possible changes as a function of stellar mass with unprecedented resolution. In addition, astrometric measurements with Gaia of
thousands of metalpoor stars and hundreds of young stars will instead allow us to probe specific predictions on giant planet formation time-scales and the role of varying metal content in the protoplanetary disk. High-precision astrometry can provide meaningful estimates of the full three-dimensional geometry of any planetary system (without restrictions on the orbital alignment with respect to the line of sight) by measuring the mutual inclination angle between pairs of planetary orbits. Coplanarity tests for hundreds of multiple-planet systems will be carried out with Gaia, and VLTI/PRIMA, and this, in combination with data available from Doppler measurements and transit timing, could allow us to discriminate between various proposed mechanisms for eccentricity excitation, thus significantly improving our comprehension of the role of dynamical interactions in the early as well as long-term evolution of planetary systems. Finally, accurate knowledge of all orbital parameters and actual mass are essential for understanding the thermophysical conditions on a planet and for determining its visibility. High-precision astrometric measurements (with Gaia, and VLTI/PRIMA) could then provide important supplementary data to aid in the understanding of direct detections of wide-separation extrasolar giant planets.

2.4 The direct imaging method

The planets are extremely faint light sources compared to stars and what little light comes from them tends to be lost in the glare from their parent star. So in general, it is very difficult to detect and resolve them directly from their host star. Planets orbiting far enough from stars to be resolved reflect very little starlight so planets are detected through their thermal emission instead. It is easier to obtain images when the star system is relatively near to the Sun, the planet is especially large (considerably larger than Jupiter), widely separated from its parent star, and hot so that it emits intense infrared radiation; the images have then been made at infrared where the planet is brighter than it is at visible wavelengths. Coronagraphs are used to block light from the star while leaving the planet visible.

Direct imaging can give only loose constraints of the planet’s mass which is derived from the age of the star and the temperature of the planet. Mass can vary considerably as planets can form several million years after the star has formed. The cooler the planet is, the less the planet’s mass needs to be. In some cases it is possible to give reasonable constraints to the radius of a planet based on planet’s temperature and apparent brightness. The spectra emitted from planets does not have to be separated from the star which eases determining the chemical composition of planets.
Sometimes observations at multiple wavelengths are needed to rule out the planet being a brown dwarf. Direct imaging can be used to accurately measure the planet’s orbit around the star. Unlike majority of other methods, direct imaging works better with planets with face-on orbits rather than edge-on orbits as face-on orbit is more easily observable during the entirety of the planet’s orbit while planets with edge-on orbits are more easily observable during their largest apparent separation from the parent star.

The planets detected through direct imaging currently fall into two categories. First ones are found around stars more massive than the Sun which are young enough to have protoplanetary disks. The second ones are possible sub-brown dwarfs found around very dim stars or brown dwarfs and are at least 100 AU away from their parent stars.

Planetary-mass objects not gravitationally bound to a star are found through direct imaging as well.

### 2.5 The gravitational microlensing method

Gravitational microlensing occurs when the gravitational field of a star acts like a lens, magnifying the light of a distant background star. This effect occurs only when the two stars are almost exactly aligned. Lensing events are brief, lasting for weeks or days, as the two stars and Earth are all moving relative to each other. More than a thousand such events have been observed over the past ten years.

If the foreground lensing star has a planet, then that planet’s own gravitational field can make a detectable contribution to the lensing effect. Since that requires a highly improbable alignment, a very large number of distant stars must be continuously monitored in order to detect planetary microlensing contributions at a reasonable rate. This method is most fruitful for planets between Earth and the center of the galaxy, as the galactic center provides a large number of background stars.

Unlike most other methods which have detection bias towards planets with small (or for resolved imaging, large) orbits, microlensing method is most sensitive to detecting planets around 1-10 astronomical units away from Sun-like stars.

A notable disadvantage of the method is that the lensing cannot be repeated because the chance alignment never occurs again. Also, the detected planets will tend to be several kiloparsecs away, so follow-up observations with other methods are usually impossible. In addition, the only physical characteristic that can be determined by microlensing are loose constraints of a mass of the planet. Orbital properties also tend to be unclear as the only orbital characteristic that can be directly determined is its current semi-major
axis from the parent star which can be misleading if the planet follows an eccentric orbit. When the planet is far away from its star, it spends only a tiny portion of its orbit in a state where it is detectable with this method so orbital period of the planet cannot be easily determined. It is also easier to detect planets around low-mass stars as it increases planet-to-star mass ratio and thus gravitational microlensing effect is greater.

The main advantages of gravitational microlensing method are that it can detect planets with face-on orbits from Earth’s viewpoint and it can detect planets around very distant stars. When enough background stars can be observed with enough accuracy then the method should eventually reveal how common earth-like planets are in the galaxy.
Chapter 3

The physical properties of extrasolar planets

3.1 Introduction

The end of the 20th century saw a revolution in our knowledge of planetary systems. The discovery of the first extra-solar planet in 1995 (Mayor & Queloz, 1995 [82]) marked the beginning of a modern era and a change of our perception of planets. The discoveries continue apace and reveal an extraordinary diversity of planetary systems and exoplanet physical properties, raising new questions in the field of planetary science. More than 1000 exoplanets have now been unveiled by radial velocity measurements and photometric transit observations. They span a wide range of masses from a few Earth masses to a few tens of Jupiter masses.

Crucial constraints on their structure are revealed by photometric transit and Doppler follow-up techniques, which provide a measure of their mass and radius. Information on the mean density and bulk composition of several exoplanets is thus now available, drastically extending the knowledge of planetary structures restricted till recently to our four giant planets. Atmospheric properties of exoplanets are also starting to be measured and the first constraints on temperature structure, composition and dynamics are now available (see Chapter 2).

In this Chapter, I summarize our knowledge, from a theoretical point of view, about formation, physical properties and evolution for extrasolar planets. I spent these few pages, not strictly related to my work, in order to put in a solid context the APACHE survey and all work performed to detecting an extrasolar planets.
3.2 A brief overview of observations

In this section, I present observed properties of solar and extra-solar planets which are relevant for the understanding of exoplanet physical properties. For more details, I invite the reader to refer to the reviews by Baraffe et al., 2010 [83] and Udry & Santos, 2007 [84].

3.2.1 Lessons from our Solar System

The understanding of planetary structure starts with the extensive works conducted on our Solar System giant planets. Important constraints on interior structures of our four giant planets are provided by measurement of their gravity field through analysis of the trajectories of the space missions Voyager and Pioneer. Our giant planets are fast rotators, with rotation periods of about 10 h for Jupiter and Saturn and about 17 h for Neptune and Uranus. Rotation modifies the internal structure of a fluid body and yields departure of the gravitational potential from a spherically symmetric potential. Within the framework of a perturbation theory largely developed for rotating stars and planets, the gravitational potential (for axysymmetric bodies) is strongly related to the
inner density profile of the rotating object. For Jupiter and Saturn, models assume that the planetary interior consists of a central rocky and/or icy core, an inner ionized helium and hydrogen envelope, and an outer neutral $He$ and molecular $H_2$ envelope. The lighter giant planets, Uranus and Neptune, are more enriched in heavy elements than their massive companions. A wide variety of models can match the mass/radius of these planets. Three-layer models with a central rocky core, an ice layer and an outer H/He envelope suggest an overall composition of 25% by mass of rocks, 60-70% of ices and 5-15% of gaseous H/He. The analysis of the atmospheric composition of our giant planets also shows a significant enrichment in heavy elements. Abundances of several elements (C, N, S, Ar, Kr, Xe) have been measured *in situ* by the Galileo probe for Jupiter and they show a global enrichment compared with solar values of about a factor 3. For Saturn, spectroscopic detections of methane and ammonia suggest significant enrichment of C and N, although with large uncertainties for the latter element. For Uranus and Neptune, carbon is significantly enriched, while the abundance of N/H is comparable to that of Jupiter and Saturn. The interior and atmospheric properties of our giant planets bear important consequences on our general perception of planetary structure. The observational evidence that our giant planets are enriched in heavy material supports our general understanding of planet formation and guide the development of a general theory for exoplanets.

### 3.2.2 Observed properties of exoplanets

The description of exoplanet physical properties must encompass the wide variety of planetary masses and orbital separations yet discovered, as illustrated in Figure 3.2.

About 30% of exoplanets have an orbital separation a less than 0.1 AU. Irradiation effects from their parent star must thus be accounted for for a correct description of their structural and evolutionary properties. Another compelling property of exoplanetary systems is the correlation between planet–host star metallicity and frequency of planets. The probability of finding giant planets is a strong function of the parent star metallicity, indicating that an environment enriched in heavy material favours planet formation. This correlation, however, is not observed for light Neptune-mass planets (see e.g. Udry & Santos, 2007 [84]) although the statistics is still poor.

Crucial information on the interior structure and bulk composition of exoplanets are unveiled by objects transiting in front of their parent star. About 424 of these planets have yet been detected, revealing an extraordinary variety in mean planetary densities and composition. As illustrated in Figure 3.3 some exoplanets are significantly denser,

---

1 The metallicity is defined as the mass fraction of all chemical elements heavier than helium.
Chapter 3. Physical properties of exoplanets

Figure 3.2: Mass of known extra-solar planets (in Mearth) as a function of orbital distance (in AU)

thus more enriched in heavy material than our own giant planets. Another puzzling property revealed by Figure 3.3 is the abnormally large radius of a significant fraction of transiting planets.

3.3 Planet formation

While the model for terrestrial (rocky) planets formation (planetesimal accumulation) is generally shared by the scientific community, the debate is still open for the preferred mechanism of giant planets formation.

3.3.1 Core-accretion model

The current most widely accepted scenario for giant planet formation is the so-called core-accretion model. In this model, solid cores grow oligarchically in the surrounding nebula by accreting small planetesimals\(^2\), located within the protoplanet’s zone of gravitational influence, called the feeding zone, which extends over a few Hills radii.

---
\(^2\)Defined as solid objects present in protoplanetary disks.
(\(R_H = a(M_p/3M_\star)^{1/3}\), where \(M_p\) and \(M\) denote the protoplanet and star mass, respectively, and \(a\) the planet’s orbital radius). Once the solid core has grown of the order of a Mars mass (\(\sim 0.1M_\oplus\)), it starts capturing an envelope of nebular gas and the protoplanet’s growth is governed by a quasistatic balance between radiative loss and accretion energy due dominantly to planetesimals (planetesimals are supposed to sink to the planet’s central regions whereas the accreted gas remains near the surface), with a negligible contribution from the \(PdV\) contraction work.

Both solid and gas accretion rates are relatively constant during this phase, with gas accretion exceeding the planetesimal one. Above a critical mass \(M_{\text{crit}}\), a static envelope can no longer be supported. Gravitational contraction is now necessary to compensate the radiative loss, which increases in turn gas accretion, leading to a runaway process, and the core accretes a massive gas envelope, becoming a newborn giant planet. \(M_{\text{crit}}\) depends from many factor such as the envelope composition (mean molecular weight, opacity), the planetesimal accretion rate and the distance of the planet to the star.

The main problem faced by the conventional core-accretion model is that core growth takes longer than typical protoplanetary disk lifetimes, a few \(Myr\). Giant planet cores can be obtained within the appropriate time-scale either by severely increasing the disk
density or by reducing drastically the accreting envelope opacity, allowing rapid core contraction.

Taking into account migration processes in the core-accretion model has been found to speed up core growth by increasing the supply of planetesimals, avoiding the depletion of the feeding zone obtained in the in situ formation models and solving the time-scale problem of the standard core-accretion scenario. Planets, including our own Solar System giants, now form on a time-scale consistent with disk lifetimes, with the appropriate observational signatures. These models, however, have to reduce the conventional migration rate by a factor 10.

In summary, including migration in the conventional core-accretion model succeeds in forming giant planets within appropriate time-scales down to the inner edge of the disk.

These models, however, use disk surface densities or dust-to-gas ratios about 2–3 times larger than the Minimum-Mass Solar Nebula, MMSN, (suggesting that giant planet cores are unlikely to form in a MMSN) and require adequate planetesimal sizes and/or viscosity parameters in order for giant planet cores to grow rapidly before migration moves them into the star. Unfortunately, these parameters, which involve complex processes such as grain growth/fragmentation or turbulent viscosity are very uncertain and can vary over orders of magnitude.

### 3.3.2 Gravitational instability

The alternative theory for giant planet formation is direct gravitational fragmentation and collapse of a protoplanetary disc, the so-called gravitational instability (GI) scenario, originally suggested to circumvent the timescale problem of the original core-accretion scenario. The disk becomes unstable to its own gravity whenever the stabilizing influence of differential rotation or pressure is insufficient. This criterion is governed by the local density and is thus a local criterion for fragmentation. A disk might become unstable during its evolution, for instance during its formation, if mass builds up faster than it is accreted by the star or, at later times, if the outer part of the disk, where stellar radiation is negligible, becomes sufficiently cool. For fragmentation to occur, however, the disk must cool quickly enough to avoid entering a self-regulated phase. Indeed, the energy loss rate determines the effect of the instability: isothermal disks, in which energy is easily lost (and gained, in gas expansion), remain unstable and evolve violently whereas adiabatic disks tend to heat up and become more stable. Energy transport and dissipation processes are thus key issues to determine whether or not planets can form from gravitational instabilities in a disk. A detailed analysis of these conditions shows that planet formation by GIs can occur only in very massive disks and at large orbital
distances. Actually, it is far from clear that the peculiar composition and structure of our Jovian planets can be explained within the GI model. Fragmentation by GI thus remains controversial, with markedly different results from various groups, and requires that the disk detailed thermal energetics are properly taken into account.

### 3.3.3 Core-accretion versus gravitational instability

It is interesting to point out that disk instability predicts that even very young ($\lesssim 1\text{Myr}$) stars should harbor gas giant planets, whereas the formation of such planets with the core-accretion scenario requires a few $\text{Myr}$. Observational searches for the presence of genuine giant planets around $\sim 1\text{Myr}$-old stars will thus provide crucial tests for the two formation scenarios. As mentioned above, determination of the heavy element abundances of extrasolar planets also provides a definitive test: compositional similarity of planets and their parent star would strongly favor gravitational collapse whereas significant heavy element enrichment of the planet with respect to the parent star composition would rule it out. The large heavy element enrichment inferred for several transiting planets clearly supports the core-accretion model. Last but not least, the efficiency of planet formation by GI should not depend on the disk metallicity, since gravitational collapse from the protoplanetary disk is a compositionally indiscriminating process, in contrast to the core-accretion scenario. The observed clear dependence of planet frequency with the host star metallicity and the suggested trend that metal-depleted stars seem to harbor lower mass planets, i.e. the lack of massive planets around metal-poor stars and the fact that stars hosting Neptune-mass planets seem to have a flat metallicity distribution clearly suggest that metallicity plays a crucial role in planet formation, in agreement with the core-accretion model.

### 3.4 Interior structure properties

There is compelling evidence from our own solar system planets that they are substantially enriched in heavy elements (C,N,O), with a $\sim 3$ to $\sim 6$ times solar value for Jupiter and Saturn, respectively, and even larger contrasts for Uranus and Neptune, as expected from a formation by core accretion. This should apply as well to the discovered exoplanets. Indeed, the small radii of some of the observed transiting planets can only be reproduced if these objects are substantially enriched in heavy material, with a total of several tens to hundreds of Earth masses. For all these objects, including our own Jovian planets, however, uncertainties remain on (i) the total amount of heavy material, (ii) the respective fractions of ‘rocks’ (silicates) and ‘ices’ and (iii) the distribution of this heavy material in the planet’s interior. As mentioned above, the first uncertainty (i)
stems primarily from the present uncertainties on the various Equation Of States (EOS) and should decrease with further experimental and theoretical progress. Addressing the two other issues necessitate to differentiate gaseous from solid/liquid planets.

### 3.4.1 Earth-like to super-Earth planets ($< 10 M_\oplus$)

As discussed in Section 3.3, planets below about $10 M_\oplus$, usually denominated as super-Earth down to Earth-like planets, are not massive enough to enter the unstable, runaway regime leading to rapid accretion of a large gaseous envelope onto the central core. Post-formation degassing or oxidization processes can only produce a tenuous gaseous atmosphere, with no significant consequences for the planet’s contraction. Therefore, these objects consist essentially of solids or liquids rather than gases, making their structure determination from mass-radius observations less uncertain than for more massive planets. The mass-radius relationship for these low-mass planets has been parametrized as 

$$R = R_{ref}(M/M_\oplus)^\beta,$$

with $R_{ref} = (1 + 0.56\alpha)R_\oplus$ and $\beta = 0.262(1 - 0.138\alpha)$, for the rocky or ocean super-Earth planets in the mass range $1 - 10 M_\oplus$ [85], where $\alpha$ denotes the water mass fraction and $\beta = 0.3$ for planets between $10 - 2$ and $1 M_\oplus$, with a weak dependence upon the iron to silicate ratio $Fe/Si$ [86]. Note that incompressible (constant density) material corresponds to $\beta = 1/3$. This provides a sound diagnostic for transiting Earth-like planet detections and the possible identification of the so-called ‘ocean planets’, planets composed dominantly of water [87], as opposed to the terrestrial (Fe-rich) planets.

### 3.4.2 Neptune-like to super-Jupiter planets ($> 10 M_\oplus$)

For planets with a 10% by mass gaseous ($H/He$) envelope, this latter essentially governs the gravothermal evolution of the planet. For instance, a $10 M_\oplus$ planet retaining a modest 10% $H/He$ envelope is 50% larger than its pure icy counterpart [88]. Under such conditions, a variety of internal compositions, either water-rich or iron-rich can produce the same mass-radius signature [88] and detailed information about the planet’s internal structure, other than inferring its bulk properties, becomes elusive. One of the main uncertainties about the internal structure is that we do not know whether these heavy elements are predominantly concentrated into a central core or are distributed more or less homogeneously throughout the gaseous $H/He$ envelope. A summary of the main consequences of the uncertainties (i) in the EOS, (ii) in the chemical composition and (iii) in the distribution of the heavy elements for planets with masses $> 10 M_\oplus$ can be found in [88]. In particular, [88] shows that for a global metal enrichment $Z \gtrsim 15\%$, all heavy material being either gathered in a core or distributed homogeneously throughout
the envelope yields a $\gtrsim 10\%$ difference in radius after 1 Gyr for Neptune-mass planets and a $\gtrsim 4\%$ difference for Jovian-mass planets.

### 3.5 Atmospheric properties

In this section, I present the state of the art of the atmosphere science. For more details, I invite the reader to refer to the review by Seager & Deaming, 2010 [5].

The wide variety of planets both in and out of our solar system provide excellent laboratories for understanding how atmospheres evolve with time and react to their environments. Atmospheres by themselves hold a nearly endless supply of complex and interesting physical problems, but are also the primary link between observations and theory. The majority of observational techniques for studying planets involve capturing photons that have emerged from the atmosphere, placing extreme importance on our ability to successfully model atmospheric behaviour. The grouping and location of substellar mass objects (brown dwarfs and giant planets) on color-color and color-magnitude diagrams are largely due to the sculpting of the emergent spectrum by atmospheric opacity sources. Furthermore, the atmosphere regulates the release of energy from the interior and establishes the upper boundary condition for interior models.

We now turn to a summary of the most significant exoplanet atmospheric discoveries. Hot Jupiters dominate exoplanet atmosphere science because their large radii and extended atmospheric scale heights facilitate atmospheric measurements to maximize S/N. Hot Jupiters are blasted with radiation from the host star. They should therefore be kinetically hot, heated externally by the stellar irradiance. Indeed, early hot Jupiter model atmospheres already predicted temperatures exceeding 1,000 K (Seager & Sasselov 1998; Barman, Hauschildt & Allard 2001; Sudarsky, Burrows & Hubeny 2003). The first and most basic conclusion from the Spitzer secondary eclipse detections was the confirmation of this basic paradigm. The fact that the planets emit generously in the IR implies that they efficiently absorb visible light from their stars. Searches for the reflected component of their energy budget have indicated that the planets must be very dark in visible light, with geometric albedos less than about 0.2 (Rowe et al. 2008) and likely much lower.

#### 3.5.1 Biosignatures

No Earth analogs have been yet detected, but their search is one of the major goals of ground- and space-based research programs of the coming decades. Our current
understanding of planet formation suggests that terrestrial planet formation should be an efficient process. We thus expect these planets to be common, as illustrated by our Solar System which has three such planets (Earth, Mars and Venus). The search for signatures of life on exo-Earths is one of the main motivations for these programs and certainly one of the most exciting scientific inquiries at the beginning of the century. Biogeochemical activity on a planet could manifest itself through spectral features of the atmosphere. Current search strategies, as derived for DARWIN/TPF [89], are thus based on the spectroscopic detection of compounds that could not be present on a planet in the absence of life. The search for biomarkers is based on the assumption that extraterrestrial life shares fundamental characteristics with life on Earth. This later is based on carbon chemistry and requires liquid water as solvent. Other paths for life, based on a different chemistry, could perhaps exist but the signatures of the resulting life-forms are so far unknown. The need for liquid water leads to the concept of habitable zone defined as the region around a star where the surface temperature of Earth-like planets allows the presence of liquid water. This zone depends on the stellar luminosity and thus evolves in time with the star. Its definition also depends on complex processes on the planet, such as the concentration of greenhouse gases or geological activity (see [90] and references therein). Any biomarker should include the signature of $H_2O$, which is a requisite for ‘Earth-like’ life. The presence of $H_2O$, $O_2$ (or $O_3$) and $CH_4$, $NH_3$ or $CO_2$ would imply some biological activity, so that these elements are considered as favorite biomakers [89]. However, as pointed out by [91], the unique detection of one of these compounds may be ambiguous. Indeed, $O_2$, and hence $O_3$, can be produced by photochemistry. The combined detection of $O_2$ with $H_2O$ and $CO_2$, which are important for habitability, would however, provide a robust signature of biological photosynthesis. Similarly, the presence of $CH_4$ or $NH_3$ together with $O_2$ or $O_3$ would be good biomarkers, as demonstrated by the observations of the Galileo probe, as it passed near the Earth and detected simultaneously $O_2$ and $CH_4$ [92]. Note that methane and ammonia are not expected to be abiotically produced on habitable, Earth-size planets, in contrast to a common production in cold hydrogen-rich atmospheres of giant planets. The analysis of biological spectral signatures on an exoplanet is thus optimized if its physical properties, such as its mass and radius, can be also determined.

### 3.6 Star-planet interaction. Tidal effects.

The geometry of an orbiting star-planet system of respective masses $M_*$ and $M_p$, illustrating the formation and dynamical evolution of the system, is encapsulated in three main parameters: the semi-major axis $a$, associated with the mean orbital motion,
the eccentricity $e$ and the stellar obliquity $\epsilon$, defined as the angle between the angular momentum vectors of the planetary orbit and the stellar rotation axis. This latter quantity, which determines the spin-orbit angle, is an interesting diagnostic for inferring the dominant interaction mechanisms at play in a protoplanetary disk: close spin-orbit alignment (as is the case for our solar system) is expected from quiescent tidal interactions or migration processes of a planet within the disk whereas planet scattering events are prone to misalignments of planetary orbit angular momentum and stellar spins [93]. The inclination angle of the orbit relative to the sky plane is usually denoted $i$ so that the line-of-sight projected value of a quantity retains a $sini$ indetermination. A transit observation implies that $i \sim 90^\circ$. Although the stellar obliquity is generally not measurable, Doppler shift observations on the parent star throughout a transit offer the possibility to determine the angle between the sky projections of the two angular momentum vectors, i.e. the projected spin-orbit angle $\lambda$, which gives a lower limit of the true three-dimensional spin-orbit angle. The angle $\lambda$ is determined through the so-called Rossiter-McLaughlin (RM) spectroscopic effect (see Sec.2.2.2). Alignment of the stellar spin and the planetary orbital axis is a strong confirmation of planets forming in a spinning protoplanetary disk surrounding the central protostar. Spin-orbit misalignment, on the other hand, as found for the HD17156, XO-3 and WASP-14 systems, can be produced by planet scattering or Kozai mechanism due to the presence of a third body (e.g. [93]). Interestingly, this spin-orbit alignment shows that the star’s axis of rotation and the orbit angular momentum evolution are not significantly altered during the early episodes of angular momentum loss characterized by outflows and disk-star magnetic coupling. A less frequently addressed tidal parameter is the planetary obliquity, i.e. the angle between the planetary spin axis and the orbit normal. Although the planet’s obliquity is expected to be rapidly damped by tidal dissipation, a persistent nonzero planet obliquity has been suggested as a result of the capture in a Cassini 2 spin-orbit resonance state as the nebula dissipates [94]. This scenario, however, has been excluded for hot-Jupiter-like planets [95]. Indeed, the resonant equilibrium is eventually destroyed by the strong tidal torque, leading the system to leave the Cassini 2 state and spiral toward the Cassini 1 state, with negligible obliquity [95]. Therefore, close-in planets quickly evolve to a state with the planet’s spin axis nearly normal to the orbit plane, i.e. a negligibly small planet obliquity. Note that the determination of stellar companion’s obliquities would provide an interesting diagnostic to distinguish planets from brown dwarf companions, these latter being formed from the same original gravoturbulent collapse of the parent cloud as the star, leading to arbitrary spin angle distributions.
Chapter 4

The path to the APACHE survey

In the previous Chapter, I briefly reviewed the state of the art of the exoplanetary sciences. This was an appropriate step to put in context our main project, the APACHE survey. APACHE is a photometric survey in the northern hemisphere for the search of transiting extrasolar planets orbiting a sample of M-type dwarf stars in the vicinity of the solar system. The APACHE survey is designed for the detection of Super-Earth size planets and for the study of the main features related to the intrinsic variability of the M dwarf stars. At the present time, APACHE and its US-based counterpart MEarth [96] are the only photometric surveys focused on the search for transiting extrasolar planets around dM stars.

APACHE (A PAthway to the Characterization of Habitable Earths) is a collaboration between INAF-Osservatorio Astrofisico di Torino (INAF-OATo) and the Astronomical Observatory of the Autonomous Region of the Aosta Valley (OAVdA) and it utilizes an array of five 40-cm telescopes at the OAVdA site, in the western Italian Alps.

In this Chapter, I highlight the big opportunities related to the M dwarf targets. Next, I summarize the milestones of four-year long efforts undertaken to build the final setup of APACHE. This work has been published as two papers(Damasso, Giacobbe et al., 2010 [97] and Giacobbe, Damasso et al., 2012 [98]).

For this work, I carried out 50% of the photometric observations at the OAVdA telescopes, I developed the data reduction and analysis pipeline TEEPEE and successively I obtained the differential light curves. Furthermore, I developed and used the framework for the periodicity analysis of the light curves and the assessment of the transit sensitivity of the survey.
4.1 The M-dwarfs opportunity

As shown in the previous Sections, we are at the beginning the comparative planetology era. The CoRoT ([99]) and Kepler ([77]) missions revolutionized exoplanet studies. With their photometric precision and long uninterrupted time baselines, each mission was capable of detecting transiting rocky planets. Some studies ([100]; [24]) have begun the characterization of the bulk composition of super-Earths detected in transit by e.g., Kepler and with a measured mass thanks to radial-velocities obtained with ultra high-precision spectrographs such as HARPS-North. On average, if the planet radius is measured to better than 5%, combined with mass measurements to better than 10%, this would allow to distinguish between an icy or rocky composition. This is due to the fact that there is a maximum radius a rocky terrestrial planet may achieve for a given mass. Any value of the radius above this maximum terrestrial radius implies that the planet contains a large (> 10%) amount of water (ocean planet). Given the performances achieved by space-borne transit photometric surveys and the discovery of thousands of candidate transiting systems by Kepler (Borucki et al. 2011), hundreds of which in multiple systems (Lissauer et al. 2011b) that might be characterized solely based on Kepler photometry thanks to the transit-timing variation (TTV) technique (e.g., Holman et al. 2010; Lissauer et al. 2011a; Ballard et al. 2011), one could then conclude that transit work from the ground may be coming to a close.

However, despite these unmatched capabilities, ground-based transit searches are by no means obsolete, and indeed complement the space missions (for a review on the subject, see e.g., Sozzetti et al. 2010). CoRoT and Kepler in fact find planets typically orbiting solar-type stars (spectral type F-G-K) with visual magnitudes $V \sim 13$ and $V \sim 15$, respectively. The prospects of follow-up studies, such as those mentioned above, are much less favourable for such relatively faint stars. For smaller mass planets, radial velocity follow-up will require a precision significantly better than the current state-of-the-art. For these reasons, ground-based transit surveys which discover planets around the closest stars are still vital. It is important to note that the expected yields of the space missions along with those of present wide-angle ground surveys (e.g, HAT, SuperWASP, TrES, and XO), are weighted against planets around low-mass stars (M dwarfs). This is simply because M dwarfs are strongly under-represented in any magnitude-limited survey operating at optical wavelengths. The application of the transit technique to M dwarfs presents several exciting opportunities, and the advantages are especially compelling for the detection of transiting habitable, rocky planets. First, the habitable zones (HZs) of M dwarfs are drawn in close to the stars, improving the transit likelihood. A planet receiving the same stellar flux as the Earth would lie only 0.074 AU from a mid-M dwarf, and would present a 1.6% geometric probability of transiting, compared
to the 0.5% probability for the Earth-Sun system. Second, transits from the HZs of M dwarfs happen much more frequently. At 0.074 AU, a planet would transit once every 14.5 days, compared to 1 year for the Earth-Sun system. This is critical for detectability, as dramatically less observational time is required to achieve transit detection. Third, the small radii of M dwarfs lead to much deeper transits. Earth-sized, rocky planets eclipse up to 0.5% of the stellar disk area, but only 1 part in 3000 of that of the Sun. Fourth, the small masses of M dwarfs lead to up to over an order of magnitude larger induced radial velocity variations compared to the Sun. The peak-to-peak amplitude of a 7$M_\oplus$ planet would be 10$ms^{-1}$ if it orbits inside the HZ of an M5V star, whereas this would shrink to 1$ms^{-1}$ for a Sun-like primary. Fifth, a number of astrophysical false alarms that plague wide-angle transit searches, such as eclipsed giant stars or stars blended with eclipsing binaries, are far less likely or at least easier to spot for nearby and relatively well characterized M dwarfs. Sixth, planet formation models predict they hold an abundance of super-Earth sized planets, and the early evidence (Howard et al. 2011) from the Kepler mission, with increasing occurrence rates for small-size, short-period planets around later-type stars, is strongly supporting this view. Aside from these observational advantages, several developments in astrophysics point to exciting possibilities with M dwarfs.

Firstly, the growing numbers of M dwarf exoplanet discoveries suggest an abundance of sub-Neptune mass planets orbiting low-mass stars. Indeed, as of October 2011, with the new assessment of the first rocky planet within the HZ of the nearby M dwarf GJ 581 (Wordsworth et al. 2011), and the striking evidence provided by the Kepler mission (Howard et al. 2011), the notion is becoming statistically solid that the higher frequency of low-mass planets around low-mass stars with respect to that of solar-type dwarfs could be a real effect (Sousa et al. 2008), and that in general the planet population orbiting M dwarfs is probably quite different from the one orbiting G dwarfs (Bonfils et al. 2007). Indeed, this appears to be confirmed by simulations of planet population synthesis (Mordasini et al. 2007; Benz et al. 2008; Ida & Lin 2008). Most importantly, that a system harbouring a potentially habitable planet has been found this nearby, and this soon in the short history of precision RV surveys, indicates that $\eta_\oplus$, the fraction of stars with potentially habitable planets, is likely to be substantial.

Second, The ground-breaking discovery of GJ 1214b (Charbonneau et al. 2009) has, however, also highlighted some of the present limitations in our understanding of transiting systems composed of low-mass primaries and small-size planets. In particular, degeneracies in the models of the physical structures of Super-Earths (Valencia et al. 2007; Adams et al. 2008), such as objects with very different compositions having similar masses and radii, prevent one from exactly inferring their interior composition (Seager et al. 2007; Rogers & Seager 2010; Charbonneau et al. 2009). Note for example (see
Fig. 1) that the large spread in observed radii for transiting giant planets indicates that a simple division between rocky/water-dominated super Earths of $\sim 2R_\oplus$ and Neptune-mass objects with $\sim 4R_\oplus$ may constitute an oversimplification, and these populations of exoplanets may well be overlapping in radius. It is thus clear that more discoveries of transiting low-mass, small-radius planets over a range of periods around nearby, relatively bright late-type stars are highly desirable. This is of particular relevance as the further spectroscopic characterization studies of the atmospheres possible for these planets are one of the most effective ways to mitigate the degeneracies of exoplanet interior composition models (see for example Miller-Ricci et al. 2009).

Third, because they are unlikely to form in situ (Terquem & Papaloizou 2007), super-Earths necessarily require some form of migration or scattering from their formation regions (e.g., Kennedy & Kenyon 2008; Raymond et al. 2008). The emerging picture is that of planetary systems with the closest planets reduced to rocky or icy cores due to strong evaporation processes for Sun-like primaries, but significantly less so for M dwarf parent stars. Observed systems around stars of various masses can thus effectively test and inform mechanisms that form and bring planets to close-in, detectable orbits (Raymond et al. 2008). A "common place" which is now subject to change is the notion that life cannot survive on HZ-planets around M dwarfs. Previously, it had been assumed that the rotational synchronization expected of close-in HZ-planets would lead either to atmospheric collapse or to steep temperature gradients and climatic conditions not suitable for life (Huang 1960; Dole 1964). However, recently it has been argued that atmospheric heat circulation should prevent each of these barriers to habitability (Tarter et al. 2007; Scalo et al. 2007; Buccino et al. 2007; Cuntz et al. 2007). Regardless, the absence of such heat redistribution would be readily observable with precise infrared photometric monitoring as a large day-night temperature difference, while the detection of a small day-night difference would provide a strong case for the existence of a thick atmosphere. With future space-borne observatories such as the James Webb Space Telescope (JWST), atmospheric observations similar to those mentioned earlier for Hot Jupiters can be extended to habitable, Earth-sized planets orbiting M dwarfs. This possibility is brought about by the small surface areas and temperatures of M dwarfs, which lead to significantly more favourable planet-star contrast ratios. The search for transiting low-mass planets around M dwarfs is thus a worthy and exciting pursuit, due to its many potentially observable consequences for models describing the formation, evolution, internal structure and atmospheres of rocky/icy planets. However, in the context of ground-based transit searches these planets require a unique mode of discovery. Whereas current ground-based surveys can stare at fixed fields containing thousands of stars, the observing strategy of a survey targeting bright M dwarfs should
be reformulated by designing a transit search program aiming at probing individually all the sparsely spread bright, nearby M dwarfs in the sky, utilizing a cluster of telescopes.

4.2 The site characterization study

The first step toward the APACHE survey was the site characterization study that is detailed in Damasso, Giacobbe et al., 2010 [97]. The site characterization study was the site characterization of the OAVdA and it aimed at establishing the potential of the observatory to host the APACHE project.

We focused the attention on those site-dependent factors that can have the largest impact on the ultimately achievable precision of the photometric measurements, such as seeing, extinction and night-sky brightness. We then correlated them with the quality of the photometric measurements of selected target fields, monitored with 20 – 40cm class telescopes available on site (see Section 4.3), and analysed using standard techniques of differential aperture photometry using an ad hoc developed data processing and analysis pipeline (TEEPEE, see Section 5.3). Relevant data were collected in situ over a period of ∼ 4 months, and complemented by additional meteorological and photometric datasets covering a timespan of > 3 years. The main findings of this study can be summarized as follows:

- The measured zenithal V-band night-sky brightness and the extinction registered at same band are typical of that of very good, very dark observing sites. The mean V-band night-sky brightness is 21.29 ± 0.16 mag arcsec$^{-2}$, while the mean of the V-band extinction coefficient is 0.26 ± 0.13 mag.

- The median seeing over the period of in situ observations is found to be ∼ 1.7arcsec (Figure 4.1). Given the limited duration of the observations, we did not probe any possible seeing seasonal patterns, or the details of its possible dependence on other meteorological parameters, such as wind speed and direction. Moreover, our data show that the seeing at the observatory was reasonably stable during most of the nights.

- The fraction of fully clear nights per year amounts to 39%, while the total of useful nights increase to 57% assuming a typical cloud cover of not more than 50% of the night.

- During 24 good observing nights over a period of over 3 months the median photometric precision was 0.006 mag for stars with magnitude $R \sim 13$ mag. The typical nightly photometric precision appears to be uncorrelated with the seeing, whose
typical value (in principle not competitive with other highly reputed observing sites around the world) turns out not to be a limiting factor for achievement of the photometric performance required for detection of planetary transits.

- The observation of known transiting planets (those in particular of WASP-3b and HAT-P-7b) allowed us to show that our differential aperture photometry approach applied to data collected with a small, 250-mm telescope, achieves best-case rms precision of 0.002 - 0.003 mag for the brightest stars (Fig. 4.2 for an example).

![Figure 4.1: The distribution of seeing measurements over the whole period of observation. The small black stars show 7864 individual data points (many of them overlapping due to the seeing values having been approximated to the first significant digit), each corresponding to an average seeing value over a 1 min interval, rounded to 0.1. The large filled circles indicate the median seeing value for each night.](image)

The results of the site testing campaign show conclusively that OAVdA is a suitable choice for hosting a long-term photometric survey for transiting planets around cool stars in the solar neighbourhood.

### 4.3 Pilot study

The pilot study was a 1.5 year-long photometric monitoring campaign of a sample of 23 nearby \((d < 60\,\text{pc})\), bright \((J < 12)\) dM stars that is detailed in Giacobbe et al., 2012 [98]. It was the necessary preparatory step towards APACHE. This survey has
Figure 4.2: Photometric errors (RMS) VS instrumental mean magnitude in the field of the planet-hosting star WASP-3 during a good observing night (and for an exposure time of 35 s).

been carried out with the pre-existing instrumentation at OAVdA (a 25-, 40- and 81-cm telescope)(see Section 4.3.1). The goals of this study were:

- the characterization of the photometric microvariability of each target over a typical period of 2 months;
- the knowledge improvement of some astrophysical properties of our targets (e.g., activity, rotation) by combining spectroscopic information and our differential photometric measurements;
- the evaluation of the sensitivity to transiting small-size planets (> $2R_\oplus$) with periods in range 0.5 – 5 days;
- the implications assessment of the stellar variability for the detection of small-size transiting planets;
This work was mandatory in order to refine the best observing strategy, the photometric data reductions pipeline and the time-series periodicity analysis for the long-term survey.

### 4.3.1 Instrumentation and methodology

The instrumental setup used in the pilot study is very similar to the one adopted for the OAVdA site characterization study (see Section 4.2). For the pilot study, we used a small-size telescope array composed of three instruments with diameters of 81 cm, 40 cm, and 25 cm, respectively. The 40-cm and the 25-cm telescopes performed all the observations equipped with standard Johnson-Cousins $I$ filters, while the 81-cm utilized a standard Johnson-Cousins $R$ filter (observations in the $I$ filter with this instrument were affected by fringing). Naturally, the choice of filters was driven by red colours of the target sample. In Table 4.1 we summarize the main characteristics of the telescope and camera systems. This instrumental setup is very similar to the one adopted for the OAVdA site characterization study (Section 4.2).

All observations were performed with the CCDs setup at the focus of the telescope, and the exposure times were chosen at the beginning of each night of observation (without being subsequently modified) in order to guarantee an optimum signal-to-noise ratio ($SNR \gtrsim 100$) for the target, while avoiding saturation (see Table 4.2 for the typical exposure times for each monitored target). The temporal sampling, including overheads, was typically of a few minutes. We chose to observe with any given instrument one star at a time during a night, tracking it without repointing. No auto-guiding was utilized for the 25- and 40-cm telescopes. For the two instruments, we recorded typical drifts of up to $\sim 100$ pixels.

For the pilot study we selected 23 targets from the input target list of the TOPP Program (see Section 4.3.2 for details). We chose an observational strategy which would allow us to monitor each target for at least 3 hr/night without interruptions for a minimum of a dozen nights over a maximum period of $\sim 2$ months. The expected phase coverage for transiting companions within a few (1-2) days of period would exceed 70%. Two stars, LHS 1976 and LHS 534, were observed for a significantly longer period of time (LHS 1976 actually over two seasons), in order to probe our sensitivity to transiting companions on periods of up to about 1 week or so. None of the above elements of the observing strategy was optimized with the intent of reproducing the actual one to be implemented in the APACHE survey, which employ vastly improved instrumentation and an adaptive observing strategy. At the end of the pilot study (a little over 1 year of observations) the whole database comprised 76287 good images, corresponding to $\sim 500$ GB of data. A summary of the pilot study observations is provided in Table 4.2.
### Table 4.1: Summary of the main characteristics of the telescope and camera systems.

<table>
<thead>
<tr>
<th>Optical scheme</th>
<th>Telescope</th>
<th>Aperture (cm)</th>
<th>Focal ratio</th>
<th>Sensor area (pixel$^2$)</th>
<th>Pixel area (µm$^2$)</th>
<th>FoV (arcmin$^2$)</th>
<th>Plate scale (arcmin/pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflector Maksutov</td>
<td>25</td>
<td>f/3.80</td>
<td></td>
<td>2184 × 1472</td>
<td>6.8 × 6.8</td>
<td>52.10 × 35.11</td>
<td>1.43 (binning 1 × 1)</td>
</tr>
<tr>
<td>Reflector Ritchey-Chrétien</td>
<td>40</td>
<td>f/7.64</td>
<td></td>
<td>1024 × 1024</td>
<td>24 × 24</td>
<td>26.4 × 26.4</td>
<td>1.55 (binning 1 × 1)</td>
</tr>
<tr>
<td>Reflector Ritchey-Chrétien</td>
<td>81</td>
<td>f/7.90</td>
<td></td>
<td>2048 × 2048</td>
<td>15 × 15</td>
<td>16.3 × 16.3</td>
<td>1 (binning 2 × 2)</td>
</tr>
</tbody>
</table>
Table 4.2: Log of observations for the target sample.

<table>
<thead>
<tr>
<th>ID</th>
<th>LHS</th>
<th>Number of frames</th>
<th>Number of nights</th>
<th>Epoch range (days)</th>
<th>Telescope (cm)</th>
<th>Typical exposure (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1104</td>
<td>2049</td>
<td>14</td>
<td>48</td>
<td>40</td>
<td>61</td>
</tr>
<tr>
<td>2</td>
<td>1475</td>
<td>6826</td>
<td>20</td>
<td>72</td>
<td>25</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>228</td>
<td>1084</td>
<td>18</td>
<td>54</td>
<td>40</td>
<td>155</td>
</tr>
<tr>
<td>4</td>
<td>243</td>
<td>774</td>
<td>14</td>
<td>44</td>
<td>40</td>
<td>137</td>
</tr>
<tr>
<td>5</td>
<td>1976</td>
<td>24761</td>
<td>108</td>
<td>482</td>
<td>25/81</td>
<td>64/7</td>
</tr>
<tr>
<td>6</td>
<td>6158</td>
<td>2804</td>
<td>10</td>
<td>40</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>269</td>
<td>1629</td>
<td>14</td>
<td>56</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>8</td>
<td>2220</td>
<td>3980</td>
<td>16</td>
<td>58</td>
<td>40</td>
<td>23</td>
</tr>
<tr>
<td>9</td>
<td>283</td>
<td>1696</td>
<td>12</td>
<td>56</td>
<td>25</td>
<td>58</td>
</tr>
<tr>
<td>10</td>
<td>2472</td>
<td>1022</td>
<td>8</td>
<td>57</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>11</td>
<td>360</td>
<td>46</td>
<td>1</td>
<td>1</td>
<td>40</td>
<td>120</td>
</tr>
<tr>
<td>12</td>
<td>...</td>
<td>4421</td>
<td>15</td>
<td>64</td>
<td>25</td>
<td>21</td>
</tr>
<tr>
<td>13</td>
<td>417</td>
<td>4693</td>
<td>34</td>
<td>85</td>
<td>25/81</td>
<td>10/3</td>
</tr>
<tr>
<td>14</td>
<td>3343</td>
<td>2122</td>
<td>17</td>
<td>88</td>
<td>25/40</td>
<td>60/30</td>
</tr>
<tr>
<td>15</td>
<td>3445</td>
<td>2835</td>
<td>19</td>
<td>66</td>
<td>25</td>
<td>54</td>
</tr>
<tr>
<td>16</td>
<td>528</td>
<td>1697</td>
<td>13</td>
<td>40</td>
<td>25</td>
<td>62</td>
</tr>
<tr>
<td>17</td>
<td>534</td>
<td>7346</td>
<td>44</td>
<td>122</td>
<td>81</td>
<td>8</td>
</tr>
<tr>
<td>18</td>
<td>1721</td>
<td>3028</td>
<td>22</td>
<td>61</td>
<td>81</td>
<td>6</td>
</tr>
<tr>
<td>19</td>
<td>370</td>
<td>633</td>
<td>9</td>
<td>33</td>
<td>81</td>
<td>120</td>
</tr>
<tr>
<td>20</td>
<td>306</td>
<td>75</td>
<td>1</td>
<td>1</td>
<td>40</td>
<td>75</td>
</tr>
<tr>
<td>21</td>
<td>...</td>
<td>1265</td>
<td>11</td>
<td>59</td>
<td>25</td>
<td>75</td>
</tr>
<tr>
<td>22</td>
<td>2686</td>
<td>914</td>
<td>10</td>
<td>60</td>
<td>40</td>
<td>75</td>
</tr>
<tr>
<td>23</td>
<td>2719</td>
<td>587</td>
<td>7</td>
<td>20</td>
<td>40</td>
<td>90</td>
</tr>
</tbody>
</table>
The data reduction procedure utilized a first version of TEEPEE (Transiting ExoplanEtS PipElinE), described in detail in 5.3.

### 4.3.2 Stellar sample

The M dwarfs targeted during the pilot study were selected from the TOPP (TOrino Parallax Program) program input list of nearby \(d < 60\) pc cool stars. We chose stars spanning a range of spectral sub-types, approximately M0 through M6, and set an \(I\)-band magnitude limit of \(I \lesssim 14\) (thus keeping exposure times typically to within a few minutes). Most of these objects, with the exception of LHS 228, LHS 306, LHS 360, and LHS 2719, are included in the LSPM-North catalog (Lépine & Shara 2005), with either Hipparcos trigonometric parallaxes or distance moduli estimates indicating they are within 33 pc from the Sun ([102]). Furthermore, for about half of the sample Smart et al. ([101]) have published precision astrometric information, including direct distance estimates. For the purpose of this study, and particularly to derive the results presented in § 4.3.7, it is important to provide reliable estimate of mass \(M_\star\) and radius \(R_\star\) for all our program stars. For the targets presented in Smart et al., we utilized the values of \(M_\star\) therein. For the other targets, we derived absolute \(K\) magnitudes from the distance estimates and we obtained \(M_\star\) from the mass-luminosity calibration of Delfosse et al. ([103]). For all the program stars, the Bayless & Orosz ([104]) mass-radius relation was then employed to calculate \(R_\star\). For completeness, metallicity [Fe/H] estimates are provided using the Johnson & Apps ([47]) calibration in the plane \(V - K - M_K\), where applicable and only for objects with trigonometric parallaxes, along with effective temperature \(T_{\text{eff}}\) values from the Casagrande et al. ([46]) M dwarfs \((V - K) - T_{\text{eff}}\) calibration. All the above characteristics of our sample are summarized in Table 4.3. For each target, the same Table also reports relevant information on flaring activity (obtained from SIMBAD) and activity indicators, i.e., \(H\alpha\) equivalent widths measurements.

### 4.3.3 Results

#### 4.3.3.1 Photometric precision

Accurately gauging the short- and medium-term photometric precision for our red dwarfs sample, as a function of the characteristics of the instruments adopted and of the details of the methods of differential photometry utilized in the analysis, and including meaningful estimates of the degree of correlated (red) noise in the data, is of fundamental importance. The short-term precision (as defined below) is the quantity that has the most relevant impact on the probability of detecting a transit signal, and of determining
## Table 4.3: Characteristics of the M dwarf sample.

<table>
<thead>
<tr>
<th>ID</th>
<th>LHS</th>
<th>GJ</th>
<th>RA (hh:mm:ss)</th>
<th>DEC (dd:mm:ss)</th>
<th>d (pc)</th>
<th>V</th>
<th>J</th>
<th>K</th>
<th>M_V</th>
<th>M_K</th>
<th>M_⊙</th>
<th>R_⊙</th>
<th>T_eff (K)</th>
<th>[Fe/H]</th>
<th>Flare</th>
<th>Hα EW (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1104</td>
<td>...</td>
<td>00:35:53.1</td>
<td>+52:41:13.7</td>
<td>24.7±1.3</td>
<td>12.5±a</td>
<td>8.93</td>
<td>8.09</td>
<td>10.58</td>
<td>6.13</td>
<td>0.44</td>
<td>0.45</td>
<td>3317</td>
<td>0.07</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>1475</td>
<td>119</td>
<td>02:56:34.4</td>
<td>+55:26:14.0</td>
<td>16.7±1.4</td>
<td>10.4±a</td>
<td>7.43</td>
<td>6.59</td>
<td>9.37</td>
<td>5.48</td>
<td>0.56</td>
<td>0.57</td>
<td>3620</td>
<td>...</td>
<td>n</td>
<td>-0.534b</td>
</tr>
<tr>
<td>3</td>
<td>228</td>
<td>...</td>
<td>07:16:27.7</td>
<td>+23:42:10.4</td>
<td>42.4±2.0</td>
<td>15.5±a</td>
<td>12.02</td>
<td>11.30</td>
<td>12.39</td>
<td>8.17</td>
<td>0.12</td>
<td>0.14</td>
<td>3424</td>
<td>-1.27</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>4</td>
<td>243</td>
<td>...</td>
<td>08:03:06.1</td>
<td>+34:56:54.8</td>
<td>24.4±0.7</td>
<td>16.0±a</td>
<td>11.51</td>
<td>10.74</td>
<td>14.15</td>
<td>8.81</td>
<td>0.09</td>
<td>0.11</td>
<td>2986</td>
<td>-0.67</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>5</td>
<td>1976</td>
<td>...</td>
<td>08:03:19.5</td>
<td>+52:50:38.1</td>
<td>24.8±3.1</td>
<td>11.3±a</td>
<td>8.06</td>
<td>7.24</td>
<td>9.407</td>
<td>5.26</td>
<td>0.58</td>
<td>0.58</td>
<td>3472</td>
<td>0.29</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>6</td>
<td>6158</td>
<td>3522</td>
<td>08:58:56.3</td>
<td>+08:28:25.9</td>
<td>6.8±0.1</td>
<td>10.9±a</td>
<td>6.51</td>
<td>5.69</td>
<td>11.76</td>
<td>6.53</td>
<td>0.36</td>
<td>0.37</td>
<td>3023</td>
<td>0.51</td>
<td>y</td>
<td>+2.76d</td>
</tr>
<tr>
<td>7</td>
<td>269</td>
<td>...</td>
<td>09:29:11.0</td>
<td>+25:58:09.3</td>
<td>16.7±1.3</td>
<td>16.4±b</td>
<td>10.91</td>
<td>9.96</td>
<td>15.32</td>
<td>8.85</td>
<td>0.11</td>
<td>0.14</td>
<td>2697</td>
<td>...</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>8</td>
<td>2220</td>
<td>3585</td>
<td>10:06:43.8</td>
<td>+41:42:52.6</td>
<td>20.7±1.2</td>
<td>11.3±a</td>
<td>8.21</td>
<td>7.40</td>
<td>9.73</td>
<td>5.82</td>
<td>0.49</td>
<td>0.50</td>
<td>3607</td>
<td>-0.21</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>9</td>
<td>283</td>
<td>3612</td>
<td>10:35:26.9</td>
<td>+69:26:58.8</td>
<td>13.2±0.7</td>
<td>11.9±b</td>
<td>7.90</td>
<td>7.16</td>
<td>11.35</td>
<td>6.56</td>
<td>0.35</td>
<td>0.37</td>
<td>3175</td>
<td>0.13</td>
<td>n</td>
<td>-0.219b, +0.13a</td>
</tr>
<tr>
<td>10</td>
<td>2472</td>
<td>452.1</td>
<td>11:54:07.9</td>
<td>+09:48:22.8</td>
<td>11.3±0.5</td>
<td>12.8±a</td>
<td>8.70</td>
<td>7.87</td>
<td>12.54</td>
<td>7.60</td>
<td>0.20</td>
<td>0.22</td>
<td>3120</td>
<td>-0.33</td>
<td>y</td>
<td>+1.248b</td>
</tr>
<tr>
<td>11</td>
<td>360</td>
<td>...</td>
<td>13:46:55.5</td>
<td>+05:42:56.3</td>
<td>56.2±9.5</td>
<td>15.2±b</td>
<td>12.39</td>
<td>11.66</td>
<td>11.47</td>
<td>7.87</td>
<td>0.28</td>
<td>0.29</td>
<td>3858</td>
<td>...</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>12</td>
<td>...</td>
<td>569B</td>
<td>16:24:09.3</td>
<td>+48:21:10.4</td>
<td>8.1±0.08</td>
<td>10.27</td>
<td>6.64</td>
<td>5.91</td>
<td>10.73</td>
<td>6.37</td>
<td>0.38</td>
<td>0.40</td>
<td>3359</td>
<td>-0.14</td>
<td>n</td>
<td>-0.237b</td>
</tr>
<tr>
<td>13</td>
<td>417</td>
<td>623</td>
<td>17:57:50.9</td>
<td>+46:35:19.1</td>
<td>14.6±0.5</td>
<td>11.8±a</td>
<td>7.85</td>
<td>7.00</td>
<td>10.85</td>
<td>6.17</td>
<td>0.40</td>
<td>0.41</td>
<td>3219</td>
<td>0.25</td>
<td>n</td>
<td>-0.215b, +0.15d</td>
</tr>
<tr>
<td>14</td>
<td>3433</td>
<td>4040</td>
<td>19:14:39.1</td>
<td>+19:19:03.7</td>
<td>17.9±0.7</td>
<td>11.9±a</td>
<td>5.78</td>
<td>6.81</td>
<td>10.32</td>
<td>5.54</td>
<td>0.41</td>
<td>0.42</td>
<td>3179</td>
<td>0.69</td>
<td>y</td>
<td>+5.4d, +5.7i</td>
</tr>
<tr>
<td>15</td>
<td>3445</td>
<td>9652A</td>
<td>22:42:38.7</td>
<td>+17:40:09.2</td>
<td>17.4±0.7</td>
<td>11.7±a</td>
<td>8.06</td>
<td>7.18</td>
<td>10.55</td>
<td>5.98</td>
<td>0.43</td>
<td>0.44</td>
<td>3260</td>
<td>0.26</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>16</td>
<td>528</td>
<td>1271</td>
<td>23:06:35.6</td>
<td>+71:43:25.5</td>
<td>17.3±5.2</td>
<td>11.7±i</td>
<td>8.34</td>
<td>7.56</td>
<td>10.56</td>
<td>6.37</td>
<td>0.37</td>
<td>0.39</td>
<td>3445</td>
<td>...</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>17</td>
<td>534</td>
<td>4311</td>
<td>04:58:46.0</td>
<td>+50:56:37.8</td>
<td>19.3±0.9</td>
<td>11.0±i</td>
<td>7.90</td>
<td>7.04</td>
<td>9.63</td>
<td>5.61</td>
<td>0.53</td>
<td>0.54</td>
<td>3540</td>
<td>-0.01</td>
<td>n</td>
<td>-0.452b</td>
</tr>
<tr>
<td>18</td>
<td>1721</td>
<td>1074</td>
<td>14:20:53.1</td>
<td>+36:57:16.8</td>
<td>21.7±1.7</td>
<td>16.7±h</td>
<td>11.05</td>
<td>10.25</td>
<td>14.49</td>
<td>8.57</td>
<td>0.13</td>
<td>0.16</td>
<td>2829</td>
<td>...</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>19</td>
<td>370</td>
<td>...</td>
<td>11:31:08.3</td>
<td>-14:57:21.1</td>
<td>11.8±2.6</td>
<td>14.1±a</td>
<td>9.36</td>
<td>8.50</td>
<td>13.83</td>
<td>8.14</td>
<td>0.12</td>
<td>0.14</td>
<td>2889</td>
<td>-0.05</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>20</td>
<td>306</td>
<td>3668</td>
<td>13:09:34.9</td>
<td>+28:39:06.6</td>
<td>10.8±0.6</td>
<td>14.5±a</td>
<td>9.48</td>
<td>8.61</td>
<td>14.35</td>
<td>8.44</td>
<td>0.12</td>
<td>0.14</td>
<td>2832</td>
<td>-0.08</td>
<td>y</td>
<td>+4.82d</td>
</tr>
<tr>
<td>21</td>
<td>...</td>
<td>1167A</td>
<td>13:10:12.6</td>
<td>+47:45:18.4</td>
<td>10.3±0.5</td>
<td>14.1±a</td>
<td>9.58</td>
<td>8.69</td>
<td>14.09</td>
<td>8.62</td>
<td>0.09</td>
<td>0.11</td>
<td>2951</td>
<td>-0.47</td>
<td>n</td>
<td>...</td>
</tr>
<tr>
<td>22</td>
<td>2686</td>
<td>...</td>
<td>13:20:27.0</td>
<td>-03:56:14.5</td>
<td>26.9±2.5</td>
<td>15.6±a</td>
<td>11.28</td>
<td>10.48</td>
<td>13.45</td>
<td>8.33</td>
<td>0.15</td>
<td>0.17</td>
<td>3058</td>
<td>-0.58</td>
<td>n</td>
<td>...</td>
</tr>
</tbody>
</table>

a Trigonometric parallax / V magnitude after Smart et al. 2010 and references therein
b Gizis, Reid & Hawley 2002
c Trigonometric parallax / V magnitude after Henry et al. 2006
d Walkowicz & Hawley 2009
e Photometric parallax / V magnitude after Lépine & Shara 2005
f Trigonometric parallax from Hipparcos data, after Van Leeuwen 2007
g V magnitude from ASCC-2.5 catalogue, after Kharchenko 2001
h Trigonometric parallax / V magnitude after Reid & Cruz 2002
i Shkolnik et al. 2009
j TASS Mark IV photometric catalogue, version 2, after Droege et al. 2006
its statistical significance. The medium-term precision is the quantity that allows one to
gather insight on different time-scales of variability which are more likely to be intrinsic
to the target, such as surface activity. The complete definition of the photometric preci-
sion is shown in Sec.6.2. Summarizing, we operationally define the photometric precision
of a given target monitored with one of our instruments as the RMS of a differential
light curve (for one observing night or for the whole time interval of the observations)
 obtained with any of the eight methods (m1 through m4 and m1f through m4f) defined
in Section 5.3. To first order, this corresponds to the uncertainty of each data point.

4.3.3.2 Global analysis

Our first goal is to determine the precision of our photometric measurements on a nightly
basis (short-term precision) and over the whole temporal length of the observations
(medium-term precision). For each target, we determine the short-term photometric
precision of the data by computing the intra-night RMS values of eight light curves:
four derived from the nightly light curves and four extracted from the full-period light
curves (see Section 5.3). We determine the medium-term precision in our dataset by
computing the global RMS values of the four full-period light curves.

For the purpose of computing the RMS values, we utilized unbinned light curves for
the 25-cm and 40-cm telescopes, while we utilized binned light curves (at 1.5 min) for
the 81-cm telescope. The binning of the 81-cm telescope data is necessary in order to
homogenize them with the typical time sampling of the other two telescopes, and more
importantly to suppress the effects of atmospheric scintillation, which usually affects to
a significant degree very short (few seconds) exposures.

In the three panels of Figure 4.3 we show the distributions of the mean short- and
medium-term RMS values for each target, for the various methods. Visual inspection of
the plots already shows qualitatively two relevant results: the apparently good agreement
between different methods for differential photometry (with or without detrending), both
in the short- and medium-term regimes, and the fact that the medium-term photometric
precision appears degraded with respect to the short-term precision. We put the discus-
sion on more firm statistical grounds by performing a comparison between the various
RMS distributions based on the Kolmogorov-Smirnov (K-S) test, which measures the
probability \( P_{K-S} \) of any two distributions being statistically indistinguishable.

First, we perform the K-S test between the four distributions within each panel of
Figure 4.3, to gauge statistical differences in the RMS distributions between the m1
through m4 analysis methods, separately for the short- and medium-term regimes. The
results, expressed in terms of \( P_{K-S} \), are summarized in Table 4.4.
Figure 4.3: Distribution of the mean RMS for each target. For each of the four differential photometry methods, the panels show: the mean intra-night RMS distribution IN1 for the sample (top), the mean intra-night RMS distribution IN2 extracted from a full-period light curve for each target (center), and the full-period RMS distribution FP for the sample (bottom).
Next, the K-S test is performed to evaluate the differences between the RMS distributions in the short- and medium-term regimes. Given the results shown in Figure 4.3 (i.e., methods $m_1$ through $m_4$ provide the same answers), we show here only for simplicity the comparison between the RMS distributions of method $m_1$ from one panel with all other distributions in the other panels. The corresponding values of $P_{K-S}$ are summarized in Table 4.5.

Finally, the K-S test is performed to evaluate the differences between the RMS distributions from data obtained with different telescopes. The corresponding values of $P_{K-S}$ are summarized in Table 4.6.

The results from the global short-term and medium-term photometric RMS analysis based on the K-S test highlight the following:

1) the use of a range of instrumental setups does not impact significantly the typical performance in the photometry;

2) the differential photometry methods (without detrending) adopted within the pipeline provide consistently similar results;
3) the similarity between the different light curve RMS values with and without detrending indicate that our data are not particularly affected by linear systematic trends that can be easily identified and removed, such as atmospheric extinction, detector efficiency, or PSF changes over the detector, for which algorithms like SysRem are very effective;

4) the differences between the short- and medium-term photometric precision suggest that correlated noise (astrophysical in nature, or not) becomes important on timescales longer than those of one observing session (one night). For example, the degradation in precision seen in the bottom panel of Figure 4.3 could be ascribed in part to flat-fielding errors and/or to pointing errors and telescope drifts, the latter effectively reducing the number of reference objects used to perform differential photometry on all the nights at once (we recall that an object can be used as comparison star only if it is detected in every frame, and the probability that a star is lost in one or more frames increases with the time span covered by the observations).

In summary, the typical short-term photometric precision is $\sim 5$ mmag and the typical medium-term photometric precision is $\sim 9$ mmag. The former result for the M dwarf sample is in excellent agreement with the findings of Damasso et al. (2010) [97], who had focused their analysis on a small number of stars more similar to the Sun.

### 4.3.3.3 Correlated (red) noise analysis

The RMS of the whole light curve can be assumed to represent the average uncertainty of each data point under a fundamental hypothesis: the photometric measurements are assumed to be uncorrelated (white noise regime). At the millimag level and in a relatively high-cadence time series this is generally untrue. There are many effects that can produce a correlated photometric measurements (red noise regime): changing airmass, atmospheric conditions, telescope tracking (and relative flat field errors) and the intrinsic variability of the targets. These effects introduce some covariance between data points.

The presence of red noise can have a rather significant impact on the statistical analysis of the data. As an example, let us consider a light curve, with no evident variability, consisting on $N$ flux measurements $f_i$ in a fixed time interval, with uncertainties for each data point $\sigma_i$ equal to the RMS of the whole light curve $\sigma_0$. We calculate the mean of the $N$ flux measurements $f_{\text{mean}}$. The uncertainty on $f_{\text{mean}}$, $\sigma_{f_{\text{mean}}}$, is then the error of the mean of $f_i$. Using the expression for the standard deviation of the mean under the assumption of white noise ($\sigma_w$), this uncertainty is:
We note as the uncertainty on $f_{\text{mean}}$ decreases with the square root of the number of points $N$.

The equivalent of Equation 4.1 in presence of red noise is:

$$\sigma_{f_{\text{mean}}} \equiv \sigma_t = \frac{\sigma_0}{\sqrt{N}} + \sqrt{\frac{1}{N^2} \sum_{i \neq j} C_{ij}},$$

(4.2)

where the $C_{ij}$’s are the covariance coefficients between the $i$-th and $j$-th measurements.

Now, we can estimated $\sigma_t$ from the light curves, taking in account the red noise, following the methodology described in Pont et al. 2006 [105], and then determine the red noise component from Equation 4.2:

$$\sigma_r = \sqrt{\frac{1}{N^2} \sum_{i \neq j} C_{ij}} = \sqrt{\sigma_t^2 - \frac{\sigma_0^2}{N}}.$$  
(4.3)

For each star in our sample, we computed $\sigma_t$ over an interval $\delta t = 30$ min, and repeated the calculations using the light curves from the four intra-night methods. We decided to investigate the red noise contribution to the error budget over this time interval, as an illustrative example. Note that this interval actually corresponds to a significant fraction of the typical transit duration for few-days period planets orbiting mid and late M dwarfs, and thus it’s of particular interest for the purpose of this study. We show in Figure 4.4 the various contributions to the photometric error budget in our dataset, expressed as a function of the target $J$ mag. The results shown correspond to the error analysis for the light curves obtained using method $m_1$. The outcome of this specific study is only marginally dependent on the chosen method for differential photometry, as we discuss below. Panel $a$ of Figure 4.4 shows the expected trend of increasing $\sigma_0$ with magnitude in the data. Similarly, panel $b$ emphasizes a positive correlation of $\sigma_w$ with magnitude, obviously related to photon noise. This effect is also expected, as the exposure times were chosen so as to guarantee a high $S/N > 100$ for each object, based on the theoretical noise estimate in Equation 5.2. Panel $c$ shows a weak inverse trend of $\sigma_r$ with $J$ mag, with decreasing red noise contribution for fainter objects. Again, this is expected because for faint objects $\sigma_w$ dominates the total error budget, and $\sigma_r$ becomes increasingly more difficult to determine at the faint end. Finally, panel $d$ shows how the
total photometric error $\sigma_t$ in the binned data behaves as a function of $J$ mag due to the combination of correlated and uncorrelated noise terms.

Figure 4.5 shows, for the whole sample, the behaviour of $\sigma_t$ (triangle) and $\sigma_w$ (square) as a function of the number of flux points in a bin (or equivalently, time length of the interval $\delta t$). In this case the range of points corresponds to $\delta t$ ranging from 5 to 90 minutes. The plot highlights that the total noise is influenced by the correlation of the data points mainly on short time scale (small $N$), while the red component significantly decreases at longer time scales. The noise $\sigma_t$, globally, follow a $1/\sqrt{N}$ relation, as it is expected in a regime where the white noise is dominant.

It is furthermore worth mentioning how the typical uncertainties over the average of the time intervals considered here, taking into account the red noise, are 1.4 mmag, 1.6 mmag, 1.2 mmag, and 1.4 mmag for photometric methods $m_1$ through $m_4$, respectively. This corresponds to values which are 1.3 times, 1.5 times, 1.2 times, and 1.3 times larger than a pure white noise regime. Based on a K-S test, such differences are deemed significant with confidence levels of 94.0%, 99.1%, 95.4%, and 99.9%, respectively. As opposed to the results obtained in Section 4.3.3.2, where we simply used the global light curve RMS as a comparison metric, this more in-depth analysis underlines how the detrending algorithms are actually useful for the purpose of suppressing some of the correlations present in our data.

Finally, we note how the red noise analysis can also be used in principle to investigate the intrinsic variability of a target at the millimag level. The points framed by a square in Figure 4.4 are the three stars in our sample showing clear signs of activity, for which we have observed either flare events or determined their likely photometric rotation period based on the presence of surface inhomogeneities, or both (see Section 4.3.6 and Section 4.3.5). In one case (GJ 1167A), the high levels of red noise could also be seen as an indication of strong activity (i.e. photometric variability) at the millimag level, and on short timescales. Of course, in order to systematically apply this type of analysis to describe stellar activity robust calibration procedures on a statistically significant sample of stars are necessary, together with the help of spectroscopic measurements.

4.3.3.4 On the choice of the comparison stars

When characterizing the sources of correlated and uncorrelated noise in our photometric data, it is reasonable to expect them to at least in part be due to details in the selection of the reference objects. This is all the more so in the case of differential photometry of late-type targets observed with red filters (e.g., Bailer-Jones & Lamm 2003 [106]). To test this possibility, for each red dwarf in our sample we calculated an ‘average’
Figure 4.4: Panel a: mean RMS of the unbinned light curves as a function of $J$ mag. Panel b: white noise term as a function of $J$ mag. Panel c: red noise term as a function of $J$ mag. Panel d: the total error term for the binned light curves as a function of $J$ mag. Stars indicated by asterisks framed by a square are objects showing clear signs of activity (see § 4.3.6 and § 4.3.5).
Figure 4.5: The behaviour of $\sigma_t$ and $\sigma_w$ vs number of points in time bins

colour index $V - K$ of the references (using the magnitudes of the NOMAD catalog) and we investigated whether the total intra-night light curve RMS, the white noise, and the red noise contributions showed any significant dependence on colour differences between the red M dwarfs and the typically bluer sets of comparison stars. In parallel, we investigated the possible dependence of correlated and random noise on the average number of comparison stars used for each target, which could also play a role to some degree (especially if one or more of the comparison stars were to show variability).

The left three panels of Fig. 4.6 show the average red noise (top), white noise (middle), and total RMS (bottom) for the sample as a function of the difference between the colour index of the target and the average color index of the corresponding comparison stars $(V_r-K_r)-(V-K)$. The data are binned along the X-axis and error-bars correspond to the dispersion of the values in a bin (the right-most bin contains only two stars). No evident trends with colour index can be found. In the right three panels of Fig. 4.6 the same quantities are plotted as a function of the average number of comparison stars for each red dwarf in the sample. Even in this case, there is no evident correlation. On the time scale of one observing night, the number of comparison stars or colour-dependent effects such as atmospheric extinction do not appear to impact to any significant extent our measurements. This is in line with, e.g., the analysis of Bailer-Jones & Lamm (2003) [106] and Irwin et al. (2011) [107]. The same considerations may not hold in case
Chapter 4. *Apache survey milestones*

Figure 4.6: Left: Red noise (top), white noise (middle), and total RMS (bottom) for the sample as a function of the difference \((V_r-K_r)-(V-K)\) between the colour index of the target and the average color index of the corresponding comparison stars. Right: The same quantity is plotted as a function of the average number of comparison stars for each red dwarf in the sample.

of our full-period photometry, which is more prone to being affected by other sources of systematic errors, including possible dependency on colour-related effects (e.g., flat-fielding errors), as already mentioned in Section 4.3.3.2.

We also looked at the possibility of evaluating the impact on the noise levels by forcing the choice of red stars as comparisons. We analysed the fields of GJ 1167A and GJ 9652A, with common proper motion companions at \(\sim 3\)′ and \(\sim 42\)″, respectively. Due to their faintness (at V-band they are 2-3 mag fainter than the primaries), GJ 1167B and GJ 9652B are not selected as comparison objects by any of the four intra-night differential photometry methods. By forcing their inclusion in the reference sets for both targets, no appreciable differences in either correlated or random noise levels can be observed, due to the significant number (> 10) of much brighter references selected by the pipeline in all cases.
4.3.4 Photometric variability: periodicity analysis

After the characterization of the photometric noise properties for the M dwarfs observed during the pilot study, we describe here the key elements of the analysis segment devoted to the astrophysical characterization of the stars themselves. We look for signals in the data which are periodic in nature, and which would indicate either the presence of a transiting companion or which could be interpreted as intrinsic to the target (e.g., due to chromospheric activity). The study of the BLS spectra (see Section 6.3.1 for BLS description), does not show any significant periodicity within the period range investigated (0.1 – 5 days), for all 23 objects in our sample. This is not unexpected, given the very small number of targets included in the pilot study, due to standard considerations of geometric transit probability. In the next section, we focus our attention over the periodic signal interpreted as intrinsic to the targets.

4.3.4.1 Characterizing stellar rotation

For most of our program stars, the photometric data collected during our one-year long observing campaign, covering a typical timespan of ∼2 months, can be used to directly measure the stellar rotation periods if quasi sinusoidal variations in the broad-band photometric signal are detected, under the assumption that it is produced by short- and medium-lived spots on the stellar photosphere.

In order to reveal the presence of presumably starspots-induced rotational modulation in the full-period differential photometric datasets (extracted with the m2 and m4 methods) of our targets, we looked for agreement between two different periodicity search algorithms (and limiting our analysis to the determination of the most significant frequency). Both algorithms produced essentially the same results when applied on each of the two light curves, making our conclusions more robust. In the following discussion we present the results of the period-search analysis as applied to the light-curve obtained with the m4 method.

The first tool utilized is the PERIOD04\(^1\) software [108], which performs Discrete Fourier Transforms (DFT) of a time series and is used routinely in asteroseismology work (e.g., [109]). The second method solves directly a linear Least Squares problem: the data are folded according to a grid of different trial periods and fit to a sine function, and at each step the reduced chi-square $\chi^2$ is evaluated. While in the Fourier analysis the significant periods correspond to peaks in the amplitude spectrum, in a periodogram obtained with the second method the best-fit period will correspond to the trial value which minimizes

\(^{1}\)http://www.univie.ac.at/tops/Period04/
Figure 4.7: The search for periodicities in the full set of the photometric data for the dM star LHS 3445 revealed the existence of a signal with period $P=0.4752812$ days (corresponding to the absolute minimum in this periodogram).

$\chi^2$. This approach has been adopted by Irwin et al. (2011) [107] to search for rotation periods in the M dwarf sample of the MEarth transit survey.

We found that our program stars can be divided in three main groups: a) targets which do not show any significant periodicity, b) targets for which the power spectra show the existence of several possible periodic signals, but with low significance, and c) stars showing one significant frequency in the periodogram which can be reliably interpreted as the star spinning frequency (2 objects in particular, $\sim 10\%$ of the targets).

Here we present the results concerning the determination of the rotation periods of these two targets, LHS 3445 (a.k.a. GJ 9652A) and GJ 1167A. In order to gather independent circumstantial evidence for their periodicity being due to rotation, as observed in our photometric data. We downloaded archival high-resolution spectra of both targets obtained by Shkolnik et al. (2009) [52] with the High Resolution Echelle Spectrograph (HIRES) on the Keck 1 telescope, and we analyzed them to provide an estimate of the
projected rotational velocity $v \sin i$. In short, the $v \sin i$ values for the stars are measured from the stellar line widths via a cross-correlation technique which employs as a template a high-resolution spectrum of a slow rotating star of similar spectral type and known projected rotational velocity (see, e.g., [110]). The width of the peak in the cross-correlation function is dependent on the line profiles of both template and target object. We used a HIRES spectrum of the star GJ 402 (spectral type dM5) as the slowly rotating template. This spectrum is used to create artificially broadened profiles, with an IDL code that implements the prescription of Gray (2008) [111], and the width of the cross-correlation function (CCF) peak between the unbroadened template and its broadened version is calibrated for a set of rotation speeds. The maximum in a CCF peak vs. $v \sin i$ plot will correspond to the best-fit projected rotational velocity for the target. For our analysis we chose the echelle order covering the spectral range 7370-7490 Å, which is free of TiO bands and telluric lines which would make the analysis difficult.

The star LHS 3445 is included in the samples analyzed by Shkolnik et al. (2009) [52] and Walkowicz & Hawley (2009) [112], and it is known to be an X-ray active star (it has...
Figure 4.9: Cross-correlation function (CCF) peaks as a function of the projected rotational velocity $v \sin(i)$ for two of our dM targets, LHS 3445 and GJ 1167A, and other M dwarfs used as a template (GJ 402) and check stars used to validate the procedure. The maximum values of the CCF peak functions correspond to the projected rotation velocities measured from the spectra taken with the Keck/HIRES echelle spectrograph. LHS 3445 and GJ 1167A appear to be rapid rotators.

been identified in ROSAT data) and a flare star (see table 4.3). We observed this target flaring three times, as described in Section 4.3.6. LHS 3445 is also considered a young star: Smart et al. (2010) [101] indicate an estimated age in the range 30-50 Myr, while Shkolnik et al. (2009) [52] suggest an age between 60 and 300 Myr. Figure 4.7 displays the calculated periodogram for LHS 3445, showing a minimum occurring at $\sim 0.47$ days, and this is confirmed by the DFT analysis. This is then to be considered the most probable rotation period. A second minimum is observed very close to the period of 1 day and we discard it as an alias on the basis of the window function. Figure 4.8 shows our photometric data folded according to the period $P = 0.4752812$ days, revealing a clear sinusoidal-like shape probably due to star spot modulation. A single-spot model (see Section 4.3.5) is overplotted to highlight the consistency of this hypothesis. As shown in Figure 4.9, the analysis of the Keck spectrum leads to the best estimation for
The search for periodicities in the full set of the photometric data for the dM star GJ 1167A revealed the existence of a signal with period $P=0.2151997$ days (corresponding to the absolute minimum in this periodogram).

The projected rotational velocity $v\sin i \sim 25\text{km/s}$. Note in Figure 4.9 that the slow rotating star GJ 317 ($v\sin i < 2.5 \text{ km/s}; [113]$) is used as a check for the goodness of the results for our targets (i.e. we were able to recover its low projected velocity analyzing the HIRES spectrum) together with the star GJ 490 B, known to be a fast rotator ($v\sin i = 41 \text{ km/s}; [114]$), for which we recover a high value of $v\sin i$ in agreement with the one found in the literature. Using the stellar radius indicated in Table 4.3 and the rotation period estimated photometrically, the mean rotational velocity of LHS 3445 results to be 44.7 km/s, which compared to the projected velocity leads us to conclude that the rotation axis is inclined by $\sim 34$ deg with respect to the line of sight.  

The second target that shows a clear low-amplitude sinusoidal modulations in the light curve is GJ 1167A. GJ 1167A is an active flare star (but no flares were observed over the timespan of our observations), included in the samples analysed by Shkolnik et al.

---

\footnote{Note that LHS 3445 has an M dwarf common proper motion companion, the star LHS 3446. We analysed its light curve, which appears to be not at all variable over the timescale of our observations.}
In Smart et al. (2010) [101] GJ 1167A is reported being a dM4.0 star with a rather uncertain (0.12−10 Gyr) age. Both the DFT and PDM analysis of our photometric data reveals a clear sinusoidal periodic signal corresponding to P∼0.22 days (the PDM results are shown in Figure 4.10). Figure 4.11 shows the light curve folded according to this period, with superposed a single-spot model as for LHS 3445 (see Section 4.3.5). The fact that GJ 1167A is a very fast rotator is confirmed by the line broadening analysis of the HIRES spectrum, as shown in Figure 4.9. We estimate $v \sin i \sim 33$ km/s. Assuming for this star a radius $R = 0.14R_\odot$, our measured photometric rotation period implies a true rotational velocity of 32.9 km/s, which means we are looking at GJ 1167A virtually perpendicularly to its rotation axis.

Given the high rotational velocities inferred for both LHS 3445 and GJ 1167A, this possibly implies relatively young ages for these stars. Alternatively, the rapid spin period could also be due to tidal synchronization effects if these stars were to be very short-period binaries. Indeed, Smart et al. (2010) hint at the possible binarity of LHS
3445 based on kinematics considerations. We checked the cross-correlation functions obtained from the archival high-resolution Keck/HIRES spectra, looking for evidence of double peaks, which would indicate the presence of the secondary spectrum. No such evidence is found in both cases, thus we tentatively conclude that LHS 3445 and GJ 1167A do not harbor very short-period companions of comparable mass. Naturally, a more aggressive analysis could be carried out, based for example on detailed modeling using improved cross-correlation techniques such as TODCOR [115], aimed at faint companions detection. We plan to pursue this issue further in future work.

4.3.5 Photometric variability: starspots analysis

On the basis of the results described in the previous paragraph, we attempted at modeling the observed photometric variability of LHS 3445 and GJ 1167A assuming it is due to a rotating spot on the stellar photosphere. This work must be intended as the first step towards a more detailed model for star spot distribution to be developed in the course of the upcoming long-term photometric monitoring program, which will become a particularly valuable tool to evaluate the impact on the low-mass planet detection thresholds (e.g. Barnes et al. 2011 [6]).

The model we developed is based on the work of Makarov et al. (2009) [66]. This three-dimensional model describes the flux modulation due to a small circular spot rotating on the stellar surface. We modelled a single circular spot of radius $r$ with its center located at latitude $b$ and longitude $l$ in the frame of reference of the star. The star is assumed rotating around its axis with a differential angular velocity $\vec{\omega}(b)$ which depends on the latitude on the stellar disk, and the spot is assumed rotating at a fixed latitude. The rotation axis is considered tilted by an angle $i$ measured starting from the line of sight ($i = 0$ if the axis is pointing toward the observer; $i = \pi/2$ when the axis is perpendicular to the line of sight). In the model, the flux modulation can be expressed as $\frac{\Delta F}{F} = -f_s r^2 \frac{I(\theta) \cos \theta}{I_{tot}}$, where $f_s$ is the spot contrast with respect to the local surface brightness, $r$ is the spot size in units of the stellar radius ($r \ll 1$), $I_{tot}$ is the integrated intensity flux from the stellar disk, $\theta = \theta(b, l(t))$ is the angle between the line of sight and the perpendicular to the star surface passing through the center of the spot (depending upon the epoch of observation), and $I(\theta)$ is the intensity flux emitted at the spot location. Using the Stefan-Boltzmann law $I = \sigma T^4$, the contrast can be approximated by the relation $f_s = 1 - \frac{T_s^4}{T_{ph}^4}$, where $T_s$ and $T_{ph}$ are the mean temperatures of the spot and photosphere respectively. Using literature prescriptions (Barnes et al. 2011), we set the contrast to vary in the range 0.2-0.8.
Table 4.7: Results of the star spot model applied to the fast rotators LHS 3445 and GJ 1167A

<table>
<thead>
<tr>
<th>Star</th>
<th>(i) (\degree)</th>
<th>(b) (\degree)</th>
<th>(f_s r^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHS 3445</td>
<td>26.7</td>
<td>20.3</td>
<td>0.00027</td>
</tr>
<tr>
<td>GJ 1167A</td>
<td>89.9</td>
<td>42.1</td>
<td>0.000007</td>
</tr>
</tbody>
</table>

The parameters kept fixed in the simulation are the \(I\)-band limb-darkening coefficients (a quadratic limb-darkening law from Claret (2000) [116] appropriate for M dwarf stars is used to describe the intensity flux distribution on the stellar photosphere) and the orbital period. The free parameters are the inclination angle of the rotation axis, the spot latitude and the initial longitude, and the scalar \(f_s r^2\) that, combining the spot contrast with its surface area, quantifies the dimming in the light intensity due to the transit of the star spot in front of the observer.

When applied to the dM stars LHS 3445 and GJ 1167A, the model reproduces well in both cases the inclination of the rotation axis, as found from spectro-photometric data: \(i = 26.7 \degree\) deg for LHS 3445, and \(i = 89.9 \degree\) deg for GJ 1167A. The post-fit residuals, having RMS of 4.2 mmag and 6.4 mmag for LHS 3445 and GJ 1167A, respectively, are comparable to the typical RMS precision achieved for the sample (see Section 4.3.3.2), and do not show any additional significant evidence for periodic signals. All best-fit spot model parameters for the two stars are summarized in Table 4.7, and the corresponding flux modulations are overplotted to the phased photometric data for the two stars in Figure 4.8 and Figure 4.11.

We note that our choice to fit for the scalar \(f_s r^2\) rather than the individual values for the brightness contrast \(f_s\) and the spot radius \(r\) is due to the fact that in the Makarov et al. (2009) [66] model these two parameters appear strongly correlated, and attempting to fit for them separately results in very loose constraints on their actual values, and poor overall convergence of the spot model. This effect can be easily understood, as in the disk-integrated stellar flux data the effect of a cold, small spot is difficult to distinguish from the signal produced by one which is warmer and larger. The price to pay in this case, in that we don’t directly estimate the spot size and temperature, is minimal, as this exercise, for the purpose of this study, must be intended solely as a consistency check of the interpretation of the rotational modulation for the two stars based on the photometric and spectroscopic data presented above. Possible future developments include the implementation of more sophisticated (single and multiple) spot models, and the detailed assessment of their uncertainties through, e.g., Bayesian statistics.
4.3.6 Photometric variability: flares analysis

Flaring events, a powerful indicator of the presence of strong stellar magnetic fields, are known to be relatively common among late-type stars, typically lasting between a few minutes up to several hours, and producing increases in the observed flux of up to several magnitudes. While ground-based as well as space-borne long time-series photometric studies of open clusters and in the field over a range of wavelengths (Moffett 1974 [117]; Lacy et al. 1976 [118]; Audard et al. 2000 [119]; Güdel et al. 2003 [120]) have begun unveiling some important correlations between flare occurrence rates and stellar characteristics such as mass, age, and activity levels (e.g., Ambartsumyan et al. 1970 [121]; Mirzoyan et al. 1989 [122]; Kowalski et al. 2009 [123]), the physics of flares is still far from being fully understood. Several open questions still need to be properly addressed, which include the details of the energy release, the mechanisms for producing the atmospheric emission, and the understanding of flares on a global scale – how do flare properties (occurrence rates, emission strength, timescales, frequency) correlate with stellar characteristics (mass, age, activity levels)?

Upcoming ground-based photometric surveys such as Pan-STARRS [124], PTF [125], and LSST will certainly provide the opportunity to gather flare data for large numbers (> \(10^6\)) of stars, sampling wide ranges of flare amplitudes and timescales. However, a potentially important niche for 'classical' flare studies (i.e., those based on the continuous monitoring of single objects) will come as a by-product of those photometric programs, such as MEarth [96] or APACHE 6.3.3, targeting large numbers of relatively bright late K and M dwarfs in search of transiting planets. For example, the detailed characterization of the flaring behavior for K and M dwarfs objectives of targeted searches for transiting planets is an important ingredient towards the thorough understanding of the impact a star and its environment might have through time on the habitability of any planet it may harbor which could sustain the presence of liquid water on its surface (e.g., [4]; [126]).

Two of the stars included in our pilot study, LHS 3445 and LHS 2686, showed flaring events in the photometry gathered over a period of 66 days and 60 days, respectively. LHS 3445 is classified as UV Cet type flare star ([127]), while no information on flaring is available in the literature for LHS 2686. Given the sampling rate of the two light-curves (54 sec and 75 sec, respectively), we were sensitive to flares with decay times larger than \(\sim 1\) min. Visual inspection of the light-curves allowed us to infer decay times of a few minutes (‘impulsive’ flares. See, e.g., [128]). As an illustrative exercise of the type of studies that will be possible once observations will be gathered in survey mode, we characterized the events following the approach of Hartman et al. 2011 [129], i.e. by solving a non-linear Least Squares problem with a set of consecutive photometric measurements.
starting from the time of recorded maximum brightness \( t_0 \) and an exponential model function of the type:

\[
m_d(t) = A \times \exp \left( -\frac{(t - t_0)}{\tau} \right) + m_{d,0},
\]

with adjustable parameters \( A \), the peak magnitude of the flare relative to the non-flaring magnitude, \( \tau \), the decay timescale, and \( m_{d,0} \), the differential normalized magnitude of the star before the flare. The Least Squares solution was obtained with an IDL implementation of the Levenberg-Marquardt method (Levenberg 1944; Marquardt 1963), \texttt{MPFIT} by Craig Markwardt\footnote{Available at http://purl.com/net/mpfit. \texttt{MPFIT} is a port of MINPACK-1 from FORTRAN, and is also available in C and Python}, and starting guesses for the model parameters \( A \) equal to the peak magnitude, \( \tau = 0.001 \) d, and \( m_{d,0} \) equal to the average differential magnitude of each time series. The results are shown in Figure 4.12. For LHS 3445, two of the three recorded flares occurred within a timespan of 1.2 hrs during the same night, possibly a case of homologous flares (e.g., [131]; [132]; [133]).

### 4.3.7 Limits to transiting companions

The photometric measurements obtained for our sample during the pilot study can be utilized to carry out simulations aimed at determining what sensitivity to transiting companions (of given radius and period) we achieved on a star-by-star basis, expressed in terms of easy-to-interpret comparison metrics, such as detection probabilities and phase coverage. These simulation tools whose application to the data collected during the pilot study we illustrate here, will be of great use in the careful prioritization and for the optimization of the scheduling of the targets during survey operations.

In order to evaluate the sensitivity of each of the M dwarfs in our sample to transiting companions of given period and given radius, provided the primary size from Table 4.3 (or equivalently inducing a transit signal of given depth), a large-scale simulation was performed by injecting synthetic transit signals into the differential light curves (both intra-night and full-period) of our targets (without \( \sigma \)-clipping as in Section 6.3.1). We utilized these data under the hypothesis that the procedure adopted to extract it from the raw images does not affect the signal. Furthermore, transit signal injection was performed on all light curves obtained with the four methods for differential photometry described in Section 4.3.1. We are aware that the specific choices for these elements of the simulation setup can in principle impact our findings to some extent. For example, we are not in a position to estimate quantitatively the effects on our capability to recover transit signals due to the application of trend-filtering algorithms during the production
of the differential photometric data in pipeline mode. However, the goal of the present
analysis is simply to gauge our ability to detect transit events as a function of the system
parameters, given an observed distribution of RMS values for our stellar sample.

The input parameters (and their ranges) to the simulation, which was entirely carried
out in IDL, we determined based on considerations taking into account the temporal
sampling, total timespan, and typical photometric precision of our data. We generated:

- 1000 random, uniformly distributed periods in the range $0.5 - 5$ days;
- 100 random, uniformly distributed phases for each period;
- 4 amplitudes of the transit signal depth $t_d$ (in flux units: 0.02, 0.015, 0.01, 0.005)
  for each period and phase. The four values of transit depth simulated correspond to

\[\text{Normalized } \Delta \text{mag} = \begin{cases} 
1.00, & T-t_{\text{max}} \leq 0 \\
1.00 + (1.04 - 1.00) \left( \frac{T-t_{\text{max}}}{4.5 \text{ min}} \right), & 0 < T-t_{\text{max}} \leq 4.5 \text{ min} \\
1.04, & T-t_{\text{max}} > 4.5 \text{ min}
\end{cases}\]
companion radii in the ranges 1.5-9.1, 1.3-7.9, 1.1-6.4, and 0.8-4.5 $R_{\oplus}$, respectively, given the sizes of the primaries from Table 4.3;

- a fixed orbital inclination $i = 90^\circ$, and perfectly circular orbits ($e = 0.0$);

Consequently, the orbital radius parameter was derived from $M_*$ and from Kepler’s third law under the assumption $M_p \ll M_*$ [134]. We recognize that allowing the inclination to float might change to some extent the details of the detection probability results. A detailed study of the sensitivity to non-central (eventually in the limit of quasi-grazing) transits is left for the future.

With the above simulation setup, 400 000 synthetic transit light curves were generated for each target using the Mandel & Agol (2002) algorithm [135]. For the purpose of this analysis limb-darkening effects are essentially irrelevant and in order to speed up execution they were turned off. For each target, the synthetic transit signals were then injected on both nightly and full-period light curves. The choice of running the simulation for the whole suite of differential photometry methods was driven by the findings of Section 4.3.3.2 and Section 4.3.3.3, thus allowing us to gauge the sensitivity of the transit detection probability to variable photometric precision. This was done in practice by comparing the results obtained by running the BLS algorithm on the synthetic, ‘noise free’ datasets as well as on the combined light curves, with the prescription of discarding datasets in which a transit would have occurred only once.

Figure 4.13 summarizes the process of injection of a transit signal in a differential light curve. The top panel shows the synthetic folded light curve, without limb-darkening, produced by the Mandel & Agol (2002) [135] formalism for a transit with a fractional depth of 2% and a period of 1.8468372 days. The center panel shows the actual phased light curve of LHS 1976, while the bottom panel shows the combined light curve in phase. The datasets shown in the upper and lower panels are both fed to BLS.

In the analysis with the transit search algorithm of a given light curve, BLS was run using 500 period steps in the range 0.5 – 5 days, the folded time series was divided into 100 bins, and the signal residue ($SR$) of the time series (see [136]) was determined using these binned values. We fixed the fractional transit length to lie in the range 0.1 – 0.01.

As mentioned above, the main results of the simulation are expressed in terms of two simple comparison metrics, which we identified to be the phase coverage and the transit detection probability.

Here we calculate the phase coverage, in a period range of 0.5 to 5 days, for 10000 trial periods. We divide the phased light curve in bins corresponding to 20 min in length (obviously the number of bins changes with the trial period). A single bin is flagged as
Figure 4.13: Top: synthetic folded light curve with $t_d = 2\%$ and a period of 1.8468372 days. The times of observation are those of LHS 1976. Center: the actual phased light curve of LHS 1976. Bottom: the combined light curve for LHS 1976.

‘filled’ (i.e. containing a satisfying number of data points) if at least five points coming from three different nights fall within. The phase coverage (expressed as percentile) represents the relative number of bins that satisfy this condition with respect to the total number of bins.

We define the detection probability as the relative number of periods that are detected by BLS with respect to the total number of injected periods. We considered as detections all those signals for which BLS returned a period $T'$ such that $|T' - T_{in}|/T_{in} < 0.01$, where $T_{in}$ was either the actual period of the injected signal, or half that period, or twice that period. We are aware of the fact that this requirement might be somewhat stringent: In practice we do not consider as detections transits identified with the correct depth, but with an incorrect period due to, e.g., the sampling properties of the light curve. We adopted this more conservative approach to the definition of detection probabilities as a way to partly compensate for the caveats and assumptions of the simulation scenario listed at the beginning of this section.

Figure 4.14 shows the phase coverage as a function of injected period for a signal with $t_d = 0.02$ for three representative objects in our sample, LHS 1976, LHS 417, and LHS 3343, with very different (from poor to excellent) levels of phase coverage.
Figure 4.14: Phase coverage as a function of period of the injected transit signal ($t_d = 2\%$). Top: LHS 1976. Center: LHS 417. Bottom: LHS 3343

For the same three stars, and in the case of differential light curves obtained with method $m3$, Figure 4.15, Figure 4.16, and Figure 4.17 give an overview of all the fundamental ingredients provided in output from the simulation. In every Figure, the top-most four sub-panels (a) show, for each of the values of $t_d$ simulated, the original distribution of injected transit signals (solid line) and that of those signals whose corresponding light curves (synthetic as well as combined) are subsequently fed to BLS (dashed line), both expressed as a function of period. The second set of sub-panels (b) shows the fraction of BLS-detected signals in the synthetic datasets (solid line) and in the combined light curves (dashed line). The third set of sub-panels (c) shows the distribution of the signal detection efficiency ($SDE$) parameter both in the case of synthetic light curves (solid line) and for the combined light curves (dashed line). We remind the reader that a value for $SDE$, which quantifies the statistical robustness of a detected transit-like periodic signal in the language of BLS, is computed as:

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

where $SR_{peak}$ is the value of the peak in the signal residue distribution, while $\langle SR \rangle$ and $sd(SR)$ are the mean and the standard deviation of $SR$ over the frequency band tested.
Chapter 4. *Apache survey milestones* 64

(see Kovács et al. 2002 for details). Finally, the last set of sub-panels (d) shows the distribution of the transit depths of the detected signals as evaluated by BLS from the synthetic light curves (solid line) and from the combined light curves (dashed line).

For all three Figures, sub-panels a and b clearly illustrate the dependency of transit detection probability on period. As expected, this behaviour is very similar to the trend of phase coverage with period (see Figure 4.14). Panels b also quantify the amount of degradation in detection probability, with respect to the 'noise-free' case, when the transit depth approaches the typical photometric precision, even when the phase coverage is very good.

The three sets of sub-panels c illustrate how the statistical significance of a detection, as measured by the $SDE$ values, changes with phase coverage and signal depth. The values of $SDE$ can also be used to quantify the probability of a detection being a false positive. The actual behaviour of $SDE$ depends on many factors, some intrinsic to the algorithm (e.g. number of bins, number of trial periods) and others depending on the quality of the data and of the signal (e.g. $S/N$), however the statistical studies Kovács et al. (2002) [136] indicate a value of $SDE \sim 5$ above which the false positive probability is low ($< 10\%$). Above such value the $SDE$ does not depend on the input parameters of BLS (Figure 5 of Kovács et al. 2002 [136]). The plots in Figure 4.15 show how the typical value of $SDE$ decreases with the amplitude of the signal and with the phase coverage. In Figures 4.16 and 4.17 this trend is less evident. This is due to the $SDE$ values fluctuating within a regime where the false positive probability is very high. The fact that many signals are detected with a low $SDE$ highlights the robustness of the BLS algorithm, which is capable of uncovering weak and noisy signals albeit with low statistical confidence.

The analysis of the transit depth distribution, as shown in panel d, is a further test of the BLS reliability. Generally BLS provides a good estimate of the signal amplitude, but looking at the three figures an underestimation of the signal amplitude can be noticed. This is due to the initial parameters (number of bins, number of trial periods) of BLS that, in this case (in order to reduce the CPU time calculation), are insufficient to define precisely the ‘border’ of the signal. This effect appears to increase with the phase coverage. This should be explained taking in account that the number of "spurious" points increases with such phase coverage. To verify that this underestimate depends on the input parameters of BLS, we ran the algorithm on our archive data for the transiting planet WASP-3b (Damasso et al. 2010) increasing the number of trial periods to 10000 (in the same range of the simulation) and the number of bins to 300, and we recovered a transit depth in good agreement with the published one (nearly 2%).
Figure 4.15: Summary of simulation output for LHS 1976 (see text for details). Panels a: number of signals vs period; Panels b: detection probability vs period; Panels c: SDE distributions; Panels d: distributions of transit depths of the detected signals.
Figure 4.16: Summary of simulation output for LHS 417 (see text for details). Panels a: number of signals vs period; Panels b: detection probability vs period; Panels c: SDE distributions; Panels d: distributions of transit depths of the detected signals.
Chapter 4. Apache survey milestones

Figure 4.17: Summary of simulation output for LHS 3343 (see text for details). Panels a: number of signals vs period; Panels b: detection probability vs period; Panels c: SDE distributions; Panels d: distributions of transit depths of the detected signals.
Chapter 4. *Apache survey milestones*

68

The previous considerations allow us to relate directly the two comparison metrics whose properties we have analysed here, i.e. transit detection probability and phase coverage. We show in Figure 4.18, for the full stellar sample under investigation and using the simulation results based on light curves obtained with method m3, the detection probability as a function of the phase coverage, both averaged over the whole period range 0.5 − 5.0 days. Different symbols are used to show the trend of detection probability for the four regimes of transit depth simulated. For the definitions of transit detection probability and phase coverage provided above, the plot suggests a limit to the probability of detecting a transit, at least utilizing a real-life detection algorithm such as BLS, even when the phase coverage is almost 100%, a limit which becomes severe when the magnitude of the signal approaches the typical precision of the photometry. We believe that this effect is due to several factors, which include (but are not limited to): a) the way we actually define a detection: As discussed above, we adopt here a conservative definition, likely contributing to missing some of the correctly identified transit depths simply because the period did not fall within the stringent agreement constraints required; b) the fact that detection probabilities are averaged over period ranges which include values (close to integer and half-integer days) not optimally sampled in a ground-based, single-site campaign (see Figure 4.14); and c) the way we actually define the phase coverage: The choice of bin size and the prescription for the number of points in a bin required to define a specific phase covered can also impact the likelihood that the transit will be detected at that specific phase.

Similarly, Figure 4.19 shows the trend of detection probability as a function of phase coverage when limiting ourselves to the period range 0.5-1.0 days. Overall, for objects with good (> 50%) phase coverage, we would have had > 80% chances of detecting transiting companions with transit depths in the range 0.5% < td ≤ 2%. Given the estimated stellar radii for these stars, this translates in a sensitivity to companions with minimum radii in the range ∼ 1.0 – 2.2 R⊕. Figure 4.19 also shows how, when enough transits (≫ 3) are observed, even signals of magnitude comparable to the photometric precision can be reliably retrieved in our data.

As mentioned earlier, the results shown here cover the analysis of the simulation based on light curves derived with the m3 photometric analysis method. The results based on the other three intra-night methods (not shown here) are very similar while they worsen significantly when the full-period light curves are considered. This is an expected result because those methods produce more noisy light curves (see Section 4.3.3.1). However, such an effect is cause of no significant worries, as operationally transit-like events, whose duration for objects in short periods does not exceed the typical length of observations during one single night, are to be searched for in nightly-reduced differential photometric datasets.
Chapter 4. Apache survey milestones

To conclude, two points are worth mentioning further here. First, in a large-scale survey of thousands of M dwarfs, while some degree of prioritization of the targets based on their measured and/or inferred variability properties (based, e.g., on activity indicators and rotation information) will be possible, for many stars disregarding variability will simply not be an option. As for the two stars for which we successfully determined rotation periods, the probability of detection of a transit is not affected by the presence of the rotational modulation effect, as in both cases the time-scales of central transit events probed in our simulations (between 0.8 hr and 1.8 hr for LHS 3445 and between 0.4 hr and 0.9 hr for GJ 1167A) are significantly shorter than the observed rotational periods. Certainly, for much longer transit durations, or much shorter rotational periods, variability can instead become a matter of concern as it would directly interfere with a transit search. Rotational modulation on time-scales of 0.1 days is indeed observed (e.g., Irwin et al. 2011), and such issue will be addressed and dealt with in detail in the future. Second, variable degrees of correlated noise, either instrumental (tracking, flat-fielding), environmental (airmass, absorption, seeing) or astrophysical (colour, variability) in nature, can also affect our capability to detect transits. As shown by Pont et al. (2006) [105], red noise primarily impacts the significance thresholds with which

Figure 4.18: Average detection probability as a function of average phase coverage over the period range 0.5-5.0 days. Different symbols are used to show the trend of detection probability for the four regimes of transit depth simulated ($t_d = 2.0\%$: diamonds; $t_d = 1.5\%$: triangles; $t_d = 1.0\%$: squares; $t_d = 0.5\%$: crosses).
a transit event can be recovered from the data in a survey. As discussed in § 4.3.3.3, it constitutes a challenge to be able to fully characterize the relative role played by correlated noise sources on the potential of a photometric survey to detect statistically significant transit events, as this analysis will also depend on the details of the observed target populations.

4.3.8 Summary and conclusions

In this Section, we report results of a one-year long photometric monitoring campaign of a sample of 23 nearby ($d < 60$ pc), bright ($J < 12$) M dwarfs carried out in Italy at the Astronomical Observatory of the Autonomous Region of the Aosta Valley (OAVdA), using small-size (< 1-m class) telescopes. This survey was conceived as a necessary preparatory step towards a long-term search for transiting, small-radius planets around thousands of dM stars, which we are conducting at OAVdA with an array of five automated 40-cm telescopes. This ‘pilot study’ was designed to achieve two goals: 1) demonstrate the sensitivity to $< 4 R_\oplus$ transiting planets with periods of a few days around our program stars, through a two-fold approach that combines a characterization of the statistical noise properties of our photometry with the determination of transit
detection probabilities via simulations, and b) where possible, improve our knowledge of some astrophysical properties (e.g., activity, rotation) of our targets through a combination of spectroscopic information and our differential photometric measurements. At a technical level, the results we obtained during the pilot study are instrumental to the accurate design and fine tuning of several aspects of our upcoming photometric survey, such as the definition of the best observational strategy, the optimization of the target list, and the identification of improvements to be carried out on the pipeline for the photometric data reduction and time-series periodicity analysis.
Chapter 5

The APACHE survey architecture

In this Chapter I describe the architecture of the APACHE survey. The ’night’ operations (the actual data acquisitions) have officially started in July 2012 at the OAVdA site. The setup of the first year of operations closely follows the architecture described below.

For this work, I developed the majority of software setup (TEEPEE reduction and analysis pipeline, database, major improvements to the interface between RTS-2 and the APACHE telescope array).

5.1 Instrumentation and methodology: hardware setup

APACHE is composed of an array of five identical PRO RC 400 Officina Stellare 1 Carbon Truss 400-mm $f/8.4$ Ritchey-Chrétien telescopes, with a GM2000 10-MICRON 2 German Equatorial mount. All telescopes are equipped with a front-illuminated Charged Coupled Device (CCD) camera FLI 3 Proline PL1001E-2 (sensor Kodak KAF-1001E Grade 2, area $1024 \times 1024$ pixel$^2$, pixel area $24 \times 24$ $\mu m^2$, Quantum Efficiency $\sim 50\%$ in the I band) and a FLI filter wheel with a Johnson-Cousin standard set of filters. The CCD cameras are operated with a nominal gain of $2e^-/ADU$ (in $1 \times 1$ binning mode), with nominal read noise $NR = 9e^-/pixel^{-1}$, dark current $ND = 0.2e^-/pixel^{-1}s^{-1}$ at a temperature of $-40$ deg C and a full frame download time of $\sim 8.4$s through a USB 2.0 port. The CCD chip is cooled thermoelectrically by a two-stages system of Peltier

1 www.officinastellare.com/
2 www.10micron.com/
3 www.flicamera.com/
modules which allows to reach a maximum temperature gradient $\Delta T \simeq 70 \pm 0.1 \text{deg C}$ with respect to the ambient temperature. The chip temperature, recorded in the fits file headers, is typically stable to better than 0.5 deg C. During survey operations, the CCD chips are cooled to temperatures of $-40 \text{deg C}$. The typical values of thermal noise for our CCD camera remain well below $1e^{-}\text{pixel}^{-1}\text{s}^{-1}$ as far as the cooling temperature is set to $-25 \text{deg C}$. This telescope-CCD configurations have a Field of View (FoV) of $26 \times 26\text{arcmin}^2$ and a plate scale of $1.52\text{arcsec}\text{pixel}^{-1}$. In Table 5.1 we summarize the main characteristics of the telescope and camera systems.

5.2 Instrumentation and methodology: software interface for the hardware setup

The open source observatory manager RTS-2 (Remote Telescope Control, Kubanek 2010 [137]) constitutes the core for the high-level software control of the five-telescope system. Around RTS-2 we developed all the software necessary to manage a single night of observation in an automatic way. This means managing all aspects of the data acquisition (including dynamic scheduling of the observations), pre-processing and backup. Provisions have been made for the hardware of relevance for the APACHE operations control, data acquisition, processing centre and backup, which is necessary to accommodate the high data rate: two HP Z600 Workstation and one HP EliteBook 8740w Mobile Workstation as control PCs and workstations for data analysis, one HP Storage Works X1600 12TB as Network Attached Storage (NAS), connected together via Intranet 1 GB. Each observing station is equipped with a pc which is synchronized every 15 minutes via Internet to the Network Time Protocol (NTP) server of the Italian National Institute of Metrological Research. In the header of each frame, time is saved as Heliocentric Julian Date (HJD) with six significant digits. A single day of RTS-2 operations is divided into five parts: day, evening, evening dusk, night, morning dusk. Here I summarize the unsupervised tasks performed by RTS-2 in each step during a typical day of observations.

- The day starts at the end of the morning dusk part and it ends two hours before the evening dusk part. Obviously, during the day the dome is closed and the instruments are offline.

- The evening starts two hours before the evening dusk part. During the evening the dome is opened (in order to thermalize the hardware), the instruments are online and the CCD is cooled at the $-40 \text{deg C}$ temperature.
### Table 5.1: Summary of the main characteristics of the telescope and camera systems.

<table>
<thead>
<tr>
<th>Optical scheme</th>
<th>Telescope</th>
<th>Aperture (cm)</th>
<th>Focal ratio</th>
<th>FoV (arcmin²)</th>
<th>Plate scale (&quot;/pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflector Ritchey-Chrétien</td>
<td>f/8.4</td>
<td>40</td>
<td>1024 × 1024</td>
<td>21 × 21</td>
<td>1.52 (binning 1 × 1)</td>
</tr>
</tbody>
</table>
- The *evening dusk* starts when the altitude of the sun below the horizon is $0\,\text{deg}$ and it ends when the altitude of the sun below the horizon is $-10\,\text{deg}$. During the *evening dusk* calibration exposures (flat fields, darks) are executed. We acquire dark-frames many as possible in the dusk sessions at the exposures of 0, 5, 25 and 60 seconds. For the flat-fields acquisition, the system determined the best zone of sky respect to the sun position. The system choose the best exposure, exploring a temporal range between 3 and 10 seconds, in order to have $\sim 32500$ average ADU in the frame, for each filter.

- The *night* start at the end of the *evening dusk* and it ends when the altitude of the sun below the horizon is $-10\,\text{deg}$, before the *morning dusk*. The first task to be performed is the telescopes' focusing. This is made in an automatic way looking for the minimum of the parabolic function that described the behaviour of the FWHM of a single star as a function of the focuser’s position. The focusing procedure spend few minutes, after the scientific frames acquisition can start, according to the observing strategy. For each scheduled object we acquired three frames every $\sim 20$ minutes for at least 3 hours per night (it depends from the observability of the target). The genesis of this observing strategy is described in section 5.2.3. The exposure of each target is fixed a priori and it was calibrated, at the beginning of the survey, as a function of the spectral type and J magnitude. Every pointing (one for each set of three frames) is followed by a precise ($\pm 5\,\text{pixel}$) field centering procedure. The system uses the acquisition of a single 5 seconds frame in order to find the astrometric solution of the field, using the Astrometry.net\footnote{http://www.astrometry.net} software, and, consequently, to find the right pointing offsets.

Each scientific image is on-line pre-processed by the *watchdog* routine that performs an update of the image header, adding astrometry, time, instrumental setups and weather keywords, and make a storage copy on the NAS.

- The *morning dusk* is completely specular to the *evening dusk*. At the end of the *morning dusk* the system and all instruments are shut down and the dome is closed.

The procedures described above are robotic and fully managed by the system. Basically, the night observer initializes the RTS-2 system, opens and closes the dome of telescopes.

All scientific images are performed in the I-band. Naturally, the choice of filter was driven by red colours of the target sample

At the end of the survey first’ year the whole database comprised 238207 good images, corresponding to $\sim 2000$ GB of data.
5.2.1 Weather Monitoring

Each telescope has on-board temperature probes to monitor environmental and optical surfaces temperature and an hygrometer in order to avoid the dew formation on mirrors. Thermalization of mirrors is made through heaters and fans managed in a closed loop by the ATC electronic system. Observations limits are fixed at 90% of humidity for decreasing probability of electronic damages and keeping an high level photometry standard. We set up an external weather probe (AAG Cloud Watcher) for monitoring cloud coverage by an infrared sensor that measures the temperature of the sky, an anemometer and a variable capacitor to determine the presence of rain. These devices are connected with a custom alarm that prevents bad data taking or instrumental damages.

5.2.2 Operations’ monitoring

In order to maximize the data acquisition efficiency and avoid loss of observing time, we monitor the correct execution of operations for each telescope. This could be intended like a redundant check on telescope operations but it minimize errors provided by lack of communications between the various hardware and software components that, obviously, are not well managed by RTS-2. Basically, on any target acquisition, the system is checked for good pointing, right frames acquisition and well stored images, independently by the nominal procedures.

5.2.3 Observational strategy

One of the pilot study results was the assessment of sensitivity to transiting companions (of given radius and period) expressed in terms of detection probabilities and phase coverage (see Sec. 4.3.7. But the success probability of a survey depends from many other factors, such as the per-star planet frequency and the geometrical transit probability. All these factors make a detection really difficult and rare. In this context, the observational strategy plays a key role. What we wont by a good observational strategy is the right balance between the detection probability of a transit like signal (as we seen it depends from photometric precision, phase coverage, time sampling, ecc) and the number of objects that we are able to observe in a single session and consequentially, in a year. In other words, we need to minimize (respect the pilot study strategy) the losses in terms of photometric precision and phase coverage, maximizing the number of objects observed. To choose quantitatively the best observing strategy, we use the simulation tools developed during the pilot study (see Sec. 4.3.7). We evaluated the transit detection probabilities as a function the average phase coverage (in the periods
Figure 5.1: Detection probability VS Temporal Sampling. Different symbols are used to show the trend of detection probability for the four regimes of transit depth simulated: $t_d = 2.0\%$: diamonds; $t_d = 1.5\%$: triangles; $t_d = 1.0\%$: squares; $t_d = 0.5\%$: crosses.

Figure 5.1 shows the detection probability as a function of different temporal samplings and consecutive exposures, for a single object.

We tested some different temporal samplings (from 20 to 50 minutes) and some different numbers of consecutive exposures (from 1 to 5). The temporal sampling is here defined as the time interval between two consecutive pointings of the telescope at the same target while the number of consecutive exposures are the scientific images taken for each pointing. The Figure5.1 shows the detection probability as a function of different temporal samplings and consecutive exposures, for a single object.

The results for the whole sample depends from the phase coverage accordingly with those found in the pilot study.
Although it seems there are several equivalent configurations, now we are adopting the 3 seconds exposures every 20 minutes strategy, which will be eventually relaxed after the first season of the survey.

5.3 Instrumentation and methodology: data reduction

TEEPEE (Transiting ExoplanEts PipElinE) is the data reduction pipeline, written by the author, for the data reduction. TEEPEE is a software package written in IDL\(^5\), which utilizes publicly available software from the Astronomy Users’ Library\(^6\) as well as external contributed FORTRAN and C++ routines and it is devoted to automatically carry out ensemble differential aperture photometry on an user-specified target. The develop of TEEPEE is started during the site characterization study and, in particular, during the pilot study. Here, I describe the operating versions of TEEPEE, highlighting the changes derived from the work done in the pilot study.

TEEPEE is organized in three main, sequential modules:

- image calibration (including dark and bias subtraction, and flat fielding)
- image alignment and photometric processing (aperture photometry)
- differential photometry and de-trending of the target with respect to a chosen set of reference stars

The heart of TEEPEE is the third block. It performs those operations which are necessary to correct, to a high degree of reliability, for systematic effects which cause the degradation of the precision of the photometric measurements, and it produces photometric light curves for hundreds of stars detected in the field.

5.3.1 Image calibration

The scientific image produced by the CCD is normally affected by a series of systematic errors: the readout noise, the bias, the dark current and the inhomogeneity of the sensor.

The readout noise is a consequence of the imperfect operation of physical electronic devices and it is generated in phase of the sensor reading An estimate of the read noise is declared by the manufacturer. In particular, the amplifier can’t do a perfect job of

\(^5\)IDL is a commercial programming language and environment by ITT Visual Information Solutions. http://www.ittvis.com/idl/

\(^6\)http://idlastro.gsfc.nasa.gov/contents.html
measuring the charge in the clump of electrons generated by the photons on sensor. Typically, it gives the right value on average, but with some random scatter. "Readout noise" is simply the scatter around the true value. It is not possible correct this effect in a single image, but it is possible to taking it into account in the errors' model (see Section 5.3.2) using an estimate of the read noise declared by the manufacturer.

The bias is caused by a constant voltage applied to the CCD detector; ideally it should appear as a constant value added to the image also with a exposure of 0 seconds. Furthermore, even when left in the most complete darkness, CCD will accumulate electrons that will eventually be counted a "light" when the image is read. This is a temperature effect: electrons can be kicked out of place by the heat from the CCD itself; this is the dark current. The dark current is not negligible in APACHE data: although the overall level is low, there are some hot pixels and the hot pixels do not seem to be completely stable from night to night. The dark-frame subtraction is the way to minimize image noise from bias and dark current. This is possible because the component of image noise, known as fixed-pattern noise, is the same from shot to shot: noise from the sensor, dead or hot pixels. It works by taking some image with the shutter closed. The master dark frame is an average of several dark frames and it can then be subtracted from subsequent images to correct for fixed-pattern noise. Dark frames are taken every night with four fixed exposure time (see Section 5.2). For each scientific image, TEEPEE chose the master dark frames with the closest exposure.

All the pixels making the CCD surface are almost similar; however, they present some variations of sensitivity: some of them will convert the light photons more efficiently into electrons than others. Also, the filters and lenses that are between the sky and the CCD can accumulate some dust particle, that will cast a shadow on the CCD. The result of all these effects is that the image of a uniformly illuminated target (like the sky during the day or twilight, or the inside of the dome) will not appear uniform on the CCD. To correct this effect, we take several, much as possible during the twilight, images of a flat field (see Section 5.2). The twilight flat fields are taken on alternating sides of the meridian to mitigate most large-scale illumination effects (the detector rotates by 180 degrees relative to the sky upon crossing the meridian, so gradients tend to average out). Each flat image is normalized and then 3-sigma averaged in order to obtain a single frame, the "master flat". The master flat map the sensitivity variations across the CCD. Our photometric analysis shows that this standard method of taking twilight flats produces a result that is not uniform to better than the several percent level.
The dark-subtracted scientific images are then divided by the flat-field, leading to the reduced image.

\[
(Reduced\ image) = \frac{\text{Rawimage} - \text{Masterdark}}{\text{Masterskyflat}}
\]  
(5.1)

### 5.3.2 Image aperture photometry

The aperture photometry of digitized scientific image is the procedure to measure the brightness of an object without including possible contributions from contaminating sources such as sky background and other stars. It basically consist to put down a computer generated aperture over the grid of data and add up the counts within the aperture. The usual method to estimate the sky background is meaning the count from other regions close to the object. TEEPEE performs three main step in order to build a catalogue of stars from an astronomical image. First, it founds the \((x, y)\) photocenters’ coordinates of each star recorded by the CCD sensor. This operation is performed by SExtractor\(^7\) (Source Extractor). SExtractor\(^8\) is a program that builds a catalogue of objects from an astronomical image. The \((x, y)\) coordinates are the necessary output to the Aper\(^9\) routine that computed aperture photometry in a series of twelve concentric apertures starting from a base aperture radius \(r\) of 4 pixels to a radius \(r\) of 8 pixels for each star. At the end of these process, TEEPEE combine magnitude, coordinates and astrometry informations to align each image to a reference frame and to make and indexes’ catalogue of each good star in the frame. We build the virtual reference frame from the coordinates and magnitude of the 2MASS catalogue. The last task of this section is the photometric error estimation for each star in the field and for each frame. We calculate a theoretical uncertainty estimate \(\sigma_{th}\), in magnitudes:

\[
\sigma_{th} = \frac{2.5 \ln 10}{N_\ast} \times \sqrt{N_\ast + \sigma_{scint}^2 + n_{pix} \cdot (N_{sky} + \sigma_{ND}^2 + \sigma_{NR}^2)}
\]  
(5.2)

where \(N_\ast\) is the number of photoelectrons from the source collected by the sensor, \(N_{sky}\) is the number of photons from the sky background per pixel, \(\sigma_{scint}\) is the anticipated scintillation noise (Young 1967), \(\sigma_{NR}\) is the noise component from readout noise and \(\sigma_{ND}\) is the noise component from the dark current (See Sec.5.1). The \(n_{pix}\) term take into account the number of pixels in the apertures. The \(\sigma_{scint}\) term is defined by:

\[
\sigma_{scint} = 0.09 \frac{AM^{3/2}}{D^{2/3} \sqrt{2t}} e^{-\frac{t}{8}}
\]  
(5.3)

---

\(^7\)http://www.astromatic.net/

\(^8\)http://www.astromatic.net/

\(^9\)http://idlastro.gsfc.nasa.gov/
where $AM$ is the air-mass, $D$ the diameter of the telescope in cm, $t$ the exposure time in seconds and $h$ the altitude of the observatory in km.

All these procedures are well suited for moderately crowded stars fields, where the probability of mixing light from superposed (blended) stars of comparable magnitude is low.

### 5.3.3 Differential photometry

With the aperture photometry (see Sec. 5.3.2) we basically obtain a raw measurements of the light flux. We are not interested to know an absolute measurements of the star light in a precise moment but its behaviour in time. This behaviour is called ”light curve”. The raw light curve, so called by utilizing the raw flux measurements, is heavily affected by a lot of systematic effects as the change of airmass, sky condition (e.g. clouds and veils), sky transparency, seeing, etc. One of the best technique to avoid these problems is the differential photometry. It basically consists to compare the raw light curve of the target star with a set of reference stars, following the simple idea that, at the first order, the systematic effects afflict all stars of the fields in the same way.

Our differential photometric method performs as follows. For each frame $i$, we use as reference magnitude $M_{rif}^i$ the average magnitude of the $n$ reference stars:

$$M_{rif}^i = \frac{\sum_{k=0}^{n} M_k^i}{n}.$$  

(5.4)

$M_{rif}^i$ is then subtracted to the magnitude of the user-defined target $M_{target}^i$, obtaining the difference $\Delta M^i = M_{rif}^i - M_{target}^i$. The procedure is iteratively repeated for all the reference stars, using as new references the remaining $n - 1$ stars. This procedure is also iteratively repeated for all detected field stars using as references the $n$ reference stars chosen for the target.

Obviously, the selections of the set of reference stars plays a key role. Here, I describe our procedure used to select the best reference set, in an automatic way. We first take care of picking up reference objects on a CCD sub-frame, avoiding the chip edges, affected by vignetting which is not fully corrected for during flat-fielding. Second, we use two efficient methods, based on the Burke et al. (2006) [138] prescription, for choosing the appropriate set of references for the target. The first method ($m_1$) selects the subset of reference stars which minimizes the RMS of the differential light curve of the target; the second method ($m_2$) selects the subset of references which minimizes the RMS of the differential light curve of each potential reference star. Both methods are then applied to all twelve apertures measurements in order to choose the optimal one, on the basis of a
minimum-RMS prescription for the target light curve. During the pilot study, we tested both methods that were statistically consistent (see pilot study results Sec. 4.3.3.2).

Next, we apply the SysRem trend filtering algorithm (Tamuz et al. 2005), to correct for unknown linear systematic effects in the light curves produced with methods \( m_1 \) and \( m_2 \). The two resulting light curves are dubbed as produced with methods \( m_3 \) and \( m_4 \), respectively. The whole procedure thus produces a total of four light curves for each target, for each night.

With a similar process, four additional light curves are produced for each target over the whole timespan of the observations, which we dub ‘full-period’ light curves (\( m_{1f}, m_{2f}, m_{3f}, m_{4f} \)). A single set of reference stars is chosen over the entire observation window. In place of SysRem, a different trend filtering algorithm is utilized, TFA (Kovács et al. 2005), which after some experimenting was found to be better suited for the full-period light curve analysis.

### 5.4 Input catalogue

The APACHE Input Catalogue (AIC) is composed of 3323 potential stars that we have defined with care first by applying some constraints on the target visibility and taking into consideration the properties of the telescopes’ field of view. Then, through the cross-correlation with several catalogues, used to collect as much as information possible concerning the physical properties of the stars, a ranking list of the selected targets has been produced, with each star receiving a score which determines its observing priority. Here we shortly describe the procedures and criteria followed for building the AIC and assigning the observing priorities to each star. The AIC is an ‘open’ catalogue, with the observing priorities of the targets that are constantly updated on the basis of new results published in literature which provide fresh information about the physical properties of the stars. Infact, being a targeted survey, as opposite to a wide-field survey, a point of strength of APACHE is the a-priori knowledge of as much details as possible of the stars of interest. All the M dwarfs in the AIC come from the catalogue of 8889 bright cool stars compiled by [139]. This high number of stars does not impose us with many restrictions in defining the criteria for selecting a reasonable number of APACHE targets. Taking into account the geographical coordinates of the OAVdA and some environmental constrains, we first selected a sub-list of targets characterized by 1) positive Declination, 2) at least 45 days in a row of visibility and 3) which are observable for at least 3 hours/night in a row, i.e. with an altitude > 40 degrees during this time range. We consider the timespan of 1.5 months as the minimum requirement for monitoring a target during an observing season, in order to increase the chance of detecting possibly more
than one single transit and confirm the existence of the event. For a similar reason, by observing a target for at least 3 hours per night there is the chance of observing a complete or partial transit event. By applying these constrains, the number of potentially good stars reduces to 4806. In a second step, we discarded 378 stars for which a very bright object falls in the field of view of the APACHE telescopes. This avoids blooming effects which are not predictable an could contaminate the flux in the circular apertures used for performing the photometry of the targets and comparison stars. The selection has been made by cross-correlating the coordinates of the potential targets with those of the objects listed in the Bright Star Catalog\(^\text{10}\), used as a source of basic astronomical and astrophysical data for stars brighter than magnitude 6.5, considering a search radius of 11 arcmin (i.e. placing the target in the center of the FoV, as it actually happens during the observations). The third and final selection filter we have applied concerns the need of guarantying a minimum number of potentially good comparison stars to be used in the ensemble differential photometry of the targets. APACHE will observe hundreds of fields for which it is not possible to identify a-priori those stars bright enough and with a low level of photometric variability that could be safely used as comparison. For this reason, we have selected only those stellar fields which contain at least three potentially good reference stars, identified on the basis of their flux in the J band from the 2MASS catalogue relative to that of the target. Defining \(F_J\) the flux of a target M dwarf in band J, our requirement for a field star to be considered a possible comparison is that its J-flux is at least 0.3\(F_J\) and less than 1.5\(F_J\). This selection criteria filters out \(\sim1100\) M dwarfs, leading to a final list of 3298 possible targets. As expected, much of the M dwarfs discarded are among the brightest in the original list and with a FoV poor of stars, for which there is not a suitable number of comparison stars. To create a ranking list, we searched for published information about several physical properties of the selected stars. First, it is important to have an accurate spectral classification of each star, to be quite confident that we are observing an early-to-mid type M dwarf, which represents the spectral range of interest for APACHE. [139] provide a tentative classification based on photometric information, which is necessarily only approximative and subject to a non-negligible uncertainty. In a following work, [140] improved the spectral classification of the brightest stars (J<9) of the [139] catalogue by analysing newly acquired low-resolution spectra. In addition to this work, we cross-checked the PMSU catalogues (Reid et al. [141], Hawley et al. [142], Gizis et al. [143]), those of [60], [144], [52], [145], [146] and the list of M dwarfs published in the DwarfArchives.org Web site\(^\text{11}\) (and references therein). M dwarfs with a reliable spectral classification have a higher position in the ranking list. The observing priority rises if the star is known to be a slow rotator, based on measurements of its projected rotational velocity.

\(\text{10}\)http://tdc-www.harvard.edu/catalogs/bsc5.html

\(\text{11}\)http://spider.ipac.caltech.edu/staff/davy/ARCHIVE/index.shtml
$v\sin i$, if it is has a low chromospheric activity, as measured by the equivalent width of the H$\alpha$ line (positive: emission; negative: absorption), and it is not known as a bright X-ray source, by verifying its inclusion in the ROSAT bright source all-sky catalogue [147]. A slow rotator and a star with a low activity level has to be preferred because, for the spectroscopic follow-up required to measure the mass of a transiting planetary candidate, the measurements of the radial velocity variations are less affected by the jitter due to stellar activity. To get information about $v\sin i$, we have cross-checked the catalogues of [148], [145] and [146]. We have flagged as slow rotators stars with a measured $v\sin i < 15$ m/s. The information about the equivalent width of the H$\alpha$ line (sign and values) has been get from the PMSU catalogues (Reid et al. [141], Hawley et al. [142], Gizis et al. [143]), and those of [60], [144], [58], and [52]. Very recently, we have also performed a cross-check with the spectroscopic catalog of M dwarfs from the LAMOST pilot survey (Yi et al. [149]). For each spectrum of this catalogue, spectral subtype and equivalent width of the H$\alpha$ line are provided. To favour those stars not yet monitored by other observing campaigns, we checked the inclusion of the M dwarfs in past or ongoing high-precision Doppler surveys (Wright et al. [50], Bonfils et al. [150]) excluding those already known to host planets that do not show transits. Moreover, because APACHE puts first the observation of stars with possibly no companions, which could represent a complication for a RV follow-up in case of transit detection, we have excluded from the schedule spectroscopic binaries (e.g. Shkolnik et al. [52]). To exclude the presence of very close visual (and possibly physical) companions, we have also taken into consideration the results of the AstraLux M dwarf multiplicity survey (Janson et al. [151]), which uses high-resolution imaging to look for multiplicity in the range 0.08-6 arcsec from the star. Finally, by cross-checking the 2MASS catalogue we have flagged those targets that have at least one field star of similar magnitude within a radius of 15 arcsec. Taking into consideration the typical seeing at our site (median value 1.7 arcsec) and the pixel scale of the CCD cameras (1.5 arcsec/pixel), in some circumstances such a close object could blend with the target and represents a source of complications for the aperture photometry. we do not exclude a-priori these problematic fields but, if selected for the observations, they are visually inspected and considered for rejection. A final parameter that we have considered is the expected number of observations of each target by the Gaia satellite, which is expected to guarantee a very relevant contribution to the extrasolar science (Sozzetti et al. [152], Sozzetti [153]). According to the scanning law of Gaia, it is possible to calculate the number of expected transits over a certain field after the nominal 5 years of mission. Targets with a number of Gaia observations greater than 100 must have higher priorities when building the APACHE schedule because they will be characterized in particular by a very accurate trigonometric parallax, from which it will be possible to measure the stellar intrinsic luminosity, mass and radius. Should APACHE discover transiting planets around bright targets, these updated values for the host stars
(especially the radius) imply a precise determination of the planet radius. Combined with exquisite RV data (e.g. those collected with the HARPS-N spectrograph), the bulk density of the planet can be precisely determined (e.g. Anglada-Escudé et al. [154]). The final number of targets in the AIC has been increase from 3298 to 3323 by including stars in common with the GAPS project, which uses the high-resolution HARPS-N spectrograph mounted at the Italian Telescopio Nazionale Galileo (TNG) (e.g. [155] [156]). GAPS is a large-programme which address several topics of the science of the extrasolar planets, in particular a large sample of M dwarfs is monitored with HARPS-N to search for small mass planets. Since Summer 2012 a strategic synergy has been established between APACHE and GAPS, with the result that dozens of M dwarfs are observed both photometrically and spectroscopically during the same epochs. We have attributed to this shared sample the highest observing priorities. Because GAPS gives priority to bright M dwarfs, we have included in the AIC also stars that were initially discarded, and to increase the number of their potential comparison stars, we have decided to observe these targets in V-band.

5.4.1 Mass and radius estimation

The mass and radius estimation plays a key role in the interpretation of a transit event and/or radial velocity measurement. A good estimate of the planet mass and radius inevitably depends to the estimate of star mass and radius. To work out the radii of the stars of the AIC, we resorted to the Stefan-Boltzmann Law

\[ L_{\text{star}} = 4\pi R_{\star}^2 \times \sigma T_{\text{eff}}^4 \]

where \( L_{\text{star}} \) is the luminosity, \( R_{\star} \) is the stellar radius, \( \sigma \) is the Stefan-Boltzman constant and \( T_{\text{eff}} \) is the effective temperature. First of all, we had to address the issue that only 170 stars out of 3323 have well-measured trigonometric parallax [139], whilst 3153 have estimates for visible (V band) and near-infrared (J, H, K bands) indexes. First, we calculated the \( T_{\text{eff}} \) (Effective Temperature) for these 3153 stars via \( T_{\text{eff}} \) and color indexes relations provided by Casagrande et al. ([46]). These relations are reckoned to be fairly metallicity-independent for disk stars. Different approach like that one based upon stellar spectra put forth by Lépine et al. ([140]) are less viable because spectroscopic data for our stars are in shorter supply. Anyway, we compared the \( T_{\text{eff}} \) derived by the two methods for a subset of 390 stars, for which both photometric and spectroscopic data are available, finding an overall good agreement (see Figure 5.2).

The intrinsic brightness was then derived for the 170 stars with the good parallax measures, exploiting the empirical bolometric corrections given by Casagrande et al. 2008 [46]. Subsequently, we worked out an empirical \( T_{\text{eff}} \) vs \( L_{\text{star}} \) relation fitting these data. We utilized this relation to get a \( L \)-value for the subset of 3153 stars. Finally, we calculated the radii inverting the Stefan-Boltzman law.
The masses were calculated using the inverted mass-radius relation supplied for low-mass stars by Bayless and Orosz 2006 [104].

In the two panels of Figure 5.3, we show the distributions in radius $R_{\text{Sun}}$ and mass $M_{\text{Sun}}$ for the Apache Input Catalogue in Sun radii and Sun masses, respectively.

The Figure 5.3 highlights the presence of late K stars in our sample. Taking into account the size of the Input Catalogue, we decide to keep these objects as low priority targets for two reasons: first, the uncertainty over the mass radius estimations; second, the potentially goodness of these targets to study the transition regime from neptune-size to earth-size planets.
Figure 5.3: Up: radius distribution of the AIC. Down: mass distribution of the AIC
Chapter 6

Results from the first year of APACHE survey

In this Chapter, I present data and results from the first year of observations of the survey APACHE.

The goals of this first analysis were:

- the evaluation of the photometric precision for each target over a typical data taking period of 2 months;
- the periodicity analysis in search of transit-like signals;
- the periodicity analysis in search of sinusoidal-like signals induced by the presence of dark spots (and/or bright active regions) unevenly distributed on the stellar photosphere;
- the evaluation of the APACHE ensemble sensitivity to the transit of small-size planets (> $2R_\oplus$) with periods in the range 0.5 − 5 days;
- the survey very first evaluation of the planet occurrence rate for periods in the range 0.5 − 5 days.

This work also carries the double benefit of providing: (i) the quasi realtime monitoring of the data quality and (ii) the opportunity for continuous improvements to data acquisitions/operations and to data reduction/analysis algorithms.

For the results presented in this Chapter, I carried out ~50% of the photometric observations taken with the OAVdA telescopes, developed and constantly upgraded the data reduction and analysis pipeline TEEPEE, and finally obtained the differential light...
Chapter 6. First results

6.1 Summary of observations

At the end of the first year of observations the whole database comprised 238207 good images, corresponding to $\sim 2$ TB of raw data (i.e. CCD images on the targets and excluding calibration frames). In the four panels of Fig. 6.1, we summarize the main results in terms of number of frames, observing nights, time span and phase coverage for each of the 241 observed targets.

During the first 369 days of operations we observed for 183 nights, i.e. 49.6\% of photometric usable nights.

With reference to our best observing strategy (see Section 5.2.3) we were able to predict the performance after the first year. We estimated the observations of 227 targets of which $\sim 40\%$ with a phase coverage $> 50\%$ for the period range 0.5-5 days. After the first year, we actually observed 241 targets of which only $\sim 20\%$ with a phase coverage $> 50\%$, in the period range 0.5-5 days. The result is encouraging but some changes in the scheduling are needed in order to increase the mean phase coverage over the sample.
6.2 Photometric precision

The single point photometric precision of a light curve is the quantity that has the most relevant impact on the probability of detecting an astrophysical signal (such as a transit event), and of determining its statistical significance. We expect, as shown in the eq.5.2, that the single point uncertainty is driven by the Poisson distribution in the absence of any kind of correlated noise (white noise regime). On the other end, we can estimate the single point photometric uncertainty as the RMS (Root Mean Square) of a photometric light curve. The Central Limit Theorem expresses the fact that a sum of many independent (or weakly dependent) and identically distributed random variables (in our case the photometric points) will tend to be distributed according to the normal (Gaussian) distribution, independently of the actual form of the probability distribution of the single measurement. Therefore, the RMS is the estimation of a mean single point uncertainty and can directly be compared with the individual predictions derived from eq.5.2.

6.2.1 Global analysis

For an efficient analysis, we decided to take into account, for each object, only the $m4f$ light curve (see Section 5.3 for the light curves extraction). This is a reasonable choice for the following reasons:

- One of the results (see Section 4.3.3.2 for details) of the pilot study shows that the two methods ($m1/m1f$ or $m2/m2f$) for the selection of reference stars are statistically equivalent. The main difference is on the number $n$ of reference stars usually selected by the two methods: $m1/m1f \rightarrow n1 < n2 \leftarrow m2/m2f$.

- During a typical observing night, the stars sample different regions of the detector when observed at positive or negative hour angles because of APACHE’s German Equatorial mounts. In other words, the mounts flip when the targets cross the meridian. This affects the photometric accuracy of the reference stars or, in general, the anonymous stars in the imaged field of view. The resulting effect is a systematic in the light curves that has the shape of a step function. The same systematic effect affects the MEarth survey that utilizes a similar setup. The problem is described in various papers (Irwin et al, 2011 [107], Berta et al., 2013 [157]) and it is addressed with an ad hoc solution: the fitting of the light curves with a step function. On the contrary, we try to remove this systematic using a detrending algorithm taking into account the presence of such effect on most of the field stars’.

The SysRem detrending algorithm (see Section 5.3.3) is not able to correct for
such types of trend, so $m3f$ is not really useful in this context. TFA (see Section 5.3.3) is better suited because the algorithm looks for systematic effects (regardless of their shape) that affects also, better only, the selected comparison stars. In this case the systematic is induced by the target-reference-star comparisons as the target is always pointed within few pixel from the flipping center. The efficiency of TFA strictly depends on the number of comparison stars and the number of points of the light curve. For this reason TFA is better suited for the $m2/m2f$ method. Anyway, we tested both approaches fitting $m2f/m4f$ (without detrending/with detrending) light curves with the step function. We found a typical step of $\sim 7$ mmag for yhr $m2f$ light curves that decreased to $\sim 2$ mmag for the $m4f$ light curves.

Taking these considerations into account, we consider the $m4f$ light curves as those most reliable for subsequent analysis. Before the RMS computation, the $m4f$ light curve is filtered with a sliding median filter and 5–sigma clipped. In Section 6.4.3.1 we discuss the marginal impact of these filters over the signals detection.

\begin{figure}[h]
\centering
\includegraphics[width=0.6\textwidth]{ensemble_rms_distribution.png}
\caption{Distribution of the RMS for each target. The continuous line shows the filtered full period RMS distribution, while the dotted line shows the photometric uncertainty distribution as predicted by our error model.}
\end{figure}

In Figure 6.2 we show the distributions of the RMS values and the photometric uncertainty distribution as predicted by our error model for each individual target. Visual
inspection of the figure already shows qualitatively the generally good agreement between the error model and the empirical estimation of it. This is confirmed by comparing our median/mean photometric precision of $\sim 7/\sim 8$ mmag with our median/mean theoretical single point uncertainties of $\sim 6/\sim 7$ mmag. In the next section, we investigate the nature of the excessive noise present in some of our targets.

### 6.2.2 Correlated (red) noise analysis

As mentioned above, the RMS of the whole light curve can be assumed to represent the average uncertainty of each data point under a fundamental hypothesis: the photometric measurements are assumed to be uncorrelated (white noise regime). At the millimag level and in a relatively high-cadence time series this is generally untrue. There are many effects that can produce a correlated photometric measurements (red noise regime): changing airmass, atmospheric conditions, telescope tracking (and relative flat field errors) and the intrinsic variability of the targets. These effects introduce some covariance between data points.

The presence of red noise can have a rather significant impact on the statistical analysis of the data. As an example, let us consider a light curve, with no evident variability, consisting on $N$ flux measurements $f_i$ in a fixed time interval, with uncertainties for each data point $\sigma_i$ equal to the RMS of the whole light curve $\sigma_0$. We calculate the mean of the $N$ flux measurements $f_{\text{mean}}$. The uncertainty on $f_{\text{mean}}$, $\sigma_{f_{\text{mean}}}$, is then the error of the mean of $f_i$. Using the expression for the standard deviation of the mean under the assumption of white noise ($\sigma_w$), this uncertainty is:

$$
\sigma_{f_{\text{mean}}} \equiv \sigma_w = \frac{\sigma_0}{\sqrt{N}}. \quad (6.1)
$$

We note as the uncertainty on $f_{\text{mean}}$ decreases with the square root of the number of points $N$.

The equivalent of Eq. 6.1 in presence of red noise is:

$$
\sigma_{f_{\text{mean}}} \equiv \sigma_t = \sqrt{\sigma_w^2 + \sigma_r^2} = \sqrt{\frac{\sigma_0^2}{N} + \frac{1}{N^2} \sum_{i \neq j} C_{ij}}, \quad (6.2)
$$

where the $C_{ij}$’s are the covariance coefficients between the $i$-th and $j$-th measurements (e.g., [105]).
Now, we can estimate $\sigma_t$ from the light curves, taking into account the red noise, following the methodology described in [105], and then determine the red noise component from Equation 6.2:

$$
\sigma_r = \sqrt{\frac{1}{N^2} \sum_{i \neq j} C_{ij}} = \sqrt{\sigma_t^2 - \frac{\sigma_0^2}{N}}.
$$

(6.3)

For each star in our sample, we computed $\sigma_t$ over an interval $\delta t = 90$ minutes. We decided to investigate the red noise contribution to the error budget over this time interval, as an illustrative example. Note that this interval actually corresponds to a typical transit duration for few-days period planets orbiting mid and early M dwarfs, and thus it’s of particular interest for the purpose of this study.

We show in Figure 6.3 the various contributions to the photometric error budget in our dataset, expressed as a function of the target $J$ mag. The results shown correspond to the error analysis for the light curves obtained using method $m4f$. The outcome of this specific study is only marginally dependent on the chosen method for differential photometry, as we discuss in 4.3.3.3.

The top left panel of Figure 6.3 shows the RMS as a function of the $J$ magnitude. We expected a trend of increasing RMS with magnitude in the data, while the plot highlights an opposite situation. This scenario is due to the effect of scintillation that becomes significant for very short exposure time, i.e. of a few seconds (see Equation 5.3). The exposure time for each target was selected in order to keep the SNR approximately constant, those preventing saturation of the objects.

The top right panel of Figure 6.3 shows the theoretical single point uncertainty as a function of the $J$ magnitude and shows well the effect of scintillation in our sample, as described above. Similarly, the bottom right panel of Figure 6.3 confirms a negative correlation of $\sigma_w$ with magnitude, definitely excluding correlated noise as a possible cause of this extra noise.

The bottom left panel shows $\sigma_r$ as a function of the $J$ magnitude. It reports only targets with $\sigma_r > 0.001$ mag. It shows a weak trend, if any, decreasing with $J$ mag. This is not unexpected because for faint objects $\sigma_w$ dominates the total error budget, and $\sigma_r$ becomes increasingly more difficult to determine at the faint end.

We can consider $\sigma_r$ as a extra dispersion of a set of points over the interval $\delta t$. This dispersion is the amplitude of the systematics that will remain in the signal regardless of the number of points measured during the interval and it is a function of the dominant wavelengths of the systematics with a peak at the corresponding wavelengths slightly
above twice the duration of $\delta t$ (see Pont et al. 2006 [105]). Schematically, the correlated noise can be thought of as consisting of three components: short frequencies that will tend to average out over the duration of the $\delta t$, long frequencies that will not vary between the $\delta t$ and the neighbouring measurements, and frequencies around $2\delta t$ that will introduce strong residuals. Our studies confirm this scenario. For example, among the intrinsically variable stars in our sample (the points framed by a square in the bottom left panel in Figure 6.3; see also Section

Summarizing, 50% of our sample follows the typical white noise regime, with photon noise dominating except at the bright end where the scintillation becomes important. For the rest of the sample, the red noise has a mean value of $\sim 3\text{mmag}$ with weak dependence on the $J$ magnitude.

### 6.3 Photometric variability: periodicity analysis

In this section, I describe the key elements of the analysis segment devoted to the astrophysical characterization of the stars themselves. We look for signals in the data
which are periodic in nature, and which would indicate either the presence of a transiting companion or which could be interpreted as intrinsic to the target (e.g., due to chromospheric activity).

6.3.1 Searching for transit-like events

While showing an exact periodicity, a highly non-sinusoidal transit event cannot be modelled efficiently using standard approaches based on finite sums of sinusoidal components, such as the Discrete Fourier Transform method (e.g., [158]), or on period-finding techniques which minimize the scatter in smoothed light curves, such as the Phase Dispersion Minimization algorithm (e.g., [159]). As is now common practice, in order to detect periodic transit events in our data we use a method of least squares fits of step functions to a folded signal corresponding to a grid of trial periods, as realized in the Box-fitting Least Squares (BLS) algorithm ([136]).

In the analysis (performed in IDL with the exception of a C++ implementation of the BLS period search), we utilized the m4f full period light curves of each target. For each light curve we apply a high-pass median filter with a sliding temporal bin of 9.6 hours and a 5σ-clipping. All light curves were inspected for transit-like signals using a dense grid of 10000 trial periods in the range 0.4-5.0 days. We divided the folded time series into 300 bins and we evaluated the signal residue SR (i.e., the BLS power spectrum) of the time series using these binned values at any given trial period. We fixed the fractional transit length in the range of 0.1 − 0.01.

To characterize the Signal Detection Efficiency (SDE) we introduce:

\[
SDE = \frac{SR_{\text{peak}} - \langle SR \rangle}{sd(SR)},
\]

where \(SR_{\text{peak}}\) is the SR at the highest peak, \(\langle SR \rangle\) is the average, and \(sd(SR)\) is the standard deviation of SR over the frequency band tested. The values of SDE can also be used to quantify the probability of a detection being a false positive. The actual behaviour of SDE depends on many factors, some intrinsic to the algorithm (e.g. number of bins, number of trial periods) and others depending on the quality of the data and of the signal (e.g. S/N), however the statistical studies Kovács et al. (2002) [136] indicate a value of \(SDE \sim 5\) above which the false positive probability is low (< 10%). Above such value the SDE does not depend on the input parameters of BLS (Figure~5 of Kovács et al. 2002 [136]). The plot in Fig.6.4 shows the SDE distribution of a subsample of observed target. We select all those objects with phase coverage > 10%. The black filled part of the histogram is referred to those periods related to 0.25, 0.5, 1 days alias.
The remaining objects with $SDE > 5$ were visual inspected. From this inspection emerges that these signals are principally related to correlated noise induced by sky background variations (see e.g. Berta et al. 2012 [160] for similar behaviours). This is expected because BLS, by assuming a flat out-of-transit light curve, have a tendency to fold up any (non-planetary) time-correlated structures into seemingly significant candidates. Probably, the adopted detrending techniques (see Section 5.3) doesn’t remove these systematic effects due by the substantial spectral type difference between our targets and comparison stars (see e.g. Berta et al. 2012 [160] and reference therein).

Basically, we conclude that the study of the BLS spectra does not show any significant periodicity within the period range investigated, for all objects in our sample.

In the Sec6.4, we aim to interpret APACHE’s zero planet detection in the light of the our planet ensemble sensitivity, the geometric transit probability and, finally, the statistical rate of planets occurrence around the dM stars.

### 6.3.2 Characterizing stellar rotation

By observing the same star over a timespan of several weeks, one relevant parameter that can be searched for by analysing the photometric light curve is its rotational period. This
can be derived if the light curve of the star shows a periodical sinusoidal-like modulation that can be ascribed to the presence of dark spots (and/or bright active regions) unevenly distributed on the stellar photosphere. Looking for evidence of rotational modulations in the photometric data of the APACHE M dwarfs is an important task to be carried out for two main reasons. First, it is important for characterizing the astrophysical behaviour of cool stars by increasing the existing statistics in literature and, taking into account the main goal of the APACHE survey, it is also necessary for possibly discard from further observations those stars showing short rotation periods. In fact stars which are fast rotators are expected to show high activity levels that represent an important source of noise in the RV measurements that can complicate the detection of a spectroscopic signal produced by a planetary companion. We searched for sinusoidal-like periodicities in the light curve of every M dwarfs by first determining the Lomb-Scargle (L-S) [161], [162] and the CLEAN [163] periodograms of binned (3 consecutive points per bin) data through the use of publicly available IDL codes of these algorithms.

The CLEAN procedure deconvolves the spectral window function from the ”dirty” discrete Fourier Transform, removing the couplings between physical frequencies and their aliases or pseudo-aliases. We then considered for folding the light curve up to ten periods with the highest peak in the periodogram and with a False Alarm Probability (FAP) < 0.01 calculated according to the definition in [164]. After a visual inspection of the folded data, we assess the statistical significance of a ’good candidate’ rotational period performing a classical bootstrap analysis of the data, as described in [165] and used in several other studies (e.g. [166], [167], [168]). 10000 fake light curves are generated by shuffling (with replacement) the differential magnitudes over the total timespan and keeping fixed the timestamps. A L-S periodogram is then calculated for each bootstrapped light curve, using the same settings adopted for the original photometric time-series. The number of fake light curves for which the maximum peak in the periodogram, for any tested frequency, has a power greater than or equal to that of the most significant period found in the original data, defines the probability that the detected modulation is not real: every random permutation of the data should remove any existing periodicity which is not due to pure white noise, leaving unaffected the frequency contributions due to noise and aliasing. Thus, if none of the 10000 permutations results in a periodogram with power peaks greater than that of the most significant period found in the original light curve, the astrophysical origin of the periodic signal can be considered having a FAP < $10^{-4}$. It should be reminded that an appropriate definition of the FAP as calculated via bootstrap analysis is not a simple task. A conservative interpretation is that there is 1% of probability that the tested period is not of astrophysical origin, by assuming a relative uncertainty of $1/\sqrt{10000}$. Furthermore, we have to consider that this
bootstrap method avoid the bad interpretation of simply white noise data but nothing
said about the presence of correlated noise that could produce false positive signals.

All of the five APACHE light curves discussed here have been analysed and processed
as described, confirming the existence of a periodic signal that can be attributed to a
rotational modulation. Figures 6.5 - 6.7 show the light curves of these targets, folded
according the best periods found.

The star GJ 3628B (APACHE 1551) (Fig. 6.5) has been monitored for slightly more
than 5 months, from 2013/1/22 to 2013/6/2 for a total of 37 nights, collecting 1328 use-
ful data points. The star is classified as M3.5V according to the Palomar/MSU nearby
star spectroscopic survey [141, 142], while M5V according to the tentative classification
of [139]. For this star we propose P = 6.5487847 days as the candidate rotation pe-
riod, as found by applying the CLEAN algorithm. The bootstrap analysis produced a
FAP=0.0001, which validate at the highest confidence level the found periodicity. The
scatter of the residuals after the subtraction of a sinusoidal fit is sensibly reduced, but
looking at the folded curve it can be noticed that the nights between phases 0.3-06 are
not well fitted by a simple sinusoid. This could be the evidence of rapidly evolving
starpots/active regions on the stellar photosphere, that changed the shape of the light
curve during the timespan of the APACHE observations.

The target I22006+2715 (APACHE 2961), confirmed as M0V according the new spec-
troscopic catalog of bright M dwarfs of [140], is particularly interesting (Fig. 6.6). Its
data showed quite clearly evidence for the existence of a short-period modulation visu-
ally inspecting the light curves produced on nightly basis by the TSE pipeline. This
finding pushed me to process the data collected over the timespan (from 2013/6/15 to
2013/8/16, for a total of 18 nights and 440 useful data points) more quickly than for
other stars, revealing the nature of fast-rotator for this target, with a rotation period
of 0.524357 days found by the CLEAN algorithm and a folded light curve with a shape
very close to a sinusoid. W have then applied a bootstrap analysis to asses the signif-
icance of this periodic signal, which resulted in a $FAP = 0.0001$. This circumstance
led us to promptly discard the star from further observations aimed at searching for
transiting planets. Moreover, this star was selected for spectroscopic observations with
the spectrograph HARPS-N within the GAPS programme (reference?), but not yet
observed. Following the finding in the APACHE data, it has been removed from the
GAPS schedule because it could reveal a high level of jitter in the RV data difficult to
remove and then preventing from detecting variations due to a small-mass planetary
companion.

Finally, I23513+2344 (APACHE 3276) shows a clear rotational modulation with a period
P = 3.180986 days, as determined by a L-S analysisys (Fig. 6.7). This target, classified
Chapter 6. First results

Figure 6.5: Light curve of the star APACHE 1551 folded according the best period found (P \( \sim \)6.5 days) that can be interpreted as the rotational period of the star. Superposed (red points) is the best-fit sinusoidal curve with the same periodicity. The residuals after the removal of the sinusoidal function are shown in the nested, small plot.

as M4V by [144], has been observed for 29 nights over a period of \( \sim \)2 months (from 2012/8/26 to 2012/10/24), collecting a total of 1682 useful data points. The bootstrap analysis resulted in a FAP \( \sim 10^{-4} \) for this period. Even if a sinusoidal fit well represents the global variation of the star flux, it can be noticed (as for GJ 3628B) that the residuals of the sine fit show an irregular pattern (despite their scatter is nearly half of the original light curve. Even in this circumstance, this can be explained by the presence of one or more starspots/active regions rapidly changing their size.

6.3.3 Photometric variability: flares analysis

Flaring events, a powerful indicator of the presence of strong stellar magnetic fields, are known to be relatively common among late-type stars, typically lasting between a few minutes up to several hours, and producing increases in the observed flux of up to several magnitudes. Considering the sampling at which the APACHE telescopes work (i.e. the same target is re-pointed every \( \sim \)20 minutes), it was not clear a-priori if the survey could be suited for detecting stellar flares, which usually have decay times of few minutes. This happened during the pilot study, when the small sample of M dwarfs was monitored with a much shorter time-sampling (see Sec.4.3.6). After the first season, we were able to detect stellar flares for 3 different targets. The stars involved are GJ 3143
Figure 6.6: Light curve of the star APACHE 2961 folded according the best period found (P \(\sim\) 0.52 days) that can be interpreted as the rotational period of the star. Superposed (red points) is the best-fit sinusoidal curve with the same periodicity. The residuals after the removal of the sinusoidal function are shown in the nested, small plot.

Figure 6.7: Light curve of the star APACHE 3276 folded according the best period found (P \(\sim\) 3.18 days) that can be interpreted as the rotational period of the star. Superposed (grey points) is the best-fit sinusoidal curve with the same periodicity. The residuals after the removal of the sinusoidal function are shown in the nested, small plot.
Chapter 6. First results

(APACHE 331), an M3.5V star according to the Palomar/MSU survey, I09108+3127 (APACHE 1427), an M4V star according to [139], and GJ 4196 (APACHE 2864), that [140] classify as an M1.5V, in agreement with the Palomar/MSU classification as M1V. The photometric time-series displaying the sudden increase in the stellar luminosity typical of a flare are showed in Fig. 6.8. Each plot refers to a single night, and each data point in the light curve is the average of three consecutive exposures. We did not detect more than on flare episode for each star. Each flaring event has been characterized following the approach of [129], that is by solving a non-linear least-squares problem with a set of consecutive photometric measurements starting from the time of recorded maximum brightness $t_0$ and an exponential model function of the type:

$$\text{mag}(t) = A\times\exp\left[-\frac{(t-t_0)}{\tau}\right] + \text{mag}_0$$

with adjustable parameters $A$, the peak magnitude of the flare relative to the non-flaring magnitude, $\tau$, the decay time-scale, and $\text{mag}_0$, the differential magnitude of the star before the flare. The least-squares solution was obtained using the IDL code curvefit, which is an implementation of the Levenberg-Marquardt method [169], by looking for the best-fit function of the type $f(t) = A\times\exp(B\times t)+C$, after providing guesses for the model parameters $A$ equal to the peak magnitude, $B$ in the range 50-85 days$^{-1}$, and $\text{mag}_0$ equal to the average differential magnitude of the post-flare event. The best-fit solutions are overplotted in each panel of Fig. 6.8. The flare-events are characterized by decay times which fall in the range 17-26 minutes, comparable to the average repointing time interval of the APACHE telescopes. GJ 4196 (APACHE 2864) appears also in the target list of the GAPS programme, and the information about its activity levels has been considered by the GAPS responsibles for a possible exclusion of the target from further spectroscopic investigations. This star was reported to have the H-alpha spectral line in absorpion [143], with an average equivalent line width of -0.0361 nm, but in contradiction it is also classified as a variable of BY Draconis-type by the AAVSO International Variable Star Index VSX. BY Draconis-type variables, as stated in the VSX document describing the variable star type designations, ’are emission-line dwarfs of dKe-dMe spectral type showing quasi-periodic light changes with periods from a fraction of a day to 120 days and amplitudes from several hundredths to 0.5 mag in V. The light variability is caused by axial rotation of a star with a variable degree of non-uniformity of the surface brightness (spots) and chromospheric activity. Some of these stars also show flares similar to those of UV Cet stars, and in those cases they also belong to the latter type and are simultaneously considered eruptive variables’. Our finding supports the latter classification. For GJ 4196 (APACHE 2864) we have analysed 1254 useful data collected in 36 nights and over a period of 3 months, without revealing any detectable periodical light changes.
Figure 6.8: The three panels show the light curve for the APACHE targets 331, 1427 and 2864 (from top to bottom, black '+' symbols), observed on the nights 2013/09/13, 2013/04/13 and 2013/07/29 respectively. There are clearly visible impulsive, flaring events detected in the curse of the night. Red '*' symbols represent the best-fit solution after applying a flare-model (see text for details), with the resulting best-fit function overplotted.
6.4 Limits to transiting companions

While no transit event has been recorded in our dataset, as reported in Section 6.3.1, the photometric measurements obtained for our sample can be utilized to carry out simulations aimed at determining what sensitivity to transiting companions (of given radius and period) we achieved on a star-by-star basis.

6.4.1 APACHE’s Ensemble Sensitivity

We define APACHE’s ensemble sensitivity, $S(R, P)$, as the number of planets with a particular radius $R$ and orbital period $P$ that APACHE should have found if all stars in our sample had one planet with that radius and period. The formalism adopted is the same of Berta et al., 2013 [157]. This is a reasonable choice driven by the similarity of the two experiments and the possibility to a ”quick and easy” comparisons of the results. We calculate the ensemble sensitivity as

$$S(R, P) = \sum_{i=1}^{N_*} \eta_{\text{tra},i}(R, P) \times \eta_{\text{det},i}(R, P)$$

(6.5)

where the functions $\eta_{\text{tra},i}(R, P)$ and $\eta_{\text{det},i}(R, P)$ are the per-star transit probability and the per-star detection efficiency, respectively. Our targets span a factor of four in stellar radius, and each star has its own observational coverage and light curve properties. As such, we calculate each of these functions individually on a star-by-star basis. The sum is over the $N_* = 150$ stars in the sample with phase coverage greater than 10%.

6.4.2 Transit Probabilities = $\eta_{\text{tra},i}(R, P)$

The first factor in eq.6.5, $\eta_{\text{tra},i}(R, P)$, is the probability that a planet is geometrically aligned such that transits occur in the system. We take $\eta_{\text{tra},i} = R_*/a$, where $R_*$ is the stellar radius and $a$ is the semimajor axis; this is the probability that a planet in a circular orbit would have an impact parameter $b < 1$. From Kepler’s third law and assuming the planet’s mass is negligible, $R_{\text{star}}/a = [3\pi/(GP^2\rho_*)]^{1/3}$ where $G$ is Newton’s constant. We estimate the mean stellar density $\rho_*$ from the mass and radius, derived for all targets in the APACHE Input Catalogue (see Sec. 5.4.1). Non-zero eccentricity can increase (or decrease) transit probabilities for planets with Earthward periastrons (or apoastrons), but the net effect for blind short periods transit surveys is small on average.


6.4.3 Transit Detection Efficiencies $\eta_{\text{det},i}(R,P)$

The second factor, $\eta_{\text{det},i}(R,P)$, is the probability that a planet aligned to transit would be detected by the APACHE survey. This probability strongly depends from the cadence of observations and the noise properties.

In order to evaluate the sensitivity of each of the M dwarfs in our sample to transiting companions of given period and given radius, provided the primary size from Sec. 5.4.1 (or equivalently inducing a transit signal of given depth), a large-scale simulation was performed by injecting synthetic transit signals into the differential light curves of our targets. We utilized these data under the hypothesis that the procedure adopted to extract it from the raw images does not affect significantly the signal. This simplificatory assumption is often adopted when dealing with such issues (e.g. Berta et al., 2013 [157]). The transit signal injection was performed on the $m_4f$ sigma clipped and median filtered light curve obtained with the procedure for differential photometry described in § 4.3.1.

We are aware that the specific choices for these elements of the simulation setup can in principle impact our findings to some extent. For this reason, we estimate quantitatively the effects on our capability to recover transit signals due to the application of trend- and median- filtering algorithms with a specific simulation over a sub-sample of 18 targets (see Sec. 6.4.3.1).

The goal of the present global analysis is simply to gauge our ability to detect transit events as a function of the system parameters.

The input parameters (and their ranges) to the simulation, which was entirely carried out in IDL, we determined based on considerations taking into account the temporal sampling, total timespan, and typical photometric precision of our data. We generated:

- 100 random, uniformly distributed periods in the range $0.5 - 5$ days;
- 10 random, uniformly distributed phases for each period;
- 3 amplitudes of the transit signal depth $t_d$ (in flux units: 0.015, 0.01, 0.005) for each period and phase. The four values of transit depth simulated correspond to companion radii in the ranges $2.3-9.5$, $1.9-7.8$, and $1.3-5.5 \, R_{\oplus}$, respectively, given the sizes of the primaries.
- two fixed orbital inclination $i = 90^\circ$ and $i < 90^\circ$ for impact parameter $b = 1$
- perfectly circular orbits ($e = 0.0$);

Consequently, the orbital radius parameter was derived from $M_*$ and from Kepler’s third law under the assumption $M_p \ll M_*$ ([134]). We recognize that allowing two orbital
inclination might change to some extent the details of the detection probability results. This allow us to study our sensitivity to non-central (in the limit of quasi-grazing) transits.

With the above simulation setup, 6000 synthetic transit light curves were generated for all targets with phase coverage $> 10\%$ ($N_{\text{star}} = 150$ targets) using the Mandel & Agol ([135]) algorithm. For the purpose of this analysis limb-darkening effects are essentially irrelevant and in order to speed up execution they were turned off.

Figure 6.9 summarizes the process of injection of a transit signal in a differential light curve. The top panel shows the synthetic folded light curve, without limb-darkening, produced by the Mandel & Agol ([135]) formalism for a transit with a fractional depth of 1.5$\%$ and a period of 0.984718 days. The center panel shows the actual phased light curve of the APACHE’s target 1820, while the bottom panel shows the combined light curve in phase. The datasets shown in the upper and lower panels are both fed to BLS.

In the analysis with the transit search algorithm of a given light curve, BLS was run using 10000 period steps in the range $0.5 - 5$ days, the folded time series was divided into 300 bins, and the signal residue ($SR$) of the time series (see [136]) was determined using these binned values. We fixed the fractional transit length to lie in the range $0.1 - 0.01$. We considered as detections all those signals for which BLS returned a period $T'$ such that $|T' - T_{\text{in}}|/T_{\text{in}} < 0.01$, where $T_{\text{in}}$ was either the actual period of

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.9.png}
\caption{Top: synthetic folded light curve with $d = 1.5\%$ and a period of 0.984718 days. The times of observation are those of the APACHE’s target 1820. Center: the actual phased light curve of the APACHE’s target 1820. Bottom: the combined light curve for the APACHE’s target 1820.}
\end{figure}
the injected signal, or half that period, or twice that period. We are aware of the fact that this requirement might be somewhat stringent: In practice we do not consider as detections transits identified with the correct depth, but with an incorrect period due to, e.g., the sampling properties of the light curve. We adopted this more conservative approach because we are not completely safe that a single transit detection scheme ([160]) guarantees a suitable relationship between detection probability and observing time of follow-up, taking into account only the APACHE follow-up facilities.

We define the Transit Detection Efficiencies $\eta_{\text{det},i}(R,P)$ as the relative number of periods that are detected by BLS with respect to the total number of injected periods. This sensitivity highly depends on how well we understand the physical masses and radii of the stars in our sample. As mentioned above, we adopt the stellar parameters as derived in Sec.5.4.1. These parameter estimates are not necessarily the best estimates currently possible, and introduce uncertainties into our statistical analysis. The considerations in [139] ensure that the sample is free of giants, but the roughly 30% uncertainties in stellar radii directly translate into comparably large uncertainties in the planetary radius to which a given transit depth corresponds. Furthermore, some stars may be contaminated with unresolved binaries. The additional flux from a binary companion dilutes transit depths, as well as perturbing the stellar parameters that we infer from (spectro)photometry. These two effects have impacts on both $\eta_{\text{tra},i}$ and $\eta_{\text{det},i}$ in this discussion we do not try to quantify these effects because an optimal statistical description it is not necessary for the first year of data, far away to be a significant sample.

### 6.4.3.1 Transit Detection Efficiencies from the raw light curves

As mentioned above, injecting synthetic transit signals into the differential light curves is a simplificatory assumption because we are not able to a priory quantitatively estimate the impact over the signal of the algorithms used to obtain the differential light curve. Therefore, we injected the synthetic signals directly in the raw data to understand in what way the procedure adopted to extract the differential magnitude from the raw images, the trend-filtering algorithm, the median filter and the sigma-clipping affect the signals. The goals of the present analysis are not simply to gauge our ability to detect transit events as a function of the system parameters but also to evaluate the quality of the calibration procedures against transit signal detection thresholds. The choice of running the simulation for only eighteen objects is driven by the high computational costs necessary to extract the differential photometry six thousand times per object. We select the sub-sample looking for object with different phase coverage and photometric precision, thus allowing us to gauge the sensitivity of the transit detection probability to
different photometric regime. The simulation setup (e.g. the input parameters) was the same of that described above as well as the definition of the Transit Detection Efficiencies $\eta_{\text{det,}i}(R, P)$ (the relative number of periods that are detected by BLS with respect to the total number of injected periods). Therefore, it is natural defining the mean loss rate as the mean difference between the two Transit Detection Efficiencies. We found that this mean loss rate is $\sim 10\%$ with a weak dependence from the input parameters $R$ and $P$ and the characteristics of the targets, such as the $RMS$ and the phase coverage.

### 6.4.4 APACHE’s Ensemble Sensitivity: results

The final sensitivity estimates as a function of orbital period $[S(R, P)]$ are shown in Figure 6.10.

![Figure 6.10](image.png)

**Figure 6.10**: APACHE’s ensemble sensitivity as a function of planet radius and orbital period. These sensitivity curves include both the geometrical probability that a planet would transit and the probability that we would detect transits that do occur. The window function (solid line) shows how sensitive APACHE would be if all that was required to detect a planet were at least three in-transit points from at least three transit events.

APACHE’s ensemble sensitivity is a function of planet radius and orbital period as given by Eq. 22. These sensitivity curves include both the geometrical probability that a planet would transit and the probability that we would detect transits that do occur. The window function (solid line) shows how sensitive APACHE would be if all that was
required to detect a planet were at least three in-transit points from at least three transit events. The detectability of planets near or below $2R_\oplus$ depends strongly on the inferred stellar radius of each star. In contrast, $>4R_\oplus$ planets are often far above the detection threshold, so sensitivity depends weakly on the exact stellar radius.

The Figure 6.10 indicates the number of planets found by APACHE if every M dwarf hosted exactly one planet. Clearly, this is an unrealistic statement. By inverting a Binomial distribution, we can translate this statement into a 95% confidence upper limit of the planet occurrence ($planetstar^{-1}$).

The planet occurrence estimates as a function of orbital period $[S(R, P)]$ are shown in Figure 6.11.

![Figure 6.11: Planet occurrence upper limits as a function of planet radius and orbital period.](image)

A physical interpretation of these upper limits is actually not significant for many reasons. The sample is far away to be complete and, as mentioned above, the uncertainties derived from a wrong estimation of the stellar parameters are not quantified. Furthermore, our detection capability could be significantly improved if we consider as a detection, with robust statistical methods (e.g. Berta et al., 2012 [160] [157]), single transit events. Nevertheless, the results highlight very well the performances of the first year of operations. With one observational season we were sensitive for planets with radius $>6R_\oplus$ in the range 0.5 – 3 days. We were moderately sensitive for planets with radius...
in the range 0.5 – 1.5 days and we were poorly sensitive for planets with radius $\sim 2R_\oplus$. We can use the machinery developed in this work to address how to increase the “planets/year” that APACHE could find. We discuss this in the last chapter.

### 6.5 Summary and Conclusion

We report results of the first year of the APACHE survey. At the end of the first year we investigate our sensitivity for $> 2R_\oplus$ transiting planets with periods of a few days around our program stars. To do this we followed a two-fold approach that combines a characterization of the statistical noise properties of our photometry with the determination of transit detection probabilities via simulations.

At a technical level, the results we obtained during the first year of routine operations are instrumental to the accurate updating of several aspects of our photometric survey, such as the fine-tuning of the best observational strategy, the optimization of the target list, and the identification of improvements to be carried out on the pipeline for the photometric data reduction and time-series periodicity analysis.

Our main findings can be summarized as follows:

- **Photometric precision.** We achieve a typical RMS photometric precision of $\sim 7$ mmag. We also carried out an analysis of the impact of correlated (red) noise on time-scales of $\sim 90$ min, which showed that 50% of the targets are affected by the read noise and only 30% of them are affected by a correlated noise that produce a RMS greater than 2 mmag. This result reveals that our data are poorly affected by short-term correlated systematics. Furthermore, the good agreement of the RMS photometric precision with the theoretical single point uncertainties ($\sim 6$ mmag) highlight our capability to manage the majority of the systematic effects that plague our photometry.

- **Searching for transit-like events.** We searched for periodic transit-like events in the photometric dataset for each target using the BLS algorithm. No such signal was recovered for any target.

- **Characterizing stellar rotations.** The light curves of our program stars were inspected for evidence of periodic signals of approximately sinusoidal shape, which could be interpreted as due to the presence of rotating spots on the stellar photosphere. Looking for evidence of rotational modulations in the photometric data is an important task to characterize the astrophysical behaviour of cool stars and, taking into account the main goal of the APACHE survey, to potentially discard from further observations those stars showing short rotation periods. In fact stars which are fast rotators are expected to show
high activity levels that represent an important source of noise in the RV measurements that can complicate the detection of a spectroscopic signal produced by a planetary companion. For this reasons, such studies can provide a good benchmark for the RV spectroscopic surveys, such the GAPS project.

For there stars in our sample, GJ 3628B, I22006+2715 and I23513+2344, we found clear evidence of a periodicity in the light curve ascribable to such effect.

• **Flares analysis.** The light curves of our program stars were also inspected for evidence of short-term, low-amplitude flaring events. Flaring events are a powerful indicator of the presence of strong stellar magnetic fields and are known to be relatively common among late-type stars, typically lasting between a few minutes up to several hours. Also in this case, flares events might indicate high activity levels of the stars that represent an important source of noise in the RV measurements.

We detected flaring events in the differential photometric measurements of GJ 3143, I09108+3127 and GJ 4196 with an approximately equal decay time of $\sim 20$ minutes.

• **Limits to transiting companions.** We utilized the photometric measurements obtained for our sample to carry out simulations aimed at determining what sensitivity to transiting companions (of given radius and period) $S(R,P)$ we achieved on a star-by-star basis. $S(R,P)$ depends from $\eta_{\text{tra},i}(R,P)$ and $\eta_{\text{det},i}(R,P)$ the are the per-star transit probability and the per-star detection efficiency, respectively. While $\eta_{\text{tra},i}(R,P)$ is simply determined with geometrical considerations, to determine $\eta_{\text{det},i}(R,P)$ we carried out large-scale simulations of transit signals (of periods in the range $0.5 \leftarrow 5$ days and depths in the range $0.5\% \leftarrow 1.5\%$ in flux units) injected in the actual (reduced) photometric data for all objects with a phase coverage (in the range $0.5 \leftarrow 5$ days) $> 10\%$. A total of 6000 light curves were analysed for each target using a real-life transit events search algorithm (BLS). The study of the BLS transit recovery rates and overall performance for a sub-sample of stars with good, fair, and poor phase coverage highlighted the capability of BLS to identify the correct period (when multiple transits were observed) even for signals with depth close to the typical photometric precision of the data ($\sim 5$ mmag).

We defined the Transit Detection Efficiencies $= \eta_{\text{det},i}(R,P)$ as the relative number of periods that are detected by BLS with respect to the total number of injected periods. At the end of the first observational season we were sensitive for planets with radius $> 6R_\oplus$ in the range $0.5 \leftarrow 3$ days. We were moderately sensitive for planets with radius $\sim 4R_\oplus$ in the range $0.5 \leftarrow 1.5$ days and we were poorly sensitive for planets with radius $\sim 2R_\oplus$. 
Chapter 7

Astrometric detection of giant planets around nearby M dwarfs: the Gaia potential

In this Chapter, I gauge the Gaia potential for precision astrometry of exoplanets orbiting an actual sample of thousands of known dM stars. For this purpose, we carried out detailed numerical experiments. In Section 7.2 I describe the adopted simulation setup, and in § 7.3 we present the statistical and numerical tools used to analyse the simulated datasets. Section 7.4 is devoted to the analysis of the simulation results. Finally, I summarize in Section 7.5 our findings and provide concluding remarks.

This work has been published as a paper (Sozzetti, Giacobbe, Lattanzi et al., 2013 [170]).

For this work, I was principally involved in the development of the statistical and numerical tools used to analyse the simulated datasets.

7.1 Introduction

ESA’s Cornerstone mission Gaia, launched DECEMBER 20, 2013, will carry out a magnitude limited ($V \leq 20$), all-sky astrometric survey (complemented by on-board photometric and partial spectroscopic information) that is bound to revolutionize our understanding of countless aspects of astronomy and astrophysics within our Milky Way, and beyond (e.g., Perryman et al. 2001). The global impact of Gaia micro-arcsecond-level ($\mu$as) astrometric measurements in the astrophysics of planetary systems has been addressed in the past (e.g., Lattanzi et al. 2000; Sozzetti et al. 2001, 2003; Casertano et al. 2008; Sozzetti 2011). However, those studies only provided general metrics
for gauging detectability thresholds as a function of planetary properties (orbital elements, masses), using solar-like stars as the reference primaries. In addition, only brief mentions were made of the potentially huge levels of synergy between Gaia astrometry and other ongoing and planned exoplanet search and characterization programs. The approach adopted to carry out the analysis, particularly at the level of single- and multiple-planets orbital solutions, was still affected by some caveats and simplifying assumptions (e.g., only partial treatment or complete neglect of the problem of identifying adequate starting values for the non-linear fits). Finally, the Gaia astrometric performance, described in those works through a simple Gaussian single-measurement error model, has further evolved. A more realistic error model, which takes into account e.g. the dependence on magnitude, ought to be utilized.

We revisit the topics of planet detection and characterization with Gaia relaxing some of the above assumptions, and focusing on the sample of nearby low-mass M dwarf stars for which Gaia, as one by-product of its all-sky survey, will deliver precision astrometry down to the $V = 20$ magnitude limit.

The main thrust of this work is two-fold. First, we will gauge the Gaia potential for precision astrometry of exoplanets orbiting an actual sample of thousands of known dM stars within $\sim 30$ pc from the Sun (Lépine 2005). We will then express Gaia sensitivity thresholds as a function of system parameters and in view of the latest mission profile, including the most up-to-date astrometric error model. The analysis of the simulations results will also provide insight on the capability of high-precision astrometry to reconstruct the underlying orbital element distributions and occurrence rates of the planetary companions. These results will help in evaluating the expected Gaia recovery rate of actual planet populations around late-type stars.

Second, we will investigate some elements of the synergy between the Gaia data on nearby M dwarfs and other ground-based and space-borne programs for planet detection and characterization, with a particular focus on: a) the potential for Gaia to precisely determine the orbital inclination, which might indicate the existence of transiting long-period planets; b) the ability of Gaia to accurately predict the ephemerides of (transiting and non-transiting) planets around M stars, and c) its potential to help in the precise determination of the emergent flux, for direct imaging and systematic spectroscopic characterization of their atmospheres with dedicated observatories from the ground and in space.
Figure 7.1: Top left: distance distribution of the LSPM M dwarf sub-sample. Top right: The corresponding magnitude distribution in Gaia’s G band. Bottom left: the derived stellar mass distribution. Bottom right: number of individual Gaia field transits for the sample.

7.2 Simulation Scheme

The simulation of Gaia observations follows closely the observational scenario described in Casertano et al. (2008). I refer the reader to that source for details. Here I describe and discuss the changes/upgrades made to that setup.

1) In the representation of the Gaia satellite, the latest nominal Gaia scanning law was utilized, with the two fields of view separated by a basic angle of 106.5 deg, with a spin rate of 60 arcsec s\(^{-1}\), a solar aspect angle between the direction to the Sun and the satellite’s spin axis \(\xi = 45\) deg, and a precessional period of the spin axis of 63 days. Details on the scanning geometry of the Gaia satellite can be found in, e.g., Lindegren (2010) and Lindegren et al. (2012). The nominal mission duration \((T = 5\) yr) was adopted.

2) The actual list of targets encompasses 3150 low-mass stars (in the approximate range \(0.09 - 0.6\) \(M_\odot\)) within 33 pc from the Sun (Lépine 2005) from the proper-motion limited LSPM-North Catalog (Lépine & Shara 2005). For convenience we are referring to this sample collectively as M dwarfs, even though some of them
have estimated masses more compatible with those of late K dwarfs. This subset of the LSPM catalog (dubbed LSPM sub-sample hereafter) is not complete within the identified volume limit, with as much as 32% of stars missing out to 33 pc (Lépine 2005). However, the choice of this catalog over, for example, the more recent Lépine & Gaidos (2011) all-sky catalog was driven by our interest to choose a volume-confined sample (so that distance effects in the detectability of astrometric signals can be more simply taken into account). Using visual and infrared magnitudes available for the sample, we utilized the color-magnitude conversion formulae of Jordi et al. (2010) to obtain G-band magnitudes in the Gaia broad-band photometric system. The LSPM sub-sample results to have an average \( G \approx 14.0 \) mag. The Delfosse et al. (2000) mass-luminosity relations for low-mass stars were then utilized to obtain mass estimates for all our targets. The LSPM sub-sample results to have an average \( M_\star \approx 0.30 M_\odot \). In the four panels of Fig. 7.1 we show the distributions in \( G \) mag, distance \( d \), mass \( M_\star \), and number of Gaia field transits (individual field-of-view crossings) as a function of ecliptic latitude \( \beta \) for our LSPM sub-sample. The dependence of the number of Gaia measurements with \( \beta \), with the maximum in correspondence of \( \beta = \xi \), is a result of the adopted scanning law (see e.g. Lindegren et al. 2012 for details).

3) The \( G \)-band magnitudes were then used to derive the resulting along-scan single-measurement uncertainty \( \sigma_{AL} \) using an up-to-date magnitude-dependent error model (e.g., Lindegren 2010). The presently envisioned gate scheme to avoid saturation on bright stars (affecting some 20% of \( G < 12 \) mag M dwarfs) was included in the calculation (see Fig. 7.2), but provisions were not made for the representation of charge-transfer-inefficiency effects in the measurements. Single-measurement errors are typically \( \sigma_{AL} \sim 100 \) \( \mu \)as. Given the typical magnitude of the Gaia positional uncertainties involved in the simulations, the astrometric ‘jitter’ induced by spot distributions on the stellar surface of active M dwarfs (e.g., Sozzetti 2005; Eriksson & Lindegren 2007; Makarov et al. 2009; Barnes et al. 2011) was considered to be negligible, and therefore not included in the error model.

4) The generation of planetary systems proceeded as follows. One planet was generated around each star (assumed not to be orbited by a stellar companion), with mass \( M_p = 1 M_J \), orbital periods were uniformly distributed in the range \( 0.01 \leq P \leq 15 \) yr and eccentricities were uniformly distributed in the range \( 0.0 \leq e \leq 0.6 \). The orbital semi-major axis \( a_p \) was determined using Kepler’s third law. All other orbital elements (inclination \( i \), argument of pericenter \( \omega \), ascending node \( \Omega \), and epoch of pericenter passage \( \tau \)) were uniformly distributed.

1Where available, Hipparcos parallaxes are used, photometric distance estimates are otherwise utilized using the Lépine (2005) values.
within their respective ranges (for the inclination $\cos i$ was uniformly distributed).

The resulting astrometric signature induced on the primary was calculated using the standard formula corresponding to the semi-major axis of the orbit of the primary around the barycenter of the system scaled by the distance to the observer:

$$\varsigma = \left(\frac{M_p}{M_\star}\right) \times \left(\frac{a_p}{d}\right).$$

With $a_p$ in AU, $d$ in pc, and $M_p$ and $M_\star$ in $M_\odot$, then $\varsigma$ is evaluated in arcsec. Note that $\varsigma$ corresponds to the true perturbation size only in the case of circular orbits. It is in general only an upper limit to the actual magnitude of the measured perturbation when projection and eccentricity effects are taken into account (e.g., Sozzetti et al. 2003; Reffert & Quirrenbach 2011).

### 7.3 Statistical and Numerical Analysis Tools

The tools utilized in the analysis of the simulated Gaia astrometric data have already been described elsewhere (Casertano et al. 2008). We briefly recall here their main features. First, statistically robust deviations from a single-star model, indicating the presence in the observations residuals of the perturbation due to a companion with a given level of confidence, are identified through the application of a $\chi^2$-test or $F$-test (low probabilities of $P(\chi^2)$ or $P(F)$ signifying likely planet, and unlikely false positive). Then,
Chapter 7. The GAIA potential

orbital fits to the data are carried out, using a Markov Chain Monte Carlo (MCMC)-driven global search approach to the identification of good starting guesses for the orbit fitting procedure that combines a period search with a local minimization algorithm (Levenberg-Marquardt). Details on the overall algorithm performance as applied to large datasets of synthetic Gaia observations produced within the context of the Gaia Data Processing and Analysis Consortium (DPAC) will be published elsewhere. As described in Casertano et al. (2008), the resulting Gaia observable, the one-dimensional coordinate $\psi$ in the along-scan direction of the instantaneous great circle followed by Gaia at that instant, will then be modeled as $\psi(\alpha, \delta, \mu_\alpha, \mu_\delta, \pi, A, B, F, G, P, e, \tau)$, where the five standard astrometric parameters correspond to the actual positions ($\alpha, \delta$), proper motion components ($\mu_\alpha, \mu_\delta$), and parallax ($\pi$) of each target M dwarf as provided in Lépine (2005), while $A, B, F, G$ are four of the six Thiele-Innes elements (Green 1985). Planetary masses are derived from the best-fit orbital elements assuming perfect knowledge of the stellar primary mass and utilizing the approximation of the mass-function formula (valid in the limit $M_p \ll M_\star$):

$$M_p \simeq \left( \frac{a_\star^3 M_\star^2}{\pi^3 P^2} \right)^{1/3},$$

(7.1)

with $M_\star$ in solar-mass units, $P$ in years, $\pi$ and $a_\star$ (the semimajor axis of the orbit of the central star around the barycenter) both expressed in arcseconds.

7.4 Results

7.4.1 Detection Probabilities

The probability of a planet being detected is obtained, following Lattanzi et al. (2000) and Casertano et al. (2008), via a $\chi^2$-test of the null hypothesis that a star is single. By setting a probability threshold $P(\chi^2) \leq 0.001$ (i.e., a confidence level of 99.9%), we find that 2704 stars (85% of the sample) are identified as variable. The detected systems have an `astrometric signal-to-noise ratio' $\zeta/\sigma_{AL} \gtrsim 3$. The ratio of the perturbation size to the single-measurement accuracy had already been shown to be of use by Sozzetti et al. (2002) to study the main trends in astrometric planet detection probabilities and the limits for accurate orbit determination. For comparison, both Sozzetti et al. (2002) and Casertano et al. (2008) found that $P(\chi^2) \leq 0.05$ (i.e., a confidence level of 95%) would enable detection of astrometric signatures with $\zeta/\sigma_{AL} \gtrsim 2$. As already discussed in Lattanzi et al. (2000) and Casertano et al. (2008), the standard experiment to gauge

---

2http://www.rssd.esa.int/gaia/dpac
the false alarm rate in the presence of pure white noise due to the adopted statistical threshold for detection gave the expected results (e.g., 1% of false positives for a $\chi^2$-test with a confidence level of 99%).

The behaviour of detection probabilities (at the $\varsigma/\sigma_{AL} \geq 3$ level) with stellar mass and distance for the LSPM sub-sample at hand is shown in Figure 7.3. The left panel shows the trend of detection efficiency with distance from the Sun marginalizing over all primary mass values, while the right panel presents the trend with stellar mass marginalizing over distance. We find that detection efficiency decreases linearly with distance and for the LSPM sub-sample detection rates of $\approx 60\%$ are still achieved at the distance limit of $\sim 33$ pc. At the high-mass end, this number falls below 50%. The lowest-mass stars ($M_\star < 0.1 M_\odot$) have the highest detection rate, which is not surprising given that in the LSPM sub-sample, while being the faintest, they are the closest neighbors to the Sun.

The same analysis can be made more robust in a statistical sense by using stellar population models to estimate the number of M dwarfs that Gaia will observe and estimate the rate of detection in the entire sample accessible to Gaia. We derived starcounts for M0-M9 dwarfs with $G < 20$ and within 100 pc from the Sun using the Besançon stellar population synthesis model (Robin et al. 2003), taking advantage of a recent correction (March 2013) applied on the spectral types and luminosity function of late-M dwarfs (see http://model.obs-besancon.fr/). The total sample amounts to $\sim 415,000$
stars, with effective temperatures $T_{\text{eff}}$ in the range between 2500 K and 4000 K, masses in the range between 0.08 $M_{\odot}$ and 0.63 $M_{\odot}$, absolute visual magnitudes $M_v$ in the range between 8.5 and 19.0, and colors $V-I$ in the range between 1.55 and 4.65. Figure 7.4 shows how detection efficiency (when $\varsigma/\sigma_{\text{AL}} \gtrsim 3$) varies with distance and primary mass for the simulated sample from the Besancon galaxy model (BGM sample for short henceforth). The distance horizon for 90% completeness around the lowest-mass bin is $\sim 15$ pc, and $\sim 40$ pc for M0 dwarfs. Detectability drops below 50% for M8-M9 dwarfs at $\sim 30$ pc, while for stars with $M_*>0.5 M_{\odot}$ the same detection efficiency levels extend out to $\sim 80$ pc. The above considerations can be turned into an actual number of detected giant planets across all spectral sub-types of M dwarfs, as we will see later on in § 7.4.3.

### 7.4.2 Orbit Determination

Casertano et al. (2008) presented detailed analyses of the sensitivity of Gaia astrometry to giant planets around solar-type stars. We discuss here the quality of orbit reconstruction for the case of Jupiter-mass companions orbiting the LSPM sub-sample, using as
proxies the fitted orbital periods and the derived planetary masses, and the precision with which these parameters are retrieved as a result of the orbit fitting procedure, expressed in terms of the fractional error \( (P_{\text{fitted}} - P_{\text{true}})/P_{\text{true}} \). We refer to an accurate determination of a given parameter when its fractional error is less or equal to 10%. The quality of orbit reconstruction will also be parameterized in terms of \( \varsigma/\sigma_{\text{AL}} \) (e.g., Sozzetti et al. 2002).

We show in Fig. 7.5 the variation of the fractional error on the orbital period as a function of the true simulated value of \( P \). As expected, some of the main features of this behaviour already described in Casertano et al. (2008) are recovered. For example, Fig. 7.5 highlights how Gaia sensitivity decreases significantly both for periods exceeding the mission duration as well as for short-period orbits which are under-sampled (as a direct effect of the scanning law) and translate in very low astrometric signals. On the other hand, well-sampled \((P < T)\) orbital periods can be determined with uncertainties of < 10% around the nearest sample \((d < 15 \text{ pc}, \text{approximately 450 targets})\). In the same range of periods, the precision improves if a magnitude cut-off \((G \leq 12, \text{approximately})\).

![Figure 7.5: Fractional error on the orbital period \( P \) as a function of \( P \) itself. Solid line: the full LSPM sub-sample. Dashed-dotted line: stars \( d < 15 \text{ pc} \). Dashed line: stars with \( G < 12 \text{ mag} \).](image-url)
Figure 7.6: Top: Fractional error on planetary mass as a function of orbital period. Bottom: the same, but as a function of distance from the Sun. In both cases, the large black dots represent the median value in a period or distance bin, respectively.

600 targets) is made, but not to a very significant extent. Bright objects are in fact somewhat affected (in terms of planet detectability and quality of orbit reconstruction) by the presently envisioned gate scheme to avoid saturation on bright stars (see Fig. 7.2). Instead, at least for the LSPM sub-sample under investigation, the nearest stars ($d < 15$ pc) appear to provide the most significant improvement in precision in orbital period determination. The resulting astrometric signatures are typically large enough to allow for good-accuracy orbit reconstruction even for relatively faint objects, for which the per-measurement precision is significantly degraded.

The planetary mass as derived using the mass function approximation will be affected by the uncertainty on $P$, $\pi$, and $a_*$ as obtained from the fitting procedure. Its median for the whole LSPM sub-sample is $1.19 \, M_J$, which reduces to $0.99 \, M_J$ for the sample within 20 pc from the Sun. The two panels of Fig. 7.6 show how $P$ and $d$ affect the uncertainty on $M_p$. In particular, planets orbiting stars within $\sim 20$ pc have their masses

---

1. We assume here that $M_\star$ is perfectly known. While uncertainties in stellar mass at the bottom of the main sequence can easily be on the order of 10-20% (e.g., Boyajian et al. 2012 and references therein), the uncertainties in the model parameters from orbital fits are the dominant source of error when deriving the companion mass in this analysis.
measured with typical precision of 10%, or better, while short- and long-period orbits allow for reduced precision in the derivation of $M_p$, as expected. For $P > T$, planetary masses are systematically overestimated as a result of the systematic under-estimation of $P$, an effect already shown and discussed in detail by Casertano et al. (2008).

One representation of the dependence of the quality of orbit reconstruction on $\zeta/\sigma_{AL}$ is visualized in Fig. 7.7. From the plot, one sees for example that $\zeta/\sigma_{AL} \simeq 10$ allows to measure $M_p$ to $\sim 10\%$ accuracy, while fractional errors of a few percent, or better, require $\zeta/\sigma_{AL} \gtrsim 30 - 40$. These findings are in line with those obtained by Sozzetti et al. (2002). In Fig. 7.7, the results of orbital fits for the full LSPM sub-sample are plotted, including increasingly more marginal detections ($\zeta/\sigma_{AL} < 3$). As expected, for some systems (with $1 \lesssim \zeta/\sigma_{AL} \lesssim 3$) it might still be possible to obtain reasonably accurate solutions and mass estimates (for example in cases of particularly favorable orbital sampling). However, in general, attempting to fit orbital signals with amplitudes too close to, or even below, the single-measurement precision will translate in very poor quality results.
Table 7.1 reports, for illustrative purposes, the relevant information for 8 giant planetary companions to nearby M dwarfs detected by Doppler surveys (as obtained from http://exoplanet.eu). From the the minimum $\frac{\varsigma_{\text{min}}}{\sigma_{\text{AL}}}$ values, one infers that, given the general systems characteristics (orbit, masses, distance), half of the sample of known Doppler-detected planets around M dwarfs considered here (GJ 832b, GJ 849b, GJ 179b, GJ 317b) would have accurately determined orbital parameters and masses based on Gaia astrometry alone, while one of the planets, HIP 57050b, would be essentially undetectable. Note, however, that this situation corresponds to the worst-case scenario, as the actual masses will typically be larger than the minimum value reported in Table 7.1. The two short-period, resonant giant planets orbiting GJ 876 would both be detectable by Gaia, but they are not included in this sample as our study does not cover the problem of multiple-planet detection and characterization around low-mass stars. Other M dwarf planets have not been considered as their inferred astrometric signals (due to a combination of low masses, short periods, and large distances) would be essentially undetectable in Gaia astrometry.

### 7.4.3 Expected Planet Yield

It is worthwhile providing a reference figure of merit on the number of giant planets we can expect Gaia to detect in a given interval of orbital separations, as a way of gauging, in a preliminary fashion, the ability of the survey to reconstruct the underlying orbital elements distributions and occurrence rates in the low-mass star regime. In two recent works, Johnson et al. (2010b) and Bonfils et al. (2013) have provided updated estimates of the fraction of M dwarfs (no distinction in the stellar sub-types given the small-number statistics involved) hosting giant planets within approximately 3 AU. Starting with a northern hemisphere sample observed with HIRES and a southern hemisphere sample observed with HARPS, with different minimum-mass sensitivity thresholds but comparable time baselines, they reach similar conclusions: short-period ($P < 100$ days) giants ($M_p \sin i > 0.3 \, M_J$) are quite rare around M dwarfs ($f_p < 1\%$). At wider separations (roughly, $a < 3$ AU), giants orbiting M dwarfs appear to be more frequent: Johnson et al. (2010b) report $f_p \simeq 3 \pm 1\%$ (corrected for metallicity effects), while Bonfils et al. (2013) obtain $f_p \simeq 4^{+5}_{-1}\%$, two estimates which appear consistent with each other, within the error-bars. Note that these values of $f_p$ are somewhat higher than those quoted in previous works. For example, Endl et al. (2006) derive an upper limit (at the $1 - \sigma$ confidence level) of $f_p \simeq 1.3\%$ for giant planets within 1 AU of low-mass stars, while Cumming et al. (2008) infer $f_p \simeq 2\%$ for M dwarfs orbited by gas giants within $\sim 3$ AU. It is furthermore worth pointing out how giant planet occurrence rates at intermediate

---

4We recall that for Doppler-detected planets only the lower limit $M_p \sin i$ to their mass can be determined.
separations (∼ 3 AU) around M-dwarf hosts from microlensing surveys (e.g., Gould et al. 2010) appear reasonably in agreement with the above results. Other recent microlensing and high-contrast imaging studies encompassing the range of orbital separations \( a < 20 \text{ AU} \) for gas giants provide roughly consistent numbers \( f_p \sim 9^{+3}_{-5}\% \) from Cassan et al. (2012) and \( f_p \simeq 6.5 \pm 3.0\% \) from Montet et al. (2013), respectively.

We can estimate the number of giant planets at a given distance \( d \) (in pc) whose astrometric signal could be detected by Gaia using the \( \zeta/\sigma_{AL} \gtrsim 3 \) detectability threshold (Section 7.4.1), starcounts \( N_\star \) for M dwarfs computed within a sphere of radius \( d \) centered on the Sun, and the value of \( f_p = 3\% \) derived by Johnson et al. (2010b). We can then infer that around the nearby \((d < 33 \text{ pc})\) LSPM sub-sample under consideration here, Gaia should be able to detect on the order of 100 giant planets within 3 AU, an order-of-magnitude increase with respect to the present-day yield from Doppler surveys. Given the average stellar mass of the sample, the corresponding orbital period limit would be \( P \lesssim 6 \text{ yr} \), thus this planet sample would also be one for which Gaia could deliver good-quality orbital solutions (at least for the brighter/nearer systems). Extrapolating to the BGM sample of M dwarfs within 100 pc from the Sun from the Besancon Galaxy model presented in Section 7.4.1, one then infers a total giant planet yield for Gaia of \( \approx 2600 \) new detections. Accurate orbit reconstruction will be possible for \( \approx 20\% \) of the detections (corresponding to systems with \( \zeta/\sigma_{AL} \gtrsim 10 \)), i.e. \( \sim 500 \) giant planetary companions orbiting M dwarfs within \( \sim 50 \text{ pc} \). Naturally, this estimate does not take into account any possible dependence of occurrence rates and orbital elements and mass distributions on spectral sub-type (for example, for M5 or later dwarfs, \( f_p \) is presently severely unconstrained by observations), nor do we attempt at extrapolating at wider orbital separations.

Finally, while theoretical arguments based on the core-accretion model of giant planet formation (e.g., Laughlin et al. 2004; Ida & Lin 2005; Alibert et al. 2011) clearly predict the existence of a trend of decreasing \( f_p \) with decreasing \( M_\star \), as observed (Johnson et al. 2010b), the predicted planet fractions in a given stellar mass range do not necessarily agree with the observations. For example, Kennedy & Kenyon (2008) predict \( f_p \sim 1\% \) within \( \sim 2 - 3 \text{ AU} \) of \( M_\star < 0.5 \text{ M}_\odot \) M dwarfs, a value somewhat lower than the observed fraction. Any discrepancy could point to either insufficient depth in the analysis of the observational data or to the necessity to further the theoretical understanding of planet formation processes in the low-mass star regime. However, at present any attempt to study fine structure details in the comparison between theory and observations is severely hampered by small-number statistics. In this respect, Gaia high-precision astrometry of thousands of nearby M dwarfs will likely help to shed light into the matter, as this unbiased sample screened for giant planets by Gaia will contribute to significantly reduce the uncertainties on the occurrence rate estimates. For
Table 7.1: Summary of relevant data on known Doppler-detected giant planets orbiting $M_* < 0.6 \, M_\odot$ dwarfs in the solar neighborhood.

<table>
<thead>
<tr>
<th>Name</th>
<th>$G$ (mag)</th>
<th>$\sigma_{\min}$ (µas)</th>
<th>$\sigma_{AL}$ (µas)</th>
<th>$\sigma_{\min}/\sigma_{AL}$</th>
<th>$P$ (yr)</th>
<th>$e$</th>
<th>$d$ (pc)</th>
<th>$M_*$ ($M_\odot$)</th>
<th>$M_p \sin i$ ($M_J$)</th>
<th>Discovery (Refs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GJ 832b</td>
<td>8.7</td>
<td>978.9</td>
<td>64.4</td>
<td>15.2</td>
<td>9.3</td>
<td>0.12</td>
<td>4.9</td>
<td>0.45</td>
<td>0.64</td>
<td>Bailey et al. (2009)</td>
</tr>
<tr>
<td>Gl 649b</td>
<td>9.7</td>
<td>66.7</td>
<td>17.5</td>
<td>3.8</td>
<td>1.6</td>
<td>0.30</td>
<td>10.3</td>
<td>0.54</td>
<td>0.33</td>
<td>Johnson et al. (2010a)</td>
</tr>
<tr>
<td>GJ 433c</td>
<td>9.8</td>
<td>116.1</td>
<td>18.1</td>
<td>6.4</td>
<td>10.1</td>
<td>0.17</td>
<td>9.0</td>
<td>0.48</td>
<td>0.14</td>
<td>Delfosse et al. (2013)</td>
</tr>
<tr>
<td>GJ 849b</td>
<td>10.4</td>
<td>539.5</td>
<td>17.4</td>
<td>31.0</td>
<td>5.1</td>
<td>0.04</td>
<td>8.8</td>
<td>0.49</td>
<td>0.99</td>
<td>Butler et al. (2006)</td>
</tr>
<tr>
<td>HIP 79431b</td>
<td>11.3</td>
<td>107.1</td>
<td>18.7</td>
<td>5.7</td>
<td>0.3</td>
<td>0.29</td>
<td>14.4</td>
<td>0.49</td>
<td>2.10</td>
<td>Apps et al. (2010)</td>
</tr>
<tr>
<td>HIP 57050b</td>
<td>11.9</td>
<td>13.0</td>
<td>18.9</td>
<td>0.7</td>
<td>0.1</td>
<td>0.31</td>
<td>11.0</td>
<td>0.34</td>
<td>0.30</td>
<td>Haghihipour et al. (2010)</td>
</tr>
<tr>
<td>Gl 179b</td>
<td>11.9</td>
<td>450.0</td>
<td>17.5</td>
<td>25.7</td>
<td>6.3</td>
<td>0.21</td>
<td>12.3</td>
<td>0.36</td>
<td>0.82</td>
<td>Howard et al. (2010)</td>
</tr>
<tr>
<td>GJ 317b</td>
<td>12.0</td>
<td>326.4</td>
<td>16.9</td>
<td>19.3</td>
<td>1.9</td>
<td>0.11</td>
<td>15.1</td>
<td>0.42</td>
<td>1.80</td>
<td>Johnson et al. (2007)</td>
</tr>
</tbody>
</table>

example, based on the above simulation results we can infer how precisely a value of $f_p = 3\%$ for Jupiter-mass companions within 3 AU around M0-M9 stars could be determined by using the number of detections ($n \approx 2600$) and the number of stars for which an astrometric detection was possible ($N \approx 9 \times 10^4$), the latter derived based on the detection efficiency estimates presented in Section 7.4.1 using the Besançon galaxy model. Using this non-parametric description (see e.g., Cumming et al. 2008) and by adopting the standard Poisson uncertainty limits ($1/n + 1/N)^{1/2}$ (e.g., Burgasser et al. 2003), we then obtain $f_p = 3.00 \pm 0.06$, an improvement by a factor $\sim 15$ with respect to, e.g., the Johnson et al. (2010b) estimates.

### 7.4.4 Measuring Transiting Systems Configurations

We focus here on gauging the potential of Gaia to identify astrometrically extrasolar planets in orbits compatible with transit configurations.

We show in Fig. 7.8 the fractional error in $i$ as a function of $i$ itself as determined in the simulations. The three cases correspond to the full LSPM sub-sample within 33 pc from the Sun, stars within 15 pc and with planets with $P < T$, and stars with $G < 12$ and with planets with $P < T$ (i.e., using the same selection criteria of Section 7.4.2 and the additional constraint of well-sampled orbits). The overall trend confirms the findings of Sozzetti et al. (2001), with the Gaia astrometric observations becoming less sensitive to the inclination itself as we move towards a quasi-face-on configuration ($i \approx 0$ deg), and a corresponding increase of the fractional error on this parameter. The immediate conclusion is that $i \approx 90$ deg could be determined with uncertainties of just a few degrees for Jupiter-mass companions on well-sampled orbits with $P < T$ around the nearest or brightest M dwarfs. Wider-separation ($a > 0.3$ AU) systems with close to edge-on configurations, indicating the presence of a planet that might transit and/or...
be occulted by its primary, would then become very interesting targets for follow-up photometry, to ascertain whether the prediction is verified or not.

The possibility to study a sample of transiting cold (i.e., long-period) giant planets around nearby low-mass stars is certainly intriguing, for systematic comparison with their strongly irradiated, short-period counterparts. While their typical transit depths (significantly exceeding 0.01 mag) would not pose a challenge even for modest-precision photometric systems, ground-based transit searches lack sufficient sensitivity at long periods due to the impossibility to guarantee continuous coverage over extended time baselines, a necessary prerequisite given that the infrequent transits make it difficult to build enough signal-to-noise ratio. Space-borne instruments can fulfill the requirement of uninterrupted photometric coverage, and indeed Kepler has identified transiting giant planet candidates on $\sim 1$ AU orbits (Fressin et al. 2013). However, these orbit F-G-K dwarfs at hundreds of pc from the Sun, with typical infrared magnitudes of $J \approx 12 - 13$.
Figure 7.9: Left: Degradation rate in the prediction for the value of orbital separation of the planet as a function of its orbital period. The results are averages over each 0.5 yr period bin. Right: The same, but for the position angle. In both panels, symbols are as in Figure 7.8.

mag. If any such objects were detected around low-mass stars in the solar neighborhood (tens of pc), as they would orbit much brighter primaries at infrared wavelengths, they would then constitute prime targets for atmospheric characterization via transit and occultation spectroscopy with future ground-based and particularly space-borne instrumentation.

For the planet’s disc to occult the stellar disc, the orbital inclination must satisfy:

$$a_p \cos i \leq R_* + R_p,$$

(7.2)

with $R_*$ and $R_p$ the stellar and planetary radii, respectively. In general, the geometric transit probability can be expressed as (e.g., Barnes 2007):

$$p_{tr} = 0.0045 \left( \frac{1 \, \text{AU}}{a_p} \right) \left( \frac{R_* + R_p}{R_\odot} \right) \left[ \frac{1 + e \cos(\pi/2 - \omega)}{1 - e^2} \right]$$

(7.3)

This expression allows to appreciate how favourable configurations of eccentricity and argument of pericenter can dramatically increase the transit likelihood for long-period objects (Kane & von Braun 2008). However, at large orbital distances (e.g., $a \gtrsim 1 \, \text{AU}$) the range of permitted inclination angles for transits to occur (given by $i > \arccos(p_{tr})$) is typically well within 1 deg from the central transit configuration ($i = 90$ deg), except for pathologically eccentric systems ($e \geq 0.9$) when $\omega \simeq \pi/2$ (i.e., the pericenter is aligned towards the observer). From Figure 7.8 we can see that in the case of accurately measured orbits (with $\varsigma/\sigma_\text{AL} \gtrsim 10$) a typical fractional uncertainty of $\sim 2\%$ on the inclination angle when $i$ is close to 90 deg implies formal errors of $\sim 2$ deg. That is,
a best-case accurate orbit determination by Gaia of a long-period giant planet with $i = 90 \pm 2$ deg would not allow to rule out non-transiting configurations, within the errors. At the same time, measuring an astrometric orbit with $i = 88 \pm 2$ deg would also indicate a possibly transiting companion, within the errors. Assuming orbits are isotropically distributed, we can convolve this information with the M dwarfs starcounts from the BGM sample within 100 pc and the occurrence rate $f_p = 3\%$ of giant planets within 3 AU of M dwarfs discussed in §7.4.3 to infer a) the number of intermediate-separation giants actually transiting their parent stars, and b) the number of giants that might have astrometric orbital solutions compatible with a transiting configuration within 1σ.

Based on the above considerations, for fractional uncertainties on the inclination angle of 10%, 5%, and 2%, Gaia could detect 255, 85, and 10 systems, respectively, formally compatible with transiting configurations within the 1σ error-bars. On the one hand, using the $f_p = 3\%$ estimate the expectation is that only 40 systems in the BGM sample would actually have $i$ above the critical value for transits to occur in practice. If the sample of actually transiting systems were entirely composed of systems with sufficiently high values of astrometric signals for which Gaia could deliver orbits with $i$ determined within 10% accuracy, then we find that the sample of candidate transiting planets identified by Gaia would encompass $\sim 85\%$ of false positives. On the one hand, tightening the requirements on the precision with which $i$ can be determined might allow to select a smaller sample of candidates with fewer false positives. On the other hand, it might well happen that a candidate system in transit with $i$ measured less precisely is in fact transiting, while one with $i$ more accurately measured in fact is not. This will depend in practice on the actual shapes of the period and mass distributions, on the details of planet frequency as a function of spectral sub-type, and distance from the Sun of the actual M dwarf sample that will be observed by Gaia.

### 7.4.5 Predicting Giant Planets’ Location and Brightness

As shown by Benedict et al. (2006) using HST/FGS measurements of the $\varepsilon$ Eridani system, astrometry, by determining the full orbital geometry and mass of a planetary companion, can have significant value for future direct-imaging programs and for the interpretation of emergent flux measurements. For example, by determining the times, angular separation and position angle at periastron and apoastron passage, it will be possible to predict where and when a planet will be at its brightest (this is also relevant for eccentric planets which can undergo orders of magnitude of variation in apparent brightness along the orbit), thus a) crucially helping in the optimization of direct imaging observations and b) resolving at least in part important model degeneracies in
predictions of an exoplanet apparent brightness in reflected host star light as functions of orbit geometry, companion mass, system age, orbital phase, cloud cover, scattering mechanisms, and degree of polarization (e.g., Sudarsky et al. 2005; Burrows et al. 2004; Madhusudhan & Burrows 2012).

The first element of the synergy between Gaia astrometry and future direct-imaging projects consists in being able to quantify the accuracy with which it will be possible to predict where to look around a given star, based on the companion mass and orbital parameters determination. We show in Fig. 7.9 the average rates of degradation $\Delta \varrho$ and $\Delta \vartheta$ in the estimated orbital separation $\varrho$ and position angle $\vartheta$ of the planet (expressed in mas yr$^{-1}$ and deg yr$^{-1}$, respectively) as a function of the orbital period. On average, the knowledge of the planet’s ephemeris will degrade at rates of $\Delta \varrho < 1$ mas yr$^{-1}$ and $\Delta \vartheta < 2$ deg yr$^{-1}$, for orbits with $P < T$. These numbers are over an order of magnitude smaller than the degradation levels attained by present-day ephemerides predictions based on mas-level precision HST/FGS astrometry (Benedict et al. 2006). In the present sample of intermediate-separation giant planets around M dwarfs (see Table 7.1), one could then conclude that at least two objects, GJ 832b and GJ 433c, with typical separations of 0.7\" and 0.4\", respectively, represent prime candidates for such an investigation when Gaia data will become available, particularly if combined with existing radial-velocity datasets (thus improving the accuracy of the ephemeris predictions). In this regime of orbital separations instruments such as SPHERE on the VLT (Kasper et al. 2012) and particularly PCS on the E-ELT (see https://www.eso.org/sci/facilities/eelt/instrumentation/) are in fact expected to achieve very high contrast ratios.
A second element of the above mentioned synergy relates to the effectiveness with which a precise knowledge of the companion mass and orbital geometry of the system from astrometry can be used to predict accurate times of optimal visibility for direct imaging and eventually help discriminate between different atmospheric compositions. Using as an illustrative example that of an isotropically, perfectly reflecting Lambertian surface (e.g., Burrows 2005; Madhusudhan & Burrows 2012), the left panel of Figure 7.10 shows how the error in the planetary phase function $\Phi(\lambda)$ varies as a function of the error in the derived phase angle $\lambda$ based on the orbital parameters determined for each of the $1-M_J$ companions around the LSPM M-dwarf sub-sample. In the Figure, for each target in the LSPM sub-sample the errors $\Delta \lambda$ and $\Delta \Phi(\lambda)$ are defined as the difference between the orbit-averaged derived value and the orbit-averaged true value of each quantity. For both $\lambda$ and $\Phi(\lambda)$ the median differences are very close to zero, indicating that there is little bias in both estimates. Based on the distribution of the absolute differences for the two quantities, the median uncertainty on the phase results to be $\Delta \lambda \sim 7$ deg, and that on the phase function $\Delta \Phi(\lambda) \sim 0.05$.

The rms uncertainty on the phase-averaged phase function is representative of the quality with which this quantity could be determined based on Gaia astrometry (assuming a specific atmospheric model), but not at all phases. The right panel of Figure 7.10 shows how the error in the phase function varies with $\lambda$. In the plot, each point corresponds to the median of the absolute differences between the derived and the true value of $\Phi(\lambda)$ in each 10-deg bin in $\lambda$ (considering only the subsets of the LSPM sample with planets whose orbits have been sampled in any given interval in phase angle). Note that, depending e.g., on orbit geometry and details on the atmospheric scattering properties the value of $\Phi(\lambda)$ can vary by a factor of 2 or more at a given phase angle (Sudarsky et al. 2005; Madhusudhan & Burrows 2012) for intermediate-separation giant planets. Based on the result shown in the right panel of Figure 7.10, then for a value of the phase angle of say $\lambda = 70$ deg, it would be possible to distinguish, on average, between the phase function of a Lambert sphere ($\Phi(\lambda) \sim 0.5$) and that of Jupiter ($\Phi(\lambda) \sim 0.2$) at the $2 - \sigma$ level (see e.g. Figure 3 of Sudarsky et al. 2005). However, a difference between $\Phi(\lambda) \sim 0.9$ for a Lambert sphere and $\Phi(\lambda) \sim 0.6$ for Jupiter at $\lambda = 30$ deg might still fall within the typical errors.

Based on the values of $\Phi(\lambda)$ so determined, we show in the top panel of Figure 7.11 the predicted phase-averaged planetary emergent flux in units of the stellar flux $F_p/F_\star$ as a function of orbital separation from the M dwarf primary. The simulated data are binned in 0.05 AU intervals, while the theoretical values are shown as a dashed-dotted line. The bottom panel of Figure 7.11 shows the ratio $(F_p/F_\star)_O/(F_p/F_\star)_C$ between the simulation results and the Lambert sphere model. Note the close agreement for well-sampled orbits ($a \lesssim 2$ AU). In this regime of orbital separations, good quality orbit reconstruction
implies that \( (F_p/F_\star)_O \) is on average within a factor 1.2 of \( (F_p/F_\star)_C \). At \( a \sim 3 \) AU, \( (F_p/F_\star)_O/(F_p/F_\star)_C \approx 2 \), on average. It should be kept in mind that, as shown by e.g. Burrows (2005) and Sudarsky et al. (2005), planet/star flux ratios depend strongly on the Keplerian elements of the orbit, particularly inclination and eccentricity, with variations in \( F_p/F_\star \) of up to over an order of magnitude. Thus for example (assuming the object is detectable by a direct-imaging device) a flux difference of a factor \( \sim 16 \) along the orbit of a Jupiter-mass companion on an 1.5-AU, high eccentricity (\( e = 0.6 \)) orbit could be inferred with very high significance. Similarly, a flux difference of a factor \( \sim 4 \) between a cloud-free giant planet at 2 AU and one with a large-size (100 \( \mu \)m) condensate particles atmosphere could also be detected (see e.g. Figure 16 and 17 of Sudarsky et al. 2005). At short orbital distances \( (a \lesssim 0.3 \) AU), the combination of insufficient orbit sampling and low values of \( \zeta/\sigma_{AL} \) worsens the quality of the determination of \( F_p/F_\star \). For \( a \gtrsim 3 \) AU, \( F_p/F_\star \) is systematically overestimated (by up to a factor of ten, and more) due to the systematic underestimation of the orbital period. Finally, we note from Figure 7.11 that ‘cold’ giant planets on 2-AU orbits are expected to have \( F_p/F_\star \sim 10^{-8} \).

Taking the numbers at face value, these systems thus appear difficult for an instrument such as SPHERE, nominally set to achieve contrast ratios in the range \( 10^{-6} - 10^{-7} \) (Kasper et al. 2012). The combination of Gaia astrometry for such systems might then become more effective with next-generation direct imaging devices on telescopes such as the E-ELT, for which contrast ratios on the order of \( 10^{-8} - 10^{-9} \) could be more readily achieved (e.g., Kasper et al. 2010).

### 7.5 Summary and Conclusions

In this chapter, I report results from a detailed numerical experiment designed to assess the potential of ESA’s Cornerstone mission Gaia to detect and characterize astrometrically giant planetary companions to our closest neighbours, the reservoir of cool low-mass M dwarfs within \( \sim 30 \) pc from the Sun. The work was motivated by the need to revisit and update Gaia’s planet detection potential now that we are within a few months from first data, relaxing some of the caveats and simplifying assumptions of previous analyses (e.g., using an up-to-date Gaia error model and employing an actual list of stars in input). A second aim of this work was to begin shedding light on some of the potentially relevant synergies between Gaia astrometry and other ongoing and planned planet detection and characterization programs, both from the ground and in space. The results obtained in this work have been specifically tailored to a sample of nearby, low-mass M dwarfs. The main findings in this experiment can be summarized as follows:
Figure 7.11: Top: Predicted planetary emergent flux as a function of orbital semi-major axis from the M dwarf primary. Logarithmic error bars on $F_p/F_*$ are phase-averaged values in each 0.05 AU bin obtained through error propagation of the uncertainties on $a_p$ and $\Phi(\lambda)$, assuming an error on the planetary radius ($R_p = 1 R_J$) of 5%. Theoretical values based on a Lambert sphere model are shown as a dashed-dotted line. Bottom: The ratio of $(F_p/F_*)_O$, derived from the simulated data, to $(F_p/F_*)_C$, computed from the model, as a function of orbital semi-major axis.
Chapter 7. The GAIA potential

1) The overall quality of orbit reconstruction (using $P$, $M_p$, and $\zeta/\sigma_{AL}$ as proxies) is in agreement with previous findings by e.g. Lattanzi et al. (2000), Sozzetti et al. (2002) and Casertano et al. (2008). In particular, the impact of the gate scheme to avoid saturation on bright ($G < 12$ mag) stars (that tends to worsen the positional accuracy of Gaia in this magnitude range) is found to be relatively mild for the M-dwarf LSPM sub-sample studied (with an average $G \simeq 14$ mag). Given that the current occurrence rate estimates for giant planets orbiting within 3 AU of low-mass stars are on the order of 3-4%, we can expect $\approx 10^2$ massive planetary companions to be detected by Gaia around this sample (a ten-fold increase with respect to present-day giant exoplanet counts). Extrapolations based on M dwarfs starcounts out to 100 pc from the Sun (a reservoir of $\sim 4 \times 10^5$ stars) allow us to infer the possibility for Gaia to detect over two thousand new giant planets around low-mass stars, and to derive accurate masses and orbital parameters for as many as five hundred systems. The size of the sample will likely help to constrain $f_p$ to a very accurate degree: based on the simulations presented here, we estimate that, assuming present-day estimates of the occurrence rates, Gaia astrometry would allow to determine $f_p = 3\%$ with an error of 2%, an improvement by over an order of magnitude with respect to the most precise estimates to-date. This in turn will allow for meaningful comparisons with theoretical predictions of giant planet frequencies in the low stellar mass regime. Such systems will also become prime targets for high-precision RV follow-up from the ground, looking for the presence of additional, low-mass companions with state-of-the-art facilities both in the visible (e.g., HARPS, HARPS-N) and in the near infrared (e.g., CARMENES, HPF);

2) for detected giant planets with periods in the range $0.2 - 5$ yr (i.e., with accurately determined masses and orbits), inclination angles corresponding to quasi-edge-on configurations will be determined with enough precision (a few percent) so that it will be possible to identify candidate transiting planets in a regime of orbital separations which is inaccessible from the ground and only marginally probed from space by dedicated transit discovery missions such as CoRoT and Kepler. Based on the BGM sample results, Gaia might be able to measure accurately the orbits of 10 potentially transiting intermediate-separation giants around nearby M dwarfs. Considering inclination angles determined with 10% accuracy, the sample of ‘astrometric’ candidate long-period transiting planets might encompass more than 250 systems. However, the majority of these candidates ($\sim 85\%$) would be likely false positives. Ground-based monitoring campaigns will be instrumental in unveiling the true nature of the systems;

3) for well-sampled orbits ($P < T$), the uncertainties on planetary ephemerides, separation $\rho$ and position angle $\vartheta$, will degrade at typical rates of $\Delta \rho < 1$ mas yr$^{-1}$
and $\Delta \vartheta < 2 \, \text{deg yr}^{-1}$, respectively. These are over an order of magnitude smaller than the degradation levels attained by present-day ephemerides predictions based on mas-level precision astrometry;

4) Planetary phases will be measured with typical uncertainties $\Delta \lambda$ of several degrees, resulting (under the assumption of simple purely scattering atmospheres) in phase-averaged errors on the phase function $\Delta \Phi(\lambda) \approx 0.05$, and expected phase-averaged uncertainties in the determination of the emergent flux of well-measured, intermediate-separation ($0.3 < a < 2.0 \, \text{AU}$) giant planets of $\sim 20\%$. The combination of detailed models of giant exoplanets’ systems and reliable ephemerides from Gaia astrometry could then greatly help both in the selection of good targets for direct-imaging instruments and for the physical interpretation of positive observational results.
Chapter 8

On the synergies between GAIA and transits survey

In the previous chapter, we revisited the topics of planet detection and characterization with the extremely precise measurements of the Gaia space mission. To do this, we focused on a sample of nearby low-mass M dwarf stars from the Lépine 2005 [102] catalogue.

In this chapter, we better quantify the relevance of the Gaia observations of the large sample of nearby M dwarfs in a synergetic effort to optimize the planning and interpretation of follow-up/characterization measurements of the discovered systems by means of transit photometry (e.g., APACHE) and simultaneous multi-wavelength spectroscopy (e.g., EChO, JWST).

For this work, I implemented a new simulation framework and a partially different numerical analysis tool. Furthermore, I give the first interpretation of the results.

8.1 Simulation Scheme

The simulation of Gaia observations follows closely the observational scenario described in the Chapter 7.2. Here I describe and discuss the changes/upgrades made to that setup.

1) The actual list of targets encompasses 8793 dM dwarf stars (in the approximate range 0.09 – 0.6 $M_\odot$) from the All-sky Catalog of Bright M Dwarfs (Lépine & Gaidos 2011 [139]). The choice of this catalogue was driven by our interest to choose a sample fully covering the APACHE Input Catalogue 5.4. This sample
results to have an average $G \simeq 12.3$ mag in the Gaia broad-band photometric system and results to have an average $M_\star \simeq 0.46 M_\odot$ (see Section 7.2). In the four panels of Fig. 8.1 we show the distributions in $G$ mag, distance $d$ \footnote{Where available, trigonometric parallaxes are used, photometric distance estimates are otherwise utilized using the Lepine & Gaidos 2011 [139] values.} mass $M_\star$, and number of Gaia field transits (individual field-of-view crossings) as a function of ecliptic latitude $\beta$ for our sample.

2) The generation of planetary systems proceeded as follows. One planet was generated around each star (assumed not to be orbited by a stellar companion), with mass $M_p = 1M_J$, orbital periods were uniformly distributed in the range $0.2 \leq P \leq 5$ yr and eccentricities were uniformly distributed in the range $0.0 \leq e \leq 0.6$). The orbital semi-major axis $a_p$ was determined using Kepler’s third law. The inclination $i$ of the orbits were fixed to 90 degree. All other orbital elements (argument of pericenter $\omega$, ascending node $\Omega$, and epoch of pericenter passage $\tau$) were uniformly distributed within their respective ranges. The resulting astrometric signature induced on the primary was calculated using the standard formula corresponding to the semi-major axis of the orbit of the primary around the barycentre of the
system scaled by the distance to the observer: \( \varsigma = \left( \frac{M_p}{M_*} \right) \times \left( \frac{a_p}{d} \right) \). With \( a_p \) in AU, \( d \) in pc, and \( M_p \) and \( M_* \) in \( M_\odot \), then \( \varsigma \) is evaluated in arcsec.

8.2 Statistical and Numerical Analysis Tools

As described above, the resulting Gaia observable, the one-dimensional coordinate \( \psi \) in the along-scan direction, is modelled as \( \psi(\alpha, \delta, \mu_\alpha, \mu_\delta, \pi, A, B, F, G, P, e, \tau) \), where the five standard astrometric parameters correspond to the actual positions \( (\alpha, \delta) \), proper motion components \( (\mu_\alpha, \mu_\delta) \), and parallax \( (\pi) \) of each target M dwarf as provided in the All-sky Catalog of Bright M Dwarfs([139]), while \( A, B, F, \) and \( G \) are four of the six Thiele-Innes orbit elements (Green 1985). The great advantage of this modelling is that the \( \psi \) equation becomes partially linear. First, we resolve the equation related to a single-star model (\( \psi \) depend only by the five standard astrometric parameter). Then, statistically robust deviations from this single-star model solution, indicating the presence in the observations residuals of the perturbation due to a companion with a given level of confidence, are identified through the application of a \( \chi^2 \)-test or \( F \)-test (low probabilities of \( P(\chi^2) \) or \( P(F) \) signifying likely planet, and unlikely false positive).

To resolve the equation related to a single-star plus planet model, we utilize a Hybrid Markov Chain Monte Carlo Differential Evolution approach. This hybrid approach consists of a ”two-step” solution adopting the twelve-parameters models for \( \psi \). Basically, during the first step we search for the best solution for the three non-linear parameter \( P, e, \tau \) while during the second step we solve the linear part of the observation equation. The Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution is described in Ter Braak, 2006 and it is implemented for fitting RV and transit data in the IDL EXOFAST library (Eastman et al., 2013 [171]). We utilize an adapted version of these codes.

8.3 Predicting the transit epochs

As discussed at length in the previous chapters, the class of transiting planets is of particular importance, as the simultaneous determination of their masses (via Doppler measurements) and radii (via transit photometry) provides the means to estimate their densities, a fundamental proxy for understanding their interior compositions (e.g., Charbonneau et al. 2007, and references therein). Furthermore, if the primaries are sufficiently bright, transiting planets can be further characterized using the techniques of transmission and occultation spectroscopy to determine the chemistry and dynamics of their atmospheres (e.g., Seager & Deming 2010).
On the one hand, detection of planetary transits is normally achieved via investigation of photometric lightcurves. The general prospects for transiting short-period (giant) planet detection with Gaia using its onboard photometry have recently been revisited by Dzigan & Zucker (2012). On the other hand, Gaia high-precision astrometry, by measuring directly the inclination angle of an orbit (unlike Doppler spectroscopy), can in principle allow to uncover the existence of a possibly transiting planet at wider orbital separations (typically $a > 0.5$ AU).

Figure 8.2 shows the acceptance rate as a function of the simulated period $P$. In this scheme, the acceptance rate is the fraction of targets that have passed the statistical test to assess if the residuals in the observations are due to a planetary companion. The acceptance rate behaviour clearly indicates how Gaia sensitivity decreases for short-period orbits, which translate in very low-amplitude astrometric signals.

Subsequently, we discuss the quality of orbit reconstruction using as proxies the fitted orbital periods and the derived planetary masses, and the precision with which these parameters are retrieved as a result of the orbit fitting procedure, expressed in terms of the fractional error (e.g., $(P_{\text{fitted}} - P_{\text{true}})/P_{\text{true}}$). We refer to an accurate determination of a given parameter when its fractional error is less or equal to 10%.
Figure 8.3: Fractional error on P vs Simulated Period P

We show in Figure 8.3 the variation of the fractional error on the orbital period as a function of the true simulated value of $P$. On the one hand, Figure 8.3 highlights how Gaia sensitivity decreases for periods approaching the mission duration. This a combination of two factors: the particular configuration of the edge-on orbits; the insufficient sampling of the orbits (as a direct effect of the mission duration). On the other hand, well-sampled ($P \lesssim T/2$, where $T$ is the mission duration) orbital periods can be determined with uncertainties of $< 1\%$. Overall, we obtained a median period estimation uncertainties over the sample of $\sim 2\%$.

We show in Figure 7.8 the fractional error in $i$ as a function of the simulated value of $i$. The inclination of the orbit is not directly derived as one of the twelve parameters of the Gaia observable $\psi$. To obtain $i$ we have to invert the equations of the Thiele-Innes elements (Green 1985 [172]). To do this, we utilize the methods described in Wright & Howard, 2009. We obtained a median inclination estimation uncertainty over the sample of $\sim 2\%$, which highlights our ability to recover planetary system with a very close to edge-on configurations.

As mentioned above, the astrometric measurements completely describe the star-planet system in a three dimensional space (we assume here that $M_*$ is perfectly known). Therefore it is possible to predict the transit epoch of the transiting systems. The
transit epoch depends on the period $P$, the eccentricity $e$, the periastron epoch $\tau$ and the argument of periastron $\omega$ (also derived by inverting the Thiele-Innes elements).

Figure 8.5 shows the distribution of the differences (in days) between the transit epoch ($TE$) as derived by the fitted orbital parameters and the simulated ("true") transit epoch $\Delta TE = |TE_{fitted} - TE_{true}|$. We obtained a median estimation of the uncertainty over the sample of $\sim 20[\text{days}]$.

The top panel in Figure 8.6 shows the dependence of $\Delta TE$ on the differences (in days) between the "fitted" periastron epoch $\tau_{fitted}$ and the simulated ("true") periastron epoch $\tau_{true}$. The bottom panel shows the dependence of $\Delta TE$ on the fractional error of $P$ $(P_{fitted} - P_{true})/P_{true})$. These are the parameters that most affect the final results and the dependence is clear, although quite noisy. This is an effect of non-linearity of these parameters and it will become objective of further investigations.

The immediate conclusion is that $TE$ could be determined with uncertainties of just a few weeks for Jupiter-mass companions on well-sampled orbits with $0.2 < P < 5$ years around the M dwarfs in the solar neighbourhood. These are very interesting targets for follow-up photometry, to ascertain whether the prediction is verified or not. In fact, possibly transiting giants planets uncovered astrometrically by Gaia will have to be confirmed by means of follow-up photometric observations, that could readily be carried...
out from the ground even with modest-size telescopes. In perspective, any experiment designed for this purpose will also have to keep the above caveats into consideration. For example, such studies would benefit from the availability of additional Doppler measurements aimed at improving the accuracy of the orbital solutions and the corresponding transit ephemeris predictions. It will also be important to identify the correct balance between size of the candidate sample and expectations of false positive rates, as such issues could have a significant impact on follow-up programs to verify the actual transiting nature of the detected systems. Finally, note that Gaia-detected intermediate-separation giants on orbits compatible with transit configurations might also help revisit the photometric light-curve databases of existing (e.g., MEarth, Nutzman & Charbonneau 2008; APACHE, Giacobbe et al. 2012; Sozzetti et al. 2013) and upcoming (NGTS, Wheatley et al. 2013) ground-based surveys focusing on late-type dwarfs as well as those of other successful programs, such as Super-WASP, HATNet, and HATSouth, looking for missed or uncategorised transit events.
Figure 8.6: Top: $\Delta TE \text{ VS } \tau_{\text{fitted}} - \tau_{\text{true}}$ Bottom: $\Delta TE \text{ vs } (P_{\text{fitted}} - P_{\text{true}})/P_{\text{true}}$
8.4 Summary and Conclusions

In this chapter, I report results from a numerical experiment designed to assess the potential of ESA’s Cornerstone mission Gaia to predict transit events for giant planetary companions of cool low-mass M dwarfs. The work was motivated by the need to investigate the potentially relevant synergies between Gaia astrometry and other ongoing and planned planet detection and characterization photometric programs, both from the ground and in space. The results obtained in this work have been specifically tailored to a sample of nearby, low-mass M dwarfs. The main findings in this experiment can be summarized as follows: the overall quality of orbit reconstruction (using $P$, $i$, and $Transitepoch$ as proxies) is in agreement with previous findings showed in Chapter 7. In particular, for transiting giant planets with periods in the range $0.2 - 5$ yr, transit epochs will be determined with enough precision ($\sim 20$ days) so that it will be possible to provide candidate transiting planets for follow-up studies in a regime of orbital separations which is not directly accessible from the ground.
Chapter 9

Summary and future steps

The work presented in this thesis has been motivated by the desire to achieve two main objectives: (i) to improve our understanding of the processes that shape super-Earth and Neptune-type planets; (ii) to investigate some aspects of the properties of extrasolar planets in the transition region between Neptune-type and giant planets, in connection with the predictions formulated by models of giant planet formation, evolution, and dynamical interaction. To investigate these two outstanding issues we need to identify a population of objects that allows for robust comparative studies between theory and observations using a clever detection technique that can exploit the potential of ground-based observations using small-sized telescopes. To this aim, we have focused on the sample of the nearest low-mass M dwarf stars and designed, developed and carried out a new experiment, APACHE, based on the technique of planetary transits. As we have seen in the Introduction, the application of the transit technique to M dwarfs presents several exciting opportunities, and the advantages are especially compelling for the detection of transiting habitable, rocky planets. At the same time, M dwarfs make up for the vast majority of the stellar mix within 30pc from our Sun.

The centrepiece of the work presented here consists of the preparatory efforts to build and run the APACHE photometric transit survey, including the detailed analysis and results of the first year of operations. Focusing on a sufficiently bright stellar sample, the APACHE project has the potential to play a key role with its capability to discover sub-Neptune-sized transiting planets for which direct measurements of the planet’s mass will be possible and which are amenable to spectroscopic follow-up for atmospheric characterization with future ground-based and space-borne instruments. Furthermore, we have investigated some elements of the synergy potential for improved characterization of the architecture of planetary systems across a wide range of masses between photometric...
measurements (such as those carried with APACHE) and high-precision astrometry with ESA's Cornerstone Gaia mission.

9.1 The APACHE survey

The APACHE survey has officially started in July 2012 at the OAVdA site (see Chapter 4). This date marks the end of a three-year long effort undertaken to build the final setup of APACHE. The possibility of designing APACHE based on real data collected within the context of the pilot study (Chapter 4) has allowed us to start nominal survey operations with a refined observing strategy, a dedicated photometric data reduction pipeline and many well-tested tools for time-series periodicity analysis. Based on the photometric data collected with the APACHE telescope array during the first year of the survey, we have investigated the performance of our hardware/software setup and we have evaluated our sensitivity to $> 2R_{\oplus}$ transiting planets with orbital periods of a few days. We have then used the machinery developed in this work to address issues such as how to improve our detection sensitivity and how to increase the planets-per-year yield that APACHE could deliver. Finally, our first results were compared with inferences coming from other datasets of recent publication, both from the ground and in space.

Our findings and future prospects can be summarized as follows (see Chapter 6 for details):

- **Photometric precision.** We achieve a typical RMS photometric precision of $\sim 7$ mmag. The analysis of the impact of correlated (red) noise reveals that our data are affected to a minor degree by unmodelled short-term systematics. This is highlighted by the good agreement of the RMS photometric precision with the theoretical per-measurement uncertainties ($\sim 6$ mmag). The analysis reveals an excess of noise at the bright end of our sample, principally induced by scintillation effects. To minimize those effects a variety of mitigation strategies can be adopted. One possibility is to adapt the observing strategy by fine-tuning it to the characteristics (e.g., magnitude at $I$-band) of each target.

- **Searching for transit-like events.** We searched for periodic transit-like events in the photometric dataset for each target using the BLS algorithm. No such signal was recovered for any target. We refer to the interpretation of APACHE’s zero planet detection at the next bullet. It is however worthwhile to discuss some of the limitations of the BLS algorithm. BLS identifies interesting candidates by phase-folding individual photometric observations on a grid of trial periods, and deriving the relevant model parameters (transit depth, epoch of transit center, transit duration) that maximize significance of the transit signal in a least-squares sense. By assuming a flat out-of-transit light curve,
BLS by itself can have a tendency to fold up any (non-planetary) time-correlated structures in the data into seemingly significant candidates, when applied to real, wiggly light curves. Furthermore, BLS is not able to detect a single transit event even if clearly present in the data. As also discussed in Berta et al., 2012 [160], future improvements in the analysis include the development of a method to detect single transits and robustly assess their significance. This could be decisive in a quasi-real-time attempt to change the observing strategy and, consequently, increment our sensitivity to transiting planets.

- **Characterizing stellar rotations.** In our sample, we found clear evidence of periodic signals of approximately sinusoidal shape, which could be interpreted as due to the presence of rotating spots on the stellar photosphere. Looking for evidence of rotational modulations in the photometric data is an important task to characterize the astrophysical behaviour of cool stars and, taking into account the main goal of the APACHE survey, to potentially discard from further observations those stars showing short rotation periods. In fact stars which are fast rotators are expected to show high activity levels that represent an important source of noise in the RV measurements that can complicate the detection of a spectroscopic signal produced by a planetary companion. For these reasons, such studies can provide a good benchmark for RV surveys and follow-up measurements, such as the GAPS project with the HARPS-N spectrograph on the TNG (e.g., Covino et al. 2013 [155]; Desidera et al. 2013 [156]). In the analysis carried out as part of the thesis work we focused on signals that are clearly significant given our typical photometric precision and time sampling. As a next step, we need to quantify our real sensitivity in period and amplitude. This is mandatory in order to interpret in a statistically robust sense the presence of such type of signals and evaluate the possibility to perform multi-band follow-up studies to resolve doubtful cases.

- **Limits to transiting companions.** As mentioned above, no strong transiting planet candidates were detected during the first year of the APACHE survey. We utilized the photometric measurements obtained for our sample to carry out simulations aimed at determining what sensitivity to transiting companions (of given radius and period) \( S(R, P) \) we achieved on a star-by-star basis. We considered the range of orbital periods \( P = 0.5 - 5 \) days and the range of planetary radii \( R_p = 1 - 10R_\oplus \).

We found that at the end of the first observing season we were sensitive (recovery fraction \( \geq 30\% \)) to planets with radius \( \geq 6R_\oplus \) in the range \( 0.5 - 3 \) days. We were moderately sensitive (recovery fraction \( \geq 10\% \)) to planets with radius \( \sim 4R_\oplus \) in the range \( 0.5 - 1.5 \) days and we were least sensitive (recovery fraction \( \leq 5\% \)) to planets with radius \( \sim 2R_\oplus \).

Our ensemble sensitivity (see Figure 6.10 of Chapter 6) shows the number of planets that APACHE would have found if every M dwarf hosted exactly one transiting planet. Clearly, this is an unrealistic statement. By inverting a Binomial distribution, we can
translate this statement into a 95% confidence upper limit of the planet occurrence \((\text{planet/star}^{-1})\). Taking into account the still small fraction of the APACHE sample observed so far, we can compare our upper limits with those reported in the literature. We use as an illustrative example that of a transiting Neptune-size \((4R_{\oplus})\) planet. Mearth found no Neptune-sized exoplanets (Berta et al., 2013 [157]) and put a 95% confidence upper limit of \(< 0.15\text{planets/star}^{-1}\) for \(P < 3\text{days}\). This finding is consistent with the APACHE zero detections and the consequent upper limit of \(< 0.6\text{planets/star}^{-1}\) for \(P < 1.5\text{days}\). The recent result from the Kepler program (Dressing & Charbonneau, 2013 [173]) gives a planet occurrence of \(< 0.003\text{planet \cdot star}^{-1}\) for \(P < 10\text{days}\). For comparison, on the basis of two radial velocity planet detections from HARPS, Bonfils et al. (2013) [150] report that \(3^{\pm 1}\%\) of stars in their M dwarf sample host planets with minimum masses \((m \sin i)\) between 10 and 100\(M_{\oplus}\) (roughly overlapping the regime of radii to which APACHE is sensitive at present) and orbital periods of \(1 - 10\text{days}\). On the basis of one transit detection (Berta et al., 2013 [157]), Dressing & Charbonneau (2013) [173] find that the Kepler early M dwarfs host \(0.004^{+0.006}_{-0.002}\) planets per star, for planets with radii of \(4.0 - 5.7R_{\oplus}\) and periods \(<10\text{days}\). The difference between these two numbers can be attributed to features in the densities for planets in this regime (see Wolfgang & Laughlin 2012 [174]) but a detailed investigation is beyond the scope of this work. The upper limits from APACHE (and MEarth) do not contradict either of these measurements. As mentioned above, the APACHE sample is far from being complete but these considerations highlight the need for further investigation. The relevance of the project is clear considering that the APACHE input catalogue (AIC) cover unexplored regions in terms of stars masses (and radii) with respect to the MEarth input catalogue. Furthermore, the AIC is focused on the bright \(J < 10\text{mag}\) M dwarfs in the solar neighbourhood while the M dwarfs in the Kepler Input Catalogue (KIC) are generally fainter (the mean value of \(J\) is \(\sim 13\text{mag}\)). This means that APACHE represents an opportunity to provide favourable targets for spectroscopic confirmation (mass determinations, density estimates) and atmospheric characterization studies.

### 9.2 Astrometric detection of giant planets around nearby M dwarfs: the Gaia potential

The work detailed in the last two Chapters of this thesis can be seen as a way of initially measuring the impact of Gaia astrometry in the realm of exoplanets orbiting low-mass M dwarf stars in the solar neighborhood and of investigating in a preliminary fashion some of the most relevant synergies between Gaia data on exoplanets and those that will come from other ongoing and planned planet search and characterization programs, such as APACHE. The main findings in these experiments can be summarized as follows:
1) the overall quality of orbit reconstruction (using $P$, $M_p$, and $\zeta/\sigma_{AL}$ as proxies) allow us to infer the possibility for Gaia to detect over two thousand new giant planets around low-mass stars, and to derive accurate masses and orbital parameters for as many as five hundred systems. Consequently, it will be possible to determine occurrence rates of intermediate-separation giant planets around M dwarfs with expected uncertainties of just a few percent, critically improving present-day estimates.

2) for detected giant planets with periods in the range $0.2 - 5$ yr (i.e., with accurately determined masses and orbits), inclination angles corresponding to quasi-edge-on configurations will be determined with enough precision (a few percent) so that it will be possible to identify candidate transiting planets in a regime of orbital separations which is inaccessible from the ground and only marginally probed from space. Furthermore, mid-transit epochs will be determined with enough precision ($\sim 20$ days) at orbital periods of $\approx 2$ yr so as to provide for very interesting targets for follow-up photometry and spectroscopy.

3) for well-sampled orbits ($P < T$), the uncertainties on planetary ephemerides, separation $\varrho$ and position angle $\vartheta$, will degrade at typical rates of $\Delta \varrho < 1$ mas yr$^{-1}$ and $\Delta \vartheta < 2$ deg yr$^{-1}$, respectively. These are over an order of magnitude smaller than the degradation levels attained by present-day ephemerides predictions based on mas-level precision astrometry;

4) Planetary phases will be measured with typical uncertainties $\Delta \lambda$ of several degrees. The combination of detailed models of giant exoplanets’ systems and reliable ephemerides from Gaia astrometry could then greatly help both in the selection of good targets for direct-imaging instruments and for the physical interpretation of positive observational results.

Indeed, several important issues will be worthy of future investigations, particularly now that Gaia is ongoing. For example, it would be valuable to provide an assessment of the effectiveness of the combination of Gaia data with high-precision RVs and an extension of the study presented in Chapter 7 to multiple-systems configurations also ought to be carried out. We have assumed all stars in the Lépine (2005) [102] LSPM sub-sample and Lépine & Gaidos (2011) [139] catalog used here to be single, but this is not likely to be a realistic approximation, and the problem of astrometric planet detection in the presence of orbital motion induced by a distant companion star (e.g., Sozzetti 2005 [175]) will have to be tackled eventually. The fine details of the synergy resulting by the actual combination of Gaia and direct imaging devices data for improving the interpretation of observables in reflected light (phase curves, geometric albedos, polarization parameters) of extrasolar planets in terms of the underlying scattering mechanisms and in turn
chemical and thermal properties of their atmospheres have also been left largely unexplored. Nevertheless, the results presented here help to quantify the actual relevance of the Gaia observations of the large sample of nearby M dwarfs in a synergetic effort to optimize the planning and interpretation of follow-up/characterization measurements of the discovered systems by means of transit photometry (e.g., APACHE), and upcoming and planned ground-based as well as space-borne observatories for direct imaging (e.g., VLT/SPHERE, E-ELT/PCS) and simultaneous multi-wavelength spectroscopy (e.g., JWST).
Bibliography


Bibliography


