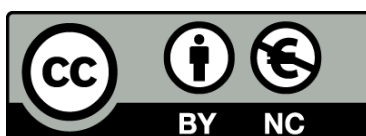




UNIVERSITAT_{DE}
BARCELONA

New contributions to algorithms and tools for the analysis of photometric and spectroscopic time-series in exoplanet searches

Albert Rosich Salgado



Aquesta tesi doctoral està subjecta a la llicència **Reconeixement- NoComercial 4.0. Espanya de Creative Commons**.

Esta tesis doctoral está sujeta a la licencia **Reconocimiento - NoComercial 4.0. España de Creative Commons**.

This doctoral thesis is licensed under the **Creative Commons Attribution-NonCommercial 4.0. Spain License**.

Tesi doctoral

*New contributions to algorithms and tools for the
analysis of photometric and spectroscopic time—series in
exoplanet searches*

Autor:

Albert Rosich Salgado

Director:

Dr. Ignasi Ribas Canudas



UNIVERSITAT_{DE}
BARCELONA

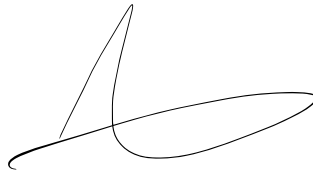
*New contributions to algorithms and tools for the
analysis of photometric and spectroscopic time—series in
exoplanet searches*

Memòria presentada per optar al grau de doctor
per la Universitat de Barcelona

Programa de doctorat en Física

Autor:

Albert Rosich Salgado



Director:

Dr. Ignasi Ribas Canudas

Tutor:

Dr. Alberto Manrique Oliva

BARCELONA, SETEMBRE 2021



UNIVERSITAT_{DE}
BARCELONA



New contributions to algorithms
and tools for the analysis of
photometric and spectroscopic
time—series in exoplanet searches

Ad astra per scientia.

Contents

Agraïments	xiii
Resum de la Tesi	xv
I Overview	1
1 Introduction	3
1.1 The exoplanet revolution	3
1.2 Time-series analysis in exoplanet science	7
1.3 Stellar activity	8
1.3.1 Multiscale approximation to stellar variability	8
1.4 Effects of stellar activity on the observables	10
1.4.1 Chromatic effects on planetary transits	10
1.4.2 Effects on RVs	11
1.4.3 Surface modelling: Turning noise into signal	12
1.4.3.1 Active star modelling: The State-of-the-Art	13
1.5 This Thesis	14
II Time series analysis	17
2 MGLS: Multidimensional Generalised Lomb Scargle Periodogram	19
2.1 Introduction	19
2.1.1 Lomb–Scargle Periodogram	20
2.1.2 Generalised Lomb–Scargle Periodogram	21
2.1.3 Normalisation	21
2.1.4 Loglikelihood periodograms and jitter	23
2.2 Iterative subtraction of signals	25
2.2.1 Prewhitening pitfalls: A toy model experiment	25

2.3	MGLS: Multidimensional Generalised Lomb Scargle	27
2.3.1	Algorithm description and mathematical model	28
2.3.2	Global optimisation approach	29
2.3.3	Solving for linear parameters	29
2.3.3.1	Uncertainties in the linear parameters	32
2.3.4	Solving for nonlinear parameters: optimal (ω_j, s_k) search	33
2.3.4.1	Uncertainties in nonlinear parameters	34
2.4	Analysis of peak significances	35
2.4.1	Analytical formulations: Scargle formula	36
2.4.1.1	Independent frequencies	37
2.4.1.2	Random fields extreme value theory	38
2.4.2	Numerical approaches	38
2.4.2.1	Bootstrap method	38
2.5	Experiments on FAP determination	39
2.5.1	Fitting Scargle's formula	39
2.5.2	Proposed procedure for MGLS	41
2.6	Notes on the numerical performance and code implementation	42
2.6.1	Numerical throughput	42
2.6.2	Code architecture and usage	43
2.7	Conclusions	44
3	MGLS performance evaluation on real multiplanetary systems	47
3.1	MGLS Testbed	47
3.1.1	HD 40307	49
3.1.1.1	HD 40307 signal search and analysis	49
3.1.2	GJ 163	55
3.1.2.1	GJ 163 signal search and analysis	55
3.1.3	GJ 667C	58
3.1.3.1	GJ 667C signal search and analysis	59
3.1.4	Teegarden's Star	60
3.1.4.1	Teegarden's star signal search and analysis	60
3.1.5	Barnard b: A super-Earth orbiting at the snow-line of Barnard's star	68
3.1.5.1	Time-series analysis and discussion	68
3.2	Conclusions	73
III	Modelling active stars	75
4	The StarSim 2 project: Fast modelling code of heterogeneous photospheres of active stars	77

4.1	StarSim modelling code	77
4.2	StarSim 2: A fast implementation of a parametric model	78
4.2.1	Physical model	78
4.2.1.1	Atmosphere models	78
4.2.1.2	Parameter system	79
4.2.1.3	Spot model	80
4.3	StarSim Forward Problem	83
4.3.1	Surface integration	83
4.3.2	Generating photometry observables	85
4.3.2.1	Transiting planet model	85
4.3.3	Generating spectroscopic observables	85
4.4	StarSim Inverse Problem	89
4.4.1	Spot map optimisation	91
4.4.1.1	Simulated Annealing	92
4.4.2	Parameter inference	93
4.5	Code architecture, usage and benchmarking	95
5	The Photometric Inverse Problem	99
5.1	Introduction	99
5.2	Analytical foundations of the inverse problem	100
5.2.1	Implications on attainable information	101
5.3	The inverse problem	103
5.3.1	Objective function	104
5.3.2	Optimising the surface distribution of active regions	104
5.3.2.1	Simulated annealing optimization	104
5.3.3	Optimising stellar parameters	105
5.4	Description and handling the degeneracies on the inverse problem	107
5.4.1	Size–latitude degeneracy	107
5.4.2	Size–temperature degeneracy	107
5.4.3	Spot–facula degeneracy	108
5.4.4	Latitude–inclination degeneracy	108
5.4.5	Zero point degeneracy	108
5.5	Toy model inversion	109
5.5.1	Model	109
5.5.2	Analysis	110
6	An activity model for WASP-52	113
6.1	Introduction	113
6.2	The WASP-52 exoplanetary system	113
6.2.1	Transmission spectroscopy early studies	113

6.3	Photometric observations of WASP-52	114
6.4	Photometric inverse problem for WASP-52	115
6.4.1	Fixed StarSim parameters	115
6.4.2	Photometric fits results	117
6.4.3	Multiband fits	121
6.4.3.1	The evolving surface of WASP-52	121
7	Chromatic spot effects on simulated transits	125
7.1	Introduction	125
7.2	Transit depth variability due to spots	126
7.2.1	Analytical formalism	127
7.2.2	Transit simulations	129
7.3	Correcting transit depth for stellar activity	133
7.3.1	Simulations	134
7.3.2	Activity attenuation factor	134
7.3.3	Comparison with alternative activity correction methods for WASP-52 b	137
7.3.4	Higher order effects: Limb darkening	138
IV	Conclusions and Future Work	141
8	Discussion and Future Work	143
8.1	On time-series analysis	143
8.1.1	Future work	144
8.2	On active surface modelling techniques	144
8.2.1	Future Work	147
8.2.1.1	Prospects of joint photometry and spectroscopic modelling	149
	List of Figures	155
	List of Tables	157
	Abbreviations	159
	Publications	161
	Bibliography	165

Agraïments

Fer una tesi doctoral no és mai una feina fàcil. Acostuma a ser un camí complicat en el que sovint un es troba amb dificultats i cruïlles en les que cal triar el camí correcte.

Al llarg del desenvolupament d'aquesta tesi m'he trobat amb persones excepcionals que m'han ajudat a orientar-me en el laberint. En primer lloc, l'Ignasi Ribas, una de les millors referències científiques que he tingut. Sense la seva encertada visió, lleial perseverança i qualitat humana aquest treball no hauria arribat a port. També als companys de grup, Kike Herrero, Manuel Perger, Juan Carlos Morales, Guillem Anglada-Escudé, Marina Lafarga i David Baroch. Sempre he pogut comptar amb la seva inestimable ajuda.

També als qui van ser els meus companys de la RACAB durant el període de la beca de col·laboració al TFRM, en Jorge Núñez, l'Octavi Fors, el Daniel del Ser i la Maite Merino.

I a la meua família, com sempre, present en els neguits de cada dia amb tot el seu suport i paciència.

A tots vosaltres moltes gràcies!

Resum de la Tesi

Noves contribucions en algorismes i eines per a l'anàlisi de sèries temporals fotomètriques i espectroscòpiques per a la cerca d'exoplanetes.

Context

Des de la descoberta del primer planeta orbitant una estrella de la seqüència principal (Mayor et al. [1995](#)), la ciència d'exoplanetes ha esdevingut un camp molt actiu i dinàmic. La detecció recent del planeta Proxima b (Anglada-Escudé et al. [2016a](#)) o el sistema multiplanetari TRAPPIST-1 (Gillon et al. [2017](#)) o la confirmació d'un planeta a l'estel de Barnard (Ribas et al. [2018](#)) han esdevingut fites molt rellevants en la ciència dels exoplanetes, pel fet de ser molt semblants a la Terra en massa, per ser molt pròxims al Sol, i els dos primers, per tenir components situats a la zona d'habitabilitat. Per aquest tipus de planetes, la possibilitat d'existència de vida, tot i que encara lluny de ser establerta, no és descartable de manera que es tracta d'uns objectes certament interessants.

Fins a la data de finalització d'aquesta tesi, han estat descoberts 5 747 exoplanetes, la majoria pel mètode del trànsit fotomètric a partir de les extraordinàries dades del satèl·lit Kepler. Malgrat la sorprenent productivitat recent dels telescopis fotomètrics en òrbita, la tècnica de la velocitat radial ha estat la més eficient —encara que no la més eficaç— per a la detecció de senyals planetaris. Donat que la probabilitat de trànsit és relativament baixa ($\sim 1\%$, per períodes curts) s'han d'estudiar un gran nombre d'estrelles simultàniament per tal d'obtenir una ràtio de descobriments raonable en una campanya observacional. Tanmateix, l'alta precisió fotomètrica de satèl·lits com Kepler possibilita la detecció de trànsits de petits planetes rocosos i fa que aquesta tècnica, des de l'espai, sigui eficaç en el sentit de nombre de descobriments.

La tècnica espectroscòpica en canvi, permet detectar moviments periòdics de l'estrella a causa del reflex gravitatori dels potencials planetes, de manera que és possible la detecció d'aquests en inclinacions orbitals més baixes per a les quals no es poden observar trànsits. A més, l'efecte gravitatori és present en qualsevol moment de l'observació, mentre que en el cas dels trànsits, s'ha de mantenir una observació continuada, ja que el trànsit és, típicament, curt en comparació amb el període orbital. Així doncs, la tècnica de trànsit ha permès moltes deteccions de planetes de curt i mitjà període, però la seva eficàcia cau per a planetes de llarg període, ja que la probabilitat de trànsit esdevé insignificant.

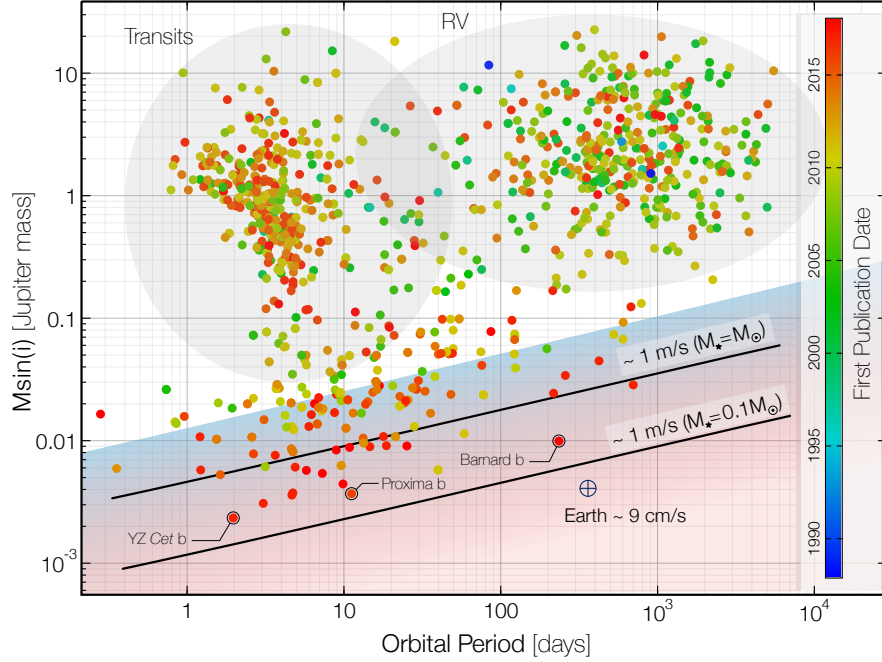


Figura 1: Descobriments d'exoplanetes pel mètode de la velocitat radial, trànsits i altres mètodes —lents gravitatòries, TTV, imatge directa, ...—. El color de la barra lateral codifica l'any de publicació. S'observen dos grups principals: a la part superior esquerra s'acumulen planetes descoberts principalment per mitjà del mètode de trànsits des de terra (programes WASP, HAT, KELT, XO, ...); mentre que a la dreta hi ha els primers descobriments pel mètode d'espectroscòpia Doppler. La línia negra contínua mostra un efecte en velocitat radial d'aproximadament 1 m/s que representaria el límit dels espectrògrafs actuals. Per sota d'aquesta, s'hi poden trobar alguns planetes descoberts per mitjà de trànsits des de l'espai com Kepler, però també per mitjà de la velocitat radial, p.ex. GJ 699 (Barnard b). La zona ombrejada inferior indica el règim on l'activitat estel·lar és comparable amb el senyal planetari, i per tant, pot esdevenir un problema per a la seva detecció. A tall d'exemple, s'indica la posició en el pla $P_{rot}-M \sin(i)$ que ocupa la Terra. Es fa evident que per a la descoberta d'anàlegs terrestres, a més d'un salt tecnològic important, és essencial també una bona comprensió dels efectes de l'activitat estel·lar sobre els observables. També s'observa una saturació en la massa mínima. Aquesta és deguda fonamentalment a dos factors: i) la tecnologia de detecció de planetes data de principis dels anys 2000, i, fins a 2017 no hi ha hagut grans canvis en espectrògrafs en servei; i en segon lloc s'ha arribat al límit de la detectabilitat. Elaborat amb dades de exoplanets.org (24/6/2020).

Estructura de la Tesi

La Tesi consisteix en el desenvolupament de tècniques i eines de codi per a la seva implementació en la recerca d'exoplanetes, i consta de dues parts ben diferenciades. La primera, versa sobre tècniques d'anàlisi de períodes en sèries temporals d'observables típics en aquest camp, p.ex. fotometria, velocitat radial, ...i es presenta el desenvolupament d'una extensió multidimensional a l'algoritme GLS (Generalised Lomb–Scargle).

En l'anàlisi de periodicitats de mesures de velocitat radial aquest algoritme ha esdevingut d'utilització gairebé omnipresent en les publicacions i treballs referents a la descoberta de planetes, així com també en altres camps de la ciència en els que es disposi de dades no-equiespaiades i on la utilització de la transformada de Fourier no és òptima.

La versió que presentem, l'anomenem MGLS (Multidimensional Generalised Lomb-Scargle) permet l'anàlisi de sèries temporals ajustant simultàniament n freqüències a diferència del mètode habitual de substracció iterativa de les senyals detectades. El mètode MGLS presenta una major robustesa al soroll especialment quan el senyal analitzat es compon de senyals d'alta amplitud i baixa amplitud amb períodes semblants. En els casos més patològics, la descomposició iterativa pot induir senyals no existents a les dades o, simplement no detectar senyals presents.

La segona part, tracta el problema general i recurrent dels efectes de l'activitat estel·lar sobre els observables planetaris. Aquests poden ser des d'observacions de velocitat radial fins a espectres de trànsits d'exoplanetes per a l'anàlisi de les seves atmosferes.

Tècniques d'anàlisi de sèries temporals

Descripció de l'algoritme MGLS

L'extensió MGLS considera un sistema d' M components d'òrbites circulars de la forma,

$$y_{i,k}(t_{i,k} | \theta, \omega) = c_k + \sum_j^{\mathcal{D}} a_j \cos(\omega_j t_{i,k}) + \sum_j^{\mathcal{D}} b_j \sin(\omega_j t_{i,k}) + \tau(t_{i,k} - t_0) \quad (1)$$

$$K_j = \sqrt{a_j^2 + b_j^2},$$

on $y_{i,k}$ és la mesura k a l'instant t_i i M és la dimensionalitat (nombre de freqüències, ω_j). a_j i b_j són els coeficients del component j , d'amplitud K_j . c_k és el punt zero del conjunt de dades k . τ és una variació lineal.

La Figura 2 mostra un dels avantatges dels periodogrames multifreqüència. En el cas de $\mathcal{D} = 2$ es pot representar gràficament els parells de freqüències que maximitzen la versemblança del model global. El cas que es mostra, correspon a l'anàlisi de gairebé 20 anys de dades de velocitat radial de l'estrella de Barnard, es veu que la configuració òptima correspon a (233, 1890) dies. No obstant, examinant el gràfic es veuen altres configuracions —tot i que subòptimes— amb un valor de $\Delta \ln \mathcal{L}$ molt semblant, i en alguns casos estadísticament equivalent. El cercle blanc indica la configuració finalment escollida en l'anàlisi (233, 6600) dies.

Al llarg del capítol 3 es demostra la major capacitat de l'esquema simultani (MGLS) per detectar senyals compostos i evitar falsos positius, respecte l'estratègia seqüencial (GLS + prewhitening), mantenint la simplicitat de l'algoritme GLS.

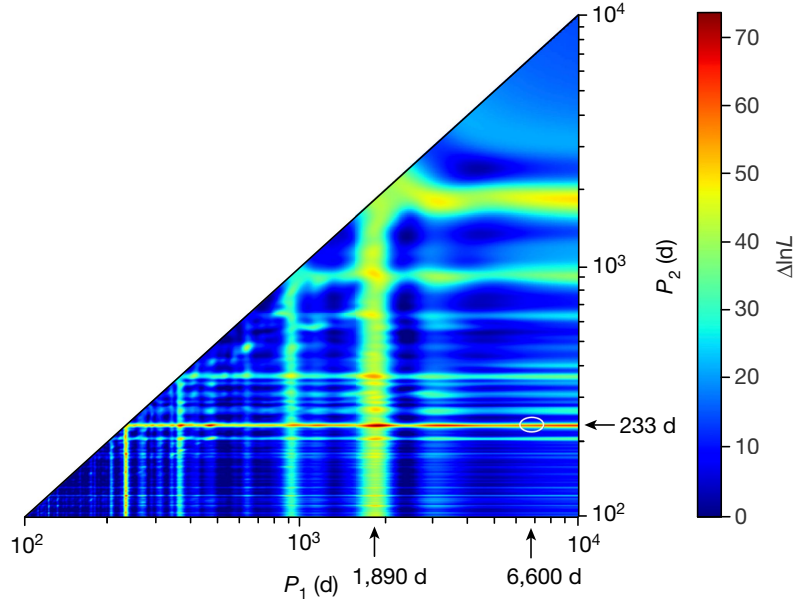


Figura 2: Mapa de color de l'escombrat MGLS bidimensional per dades de l'estel de Barnard compostes per 8 sèries observacionals de velocitat radial obtingudes durant gairebé 20 anys. L'escala de color mostra la millora de la versemblança de l'ajust en funció de la parella de períodes de prova. Es mostra clarament un període a 233 dies que combinat amb altres dóna solucions d'alta significança estadística. El valor màxim de $\ln \mathcal{L}$ es troba per (233, 1890) ($\Delta \ln \mathcal{L} = 71$).

Modelització de l'activitat estel·lar

La part que fa referència a la modelització de l'activitat estel·lar per rotació de taques, consisteix en el desenvolupament del codi de simulació StarSim 2, evolució de la versió preexistent presentada per Herrero et al. (2016) i que permet la realització del problema invers.

El projecte StarSim

StarSim genera sèries temporals dels observables de fotometria en qualsevol banda i mesures espectromètriques com la velocitat radial (RV) i índexs d'activitat (BIS, FWHM, Contrast, ...) prenent com a entrada un mapa de la superfície tacada, \mathcal{S} i un conjunt de paràmetres estel·lars, θ . El codi genera un model rotatiu de superfície sintètica heterogènia construït per integració d'elements finits en la que es divideix la superfície. Es consideren tres tipus d'elements de superfície: fotosfera immaculada (ph), taques fredes (sp) i fàcules (fc). Les seves característiques espectrals es reproduïxen a través dels espectres sintètics BT-Settl (Allard et al. 2013) generats amb el codi Phoenix. L'únic paràmetre que diferencia els tres tipus de superfície és la seva temperatura efectiva, $T_{sp} < T_{ph} < T_{fc}$.

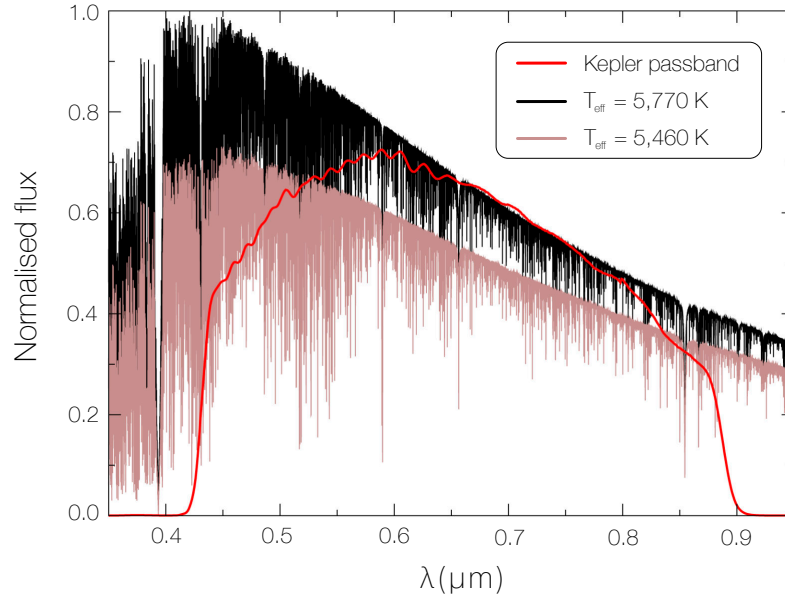


Figura 3: Exemples d'espectres sintètics BT-Settl. En negre es representa un anàleg solar ($T_{eff} = 5770$ K) i $\log g = 4.5$, comparat amb un de $\Delta T = 310$ K inferior (marró). En vermell es representa la funció de transferència del filtre fotomètric de la missió Kepler.

El model inclou tots els efectes significatius sobre les intensitats de flux i les seves distribucions espectrals. El procés de construcció de l'espectre de la superfície heterogènia consisteix en interpolar cadascun dels diferents espectres, $f_{ph}(\lambda, T_{ph})$, $f_{sp}(\lambda, T_{sp})$ i $f_{fc}(\lambda, T_{fc})$. La resta de paràmetres dels espectres que StarSim considera, $\log g$, metal·licitat i $[\alpha/Fe]$ no varien significativament al llarg de la superfície heterogènia.

Posteriorment es calculen els espectres de Kurucz (ATLAS9) (Kurucz 2017) que reprodueixen els perfils d'intensitat de cadascun dels elements de superfície al llarg del disc incloent el càlcul dels coeficients del limb-darkening en funció de $\mu = \cos \theta$,

$$LD_{ph,sp}(\lambda, \mu) = \frac{I_{ph,sp}(\lambda, \mu)}{I_{ph,sp}(\lambda, 0)} \quad (2)$$

on $I_{ph,sp}(\lambda, 0)$ representa la intensitat al centre. D'aquesta manera, l'espectre de cada element de superfície es multiplica per aquest coeficient en funció de la seva posició relativa respecte la normal de l'esfera,

$$f(\lambda, \mu) = f(\lambda) \cdot LD(\lambda, \mu) \quad (3)$$

Per obtenir les sèries temporals dels observables els fluxos per cada element de superfície es multipliquen pel projector respecte a la línia d'observació. Així, el flux de l'element j queda com,

$$f_j(\lambda, \mu_j) = f(\lambda) \cdot LD(\lambda, \mu_j) \cdot a_j \cdot \max(0, \mu_j) \quad (4)$$

on $a_j = 2 \cdot \Delta\alpha \cdot \sin(\Delta\alpha/2) \cdot \sin(\theta_j)$ és l'àrea de l'element de superfície j , i

$$\mu_j = \sin i_* \sin \theta_j \cos(\phi_j + \omega t) + \cos i_* \cos \theta_j \quad (5)$$

és el cosinus de l'angle de projecció. L'expressió de l'Equació (4) $\max(0, \mu_j)$ garanteix la condició de visibilitat de l'element de superfície ($\mu \geq 0$). i_* és la inclinació de l'eix de rotació estel·lar, θ i ϕ són les coordenades de colatitud i longitud, respectivament. En el cas d'StarSim 2, la superfície de l'element del grid a_j es substitueix directament per l'àrea d'una circumferència sobre l'esfera, $a_j = 2\pi(1 - \cos(r_j))$, on r_j és el radi angular de la taca circular considerada, i les propietats físiques de la taca es calculen al centre d'aquesta.

La quantitat $\mu_j = \mu_j(t)$ és funció explícita del temps com a resultat de considerar un sistema de coordenades (θ, ϕ) comòbil a l'esfera rotativa de velocitat angular ω . Adicionalment, hem implementat un model de rotació diferencial donat per l'expressió (Beck 2000)

$$\omega = \omega_0 - \delta_{rot}[B \sin^2(\varphi) + C \sin^4(\varphi)] \quad (6)$$

on ω_0 representa la freqüència de rotació fonamental, és a dir, la freqüència de rotació equatorial. φ és la latitud de l'element actiu i $B = 2.39 \text{ deg} \cdot \text{day}^{-1}$; $C = 1.78 \text{ deg} \cdot \text{day}^{-1}$ per a un model solar (Snodgrass et al. 1990). δ_{rot} és un factor que quantifica el grau de rotació diferencial. Per $\delta_{rot} = 1$ es recupera el model solar i per $\delta_{rot} = 0$, el model d'esfera rígida.

Per calcular el flux integrat bolomètric F a tot el disc caldrà simplement sumar sobre tots els elements de superfície

$$F_{bol}(\lambda, t) = \sum_j f_{ph}^j + \Delta f_j \quad (7)$$

on l'últim terme del sumatori és la variació de flux produït per les regions actives,

$$\begin{aligned} \Delta f_j(\lambda) = & (f_{sp} \cdot J_{sp}(\lambda) - f_{ph}(\lambda) \cdot J_{ph}(\lambda) + \\ & + f_{fc} \cdot J_{fc}(\lambda) - f_{ph}(\lambda) \cdot J_{ph}(\lambda)) \end{aligned} \quad (8)$$

on el primer terme del sumatori és el dèficit de flux produït per les taques, mentre que el segon representa el sobreflux que proporcionen les fàcules. Les expressions J_{ph} , J_{sp} , i J_{fc} recullen la contribució del limb-darkening i els efectes projectius sobre el flux d'acord amb la posició de cada element de superfície j .

$$\begin{aligned} J_{ph}^j(\lambda) &= \frac{I_{ph}(\lambda, \mu_j)}{I_{ph}(\lambda, 0)} \cdot a_j \cdot \max(0, \mu_j) \\ J_{sp}^j(\lambda) &= \frac{I_{sp}(\lambda, \mu_j)}{I_{sp}(\lambda, 0)} \cdot a_j \cdot \max(0, \mu_j) \\ J_{fc}^j(\lambda) &= c_{fc}(\mu_j) \cdot a_j \cdot \max(0, \mu_j) \end{aligned} \quad (9)$$

En el cas de les fàcules es coneix que presenten un efecte de sobreflux al limbe (Frazier et al. 1978; Berger et al. 2007). Les observacions mostren que les fàcules dominen la irradiància en estrelles tipus solar, on les regions actives presenten una brillantor ~ 1.2 vegades més gran al limbe que al centre. En el nostre codi utilitzem la formulació seguint Meunier et al. (2010a),

$$c_{fc}(\mu_j) = \left(\frac{T_{eff} + \Delta T_\mu(\mu_j)}{T_{eff} + \Delta T_{fc}} \right)^4, \quad (10)$$

on $\Delta T_\mu(\mu_j) = a_\mu + b_\mu \cdot \mu_j + c_\mu \cdot \mu_j^2$. Els coeficients utilitzats són $a_\mu = 250.9$, $b_\mu = -407.4$ i $c_\mu = 190.9$ de tal manera que $c_{fc} \sim 1$ al centre del disc i $c_{fc} \sim 1.16$ a prop del limbe per una estrella de tipus solar.

Sèries fotomètriques

Un cop hem obtingut el flux integrat a tot el disc (Eq. 7), es normalitza respecte el flux de la fotosfera immaculada i es convoluciona amb la funció de transmissivitat d'un filtre, $T_f(\lambda)$

$$F_n(\lambda, t) = \frac{\sum_j f_{ph}^j + \Delta f_j}{\sum_j f_{ph}^j} \cdot T_f(\lambda) \quad (11)$$

en aquest punt i en el cas que $T_f(\lambda) = 1$ obtenim una sèrie temporal d'espectres de flux bolomètric. Per una funció T arbitrària, i integrant per totes les longituds d'ona, s'obté el flux instrumental normalitzat i en una banda determinada, B de la superfície heterogènia de l'estrella.

$$F_n(t) = \int_\lambda F_n(\lambda, t) d\lambda \quad (12)$$

Sèries espectroscòpiques

Per simular els observables derivats de mesures espectroscòpiques —velocitat radial i índexs d'activitat— es correlaciona un espectre d'alta resolució ($R > 100\,000$) amb una màscara específica per un tipus espectral determinat, de línies prèviament seleccionades a fi de minimitzar la seva variabilitat a causa d'activitat estel·lar.

En el codi StarSim hem utilitzat els espectres sintètics d'alta resolució Phoenix (Husser et al. 2013) per calcular les funcions de correlació creuada (CCF) entre els espectres i la màscara. Donat que per tal de ser sensibles a variacions de velocitat per efecte Doppler de pocs metres per segon, la resolució dels espectres ha de ser necessàriament molt alta amb el conseqüent problema computacional que ocasiona. Per aquest motiu, el mode espectroscòpic del codi treballa a l'espai de les CCF, en lloc del seu espai natural dels espectres.

En primer lloc es correlacionen els espectres de cadascun dels tipus de superfície (ph,sp,fc) amb la màscara corresponent al tipus espectral de l'estrella que modelitzem. En els espectres de Phoenix, s'inclou una model·lització dels efectes de blueshift convectiu que volem substreure per incorporar un model més sofisticat de CIFIST 3D

(Ludwig et al. 2009). El procediment consisteix en ajustar un polinomi de grau baix al bisector de la CCF calculada amb els espectres de Phoenix, per després integrar els bisectors dels models CIFIST 3D.

La manera de procedir per integrar la superfície heterogènia en aquest mode consisteix en assignar una CCF a cada element de superfície, que s'escala d'acord amb els seus elements projectius i el seu flux espectral (obtingut de forma anàloga en el cas fotomètric, veure equacions anteriors) tenint amb compte el tipus de superfície que ocupa la cel·la. Addicionalment, a cada CCF se li afegeix el desplaçament Doppler corresponent $\lambda_j = \lambda + \Delta\lambda_j$

$$\Delta\lambda_j = 8.05 \cdot \lambda \cdot \frac{1}{c} \cdot R_{star} \cdot \omega \sin i_* \sin \theta_j \sin \phi_j \quad (13)$$

on θ_j i ϕ_j són les coordenades colatitud i longitud de l'element de superfície j , i ω la velocitat angular de rotació (Eq. 6).

La CCF resultant de la superfície integrada s'obté de la suma de totes les CCF parcials

$$CCF(v) = \sum_j CCF_j(v) \quad (14)$$

Es poden distingir dos efectes principals en la modelització de la velocitat radial. L'efecte Doppler provocat per la diferent velocitat de les cel·les que componen el reticulat del model i les característiques espectrals en funció del tipus de superfície (immaculada, taca o fàcula); i l'efecte de la convecció. La Figura 4 mostra un model d'una sola taca amb fàcula (casos $Q = 0$ (vermell)/ $Q = 3$ (blau)) que rota en una estrella de tipus solar amb $P_{rot} = 25$ dies. Les corbes puntejades mostren el senyal de RV tenint en compte només l'efecte del bloqueig convectiu de la taca. S'observa que és una funció simètrica respecte el pas de la taca pel centre del disc, que correspon al moment en que el bloqueig convectiu és màxim a causa de la geometria del camp de velocitats. Les línies discontinües mostren l'efecte Doppler, que mostra també un perfil simètric respecte el pas pel centre. La meitat esquerra de l'estrella mostra blueshift ($v < 0$) i la meitat dreta, redshift ($v > 0$). Quan en una d'aquestes meitats hi ha un element de superfície activa, el fet de tenir una temperatura diferent presenta una emissivitat diferent i per tant el balanç Doppler hemisfèric es trenca. Finalment, en línies contínues es mostra l'efecte conjunt. El resultat és un senyal asimètric —a causa del terme convectiu, que sempre és positiu¹— i assolix els extrems en un punt intermig, resultat de l'equilibri velocitat i emissivitat de la superfície radiant d'acord amb els models de limb-darkening/limb-brightening.

La Figura 5 mostra un exemple de corbes generades per StarSim 2 utilitzant els paràmetres de la Taula 1.

1. Cal mencionar que, malgrat el terme convectiu durant el trànsit de la taca dona sempre com a resultat un comportament net de blueshift ($v < 0$) l'efecte que es mostra al gràfic és sempre positiu a causa de la substracció de la velocitat sistèmica (constant) durant el pas de la taca.

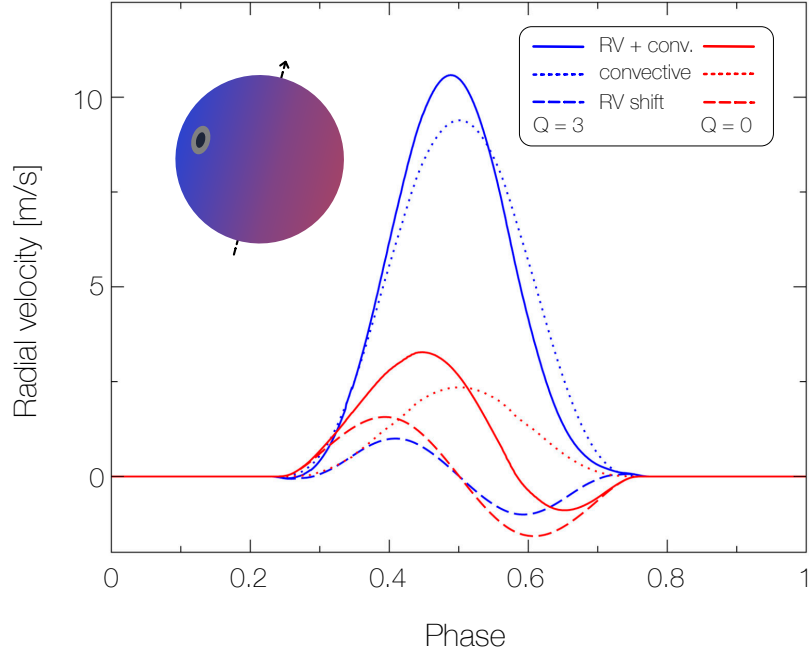


Figura 4: Senyal de velocitat radial per un sistema d'una taca equatorial i $1.6 \cdot 10^{-3}$ superfícies estel·lars per un equivalent solar ($P_{rot} = 25$ dies). En línies discontinúes es mostra exclusivament l'efecte Doppler, mentre que en línies puntejades es mostra l'efecte convectiu. L'efecte total del bloqueig de la convecció per la regió activa i el desplaçament Doppler es mostra en línies contínues. En vermell, el model d'activitat no presenta fàcula ($Q = 0$) i en blau $Q = 3$.

StarSim 2 i el problema invers

En la versió inicial de codi StarSim, la integració de la superfície té lloc per suma de les cel·les elementals de mida configurable. Aquesta implementació imposa sobre el temps càlcul un ordre al·lògmic $\sim \mathcal{O}^2$ només en l'escombrat de l'àrea ocupada per les regions actives, cosa que limita la resolució dels elements de superfície per tal que el temps d'execució del codi sigui raonable. Cal tenir en compte que el flux d'execució del codi consisteix en un triple loop: *i*) recórrer les èpoques per les quals mostrejar les corbes dels observables, *ii*) recórrer totes les taques i, *iii*) recórrer totes les longituds d'ona dels espectres dels elements de superfície. Tot i que aquesta estratègia d'integració té avantatges com la lliure disposició de píxels actius per configurar formes arbitràries de les taques, per contra no permet la generació eficient de les sèries temporals, aspecte clau per a la inversió de corbes.

El problema directe consisteix en: donat un mapa de taques \mathcal{S} i un conjunt de paràmetres físics del sistema, θ , calcular els observables fotomètrics i/o espectroscòpics,

$$\mathbf{X}(t) = \mathbf{F}(\mathcal{S}, \theta) \quad (15)$$

on $\mathbf{X}(t)$ representa la sèrie temporal dels observables. \mathbf{F} és el model StarSim.

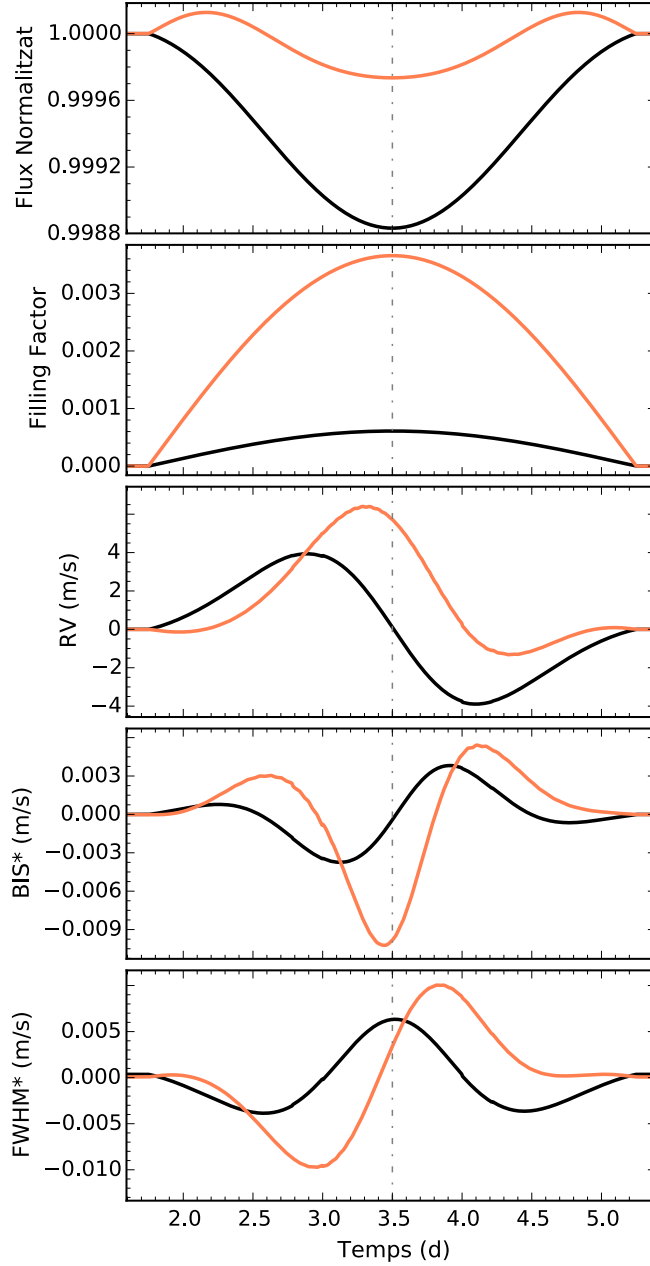


Figura 5: Exemples de corbes generades per StarSim 2 per una estrella de tipus K5V i un mapa compost per una sola taca equatorial de radi 2° .

Paràmetres: $T_{\text{eff}} = 4500 \text{ K}$, $\Delta T_{\text{sp}} = 800 \text{ K}$, $\Delta T_{\text{fc}} = 150 \text{ K}$, $P_{\text{rot}} = 7.0 \text{ d}$, $i = 90^\circ$ i sense rotació diferencial ($\delta_{\text{diff}} = 0$). En negre es representa la sortida per $Q = 0$ i en taronja per $Q = 3$, mantenint constant tots els altres paràmetres de la simulació. Els asterix (*) al bisector i FWHM indiquen quantitats respecte la mitjana de la sèrie, $\text{BIS}^* = \text{BIS} - \langle \text{BIS} \rangle$ i $\text{FWHM}^* = \text{FWHM} - \langle \text{FWHM} \rangle$. La línia discontinua indica el pas del centre de la taca pel meridià.

Taula 1: Paràmetres estel·lars i de simulació de les corbes d'exemple de la Figura 5.

Paràmetre	Valor
Rang espectral (nm)	450-900
T_{eff} (K)	4500
$\log g$	4.5
Contrast de temperatura de taques, ΔT_{sp} (K)	800
Contrast de temperatura de fàcules, ΔT_{fc} (K)	150
Ràtio fàcula/taca, Q	0 / 3
Període de rotació, P_0 (dies)	7.5
Inclinació eix rotació, i ($^\circ$)	90
Constant de rotació diferencial, δ	0
Vida mitjana taques (dies)	50
Desviació vida mitjana taques (dies)	20
Ràtio evolució de taques ($^\circ$ /dia)	1.5

El problema invers consisteix en trobar les causes que provoquen un determinat comportament en un sistema físic. En el cas que ens ocupa, el nostre interès és trobar el mapa de taques i els paràmetres estel·lars.

En un sistema simple d'una sola taca, és relativament fàcil i barat computacionalment invertir les dades i trobar els paràmetres i coordenades de l'única taca. No obstant, en sistemes més complexos i realistes, es necessiten un grup de taques per modelitzar satisfactòriament un sistema. Llavors la complexitat en el procés d'inversió augmenta considerablement, de manera que converteix el problema en un repte computacional, i molt sovint haver de tractar amb aproximacions a la solució òptima.

El codi StarSim 2 ha estat optimitzat considerant les taques com un sol element de superfície. Així, es calcula el flux de la fotosfera immaculada i s'afegeix el diferencial de flux de les regions actives, escalades a l'àrea que ocupen. Com a contrapartida, aquesta aproximació provoca aparicions i desaparicions de les taques en el limbe de forma sobtada. Tanmateix, amb sistemes de taques petites, aquestes desviacions es fan negligibles.

El problema invers fotomètric

El problema de determinar la superfície tacada juntament amb els paràmetres estel·lars que fan que el model StarSim 2 reproduïxi les variacions fotomètriques observades és l'objectiu d'aquesta part per l'importància de les seves aplicacions en la determinació dels efectes cromàtics que indueixen les taques en els trànsits planetaris en l'observació espectrofotomètrica de precisió per la caracterització d'exoatmosferes planetàries.

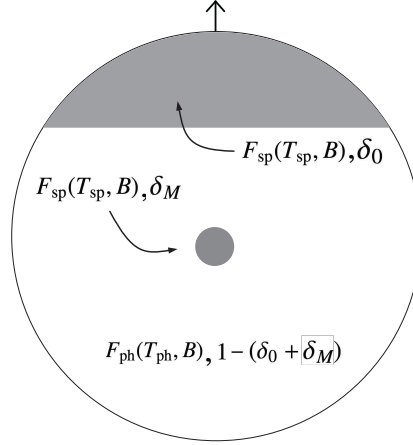


Figura 6: Esquema de fotosfera activa separant part modulante i no-modulant. El model es compon d'un casquet polar, sense contribució a la modulació de la corba de llum amb un filling factor δ_0 ; i una petita taca equatorial amb un filling factor dinàmic de valor màxim δ_M . $F(\cdot)$. La notació $\delta_M \cdot F(\cdot)$ representa la brillantor de cada element de superfície.

Fonaments analítics

A Rosich et al. (2020) es demostra la capacitat de la fotometria multibanda per la resolució del problema invers fotomètric. Si separem la cobertura de taques (filling factor) de la superfície estel·lar en una part modulante i una no-modulant obtenim una representació molt favorable per interpretar el grau d'indeterminació del sistema d'inversió.

La figura 6 mostra un esquema d'una superfície activa tipus, amb dues taques cadascuna amb un comportament dinàmic diferent. Qualsevol superfície arbitràriament tacada es pot reduir a un filling-factor modulante i no-modulant. Tot i que tots els elements de taques tenen, en realitat, un comportament modulante donat que tots traçen una trajectòria sobre l'esfera de propietats radiatives variables amb el temps, es poden definir estructures contínues de regions actives tals com casquets polars o bandes equatorials que de manera integrada el seu efecte no varia amb el temps.

En funció del nombre de bandes fotomètriques observades simultàniament d'una estrella activa obtenim la següent casuística:

- Si sols disposem d'observacions en una banda fotomètrica, el problema invers només pot determinar el filling-factor modulante. Tant la part no-modulant com la temperatura de contrast de les taques quedarà indeterminada. Aquest cas representa la gran majoria de campanyes observacionals relacionades amb exoplanetes tant des de la Terra com de l'espai. Molts autors assumeixen temperatures de taques mitjançant relacions calibrades empíriques (p.ex. Berdyugina 2005) i es negligeix l'efecte de possibles components no-modulants.
- En el cas de disposar dues bandes fotomètriques, es pot determinar tant el filling factor modulante com el contrast de temperatura de les taques ΔT_{sp} , malgrat la solució del sistema esdevé bivaluada i s'obtenen dos valors de temperatura diferents.

- Finalment, si es disposa de tres o més bandes la degeneració en la determinació del problema invers es trenca i es pot determinar unívocament les tres variables rellevants del problema (δ_M , δ_0 , and ΔT_{sp}). Tanmateix, la determinació de la component no-modulant (δ_0) és molt complicada ja que el debil efecte cromàtic roman estàtic. Anàlogament i pel mateix motiu l'efecte de contaminació cromàtica serà també petit i/o negligible. Aquesta circumstància fa que la determinació de δ_0 només sigui possible quan les bandes cobreixin regions espectrals molt distants i amb una precisió fotomètrica molt alta.

Model d'activitat de WASP-52 i efectes cromàtics en trànsits planetaris

Una de les nombroses aplicacions de la modelització física de la modulació rotacional d'estrelles actives és l'anàlisi i correcció dels efectes cromàtics que les superfícies heterogènies imprimeixen sobre les observacions espectrofotomètriques de trànsits.

Com a cas d'exemple hem analitzat sèries temporals de fotometria multibanda (BV-RI) de l'estrella WASP-52 obtingudes des dels observatoris STELLA (1.2-m, Tenerife) i el Joan Oró (0.8-m, Montsec) durant un període total de gairebé 600 dies dividits en dues campanyes d'uns 200 dies i un període intermig no observat. Aquest monitoratge de llarg termini ens ha permès determinar una estimació de la temperatura mitjana de les taques, així com una sèrie absoluta de filling factor, a més de la mesura del període de rotació estel·lar. Addicionalment, el model obtingut és compatible amb absència de fàcules.

Al llarg de la Tesi s'estudien detalladament les degeneracions inherents al problema invers. És important remarcar que malgrat amb fotometria multifiltre es puguin obtenir unívocament les components estàtica i modulant del filling-factor, a més de la temperatura de les taques, no implica l'obtenció d'un mapa de superfície únic, sinó el que només podem garantir un perfil temporal absolut de filling-factor projectat i un mapa de filling-factor longitudinal projectat (veure Figura 6.4). En el cas d'estudi considerat, hem fixat una inclinació de l'eix de rotació de WASP-52 de 90° , i per tant, tenim com a mínim la degeneració hemisfèrica a més de la degeneració latitud-mida de la taca que amb dades de precisió terrestre no es pot resoldre amb univocitat.

El principal observable dels trànsits planetaris és la seva profunditat observada en una longitud d'ona determinada (\mathcal{D}_λ). Aquesta magnitud essencialment depèn de la ràtio de radis del planeta i l'estrella, de les propietats de la pròpia estrella —limb darkening— però en estrelles actives, també de les inhomogeneïtats del fons estel·lar (Figura 7). Aquestes variacions induïdes per estructures actives poden representar una variació d'entre un 3% i un 12% de la profunditat observada, depenent l'estat de la cobertura de taques i per a WASP-52 en el període analitzat. Aquesta variabilitat té una afectació fonamental per a l'estudi d'atmosferes planetàries donat que la contribució real de les espècies atmosfèriques sobre la profunditat espectral és, típicament, de l'ordre de 10^{-3} de manera que s'imposa el coneixement de l'estat de la fotosfera en el moment del trànsit per a futurs investigacions d'espectroscòpia de transmissió amb les properes missions espacials ARIEL i James Webb Space Telescope.

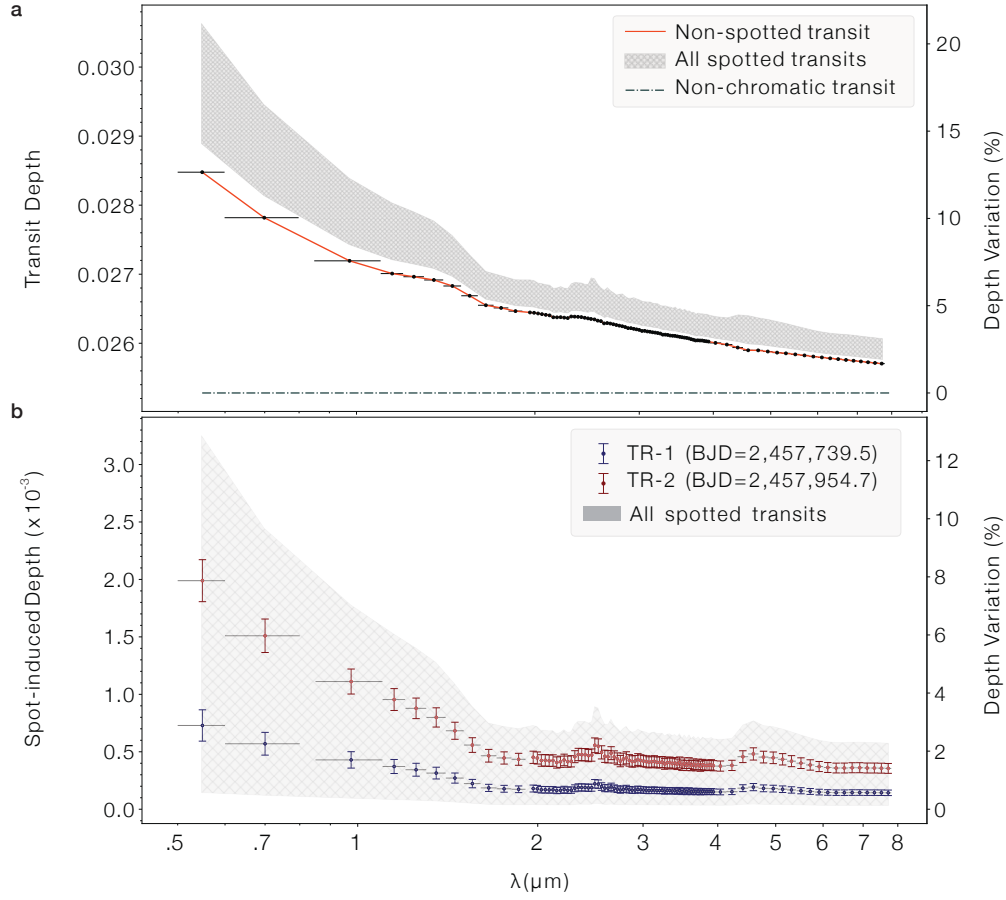


Figura 7: (a) Profunditat dels trànsits simulats en funció de λ . Assumint una fotosfera immaculada (línia vermella) i tacada (banda gris) (b) Contribució de les taques a la profunditat dels trànsits per tota la simulació de WASP-52 (gris), i per dos trànsits que exemplifiquen casos d'alta i baixa activitat.

Els procediments de modelització de superfícies estudiats, permeten corregir els efectes de les regions actives sobre la profunditat dels trànsits, assolint uns residus de poques parts en cent mil, que és el requeriment de les missions en les quals treballem i per les quals la correcció de l'activitat magnètica és essencial per al seu èxit.

Part I

Overview

Chapter 1

Introduction

1.1 The exoplanet revolution

Since the announcement of 51 Pegasi b (Mayor et al. 1995) nearly 5 000 planets have been discovered, most of them employing the photometric transit and radial velocity methods. But it was not until the discovery of the transiting HD 209458 b (Charbonneau et al. 2000) that it was possible to determine both a planetary mass and radius, and therefore, to obtain an estimation of its density and allowed the comparison among different internal structures and compositions, gaining increasingly more knowledge about these new worlds. Current statistics show that exoplanets are very common around stars to the extent that every star has, statistically at least, one planetary companion on average (for FGK stars, Kunimoto et al. 2020; Hardegree-Ullman et al. 2019, for M stars). Various techniques have been used to detect exoplanets, most of them being indirect methods because planets can generally not be imaged but, instead, a detection can be made by measuring tiny effects induced on host stars. Among the most successful techniques are the Doppler spectroscopy and transit methods.

Doppler spectroscopy was very successful in the detection and confirmation of the first exoplanets (Mayor et al. 1995; Marcy et al. 1996) and has culminated in the detection of Earth-like planets around nearby stars such as Proxima Centauri (Anglada-Escudé et al. 2016b), Barnard's star (Ribas et al. 2018) or Teegarden's Star (Zechmeister et al. 2019a). This technique is the same used for decades in the field of spectroscopic binary stars and consists of measuring the periodic shift of spectral lines due to the relative radial motion between the target star and the observer. The measured radial velocity, in astronomer-friendly units, is found to be (Duric 2003)

$$v_r [km s^{-1}] \approx 30 \frac{m}{\sqrt{M} a_0} \sin i \quad (1.1)$$

where m and M are the masses of the planet and star in solar units, and a_0 is the orbital semi-major axis in AU.

Until the advent of space-based photometric telescopes (COROT, Kepler, TESS, ...) the Doppler technique has been, by far, the most prolific method for detection and

confirmation exoplanets. This is thanks to the great stability and precision of the spectrometric instruments developed for such purpose, e.g. HARPS (Pepe et al. 2005) and CARMENES (Quirrenbach et al. 2014), the latter being particularly optimised to observe cool dwarf stars in the red and near-infrared spectral domain. Great advances in this field are expected in the near future, with the coming online of instruments like ESPRESSO (Pepe et al. 2021), which is reaching precisions ~ 10 cm/s and will open an entirely new window of opportunity to detect small and Earth-sized planets in systems comparable to our Solar system. One of the major drawbacks of this technique is that only the quantity $m \sin i$ can be inferred, therefore only a minimum mass can be determined for a planet candidate.

The transit method is based on the detection of a small drop in stellar brightness, that occurs when a planet crosses in front of the star as viewed from the Earth. The decrease in flux is typically below 1% and depends on the star-planet radius ratio. Although this method is especially powerful and a wealth of information that can be obtained, the probability of having transits for a randomly oriented planetary system is $< \sim 1\%$. The first detection of a transiting planet was announced by Charbonneau et al. (2000) and consisted of a gas giant orbiting the G0 dwarf HD 209458 in Pegasus.

Considering circular orbits, transits and eclipses will occur if the following condition is satisfied,

$$a_0 \cos i \leq R_* + R_p \quad (1.2)$$

where R_* and R_p are the radii of the star and planet, respectively. a_0 is the semi-major axis of the planet's orbit, and i is the inclination of the orbital plane. By assuming a random distribution of the inclinations, the probability of a star with a planet to be transiting is,

$$\text{Prob}_{tr} = \frac{R_* + R_p}{a_0} \approx \frac{R_*}{a_0}. \quad (1.3)$$

Just to mention some examples, the transiting probability for a system like Sun-Earth is $4.7 \cdot 10^{-3}$, for a close-in Earth-like planet in a cool M-dwarf like TRAPPIST-1 b is 4.8%, and in case of hot Jupiters, this probability typically rises up to $\sim 10\%$.

The fundamental magnitude to be taken into account is the transit depth, which also determines the photometric precision needed to significantly detect transit events. If we consider the star as a uniform disc in brightness, the relative variation in flux can be written as

$$\mathcal{D} = \frac{\Delta F}{F_0} = \frac{\pi R_p^2 \mathcal{B}_* - \pi R_p^2 \mathcal{B}_p}{\pi R_*^2 \mathcal{B}_*} = \left(\frac{R_p}{R_*} \right)^2 \left(1 - \frac{B_{pl}}{B_*} \right) \approx \left(\frac{R_p}{R_*} \right)^2, \quad (1.4)$$

where B_* and B_{pl} are the spectral flux densities of the star and the planet, respectively.

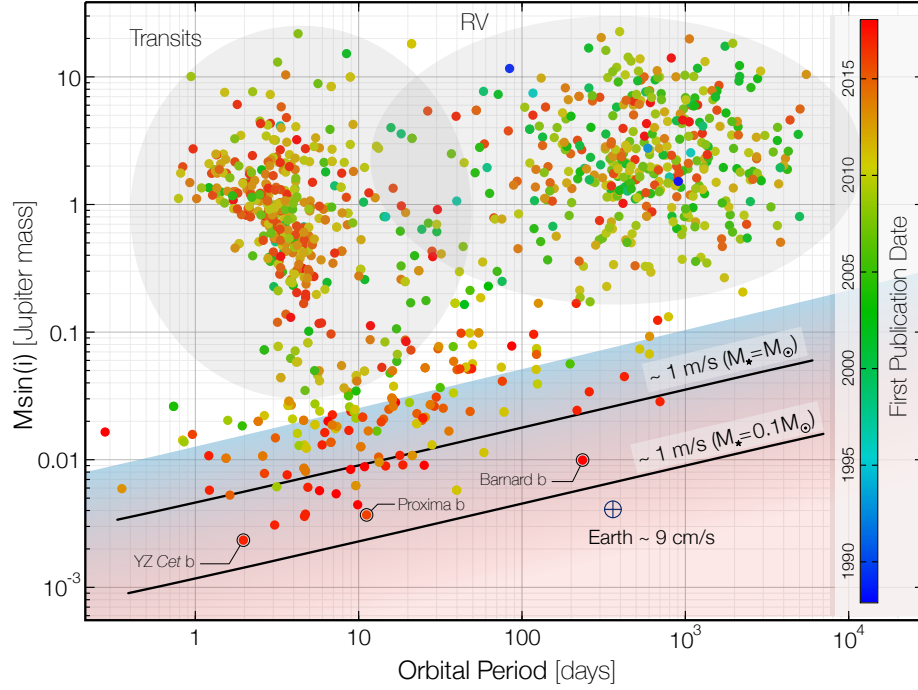


Figure 1.1: Exoplanet discoveries by radial velocity, transits, and other methods — gravitational lensing, TTV, direct imaging— The color in sidebar represents the publication year. Two main groups are observed: at the top left there is a cluster of planets discovered through ground-based transiting method (WASP, HAT, KELT, XO, ...); while at the top right are located the first discovered planets by the RV method. The solid black line shows the 1 m/s limit of most modern spectrographs. Below we find some planets found by space photometric telescopes such as Kepler, but also by RV e.g. GJ 699 b (Barnard b). The lower shaded area indicates the regime where stellar activity may induce effects comparable to the planetary signals thus representing a challenge their detection. The Earth position is marked and it becomes apparent that a proper understanding of stellar activity, together with a step forward in the capabilities of the instrumentation, will be needed for the detection of Earth analogues. Designed with data from exoplanets.org (24/6/2020).

By combining the Doppler technique with transit observations one can obtain a comprehensive description of the planetary system including a precise determination of the orbital parameters, that may allow ulterior precise measuring of transit timing variations (TTV) to infer the existence of new and non-observed planets around the star.

Another breakthrough was done soon after the discovery of the first transiting planets by observing the spectral signature across the stellar disc while the planet crosses and measuring the transit depth dependence on wavelength. This feature led to the detection and characterisation of the first exoplanetary atmospheres (Charbonneau et al. 2002) in HD 209458 b. The periodicity of transits gives the unique opportunity to observe known transiting planets in schedule to study planetary atmospheres by transmission spectroscopy technique. The core idea behind transmission spectroscopy is that the planetary transit depth has a chromatic dependence since its atmosphere absorbs selectively certain wavelengths according to its properties.

The measured transit depth as a function of wavelength constitutes the exoplanetary transmission spectrum. This has been revealed as a flourishing and acknowledged technique (Seager et al. 2000; Tinetti et al. 2007; Swain et al. 2008; Fortney et al. 2010; Burrows 2014; Sing et al. 2015; Tsiaras et al. 2019) and opens the possibility of studying the chemical (composition, abundances) and physical (thermal structure, pressure profile) properties of the atmosphere.

The generation of theoretical forward models of atmosphere transmission spectra involves radiative transfer calculations and it is a very time-consuming computational task. Therefore we have to study the atmosphere behaviour from a synoptic point of view. A magnitude that explains large features is the pressure scale height, H , which is the change in altitude where the pressure drops by a factor e . By assuming hydrostatic equilibrium and the ideal gas law, the parameter H can be written as (Kreidberg 2018),

$$H = \frac{k_B T_{eq}}{\mu g} \quad (1.5)$$

where k_B is the Boltzmann constant, T_{eq} is the equilibrium temperature, μ the mean molecular weight and g the surface gravity. The amplitude of the observed atmosphere features is given by

$$\delta_\lambda = \frac{(R_p + nH)^2}{R_*^2} - \frac{R_p^2}{R_*^2} \approx \frac{2nR_p H}{R_*^2} \quad (1.6)$$

where n is the number of scale heights over which the signal is produced. The last approximation relies on the assumption of a thin atmosphere in comparison with the planetary radius, $R_p \gg H$.

The optimal candidates for transmission spectroscopy are low-density planets with atmospheres of low mean molecular weight (hydrogen-dominated) and high equilibrium temperatures, thus favoring hot, giant gaseous planets (Stevenson 2016; Kreidberg 2018). Even for these, the observed transit depth effects induced by the planet's spectral features are of order 10^{-3} , which makes their detection very challenging, particularly in the case of active host stars, where other sources of variations are expected.

Future space-based telescopes such as the JWST (James Webb Space Telescope; Gardner et al. 2006) and the Ariel (Atmospheric Remote-sensing Infrared Exoplanet Large-survey; Tinetti et al. 2018) missions are designed for this purpose. The latter will carry out a comprehensive, large-scale survey of the chemical compositions and thermal structures of the atmospheres of ~ 1000 known transiting exoplanets in the optical and near-infrared (NIR) wavelengths (0.5 to 8 μm) following its expected launch in 2029 (Tinetti et al. 2018; Encrenaz et al. 2018).

As has been shown by many studies (e.g. Boisse et al. 2009; Lagrange et al. 2010; Oshagh et al. 2013b; Oshagh et al. 2014a; Robertson et al. 2015a; Perger et al. 2017), exoplanet search and characterisation methods may be affected by stellar activity effects, which hamper the identification of true planetary signals. Therefore, a detailed understanding of the different scales and magnitudes of these magnetic phenomena is crucial to overcome the difficulties, particularly in the new cutting-edge instruments able to offer long-term stability below 0.5 m/s in Doppler measurements and high-precision photometric space-based telescopes intended for atmosphere studies aimed at reaching precisions of a few parts per 10^5 in spectrophotometric fluxes. Such level of instrumental accuracies imposes a major challenge in the activity correction strategies.

1.2 Time-series analysis in exoplanet science

A common problem in many fields of Astronomy, consists of determining periodicities in observational data, usually with poor sampling and low signal-to-noise ratio. In these conditions, the analysis of periodic signals, which may contain not only a single periodicity but a superposition of signals coming from various physical phenomena, requires the use of specific algorithms to correctly extract the information.

The exoplanet detection by means of radial velocity method (RV), the observational RV time-series are searched for periodic signals. The signals detected are either associated with the reflex Keplerian motions caused by one or more exoplanets in orbit around the star and from stellar magnetic activity effects manifested in phenomena such as dark spots or bright faculae on the stellar surface (e.g., Queloz et al. 2001; Perger et al. 2019).

Lomb (1976) & Scargle (1982) developed a methodology, dubbed Lomb-Scargle (LS) periodogram, to identify coherent periodic signals in unevenly spaced time-series data. This technique overcomes problems caused by Fourier transformations in such data and relies on the calculation of a full sine χ^2 -fit of the form $y = a \cos(\omega t) + b \sin(\omega t) + c$ for a range of frequencies. LS and LS-based periodogram techniques have been widely employed within the astronomy community because of their simplicity and computational efficiency.

A generalisation of this method, the GLS, was presented by Zechmeister et al. (2009), and considers both uncertainties and a zero-point offset for the data. Because of its simplicity and effectiveness, the GLS is commonly used as the first tool to search for periodic signals in RV time series data (Wittenmyer et al. 2014; Robertson et al. 2015a; Robertson et al. 2015b; Perger et al. 2017).

Several authors have proposed further improvements to the LS algorithm. Cumming (2004) considers a Keplerian model as the fitting function, which provides better performance in cases of high orbital eccentricity. In such cases, the RV curve deviates significantly from a sinusoid resulting in the detection of harmonic frequencies in the classical GLS. The Bayesian-GLS (Mortier et al. 2015) employs Bayesian statistics to select the most likely period when two or more signals are of similar significance. Regarding the assessment of the significance of multiple signals, Baluev (2013b) provides a general framework for the detection of multiple sinusoidal components in noisy datasets. For a detailed comparison of the different methodologies see Süveges et al. (2015).

In hierarchical planetary systems, where each planet induces most of the RV variability in its relevant period domain, iterative periodogram computations are done via the so-called prewhitening technique. This procedure identifies the statistically most significant period in the RV data, subtracts the best-fit sinusoidal model, and then constructs the GLS periodogram from the remaining RV residuals. This process is repeated until no significant signals are left. The threshold for signals being tentative or significant is typically adopted as the 1 and 0.1% false alarm probability (FAP) either calculated by bootstrapping randomisation (Murdoch et al. 1993) or from the Horne number of independent frequencies (Horne et al. 1986). Note that the adoption of the 0.1% FAP criterion is mostly based on experience rather than on a robust mathematical framework. The prewhitening approach, however, might be doubtful in certain multisignal configurations and may lead to spurious and/or missing detections, bias or distort signal parameters as we show later.

1.3 Stellar activity

The concept of stellar activity encompasses all effects that modulate the brightness emitted by the star caused by internal magnetic processes. The coupling between stellar rotation and convective envelopes generates a variety of phenomena such as the emergence of dark spots and faculae on the photosphere or granulation, which consequently induces chromatic variations in flux and distortions in the spectral lines (Berdyugina 2005; Strassmeier 2009).

Cool (FGKM) stars with convective envelopes are affected by magnetic processes that induce a wide variety of surface phenomena (Donati et al. 2006, e.g.). The dynamo theory (Simon et al. 1987) explains the mechanism by which stellar rotation is an essential part for understanding the magnetic field generation in electrically conducting plasma in a turbulent environment. The turbulence in the magnetic flux leads to the formation of small areas in the surface where the density of magnetic field is particularly high. These areas are starspots, faculae and plages.

1.3.1 Multiscale approximation to stellar variability

We classify the types of stellar variability according to their timescale. High-frequency activity encloses stellar oscillations, pulsations and granulation, with a typical time-

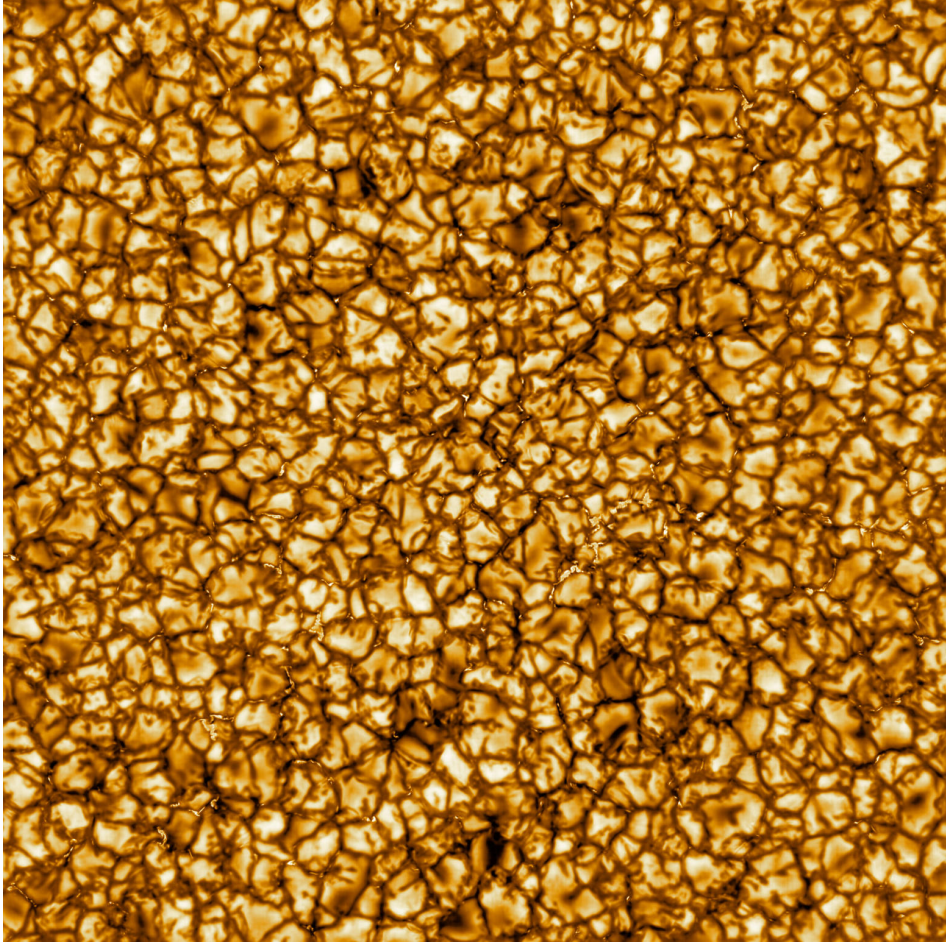


Figure 1.2: Detail image of the Sun's photosphere showing the convective cells. Brighter grains are hotter than darker ones and come out to the surface imprinting a positive radial velocity, while cooler plasma goes down giving a negative contribution, but with a lower weight due to their lower flux. Image taken by Daniel K. Inouye Solar Telescope in Haleakala Observatory (Hawaii). Credit: NSO/NFS/AURA under Creative Commons Attribution 4.0 International (CC BY 4.0).

scale of minutes. Cool stars have convective envelopes that can give rise to p-mode oscillations propagated through their stellar interiors. In spite of having typical amplitudes of $\sim \text{cm/s}$, constructive interference of several modes can raise amplitudes to the range of a few m/s (Dumusque et al. 2011b). Granulation and supergranulation phenomena are produced by convective flows due to thermal transport. Figure 1.2 shows a high-resolution image displaying convective cells. Brighter structures are hotter than darker ones and move out to the surface imprinting a positive radial velocity, while cooler plasma goes down giving a negative contribution, but with a lower weight due to their lower flux. Surface-integrated measures give residual RV amplitudes in the domain of a few m/s (Meunier et al. 2015; Cegla et al. 2018).

Low-frequency activity features are related to magnetic stellar activity and comprise heterogeneous surface rotation and magnetic cycles. This regime is the most

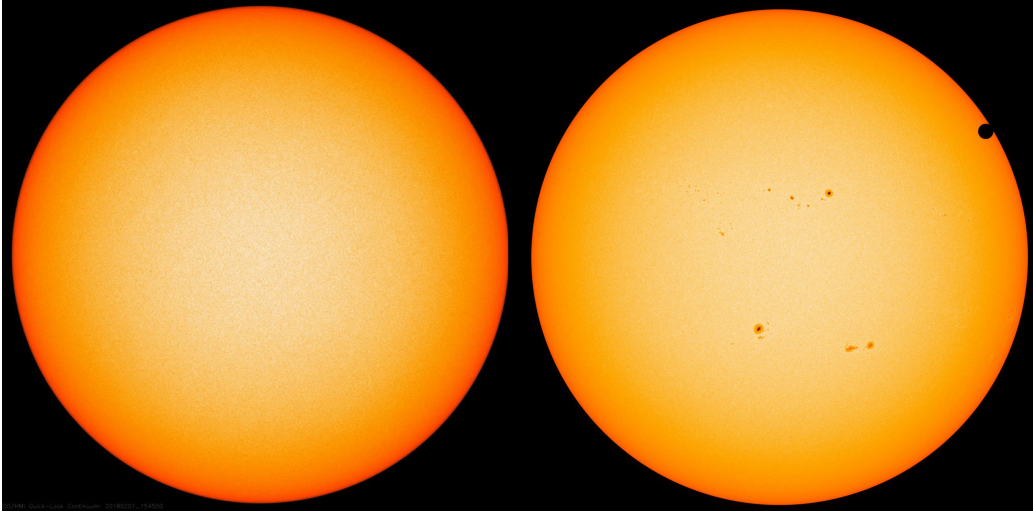


Figure 1.3: Solar Dynamics Observatory (SDO) images of an immaculate solar photosphere (*left*), and active spotted Sun during the Venus transit in 2012 (*right*). Credit: SDO (Solar Dynamics Observatory, NASA).

challenging as the frequencies and amplitudes can be confused with those planetary of the signals searched (e.g. Lanza et al. 2003; Lagrange et al. 2010; Dumusque et al. 2011a). Amplitudes in RVs of these signals range from a few m/s to hundreds of m/s for young active stars. Magnetic cycles show timescales of the order of years with amplitudes rarely above ~ 1 m/s in RV (Santos et al. 2010; Oláh et al. 2016). Besides, these surface inhomogeneities can modify the transit depth and bias the determination of planetary parameters. The presence of dark spots in a moderate activity regime may lead to anomalies in planetary radius up to 4% smaller than the real, as well as the transit duration may be modified (Oshagh et al. 2013b). In addition, depending on spot size, stellar contamination signals can be more than an order of magnitude larger than the transit depth changes expected for atmospheric features in rocky exoplanets (Kreidberg 2018). Spot-crossing events can also induce abrupt variations in the transit shapes (a simulation can be seen in Figure 1.4), as well as induce variations to the expected monotonic negative slope of transit depth curve with wavelength (Mallonn et al. 2018).

Figure 1.3 displays Solar Dynamics Observatory (SDO) images of an immaculate photosphere Sun, and active spotted Sun during the Venus transit in 2012.

1.4 Effects of stellar activity on the observables

1.4.1 Chromatic effects on planetary transits

Active regions can produce significant alterations on the measured planetary transit depths (Lagrange et al. 2010; Meunier et al. 2010b; Barros et al. 2013; Oshagh et al. 2014a). Inhomogeneities on the stellar photosphere can induce chromatic effects (Sing et al. 2015), which, in the case of unocculted spots, can be very similar to the signature of atmospheric Rayleigh scattering (Rabus et al. 2009; McCullough et al. 2014). Rackham et

al. (2018) studied their influence in M dwarfs and find it to be up to ten times larger than the effect expected from atmospheric compounds even at NIR wavelengths. Although find the influence of surface phenomena in FGK stars to be generally lower and possibly only measurable for active stars, those studies challenge the reliability of retrieved transmission spectra, at least under certain circumstances. Clouds and hazes can also produce chromatic effects, introducing a Rayleigh-like slope and grey opacity, respectively (Pinhas et al. 2017). Distinguishing the different contributions to wavelength-dependent transit depth variations is therefore of crucial importance, and this calls for sophisticated modelling of photospheric inhomogeneities.

1.4.2 Effects on RVs

Radial velocity measurements on active stars are also prone to stellar variations. Inhomogeneities on the stellar surface such as dark spots contribute to break the balance between the two halves of a rotating star. The effects that contribute the most to imbalance the flux are the Doppler shift due to the flux contrast, and the convective blueshift inhibition produced by these active regions (Dumusque et al. 2014). Other effects with a relevant contribution are the limb darkening of the quiet photosphere and limb brightening of faculae.

The RVs induced by heterogeneous surfaces are in the range of few cm/s to tens of m/s depending on the spectral type and the level of magnetic activity. These signals may hamper the detection of planetary signals, particularly at the current level of 1 m/s precision for the search of Earth-mass exoplanets. Even though for a quiet star, achieving a long-term accuracy of 0.5 m/s is mandatory to detect Earth-twins.

The effect of RV signals on the searches of small planets led to various controversial announcements (Robertson et al. 2014; Rajpaul et al. 2016) because of the statistical analysis performed or the existence of spurious signals related to stellar rotation. This is a tale of caution when analysing RV data, and emphasises the need of a robust and precise modelling of these effects, that, even with the aid of stellar activity indices derived from spectroscopic data, which in principle, are not affected by true Keplerian signals. However, it is also obvious that precise modeling of spot properties can help to disentangle and correct for stellar activity effects, thus enabling the detection of exoplanet signals that would otherwise be hidden within the stellar RV activity signals.

An interesting phenomenon and a potential rich source of information occurs during a transit when RV measurements are taken, and it is known as the Rossiter-MacLaughlin effect (RM) (Rossiter 1924; MacLaughlin 1924). When a planet transits its host star, different regions of the apparent stellar disc are occulted. This produces an anomaly in RV signal as the planet transits across the blueshifted and redshifted hemispheres (Triaud 2017). The shape of this effect depends on planet orbital parameters, and in particular on the relative positions of the stellar spin axis and the orbital angular momentum vector (Triaud et al. 2009). Furthermore, analysis of the RM effect has proved to be effective in the detection of differential stellar rotation and in determining spatially-resolved stellar spectra (Cegla et al. 2016).

The semi-amplitude of the RM effect also scales with the planetary radius. Snellen (2004) proposed a technique to retrieve transmission spectra by measuring the variation

of the RM amplitude at different wavelengths, using chromatic Rossiter–McLaughlin observations. Similarly to RV measurements, chromatic RM is also affected by surface inhomogeneities, which can mimic broadband features, such as the Rayleigh scattering slope, in their retrieved transmission spectra (Boldt et al. 2020). The planet-to-star radius ratio effect can lead to variations of up to 10%, resembling the signature of light scattering in the planetary atmosphere (Oshagh et al. 2014b).

1.4.3 Surface modelling: Turning noise into signal

Stellar activity has been considered traditionally as a source of noise since it can dilute the signals of interest such as the photometric depth of a transit or the radial velocity modulation of Keplerian motion. However, when high-precision data are available, becomes possible to infer the presence of some structure behind the generation of stellar noise. In some studies, stellar activity is treated as a black-box correlated noise process (e.g. Feroz et al. 2014) such as ARIMA or ARMA correlated noise models and it is known that noise in photometric observations in transits is often correlated (Pont et al. 2006). Coloured noise —understood as mixtures of white noise and long-term correlated noise— induce false positive detections since correlated data contain less information than in case of a pure white noise process.

Hitherto, the lack of a synoptic and systematic technique to properly debias the stellar activity affected observables has made this problem a capital issue in exoplanet research for the last decade up to now.

A good understanding of how activity affects observables is essential in the search for Earth-like planets, where $< \sim 0.5$ m/s long-term radial velocity stability is required. Even though the last generation instruments are capable of achieving these stabilities, without a thorough and effective understanding of stellar activity, the detection of these planets will not be feasible.

Starspots Starspots are created by local magnetic field variations on the surface of stars. In the regions where the magnetic flux is sufficiently strong, the convective upward motion is inhibited. This results in an area where the plasma flow, which is hotter than the surrounding region appears to be visually darker (Berdyugina 2005).

As starspots are a visible consequence of the charged plasma motion and mesoscopic behaviour of the magnetic dynamo, these active regions are the perfect tracers of stellar activity and allow the precise determination of rotation periods, differential rotation and, potentially, opens a door to deeper researches on the —yet poorly understood— magnetic structure and behaviour of energy transfer in convective stars. Most likely, all late-type stars, either fully convective or with an outer convective envelope can develop spots on their surface.

Starspots and transiting exoplanets Brightness inhomogeneities caused by magnetic effects in the form of spots or faculae modify the derived parameters of transit events and in case of transmission spectroscopy could lead to mimic spectral features found in exoplanet atmospheres (Oshagh et al. 2014a; Sing et al. 2015).

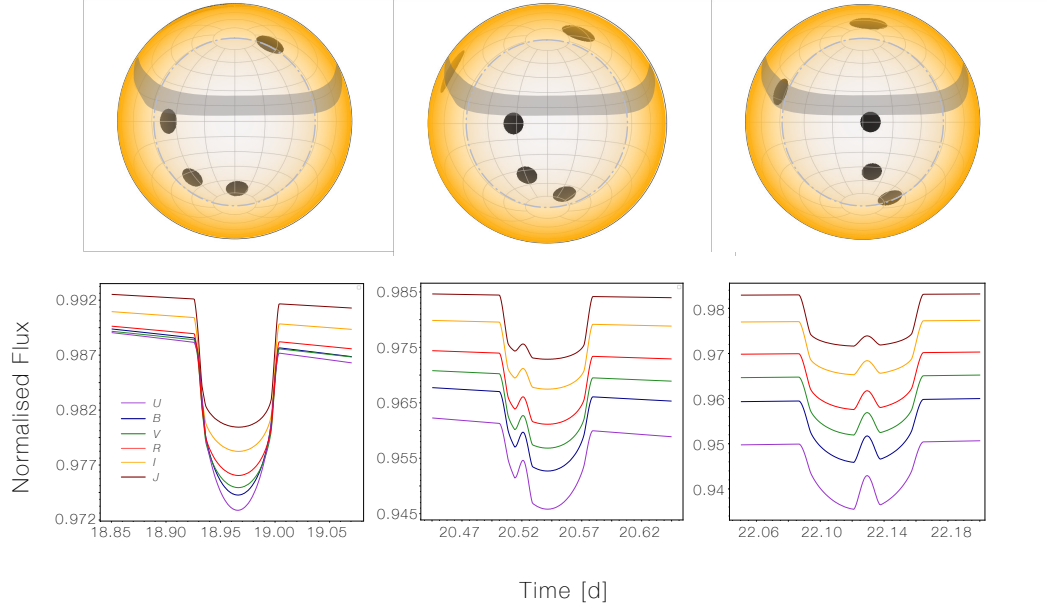


Figure 1.4: Simulated multiband transits on a spotted surface. The synthetic map contains five static (non-evolving) spots on a star $T_{\text{eff}} = 5000$ K, $\Delta T_{\text{sp}} = 700$ K and $P_{\text{rot}} = 15$ days. The gray band depicts the path of the transiting planet ($r/R_* = 0.1$, $b_{\text{imp}} = 0.15$ and $P_{\text{pl}} = 1.58$ days) and appears curved shaped due to the Lambert projection used to show the map, and therefore the most peripheral spots are not actually visible from the observer's line-of-sight. The innermost gray circle indicates the visible side of the sphere.

Inhomogeneities on the stellar surface can induce some chromatic effects, which in the case of, for example, unocculted spots can be very similar to the signature of atmospheric Rayleigh scattering (Rabus et al. 2009; McCullough et al. 2014). The effect of those chromatic effects on the determination of planetary and atmospheric parameters is strongly dependent on the mean size of the spots —filling factor— and their temperature contrast with respect to the photosphere. As a consequence of this, precise understanding of the stellar surface becomes of capital importance to retrieve unbiased transmission spectra of exoplanets.

Figure 1.4 shows a multiband simulation with StarSim 2 code of a transiting planet across an active surface. It might be noted that a bump appears when the planet crosses a spot —if the lower brightness spot area is occulted, the overall flux of the star increases temporarily with respect the rest of the transit, when the planet occults the brighter photosphere. The effect of limb-darkening is clearly seen in the bottom parts of the transits when observed in different photometric bands. In the redmost part of the spectrum (I,J Johnson bands) the effect of activity dampers and the transit appears more box-shaped.

1.4.3.1 Active star modelling: The State-of-the-Art

Hitherto, the most strategies to model stellar active surfaces may be divided into two groups: i) Analytical methods and, ii) Numerical simulators.

The first group encloses those formulations relying on the assumption of circular-shaped spots, uniform temperature contrast all across the spot. Early works of Budding (1977) and Dorren (1987) present analytical models to model dark starspots on rotating stars. Analytic codes are, in general, very fast to compute the forward problem either in photometry or radial velocity. The weaknesses are related to the predefined shape of the spots and the need of using analytical simple of limb-darkening laws.

Among the most recent codes are `macula` (Kipping 2012), which assumes circular spots, includes non-linear limb-darkening for photosphere and spots, differential rotation and single-domain analytic function; and empirical recipes, such as `FF'` (Aigrain et al. 2012), which involves the flux and its derivative, predicts the radial velocity series given the photometric variations. The latter is based on a simple spot map composed of only one active region, and the method assumes that photometry carries all the information needed to infer the spot distribution causing the observed velocity signal. One of the strong points of the `FF'` method, is its speed and also that it is, easy to implement, and only requires an estimate of the stellar radius. Thus, it is particularly useful for processing large amounts of data from missions like CoRoT, Kepler/K2, TESS to derive some statistical properties of the RV of the surveyed stars.

Among the second group of stellar simulators, we can find codes able to model the star by tessellating the surface in squared cells assigning to each one a type of surface, i.e. quiescent immaculate photosphere, spot or facula. Some examples are `SOAP/SOAP-T` and `SOAP-2` (Boisse et al. 2012; Oshagh et al. 2013a; Dumusque et al. 2014). Spot Oscillation and Planet (SOAP) is a code for estimating the effect of spots and plagues on the photometric and radial velocity time series. New issues of the SOAP code include many contributions with respect to previous versions, such as a more realistic contrast temperature of photosphere/plagues and a quadratic limb darkening law and a limb brightening effect for the plagues.

`SOAP-2` uses solar CCFs, therefore, their use might be limited to those stars which are reasonably comparable to the Sun. In addition, the performance of the code makes computing the effect of an arbitrary distribution of active regions very time-consuming, since the simulator is designed to only consider one or few spots on the photosphere.

A number of other simulator-type forward modelling codes have been released. A non-exhaustive list includes `PyTranSpot` (Juvan et al. 2018), `Allesfitter` (Günther et al. 2020), `Fleck` (Morris 2020), `Ksint` (Montalto et al. 2014). None of them allow RV simulation and only `PyTranSpot` and `Allesfitter` provide parameter estimates from a simple photometric model. The `Fleck` code is more intended for generation of light curve ensembles using simple and fast algorithm.

1.5 This Thesis

The aim of this Thesis is to produce advanced computational tools for exoplanet research. The structure is composed of two parts.

The first part describes the development of a multidimensional extension to the well-known GLS (Generalised Lomb-Scargle) algorithm. Multiplanetary systems have revealed very common around many stars, particularly those with the lowest masses.

Besides, the time series of radial velocity measurements contain numerous signals in addition to potential planetary companions. Stellar activity patterns and spurious periodicities arising from the commonly used sampling window are the most common complications in detecting exoplanets.

In Chapter 2, an extension to the classical GLS algorithm is presented, with the aim of improving the performance in multi-signal datasets where signals of small and similar amplitude are present. We demonstrate that our approach compares very well with more sophisticated techniques, such as Bayesian inference searches, and neatly overcomes the performance of the iterative prewhitening technique, while keeping the code simplicity and stability.

The aim of Chapter 3 is to test the functionality of the MGLS code in real multiplanetary systems. We re-analysed observational data from known systems to evaluate the detectability performance in comparison to the hierarchical approach (prewhitening).

The second part of the work deals with the effects of stellar activity on spectroscopic and photometric observables in the exoplanet search context. The general context involves the description of the inverse problem and two relevant applications are developed by using the code StarSim: the effect of heterogeneous surfaces on transmission spectroscopy for transits, and the activity-induced variability on observed radial velocities.

In Chapter 4, the StarSim 2 code is described in detail.

In Chapter 5, the photometric inverse problem is explored both analytically and by simulations with StarSim 2 through a toy model. An exhaustive account of degeneracies found in light curve inversion is provided, and so are the requirements needed to overcome them and in which degeneracy and what conditions are eliminated by spectral information.

In Chapter 6, all these principles are implemented, where we fit an activity model to BVRI multiband time-series photometry of the exoplanet host WASP-52 to retrieve a spot map and a set of the relevant stellar parameters involved in the inversion process (P_{rot} , ΔT_{sp} and facula area ratio).

In Chapter 7, we use the retrieved activity model fitted for WASP-52 to study the different sources of chromatic effects on the transit depths of simulated WASP-52 b planet. The importance of a good understanding of chromatic effects and how to correct them out is essential for the analysis of exoatmospheres.

Part II

Time series analysis

Chapter 2

MGLS: Multidimensional Generalised Lomb Scargle Periodogram

2.1 Introduction

Time-series analysis is a broad field of applied mathematics spanning from signal processing to statistics. Even when limited to astronomical datasets, the diversity of applications is enormous. The most common problems range from detection of variability and periodicity to treatment of non-periodic variability and searches for localised events (VanderPlas et al. [2012](#)).

This chapter focuses on periodicity detection for multicomponent signals in the context of radial velocity and photometric time-series for exoplanet searches. Doppler spectroscopy has been very successful in the detection and confirmation of the first extrasolar planets (Mayor et al. [1995](#); Marcy et al. [1996](#)) and has culminated in the detection of Earth-like planets around nearby stars such as Proxima Centauri (Anglada-Escudé et al. [2016a](#)), Barnard’s star (Ribas et al. [2018](#)) or Teegarden’s star (Zechmeister et al. [2019a](#)).

The analysis of a periodogram of time-series radial velocity data is the usual starting point to seek periodic signals that can be associated with the reflex Keplerian motion caused by an exoplanet or with stellar activity effects produced by phenomena such as dark spots or bright faculae on the stellar surface (e.g., Queloz et al. [2001](#)). In the case of multiplanetary systems such analysis has been traditionally carried out in an iterative way, known as prewhitening, and consists of subsequently subtracting the strongest periodicity until no relevant signals are left in the periodogram.

Later in this work, we show that classical prewhitening can diminish the significance of some periodic signals, especially in the case of low signal-to-noise ratio, depending on the nature of the signals (periods, amplitudes) and the data sampling. This

limitation can be solved by using a multidimensional approach to the periodogram analysis.

The concept of multifrequency periodogram is not novel. Early applications in the field of Astronomy date back to Foster (1995), who developed a statistic test in `CLEANest` search code. More recently, Baluev (2013a), proposed a multifrequency statistic using the theory of extreme values of stochastic fields to approximate false alarm probabilities.

In this chapter, we present a reliable and fast extension to the very often used Generalised Lomb–Scargle periodogram for a multidimensional frequency space using simultaneous search, along with a procedure to assess statistical significances of multifrequency analysis. Our new algorithm, dubbed MGLS, is subsequently tested in real multiplanetary systems in the following chapter.

2.1.1 Lomb–Scargle Periodogram

Lomb–Scargle periodogram techniques (Lomb 1976; Scargle 1982) are the best known and extensively employed in Astrophysics and their use has been exported to many other scientific fields such as genetics for finding periodic gene expression profiles when significant proportion of information is missing (Glynn et al. 2005); or nuclear physics for the analysis of cyclic modulations in radionuclide decay rate (Gururajan et al. 2020); or even in finance for the analysis of high-frequency data (Giampaoli et al. 2009). They are particularly suited for non-equally spaced data and their simplicity of implementation and use has made them standard in the community. LS algorithms are model-oriented and most commonly a sinusoidal wave is fit by least-squares. Although LS is motivated by Fourier techniques, it is really a least-squares fit.

The LS estimator (Lomb 1976; Scargle 1982) for a time-series with zero mean, and not considering uncertainties on data is

$$\hat{P}_{LS}(\omega) = \frac{1}{2} \frac{(\sum_i^N x_i \cos \omega(t_i - \tau_{LS}))^2}{\sum_i^N \cos^2 \omega(t_i - \tau_{LS})} + \frac{1}{2} \frac{(\sum_i^N x_i \sin \omega(t_i - \tau_{LS}))^2}{\sum_i^N \sin^2 \omega(t_i - \tau_{LS})} \quad (2.1)$$

where the parameter τ_{LS} can be obtained with

$$\tan 2\omega\tau_{LS} = \frac{\sum_i^N \sin 2\omega t_i}{\sum_i^N \cos 2\omega t_i} \quad (2.2)$$

where the parameter τ_{LS} is a time reference point ($t_i = t_i - \tau_{LS}$) which depends on the frequency and time sampling, and the χ^2 value is not affected by time shifts.

The Lomb–Scargle periodogram (LS) presents an important drawback when data are not uniformly distributed in time. As the model to fit does not include an offset term, if the observational points are not evenly distributed with respect to their mean value, the parameter estimation may lead to a largely incorrect determination of the mean and therefore cause problems such as aliasing in period determination. A simple remedy is proposed in the Generalised Lomb–Scargle (GLS) periodogram (Zechmeister

et al. 2009) by simply adding a fittable offset term. An illustrative example of poor detection performance in such case is given in VanderPlas et al. (2012) and reproduced in Figure 2.1

Several authors have proposed improvements to the LS algorithms. Cumming (2004) considers a Keplerian model as the fitting function, and this provides better performance in cases of high orbital eccentricity, where the radial velocity curve deviates significantly from a sinusoid. The BGLS (Mortier et al. 2015) employs Bayesian statistics to select the most likely period when two or more peaks are similar in height. Regarding the assessment of the significance of multiple periodicities, Baluev (2013b) provides a general framework for the detection of multiple sinusoidal components in noisy datasets.

As another limitation to take into account, the Lomb–Scargle based algorithms only fit a single harmonic to the data (Equation 2.3). For more complicated periodic signals such as that of a double-eclipsing binary stellar system, or simply a dataset with multisignal content, this single-component fit may lead to the detection of false periodicities generated from constructive interference of the real signals by convolution with the sampling.

2.1.2 Generalised Lomb–Scargle Periodogram

A generalisation of the LS method, the Generalised Lomb–Scargle (GLS), was presented by Zechmeister et al. (2009), which considers both uncertainties and a zero-point offset for the time-series. Because of its simplicity and effectiveness, the GLS is commonly used as the first tool to search for periodic signals in RV time-series data (Wittenmyer et al. 2014; Robertson et al. 2015a; Robertson et al. 2015b; Perger et al. 2017).

The Lomb–Scargle periodogram consists of optimising a merit function $P(\omega_j)$, usually χ^2 or $\ln \mathcal{L}$ based for a given range of frequencies ω_j to a pure sine model wave given by

$$y(t) = a \cos(\omega t) + b \sin(\omega t) + c \quad (2.3)$$

The parameters a, b and c can be obtained by minimising

$$\chi^2 = \sum_{i=1}^N w_i [y_i - y(t_i)]^2$$

where $w_i = \frac{1}{\sigma_i^2}$. This can be expressed as a linear system and, therefore be solved very efficiently.

2.1.3 Normalisation

The interpretation of the LS/GLS spectral power values depends on the normalisation used. Many possible normalisations exist, and while the shape of the periodogram is kept invariant, the computation of peak significances via FAPs is dependent on the used scheme. The most commonly used are, i) PSD, and ii) least-squares.

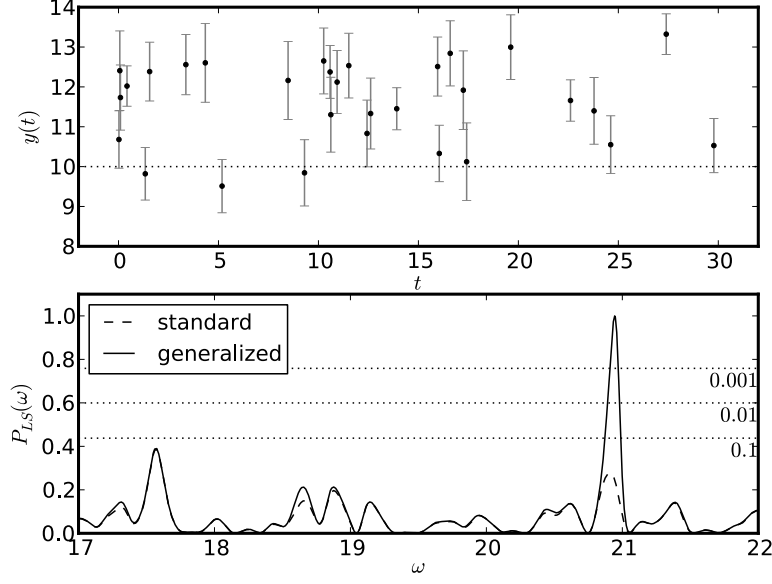


Figure 2.1: Comparison between standard LS and floating-mean GLS periodogram. A signal $y(t) = 10 + \sin(2\pi t/P)$ with $P = 0.3$ ($\omega_0 = 21$) is unevenly sampled and intentionally unbalanced around its mean. The effect of this biased phase coverage is the true period detection failure using LS, while GLS performs well. From VanderPlas et al. (2012)

When considering the LS periodogram from the perspective of Fourier analysis, the power spectral density (PSD) normalisation converges to the classic Fourier spectrum if the condition of evenly sampled data is satisfied. The PSD normalisation is (VanderPlas 2018),

$$P(\omega) = \frac{\chi_0^2(\omega) - \chi^2(\omega)}{2} \quad (2.4)$$

where χ_0^2 is referred to the null model —only a mean value is fitted—. For evenly-spaced data this periodogram equals the standard definition of the FFT-based Fourier transform of data \mathbf{X} ,

$$P(\omega) = \frac{1}{N} |\text{FFT}(\mathbf{X})|^2 \quad (2.5)$$

This last normalisation imposes for a perfect fit $\chi^2(\omega) = 0$ and, therefore, the power statistic equals $\chi_0^2/2$. The least-squares normalisation acts as an inverse fit quality measure,

$$P(\omega) = 1 - \frac{\chi^2(\omega)}{\chi_0^2} \quad (2.6)$$

This provides more intuitive power values: if the fit is perfect, $P(\omega) \rightarrow 1$, while the worst possible model yields $P(\omega) \rightarrow 0$.

The normalisations based on the χ^2 statistic show some disadvantages in accounting for additional information about the noise behaviour. A quadrature-added jitter (s) is a statistical trick for considering e.g. underestimated data error bars, an incomplete model specification, or even the presence of traces of correlated noise. Hence, the optimisation of the periodogram power is enhanced and better statistics are obtained. However, the χ^2 function with a jitter parameter, s , defined as

$$\chi^2 = \sum_{j=1}^N \frac{(X_j - \mathcal{M}(\theta))^2}{\sigma_j^2 + s^2} \quad (2.7)$$

poses a fundamental inconsistency. By $s \rightarrow \infty$, then the function $\chi^2 \rightarrow 0$, that means it is not well-behaved as the optimal fit would be reached by large jitter values. Therefore, other estimators than χ^2 should be used when including a jitter term.

2.1.4 Loglikelihood periodograms and jitter

Besides the inconsistency seen in the previous section, when a jitter term is included in χ^2 statistic, classical χ^2 periodograms present a number of issues. They are mainly related to the fact that their estimation of spectral density does not have an easy connection to relevant quantities, such as the amplitude, or a straightforward statistical significance.

An alternative computation of periodograms is done by using the maximum likelihood estimator (MLE). We may consider a general case of a data vector \mathbf{x} distributed as a multivariate Gaussian. If we assume μ to be the vector of means and Σ the covariance matrix, which accounts for the point-to-point variance, the following expression gives the total likelihood

$$p(\mathbf{x}) = \frac{1}{\sqrt{\det \Sigma} [\sqrt{2\pi}]^N} e^{-\frac{1}{2}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)} \quad (2.8)$$

where N is the dimension of the vector \mathbf{x} .

Assuming independent and identically distributed points (pure white noise scheme) Σ becomes an identity matrix with all points with the same variance $\Sigma = \sigma_0^2 \mathbf{I}$, also known as homoskedasticity. In case of heteroskedasticity (i.e. different variances in each point) Σ becomes diagonal and associates a σ_j for each measurement. Hence, the likelihood can be written as the product of Gaussian probability for each point of the dataset,

$$p(X, \theta) = \prod_{i=1}^N p(x_i | \theta) \quad (2.9)$$

After taking logarithms, the joint likelihood function (2.9) becomes a sum over the data points instead of a product,

$$\ln \mathcal{L}_\theta = -\frac{1}{2} \sum_{j=1}^N \frac{(X_j - \mathcal{M}(\theta))^2}{\sigma_j^2 + s^2} - \frac{1}{2} \langle \ln(\sigma_j^2 + s^2) \rangle + N \ln \sqrt{2\pi} \quad (2.10)$$

As introduced in the preceding section, the inclusion of a jitter term, s , may be understood as the simplest noise model to employ.

With the adoption of loglikelihood estimator, a number of features are gained that χ^2_θ least squares estimator lacks. Specifically, this form is particularly sensitive to the goodness of fit since the nature of the product operator requires that all data points have a good agreement with the best model. This results in a clear separation between signal and noise, and therefore narrower peaks (Murakami et al. 2021).

The effect of the jitter is compensated by the term in the middle, and the estimator acquires the desired properties by defining an optimal value of s . In addition, MLE has good asymptotic properties and they are *i*) consistent i.e. $\hat{\theta}$ are asymptotically unbiased; *ii*) efficient i.e. whose variances are optimised—the minimum possible, as they saturate the inequality of Cramér–Rao bound (Cramér 2016)—; and *iii*) asymptotically normal distributed. The sharpness of the peaks has also a bad counterpart when numerically search for these peaks. Gradient descent methods to optimise loglikelihood periodograms may lead to poor performance due to the difficulty of approaching the optima by following gradients. Therefore, searching algorithms in such periodograms must be carefully designed to avoid this drawback.

Peak significance Various statistical tests can be performed to assess the significance of the detected signals. Most of them are related to the likelihood–ratio tests build upon the comparison of signal likelihood and a reference value, i.e.

- Taking the likelihood to average noise ratio (Buse 1982) defined as,

$$\text{LR}_{peak} = \ln \left[\frac{\mathcal{L}(\omega_{peak})}{\langle \mathcal{L}_{noise}(\omega) \rangle} \right] = \ln[\mathcal{L}(\omega_{peak})] - \ln[\langle \mathcal{L}_{noise}(\omega) \rangle] \quad (2.11)$$

with $\langle \mathcal{L}_{noise}(\omega) \rangle$ representing the average of noise likelihoods for a range of search frequencies; and $\mathcal{L}(\omega_{peak})$ is the likelihood of the peak we want to quantify the significance. The statistical significance of the z-score statistic is given by,

$$z_{peak} = \frac{\text{LR}_{peak}}{\sigma_{LR,noise}} \quad (2.12)$$

- Computing the improvement in the logarithm of likelihood function of a model $\ln \mathcal{L}$, with respect to a less restricted model (null model) usually taken as the GLS model with no oscillatory term having to fit a model consisting just of an offset, and a statistic jitter. Hence, the statistic normalisation is,

$$\Delta \ln \mathcal{L} = \ln \mathcal{L} - \ln \mathcal{L}_0 \quad (2.13)$$

The vertical axis of the periodogram is then this last quantity, which is the natural logarithm of a probability. Therefore, an increase of one unit of loglikelihood means that the tested model is e^1 times more likely. $\Delta \ln \mathcal{L}$ statistic can also be used for false alarm probability (FAP) estimation using the likelihood–ratio test in nested models (Anglada-Escudé et al. 2013).

2.2 Iterative subtraction of signals

In hierarchical planetary systems, where each planet induces most of the RV variability in its relevant period domain, iterative periodogram computations are done through the so-called prewhitening technique. This procedure identifies the statistically most significant period in the time-series, subtracts the best-fit sinusoidal model, and then constructs the GLS periodogram from the corresponding RV residuals. This process is repeated until no significant signals are left.

The threshold for signals being tentative and significant are typically adopted as the 1% and 0.1% false alarm probability (FAP), respectively either calculated by bootstrapping randomisation (Murdoch et al. 1993) or from the Horne number of independent frequencies Horne et al. (1986). Note that the adoption of the 0.1% FAP criterion is mostly based on experience. The prewhitening approach, however, might be doubtful in certain multiplanetary configurations and may lead to spurious and/or missing detections, or bias or distort signal parameters, as we will show in the following chapter.

With the main objective of addressing multiplanetary systems, we present here the Multidimensional Generalised Lomb-Scargle (MGLS) periodogram, which is an extension to the GLS algorithm to simultaneously fit multiple sinusoidal signals and to visualize the periodogram in a two-dimensional fashion (see also, Ribas et al. 2018). The code is based on the likelihood statistics, which allows to compare directly the significance of different models and to find the dimensionality of a possible multiplanetary system.

2.2.1 Prewhitening pitfalls: A toy model experiment

To illustrate the potential drawbacks that the prewhitening procedure presents in multifrequency analysis, we define a 2-frequency toy model containing sinusoidal signals of periods 32 and 45 days with amplitudes ≈ 1.5 m/s. The uncertainty scheme adopted is heteroskedastic white Gaussian noise $\sim \mathcal{N}(0, \sigma)$ with $\sigma \sim \mathcal{U}(0.1, 0.25)$ m/s. These errors are below the limit reached by current spectrographs like HARPS or CARMENES, although not too far from the demonstrated accuracy of ESPRESSO, for example.

The joint model we will test with the prewhitening procedure is simply a sum of two oscillatory components, i.e.

$$y(t) = c + a_0 \cos(\omega_0 t) + b_0 \sin(\omega_0 t) + a_1 \cos(\omega_1 t) + b_1 \sin(\omega_1 t) \quad (2.14)$$

The time sampling is taken from a real observation campaign of red-dwarf GJ 412A with the CARMENES spectrograph. It contains 91 measurements over a timespan of 670 days and a mean separation between consecutive measurements of 7.3 days.

The top panel of Figure 2.2 shows the synthetic model generated; the solid curve being the noise-free simulation. After having subtracted the fit of a sinusoidal model ($P_1 = 45.157$ days), we repeated the same process with the residuals, fitting a second significant signal of $P_2 = 31.997$ days. The remaining residuals were analysed as well,

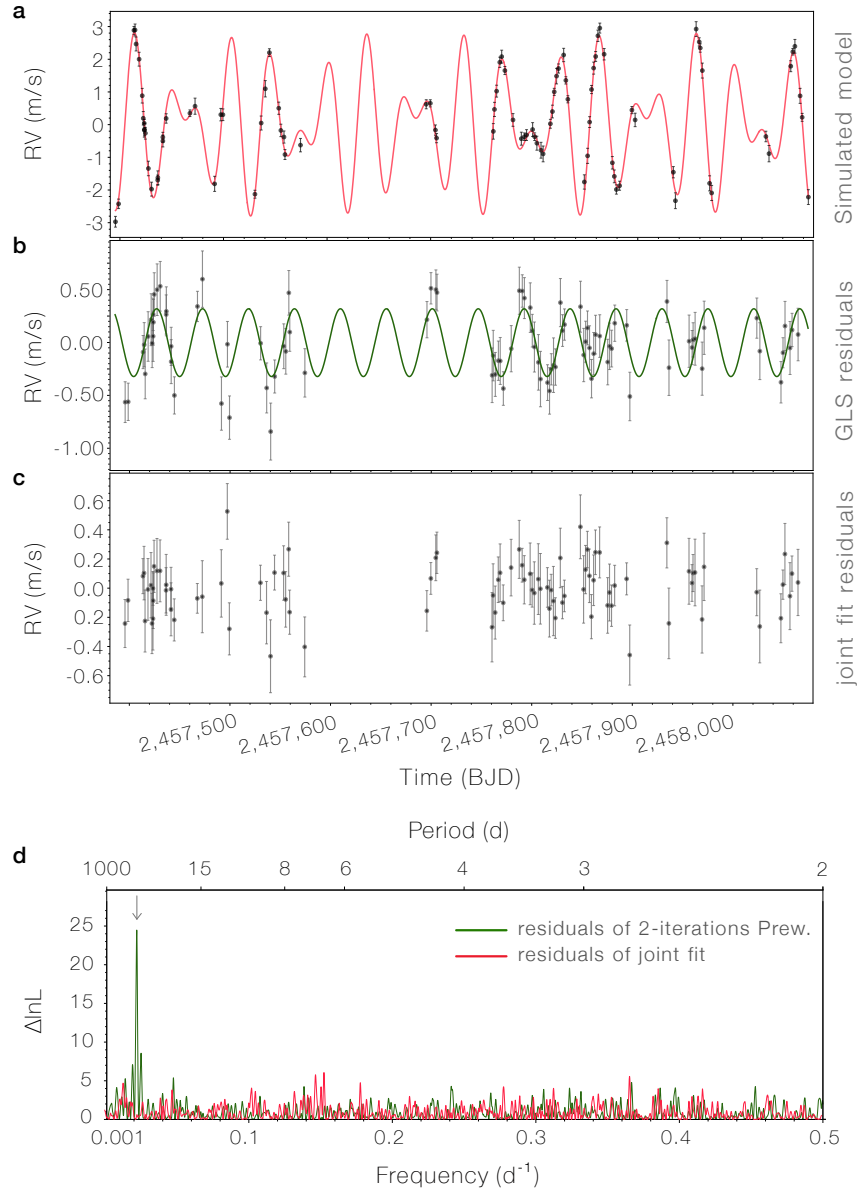


Figure 2.2: (a) Simulated two-period signal ($P_1 = 32$ and $P_2 = 45$ days) with amplitudes ≈ 1.5 m/s and additive Gaussian noise. (b) Residuals after having fitted and subtracted two signals iteratively using GLS algorithm. The green solid line shows the fit of these residuals. In this case, a significant ~ 30 cm/s signal was found, despite not being injected in the simulated data, and constitutes an artifact of spurious residual noise. (c) Residuals of a multifrequency joint fit of two sinusoids. A fit was done on these residuals and shows a very low significance ($\Delta \log \mathcal{L} = 5.61$) with respect to the null model, which is not significant. (d) GLS unidimensional periodogram in the case of iterative subtraction (in green), and simultaneous 2-signal fitting (multifrequency, in red). While in the green periodogram a strong and significant peak arises with $\Delta \log \mathcal{L} \sim 25$ at ~ 46 days, in the red one appears to be a forest of non-significant peaks, with $\max(\Delta \log \mathcal{L}) \approx 6$.

Table 2.1: Significant periodic signals found on ad-hoc toy model to demonstrate the potential flaws of time-series iterative processing.

M (dim.)	GLS (prewhitening)			MGLS		
	$\{P_j\}$	$\{K_j\}$	$\Delta \ln \mathcal{L}$	$\{P_j\}$	$\{K_j\}$	$\Delta \ln \mathcal{L}$
1	45.157	1.74 ± 0.19	42.85	45.157	1.74 ± 0.19	42.85
2	45.157	1.74 ± 0.19	42.85	44.972	1.54 ± 0.03	196.61
	31.997	1.21 ± 0.06	99.52	32.018	1.33 ± 0.03	
3	45.157	1.74 ± 0.19	42.85			
	31.997	1.21 ± 0.06	99.52			
	45.675	0.32 ± 0.04	24.52			
Injected signal	$\{P_j\}$	$\{a_j\}$	$\{b_j\}$	$\{K_j\}$		
	45	0.25	1.50	1.52		
	32	0.65	1.15	1.32		

Notes:

We give the maximum likelihood period P in days, amplitude K in ms^{-1} , and the corresponding $\ln \mathcal{L}$ value. The $1\text{-}\sigma$ uncertainties given in the linear parameters are calculated as described in Section 2.3.3.1. The left column shows the dimensionality of the fit, i.e., the number M of signals.

and the resulting fit is shown in the middle panel. Even at first sight, a good-looking and apparently significant new signal is found, and when it comes to the statistics of the fit, we obtain a $\Delta \ln \mathcal{L} = 36.72$, which is convincingly significant. However, this signal is not a $\sim 30 \text{ cm/s}$ real signal buried in the noise, but it is a spurious artifact built on the correlated residuals coming from fitting individual signals in multifrequency time-series.

In contrast, when a joint fit of two sinusoids is performed, the result is far more accurate. The bottom panel of Figure 2.2 shows the search for a signal in the residuals, and the result does not show evidence for the presence of an additional signal in the original time-series. Hence, we obtain a poor $\Delta \ln \mathcal{L} = 5.61$ improvement with respect to the null-hypothesis¹ model.

The bottom panel of Figure 2.2 shows the one-dimensional periodogram (GLS) of the residuals from the two-signal simultaneous fit, in red; and the iterative procedure, in green. In the first case, the distribution of peaks is very similar to those spectral signatures of white noise structures, with modest improvements in likelihood.

2.3 MGLS: Multidimensional Generalised Lomb Scargle

We present here MGLS² which is a newly-built as an extension of GLS to identify periodic signals from multiplanetary systems from unevenly sampled time-series. The procedure works by fitting a n -frequency model containing an oscillatory part.

1. The null-hypothesis model is a MGLS fit with zero dimensionality; i.e. no oscillatory part is considered, but just an offset (c_k) and a jitter.

2. The MGLS code is publicly available at github.com/rosich/mgls

The main idea behind MGLS is to fit simultaneously all the signals, avoiding having to iteratively subtract individual components and analyse again its residuals. As explained and demonstrated before, prewhitening could be insensitive to some planetary configurations—particularly for smallest signals—and even may lead to misdetections due to noise amplification (Foster 1995; Baluev 2013c; Anglada-Escudé et al. 2015).

2.3.1 Algorithm description and mathematical model

We consider a model composed of \mathcal{D} circular orbits expressed as a linear combination of pure sine waves with a zero point (c) plus a linear trend term,

$$y_i(t) = c + \sum_j^{\mathcal{D}} K_j \sin(\omega_j t_i + \phi_j) + \tau(t_i - t_0) \quad (2.15)$$

This equation has $2\mathcal{D}$ non-linear parameters (ω_j and ϕ_j) and $\mathcal{D} + 2$ linear parameters (K_j , offset c and τ linear trend). We can reduce the number of nonlinear parameters by transforming the last expression using a trigonometric identity $\sin(\omega t + \phi) = \sin(\omega t) \cos \phi + \cos(\omega t) \sin \phi$ and writting the wave to fit as a linear combination of sines and cosines,

$$K \sin \phi \cos(\omega t) + K \cos \phi \sin(\omega t) = a \cos(\omega t) + b \sin(\omega t) \quad (2.16)$$

identifying terms the transformation equations become

$$K = \sqrt{a^2 + b^2} \quad (2.17)$$

$$\phi = \tan^{-1} \left(\frac{a}{b} \right). \quad (2.18)$$

The multiset MGLS model consists of a set of \mathcal{D} non-linear parameters and $2\mathcal{D} + k + 1$ linear parameters, being k the number of datasets to be analysed simultaneously. This implies a substantial numerical efficiency boost since the linear part can be solved with a linear system, which is a task several orders of magnitude faster to compute in comparison to the non-linear parameters. In the particular case of $\mathcal{D} = 1$, the classical GLS algorithm is recovered and the only non-linear parameter, ω , is evaluated in a range of search frequencies, determining the coefficients a, b, c for each tested frequency.

In addition, a quadrature-added jitter can be included in each independent dataset when computing the merit function of the fit. In such case, the count is $\mathcal{D} + k$ non-linear and $2\mathcal{D} + k + 1$ linear parameters. In the case of linear trend is included. Otherwise, there are $2\mathcal{D} + k$ linear parameters.

Therefore, the model implemented in MGLS is a linear combination of sinusoidal functions in the transformed form of Equation 2.16 with additional offset term and linear trend function as follows,

$$y_{i,k}(t_{i,k} | a, b, c, \omega) = c_k + \sum_j^{\mathcal{D}} a_j \cos(\omega_j t_{i,k}) + \sum_j^{\mathcal{D}} b_j \sin(\omega_j t_{i,k}) + \tau(t_{i,k} - t_0) \quad (2.19)$$

where $y_{i,k}$ is the measured value from dataset k at time t_i . ω_j are the frequencies of each signal, c_k is zero-point (offset) corresponding to the k -dataset, and τ is a linear trend.

Proceeding in this way, we consider a multiset data frame composed of k different time-series, each one with a single offset, c_k , and a harmonic term which is common for all datasets involved in the analysis.

2.3.2 Global optimisation approach

We present in the following sections the approach used for the computation and optimisation of the MGLS periodogram. This may constitute a numerical challenge, particularly for high dimensionality, \mathcal{D} , but even for 2-dimensional MGLS the number of algorithm evaluations can be too expensive if a fine frequency grid is chosen. Assuming the computation of 1,000 samples in each dimension, the all-space MGLS computation of a \mathcal{D} dimension periodogram involves $10^{3\mathcal{D}}$ evaluations of the linear parameters. For a 4 dimensional search, it would take 10^{12} evaluations. Estimates show that the computing time would be ~ 500 days, which is clearly unaffordable for most applications. Although the computation of linear parameters after fixing a frequency tuple $\bar{\omega} = \{\omega_j; j = 1, \dots, 4\}$ is very fast, an efficient method to obtain optimal $\bar{\omega}$ is needed to avoid such prohibitive exhaustive search.

In the next sections, the strategy to compute optimal frequency tuple is developed, consisting of fixing a frequency tuple $\bar{\omega}$ and optimising linear parameters through a fast linear system solver, iteratively, following a random walk in $\bar{\omega}$. By implementing a Simulated Annealing non-linear optimiser allows obtaining good approximations of optimal $\bar{\omega}$ tuples in reasonable times, avoiding local optimals due to noisy and multimodal search space. Those approximations to the global optimals are refined by a classical gradient-descend optimisation algorithm to derive the absolute optimal. With this approach, the computation of high-dimensional MGLS spaces is affordable in reasonable times. In the next sections, the optimal finding design is explained in detail.

2.3.3 Solving for linear parameters

The general form of models like Equation 2.19) can be written as a linear combination of functions X (Numerical Recipes, Press et al. 1993),

$$y(\mathbf{x}) = \sum_{m=1}^B \theta_m X_m(\mathbf{x}) \quad (2.20)$$

where X_1, \dots, X_B are whatever type of linear or non-linear fixed functions with a multidimensional argument, \mathbf{x} . What we consider linear in these equations is the model relation with respect to the parameters θ_m . X_m are commonly referred as basis functions.

The merit function used will be the logarithm of the likelihood function

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{i=1}^N \frac{r_i^2}{\epsilon_i^2 + s^2} - \frac{1}{2} \sum_{i=1}^N \ln(\epsilon_i^2 + s^2) - \frac{1}{2} N \ln(2\pi), \quad (2.21)$$

r_i , ϵ_i and s are the residuals between model and data, the uncertainties and jitter, respectively.

The first term of the right-side of the equation is $\frac{1}{2}\chi^2$. Hence, $\ln \mathcal{L}$ can be written as a sum of a $\chi^2 = \chi^2(\theta, s)$ function which depends both on model parameters and jitter; and an additional term only dependent on jitter plus a constant.

Redefining Equation 2.21 as,

$$\ln \mathcal{L} = -\frac{1}{2}\chi^2(\theta, s) - g(s), \quad (2.22)$$

where

$$g(s) = \frac{1}{2} \sum_{i=1}^N \ln(\epsilon_i^2 + s^2) + \frac{1}{2} N \ln(2\pi) \quad (2.23)$$

θ are the set of parameters described at Equation 2.20. We are interested in finding the optimal linear parameters while having fixed all non-linear ones (ω_j, s_k). Working in likelihood logarithm, those optimal configurations θ will correspond to a maximum of the function defined in Equation 2.22. Provided this function only depends on θ in the χ^2 part, its maxima will be found when the derivative of χ^2 function with respect to each parameter θ_α , vanishes:

$$0 = \frac{\partial \ln \mathcal{L}}{\partial \theta_\alpha} = \sum_{i=1}^N \frac{1}{\sigma_i^2 + s^2} \left(y_i - \sum_{m=1}^B \theta_m X_m(\mathbf{x}_i) \right) X_\alpha(\mathbf{x}_i), \quad (2.24)$$

$$\alpha = 1, \dots, B$$

by inverting the order of summations we can write the last expression as the following matrix equation,

$$\sum_j^B \alpha_{kj} a_j = \beta_k \quad (2.25)$$

with

$$\alpha_{kj} = \sum_{i=1}^N \frac{X_j(x_i) X_k(x_i)}{\sigma_i^2}; \quad [\alpha] = A^\top A \quad (2.26)$$

where $A_{ij} = \frac{X_j(x_i)}{\sigma_i}$ and,

$$\beta_k = \sum_{i=1}^N \frac{y_i X_k(x_i)}{\sigma_i^2}; \quad [\beta] = A^\top b \quad (2.27)$$

therefore, the final matrix equation becomes

$$[\alpha] \cdot \vec{a} = [\beta]. \quad (2.28)$$

Using Equation 2.28, the final expression for the linear system becomes

$$(\mathbf{A}^\top \mathbf{A}) \theta_\omega = \mathbf{A}^\top \beta \quad (2.29)$$

where $\beta_i = y_{ik}/\sigma_{ik}$ for $i = 1, \dots, n_k$ and $k = 1, \dots, K$.

This last linear equation is indeed the MGLS function to obtain the linear parameter estimates taking as arguments the frequency and jitter tuples. The function returns the linear parameters a_j, b_j, c_k and, the $\ln \mathcal{L}$ evaluation of the best model.

$$\{\hat{\theta}, \ln \mathcal{L}\} = \text{MGLS}(\omega_0, \dots, \omega_n; s_0, \dots, s_k) \quad (2.30)$$

The i -index indicates the data point position in dataset k , while the length of dataset k is n_k . Hence, the β vector becomes a simple concatenation of all y values for all K datasets.

We define the design matrix \mathbf{A} of the fitting problem as $A_{ij} = X_j/\sigma_i$. In our case, this matrix will be,

$$\begin{pmatrix} 1/\sigma_{1,1} & 0 & \dots & 0 & \boxed{C_1(\bar{\omega})} & \boxed{S_1(\bar{\omega})} & (t_{1,1} - t_{\text{ref}})/\sigma_{1,1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1/\sigma_{n_1,1} & 0 & \dots & 0 & \vdots & \vdots & (t_{n_1,1} - t_{\text{ref}})/\sigma_{n_1,1} \\ \\ 0 & 1/\sigma_{1,2} & 0 & \dots & \boxed{C_2(\bar{\omega})} & \boxed{S_2(\bar{\omega})} & (t_{1,2} - t_{\text{ref}})/\sigma_{1,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1/\sigma_{n_k,2} & 0 & \dots & \vdots & \vdots & (t_{n_k,2} - t_{\text{ref}})/\sigma_{n_k,2} \\ \\ 0 & 0 & \dots & \dots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \dots & \vdots & \vdots & \vdots \\ \\ 0 & 0 & \dots & 1/\sigma_{1,K} & \boxed{C_K(\bar{\omega})} & \boxed{S_K(\bar{\omega})} & (t_{1,K} - t_{\text{ref}})/\sigma_{1,K} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1/\sigma_{n_K,K} & \vdots & \vdots & (t_{n_K,K} - t_{\text{ref}})/\sigma_{n_K,K} \end{pmatrix} \quad (2.31)$$

where $\bar{\omega} = (\omega_1, \omega_2, \dots, \omega_{\mathcal{D}})$ is the frequency tuple. The row blocks corresponds to each of the K datasets, which is the number of columns in the first block. The total number of rows is the sum of all data points. The next C and S matrices contain \mathcal{D} columns each one, and the last column accounts for the linear trend. The order of \mathbf{A} is

therefore, $N \times K + 2\mathcal{D} + 1$. The matrices $C_k(\bar{\omega})$ and $S_k(\bar{\omega})$ with $k = 1, K$ are constructed as follows,

$$C_k(\bar{\omega}) = \begin{pmatrix} \cos(\omega_1(t_{1,k} - t_{\text{ref}})) / \sigma_{1,k} & \cdots & \cos(\omega_{\mathcal{D}}(t_{1,k} - t_{\text{ref}})) / \sigma_{1,k} \\ \vdots & \ddots & \vdots \\ \cos(\omega_1(t_{n_k,k} - t_{\text{ref}})) / \sigma_{n_k,k} & \cdots & \cos(\omega_{\mathcal{D}}(t_{n_k,k} - t_{\text{ref}})) / \sigma_{n_k,k} \end{pmatrix}, \quad (2.32)$$

and,

$$S_k(\bar{\omega}) = \begin{pmatrix} \sin(\omega_1(t_{1,k} - t_{\text{ref}})) / \sigma_{1,k} & \cdots & \sin(\omega_{\mathcal{D}}(t_{1,k} - t_{\text{ref}})) / \sigma_{1,k} \\ \vdots & \ddots & \vdots \\ \sin(\omega_1(t_{n_k,k} - t_{\text{ref}})) / \sigma_{n_k,k} & \cdots & \sin(\omega_{\mathcal{D}}(t_{n_k,k} - t_{\text{ref}})) / \sigma_{n_k,k} \end{pmatrix} \quad (2.33)$$

The parameter vector is defined as

$$\theta = \begin{pmatrix} \mathbf{c} \\ \mathbf{a} \\ \mathbf{b} \\ \tau \end{pmatrix}, \quad (2.34)$$

where the submatrices \mathbf{c} , \mathbf{a} and \mathbf{b} are defined as

$$\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_K \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_{\mathcal{D}} \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_{\mathcal{D}} \end{pmatrix} \quad (2.35)$$

2.3.3.1 Uncertainties in the linear parameters

The inverse matrix $C_{jk} = [\alpha]_{jk}^{-1}$ is related to the standard error of the estimated parameters θ .

$$\theta_j = \sum_{k=1}^B [\alpha]_{jk}^{-1} \beta_k = \sum_{k=1}^B C_{jk} \left(\sum_{i=1}^N \frac{y_i X_k(x_i)}{\sigma_i^2} \right) \quad (2.36)$$

and the variance associated with θ_j is given by

$$\sigma^2(\theta_j) = \sum_{i=1}^N \sigma_i^2 \left(\frac{\partial \theta_j}{\partial y_i} \right)^2 \quad (2.37)$$

note that α_{jk} is independent of y_i , so that

$$\frac{\partial \theta_j}{\partial y_i} = \sum_{k=1}^B C_{jk} \frac{X_k(x_i)}{\sigma_i^2} \quad (2.38)$$

and we find,

$$\sigma^2(\theta_j) = \sum_{k=1}^B \sum_{l=1}^B C_{jk} C_{jl} \underbrace{\left(\sum_{i=1}^N \frac{X_k(x_i) X_l(x_i)}{\sigma_i^2} \right)}_{[\alpha]=C^{-1}} \quad (2.39)$$

and reduces to

$$\sigma^2(\theta_j) = C_{jj} \quad (2.40)$$

the diagonal elements are the variances and the off-diagonal elements are the covariances.

2.3.4 Solving for nonlinear parameters: optimal (ω_j, s_k) search

The best set of parameters, $\hat{\theta}_\omega$, are those that maximise the merit function $\ln \mathcal{L}_\omega$ (Equation 2.21), which contains both linear and nonlinear parameters. Hence, the optimal set of parameters for each frequency tuple $\bar{\omega}$ is,

$$\hat{\theta}_\omega = \arg \max_{\theta \in \Theta} \ln \mathcal{L}_\omega(\mathbf{x} | \theta), \quad (2.41)$$

In practice, the search for these optimal linear parameters consist of fixing a tuple of frequencies $\bar{\omega} = (\omega_1, \omega_2, \dots, \omega_{\mathcal{D}})$ and solving a linear system, as explained in the last section (2.3.3). However, a full solution consists of finding an optimal set of frequencies and its associated optimal linear parameters. Therefore, we need to perform multiple searches for varying $\bar{\omega}$ to find the full optimal solution, which is composed of a set of frequencies and their optimal linear parameters. The enormous search space and multimodality involved in this type of optimization problems leads to a complex structure of local optima, in addition to the huge volume of search space in comparison to the volume of the global optimal solution. These conditions make the task of finding the global optimal quite arduous.

To solve the nonlinear part, we implemented a Simulated Annealing (SA) optimisation approach (Kirkpatrick et al. 1983) to find approximate solutions of frequency tuples, $\bar{\omega}$. The strategy followed to obtain global optimal tuples, $\bar{\omega}^*$, starts by searching an approximation to the solution by means of the SA procedure, and refining this solution using a local optimisation algorithm, which is much faster in reaching the optimum if a sufficiently good initial hint is provided by in the previous SA-based stage. We understand that a sufficiently good starting point is such that prevents the local optimiser from getting trapped in a relative maximum.

To boost the efficiency of the search and to assure the correct convergence, MGLS implements a parallel exploration strategy consisting in initializing N coarse global searches implementing a SA algorithm to make a list of optimal candidates, which can be run in parallel to enhance the performance. Then, the list of candidates is taken as initial hint for a SLSQP optimizer (Sequential Least-Squares Programming, Kraft 1988) with the Scipy library Python implementation (Virtanen et al. 2020), which in

turn will generate the final list of global candidates. The SLSQP algorithm is known to be the most efficient computational method to solve the general nonlinear programming problem, consisting of solving for

$\min_{\mathbf{x} \in \mathfrak{R}^n} f(\mathbf{x})$ subject to restrictions $g_j(x) = 0$ and $g_i(x) \geq 0$ for $j = 1, \dots, m_e$ and $i = m_e + 1, \dots, m$. The best optimal of the candidate list is taken as the global optimum.

Hence, the flow of the search algorithm can be summarised as follows,

1. Draw a random configuration (a tuple of random frequencies).
 - Set initial inverse temperature $\Theta = \Theta_0$, where $T = 1/\Theta$ is defined as the cooling temperature of the SA algorithm.
 - Let $k = 0$ (number of temperature steps).
 - Let $n_\Theta = 0$ (number of steps for each temperature)
2. Evolve the current configuration ϕ_j to a perturbed configuration ϕ_j^p by slightly modifying one randomly selected frequency of the tuple. We define f as the function to be maximised. In this case, it is the optimal $\ln \mathcal{L}$ of the linear optimisation (Equation 2.30) with a set of nonlinear parameters as arguments.
3. If $f(\phi_j^p) > f(\phi_j)$ then
 - accept ϕ_j^p
 - else
 - accept ϕ_j^p with probability $e^{-\Theta \Delta f}$
 - $n_\Theta \leftarrow n_\Theta + 1$. If $n_\Theta \leq n_\Theta^{\text{MAX}}$ go to 2.
4. Increase the inverse temperature $\Theta \leftarrow \Theta + \delta\Theta$ according to a cooling schedule.
5. $k \leftarrow k + 1$. If $k < k_{\text{max}}$ go to 2, else go to 6.
6. Refine the approximate simulated annealing solutions using SLSQP algorithm with the last $\bar{\omega}$ solution as initial hint.
7. End.

Steps from 1 to 5 are the SA optimisation algorithm itself that provide approximate solutions to the local SLSQP optimiser.

2.3.4.1 Uncertainties in nonlinear parameters

The errors on nonlinear parameters —frequencies and jitters— may be found through two methods. Namely,

1. *MCMC procedure*— By employing the solution obtained with the SA MGLS procedure as the starting point for a Monte Carlo Markov Chain (MCMC) a search around this optimal is done to derive frequency and jitter distributions. The evaluation function of the MCMC algorithm is that in Equation 2.30, which is very efficient to compute allowing to obtain MCMC distributions in a rapid fashion. Note

that we could not apply this method directly from the beginning of the search due to the sharpness of the optimal hyperpeaks in comparison to the vast search space, making them very difficult to find by taking random initial hints. In Figure 2.3 we show an example performed by analysing the distributions of periods and jitter for Teegarden’s star CARMENES observations in search for 4 signals.

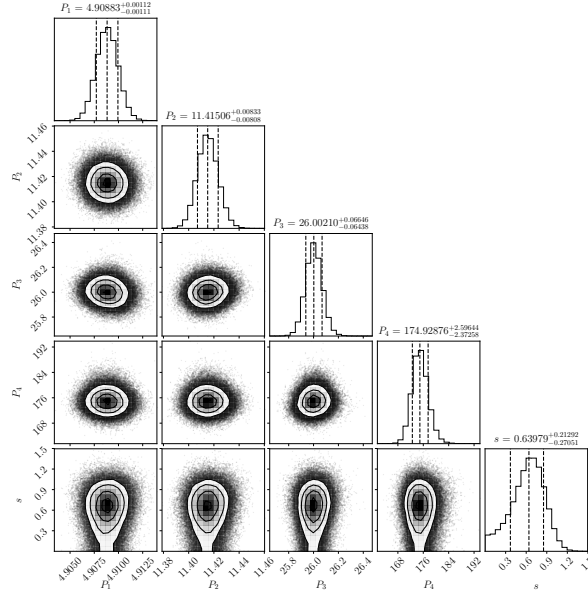


Figure 2.3: MCMC for the period and jitter distribution of Teegarden star observations (see Section 3.1.4) searching for 4 signals.

2. *Noisy data*— This approach consists of adding a zero-mean Gaussian distributed random value to each data point of the time-series, with the same standard deviation as its measurement uncertainty. Hence, a new sequence of measurements compatible with that observed and is obtained, and a new optimal search is performed. By generating typically thousands of repetitions of this process and applying a 3σ clipping to discard outlier solutions in which the search had not converged, we obtain the distribution of frequencies and jitters.

The results obtained by using both methods are in a very good agreement, however, the MCMC is much faster than noising data because, in the first method, the MCMC chain only needs to evaluate the linear function (Equation 2.30) and in the second, full optimisations are needed implying an investment of time orders of magnitude higher, making this approach unaffordable in practice.

2.4 Analysis of peak significances

The power of the periodogram provides an estimate of the statistical peak significance. All the possible normalisations offer different ways to understand the meaning of peak height, but the usefulness of a periodogram is to show the strength of each frequency to

explain the frequency content in data. Since all observational data have measurement uncertainties, these produce noisy periodograms and, therefore, many of the peaks in the periodogram are not due to real signals. The power of a peak only makes sense when compared to the power raised by pure noise: it is a measure of how likely is a signal to be generated just by chance.

The FAP evaluates the probability of a periodicity to be produced by statistical noise. Under the assumption of Gaussian uncorrelated noise, when the null hypothesis (no signal) is true, the FAP is the chance that a spurious signal of the same statistical power would be obtained by a noise effect. Very commonly, the FAP is given as a threshold of 1%, 0.1% or 0.01%, by usage convention.

2.4.1 Analytical formulations: Scargle formula

In the case of uncorrelated white noise, the spectral powers $Z_j = P(\omega_j)$ of sampling frequencies $\{\omega_j : j = 1, \dots, M\}$ are drawn according an exponential distribution with parameter $1/\sigma_X^2$, and σ_X^2 is the variance of time-series data (Scargle 1982; Cumming 2004). A comprehensive review can be found in Frescura et al. (2008).

Hence, the probability of obtaining a power $P(\omega)$ is given by

$$p_Z(z)dz = \text{Prob}(z < Z < z + dz) = \frac{1}{\sigma_X^2} \exp\left(-\frac{z}{\sigma_X^2}\right) \quad (2.42)$$

Where z is the height of the peak (spectral power). The cumulative distribution function $F_Z(z)$ of the exponential distribution is

$$P_Z(z) = \text{Prob}(Z < z) = \int_{\zeta=0}^z p_Z(\zeta)d\zeta = 1 - \exp\left(-\frac{z}{\sigma_X^2}\right) \quad (2.43)$$

When computing a unidimensional periodogram, a range of frequencies usually evenly spaced are tested to obtain a list of spectral powers $\{P(\omega) : j = 1, \dots, M\}$, where M is the number of sampled points. Therefore, the best model is that obtained with the frequency that maximises its fitting statistic,

$$\hat{Z}_M = \max(z_0, z_1, \dots, z_M) \quad (2.44)$$

If the set of variables $\{z_0 = P(\omega_0), \dots, z_M = P(\omega_M)\}$ are independently distributed, the probability of picking a value $z \leq Z_M$ is given by the product of the probability of independent events of Equation 2.43,

$$\text{Prob}(z \geq \hat{Z}_M) = \prod_j^M \text{Prob}(z_j \leq z) = \left[1 - \exp\left(-\frac{z}{\sigma_X^2}\right)\right]^M \quad (2.45)$$

We are interested in the probability that a given threshold z_0 is higher than the maximum \hat{Z}_M . The analytical FAP (false alarm probability) expression is the well-known

Scargle formula,

$$\text{FAP}(z_0) = 1 - (1 - e^{-\tilde{z}_0})^M \quad (2.46)$$

where $\tilde{z}_0 = \lambda z_0 = z_0 / \sigma_X^2$ is the power of the tested frequency. The derivative of $P^M(z)$ is the probability density function, $f(z; \lambda, M)$

$$f(z; \lambda, M) = \frac{dP^M(z)}{dz} = \lambda e^{-\lambda z} M (1 - e^{-\lambda z})^{M-1} \quad (2.47)$$

By inverting Equation 2.46 we obtain the z limiting value given a FAP level

$$z_l = -\frac{1}{\lambda} \ln \left(1 - (1 - \text{FAP})^{1/M} \right) \quad (2.48)$$

The practical interpretation of the M parameter is somewhat confusing because the maximum number of samples $P(\omega_j)$ we can draw when generating the periodogram does not correspond to the number of sampling points and it is highly dependent on the frequency range where the analysis is done, the normalisation used to compute the goodness-of-fit statistic and, the time sampling.

2.4.1.1 Independent frequencies

The Scargle test (Scargle 1982) is built under the assumption that $\omega_0, \omega_1, \dots, \omega_M$ are independent variables. This condition is not trivial to reach, mainly due to the effects that sampling, being this even or uneven, has on the computation of spectral estimation: aliasing and spectral leakage.

The case of evenly sampled observations is the only one in which independent frequencies are known, and these are given by (Priestley 1981)

$$\omega_k = 2\pi k/T \quad \text{with } k = 0, \dots, \lfloor N_0/2 \rfloor \quad (2.49)$$

where T is the data timespan, N_0 is the number of observations, and $\lfloor \cdot \rfloor$ denotes the floor operation. Therefore, the number of independent frequencies is $\lfloor N_0/2 \rfloor$.

Aliasing is the effect of arising different frequencies that become undistinguishable when sampled in certain patterns. A single real signal can appear in more than one significant peak in the periodogram, due to e.g., regular sampling. Spectral leakage is produced when the periodogram calculation creates new frequencies through a nonlinear time-invariant transformation. The most important is *windowing* effect, which also can be understood as a sampling problem: a limited timespan introduces a box-shaped signal, that vanishes outside the epochs with observations. The convolution of a real continuous signal with this artifact window results in a complex peak forest.

The quality of the predictions of Scargle's formula decreases dramatically in case of structured observational patterns, in which case the number of independent frequencies, M , can not be determined. Several analytical or hybrid strategies to estimate

M have been published (e.g. Horne et al. 1986; Koen 1990; Frescura et al. 2008; Süveges 2012)

2.4.1.2 Random fields extreme value theory

As said before, in the case of unevenly spaced observations there is no formula for the maximum number of independent frequencies, and the determination of an analytical expression for the false alarm probability is approximate since it depends on this parameter.

Baluev (2008) used the framework of extreme value theory for stochastic processes to derive an analytical formula to estimate an upper bound of the FAP even for highly structured window functions. For a given peak height z , the FAP can be approximated as,

$$\text{FAP}(z) \approx 1 - (1 - e^{-z})e^{-\tau(z)} \quad (2.50)$$

where

$$\tau(z) \approx W(1 - z)^{(N-4)/2}\sqrt{z} \quad (2.51)$$

where N is the number of observations, and

$$W = f_{\max} \sqrt{4\pi \text{VAR}(t)} \quad (2.52)$$

where f_{\max} denotes the maximum tested frequency, and $\text{VAR}(t)$ is the variance of the time array.

2.4.2 Numerical approaches

All the analytical methods presented above, have a number of assumptions that make them difficult to apply in real observational cases. In those circumstances where little or no information about the distributions of the statistics, a numerical approach consisting of data resampling may be implemented to obtain empirical FAP thresholds.

2.4.2.1 Bootstrap method

Bootstrap (Efron 1979, 1982; Kulesa et al. 2015) is a nonparametric numerical technique for drawing the distribution of maximum $\ln \mathcal{L}$ produced by pure noise (e.g. Endl et al. 2001). As a general rule, the procedure consists of preserving the epochs in the original sorting and randomly replacing each magnitude jointly with its error by another, with the possibility of repetition. For each of these bootstrap samples, the periodogram is computed and the highest peak $\ln \mathcal{L}$ value is preserved. This procedure takes into account all the effects carried by the unevenly spaced sampling and the fact of often is difficult to assure the independence of the X random variables drawn from the periodogram sampling (Equation 2.44).

The bootstrap samples are computationally expensive and even though this method has been revealed as the most straightforward and realistic in obtaining FAP levels, $\sim 100,000$ samples are necessary to assess stable and significant thresholds in the range of 0.1% or lower. Note that even when calculating this amount of samples, the expected count for a 0.1% level FAP is 100 events. Therefore, to overcome this limitation, several shortcut strategies have been developed to shorten the computational effort (e.g. Süveges 2012).

Unfortunately, the bootstrap technique can not be applied in those situations where noise correlations are not negligible.

2.5 Experiments on FAP determination

Considering the derivations of the Scargle theory (Section 2.4.1), the probability density function of maximum peaks is written as

$$f(z; M, \lambda) = \lambda M e^{-\lambda z} \left(1 - e^{-\lambda z}\right)^{M-1} \quad (2.53)$$

It has two parameters, λ and M . The first accounts for the dispersion of data and is the same parameter in the exponential distribution of sampling, i.e. the spectral power of random points in the periodogram. The second is the number of independent frequencies, which is only known in the case of uniform sampling of data (Equation 2.49). As has been discussed before, a fundamental assumption made for the development of Equation 2.53 is the independence of the samples collected to calculate the extreme value of a set of M samples. That is not the case of uneven observations, where the correlations appear in convolutions with the window function and are very difficult to develop analytically. A common solution to estimate the number of independent frequencies is to consider an effective number of independent frequencies to calculate the FAP. There are various estimators for M_{eff} , the most simple is $M_{\text{eff}} = \Delta T / \delta t$ where ΔT is the timespan and δt is the minimum data separation. Other more sophisticated methods are found in Horne et al. (1986), Schwarzenberg-Czerny (1998), Cumming (2004), and Frescura et al. (2008).

2.5.1 Fitting Scargle's formula

We propose here a method for inferring good approximations of the distribution peak values by using ~ 1000 bootstrap samples to allow us to compute tail values of the distribution without the need of having to compute $\sim 10^5$ full periodograms to obtain consistent estimations of FAP levels.

1. Generate ~ 1000 bootstrap samples shuffling data with repetition,
2. Fit the M and λ parameters of Scargle equation in 2.53 with the histogram of $\Delta \ln \mathcal{L}$ obtained in the first step,
 - (a) Generate $\sim 1,000$ samples from the Scargle distribution

- (b) Compare both sets with Kolmogorov–Smirnov (KS) test to assess whether the two groups are drawn from the same distribution
- (c) The best M and λ are taken as the effective values

Figure 2.4 shows an example of fitting the highest peak distributions using observational data of Teegarden’s star acquired with the CARMENES spectrograph (see Section 3.1.4). On the left, the periodogram sampling distribution is shown, i.e. the power at random frequencies. In red, we fitted an exponential law with density distribution,

$$\text{Exp}(x; \lambda) = \lambda e^{-\lambda(x-c)} \quad (2.54)$$

obtaining a value of $\lambda = 0.993$ and location parameter $c \approx 0$. The right part of the plot, depicts the experimental highest peak distribution—in black— obtained by bootstrapping ($\sim 10^5$ samples). The lines represent the fits of Scargle distribution (Equation 2.53) adjusting both M and λ —in light blue— and in the dashed line, fitting only M and assuming the λ value fitted in the exponential distribution of the periodogram sampling. When the assumption of white Gaussian noise is not satisfied, correlations in the observations make the λ value change from the expected value of the exponential distribution fit, as seen in Figure 2.4 and Equation 2.54.

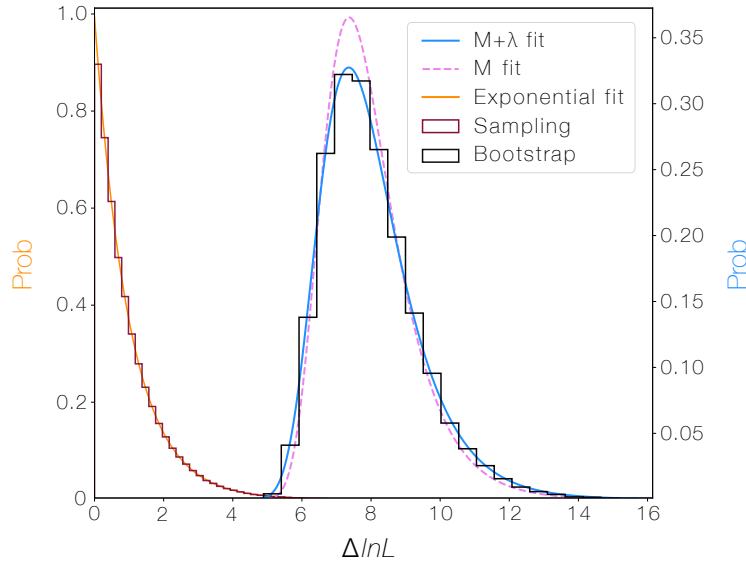


Figure 2.4: Example of bootstrap histogram fit. The plot shows an example of fitting highest peak distributions using observational data of Teegarden’s star with the CARMENES spectrograph. The left part depicts the periodogram sampling distribution, i.e. the power at random frequencies. In red, we fitted an exponential law to this data (Equation 2.54). By fitting both M and λ parameters of the probability distribution function describing the distribution of highest peaks in the noisy periodogram a significantly better result is obtained (light blue solid line). In case of taking the λ value fitted in the exponential law of the periodogram sampling, and only fitting the M parameter, the fitted Scargle law is slightly biased (dashed pink line).

The two parameter fit—solid blue line— results in $\lambda = 0.94$ and $M = 1026$. With these values we are able to compute the FAP values, which in this case yield $\text{FAP}_{0.1\%}^{\text{fit}} =$

15.12, while the numerical estimation done by bootstrapping provides a value of $\text{FAP}_{0.1\%}^{\text{boot}} = 15.31$. The advantage in using the analytical FAP functions relies on estimating FAP values with high $\Delta \ln \mathcal{L}$ or low false alarm probabilities, which in the case of bootstrap sampling would induce large statistical uncertainties for the usual sample width of 10^5 . Note that these Scargle law fits are obtained by using $\sim 1,000$ samples instead of $\sim 100,000$, resulting in typical savings in numerical effort of up to a factor of 100. Fitting the entire distribution versus the classic bootstrap procedure has the key advantage of including the information from all the computed samples instead of just counting the extreme values.

2.5.2 Proposed procedure for MGLS

The FAP determination of periodogram peaks can be performed by following a simple recipe. If the peak is higher than a given threshold computed either analytically or by bootstrap numerical simulation, the signal is significant. However, in higher dimensionality, this procedure is not enough since, e.g., a tuple of two frequencies can overtake a certain significance level obtained by bootstrapping and looking for a 2-signal solutions. The two-signal solution we search can respond to various scenarios in real data: i) None of the signals is significant, ii) Signal 1 is significant while signal 2 isn't, iii) Signal 2 is significant while signal 1 isn't, and, iv) Both signals are jointly significant.

An overall significance test is, therefore, not useful for this purpose: a single strongly significant signal can be confused as a single signal or two significant signals because performing a search for two signals the $\Delta \ln \mathcal{L}$ may exceed the two-dimensional distribution of $\Delta \ln \mathcal{L}$ for pure noise.

We describe the procedure to extract a tuple of significant frequencies.

1. Compute the null-model $\ln \mathcal{L}_0$ considering a model that fits only an offset and jitter for each dataset, with no oscillatory part.
2. Set $\mathcal{D} = 1$ search for one signal and compute $\Delta \ln \mathcal{L}$ with respect to the null model. If this value exceeds the FAP threshold: adopt $\mathcal{D} = \mathcal{D} + 1$; otherwise STOP.
3. Search for simultaneous $\mathcal{D} = \mathcal{D} + 1$ signals. Evaluate the new $\Delta \ln \mathcal{L}$ with respect the \mathcal{D} model. If this value exceeds the FAP threshold; adopt $\mathcal{D} = \mathcal{D} + 1$; otherwise STOP
4. Iterate over step 3 until termination condition occurs.

The FAP threshold is set according to the probability of failing in considering a signal to be real when it actually results from noise. The procedure described above is implemented in the next chapter to analyse and discuss multiplanetary systems with observational data and compared to the iterative prewhitening scheme to show the improvement in robustness and detectability when performing global fitting of multifrequency models.

2.6 Notes on the numerical performance and code implementation

As explained in previous sections the working mechanics of MGLS consist in setting a tuple of frequencies and jitters (one per each dataset) and compute the optimal linear parameters by using a linear system recipe. We use the general-purpose and efficient *SGESV* linear system solving routine from the LAPACK Fortran library (Anderson et al. 1999).

However, the matrix product $A^\top A$ admits Cholesky decomposition, consisting of the decomposition of a Hermitian, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose (Haddad 2009), making it the most efficient way to solve the normal equations developed previously (Equations 2.24, 2.25). In this implementation, we used the routine *SPOSV* to solve the linear system by this method. Since the linear parameter search is evaluated massively in a MGLS search run, the efficiency in computing the linear system is a key point to make feasible the MGLS search in a reasonable time.

The largest fraction of computing time is spent in arranging the matrices, by looping over the data structures. The matrix operations are performed using the highly-efficient BLAS library (Blackford et al. 2002).

2.6.1 Numerical throughput

In Table 2.2 we show the computation times of the MGLS linear parameters fit for different datasets tested in following sections. The values shown correspond to the time needed to compute the function

$$\{\hat{\theta}, \ln \mathcal{L}\} = \text{MGLS}(\omega_0, \dots, \omega_n; s_0, \dots, s_k), \quad (2.55)$$

where $\hat{\theta}$ are the linear parameters estimated from the model, $\ln \mathcal{L}$ the merit function of the fit and ω and s are the frequencies and jitters (for k datasets), respectively.

We show the results corresponding to the different number of signals searched. In the first row, a single dataset containing 91 epochs —toy model developed in Section 2.2.1—, and a complex set of data of Barnard’s star analysis in Section 3.1.5, consisting of 8 independent datasets comprising a total of 771 measurements gathered during more than 20 years. The test was done on a Linux box with Intel i5-8259U@2.3 GHz processor.

Table 2.2: MGLS running times with Intel i5-8259U@2.3 GHz CPU^(a).

computational time (core)				
dataset/dimensionality	$n_\omega = 1$	$n_\omega = 2$	$n_\omega = 3$	$n_\omega = 4$
Toy model ^(b)	30 μ s	34 μ s	41 μ s	50 μ s
Barnard's star ^(c)	204 μ s	247 μ s	316 μ s	371 μ s
computational time (full optimization)				
Toy model	7.6 s	15.6 s	28 s	43 s
Barnard's star	3.2 min	4.7 min	6.1 min	7.8 min

^a The computing times were evaluated by using a single CPU core. Large improvements in time are reached in multicore machines for the full optimisation routine.

^b 91 data points in a single dataset (see Section 2.2.1).

^c 771 measurements in 8 datasets (see Section 3.1.5).

2.6.2 Code architecture and usage

Package dependences MGLS is written in Python 2.x with precompiled libraries in Fortran 90 which are linked through `f2py` interfacier (Peterson 2009). This wrapper facilitates the integration of fast compiled Fortran code into Python environments resulting in a dramatic speed-up in those routines in where Fortran excels in speed, i.e. loops, conditional statements or basic algebra. The Fortran subroutines must be compiled before to be callable by the main Python code using this build-in bash script,

```
$ bash compile_fortran_modules.sh
```

`f2py` is a part of Python `Numpy` package and can also be installed as a standalone command line tool. In addition to this, MGLS also needs `matplotlib 1.4+` and `Lapack` library for fast linear algebra routines. To evaluate the uncertainties in frequency tuples MGLS uses the MCMC ensemble sampler `emcee` implementation for Python by (Foreman-Mackey et al. 2013).

Basic usage Next, we provide a list of the commands implemented in the MGLS code. Following the structure

```
./mgl.py <datafile_0> <datafile_1> ... <datafile_n> [--options]
```

a number of options are available:

1. `--gls`: perform of the one-dimensional GLS (default 5,000 frequency steps).
2. `--ndim=`: dimensionality of MGLS solve.
3. `--pmin=`/`--pmax=`: define the period range, P_1 to P_2 (default: 1.5 and 10,000 days).
4. `--linear_trend`: adjust an additional linear trend (see Equation 2.19).
5. `--jitter`: an additional jitter term s is included in quadrature for each set of data, $\epsilon^2 = \sigma^2 + s^2$.
6. `--mcmc=`: use the MGLS solution as starting point to compute mcmc chain to evaluate uncertainties in frequency tuples.
7. `--ncpus=`: number of computing cores.
8. `--bidim`: produce a two-dimensional periodogram plot.
9. `--inhibit_msg`: do not show informative screen messages.
10. `--bootstrapping=`: Compute a given number of bootstrap samples.
11. `--col=`: Use a specific column for time-series in data file. The errors are in the next column.
12. `--period`: plot GLS periodogram in period log-scale (only for `--gls` option).

2.7 Conclusions

We have designed a new approach in computing a Lomb–Scargle periodogram for unequally sampled time-series data based on the fit of multifrequency models, which we dub MGLS. The aim of our code is to provide an alternative quick way to find significant periodicities and to evaluate the multiplicity of a system. In the following chapter, a testbench is performed with the code on real multiplanetary systems.

The procedure we suggest has been shown to yield better results than the prewhitening technique both in the determination of the period and amplitudes of the signals. This is partly due to the fact that the MGLS is more flexible in the sense that it is not based upon a grid of frequencies but applies the best fit, but also on the simultaneous consideration of the signals. In all our tests we find the $\ln \mathcal{L}$ level and the significance of the signals increasing with the multiplicity of the systems compared to the prewhitening technique.

The MGLS approach is, of course, based on the search for sinusoidal variations. Therefore, it has the same limitations as the GLS in the detection of strongly eccentric

Keplerian orbits or signals that vary their parameters as expected for the stellar contribution to the RVs. MGLS is also not able to yield information on the nature of signals. Since the MGLS is designed for a quick evaluation of the system, it is also not able to compete with a full fit of Keplerian orbits plus noise treatments as done by, e.g., the moving average or Gaussian Process techniques.

Chapter 3

MGLS performance evaluation on real multiplanetary systems

3.1 MGLS Testbed

In this chapter, we describe the results of testing the MGLS algorithm with real RV time-series data. We evaluate its performance through the comparison with the published values by using the classical prewhitening technique, and the simultaneous search presented in the preceding chapter.

To benchmark the performance of the MGLS code, we have selected four multiplanetary RV datasets that probe a broad parameter space, such as the multiplicity of the planetary system, availability of chromatic RVs, the signal-to-RV uncertainty ratio, the ratio between different signals, the time-sampling, systems with controversial signals and. Furthermore, we explore cases where the use of bidimensional periodogram plots provide a full picture of the potential signals across the frequency plane that may be helpful for considering suboptimal combinations of frequency tuples for further analysis.

We analyse RV datasets of four multiplanetary host stars, HD 40307, GJ 163, GJ 667C, and Teegarden’s star. In addition, we included the MGLS analysis performed in Barnard’s star, where the large available dataset is used to show the capabilities of combining measurements from different instruments and the advantage that bidimensional periodogram plots offer to the comprehensive analysis. An overview of the analysed target stars is given in Table 3.1. Some specific details of our test are the following:

- i)* The FAP values at 0.1% are evaluated by fitting the Scargle distribution using a ~ 1000 bootstrap samples through the procedure described in Section 2.5.1;
- ii)* The dataset analysis is performed employing both GLS and MGLS modes of the code, and the full results are displayed in tables for each explored dimensionality (i.e. number of signals included in the model);

Table 3.1: Overview of the stellar parameters for the analysed stars.

parameter	unit	HD 40307	GJ 163	GJ 667C	Teegarden's Star
Sp.type		K2.0 V ^e	M3.5 V ^f	M1.5 V ⁿ	M7.0 V ^b
G ^a	mag	6.8437 ± 0.0003	10.6771 ± 0.0008	9.3803 ± 0.0007	12.2746 ± 0.0010
distance ^a	pc	12.9412 ± 0.0046	15.1351 ± 0.0072	39.005 ± 0.038	3.8312 ± 0.0039
α^a	J2000	05:54:04.24	04:09:15.67	17:18:57.16	02:53:00.89
δ^a	J2000	-60:01:24.5	-53:22:25.3	-34:59:23.14	+16:52:52.6
μ_α^a	mas yr ⁻¹	-52.423 ± 0.065	1046.141 ± 0.064	1129.7 ± 9.7	3429.53 ± 0.33
μ_δ^a	mas yr ⁻¹	-60.151 ± 0.060	584.130 ± 0.070	-77.0 ± 4.6	-3806.16 ± 0.31
M	M _☉	0.77	0.40 ± 0.02 ^g	0.33 ± 0.02	0.0886 ± 0.0088 ^c
R	R _☉	0.707 ^{+0.014, a} _{-0.012}			0.1067 ± 0.0038 ^c
L	L _☉	0.25833 ± 0.00016 ^a	0.0196 ^g	0.0137 ± 0.0009 ^f	0.00073 ± 0.00001 ^c
T _{eff} ^a	K	4890 ± 50	3910 ⁺⁹⁰ ₋₂₇₀	3350 ⁺⁵⁰ ₋₅₀	3296 ⁺⁶ ₋₉
Fe/H		-0.36 ^h	0.1 ± 0.1 ^g	-0.55 ± 10 ^m	-0.19 ± 0.16 ^c
log R' _{HK}	cgs	-5.02 ± 0.07 ^h , -5.00 ± 0.13 ^j	-5.37 ± 0.11 ^h	-4.996 ± 0.027 ^j	-
P _{rot}	d	31.8 ± 6.7 ^h , 47.2 ± 5.3 ^j	61.0 ± 0.3 ^h	24.1 ± 3.0 ^j	-
age	Gyr	7.08 ± 0.81 ^j	0.77 ± 0.03 ^g	> 2	> 8 ^c

- (a) Gaia Collaboration et al. (2018)
(b) Alonso-Floriano et al. (2015)
(c) Schweitzer et al. (2019)
(d) Reiners et al. (2018)
(e) Gray et al. (2006)
(f) Koen et al. (2010)
(g) Tuomi et al. (2013a)
(h) Suárez Mascareño et al. (2015a)
(i) Kopparapu (2013)
(j) Lovis et al. (2011a)
(k) Lovis et al. (2011c)
(l) van Leeuwen (2007)
(m) Anglada-Escudé et al. (2013)
(n) Geballe et al. (2002)

- iii) By using the FAP values obtained in the first step, we are able to determine the number of significant signals that can be extracted from the data, and we compare the detectability performance of the simultaneous fit with the iterative procedure.
- iv) An additional test is performed to assess the significance of simulated signals in each time-sampling. The detectability test consist of injecting a sinusoidal wave of amplitude K with randomly selected parameters a and b so that $K = \sqrt{a^2 + b^2}$ and then, use MGLS to solve the linear parameters (see Equation 2.30) and compute $\Delta \ln \mathcal{L}$ at each given point on the plane (ω, K) . In the synthetic models were added Gaussian noise with the uncertainties given by the observational error bars. The results for all the test cases analysed are shown in Figure 3.2. The period-amplitude plot is smoothed with a Gaussian filter with $\sigma = 1$ to capture the low-frequency behaviour and filter out the variations due to the added random noise in each synthetic model. This test determines a lower bound of signal detectability since assumes a white noise scheme. If correlated red noise was present, these $\Delta \ln \mathcal{L}$ values would be larger, particularly for longer periods where the chance to generate structured signals in noise is higher.

3.1.1 HD 40307

We analyse data from the metal-poor K2.5 dwarf HD 40307. Mayor et al. 2009, M09 hereafter, detected three planets and a long-term drift in 135 radial velocity epochs observed with HARPS spectrograph. Since none of the three initial announced planets (M09) is massive, it appeared to be a potentially good candidate to host additional low-mass planets.

By including additional measurements taken over three nights, reducing all the datasets using the TERRA code and a moving average noise treatment, Tuomi et al. (2013b, TA13 hereafter) claimed three more planets orbiting this star, but rejected the claim of a long-term trend. A revision of this target with still a larger number of HARPS measurements data was done by Díaz et al. (2016), who confirmed four of the previously announced planets, fitted a third order polynomial to the long-term variation, and discussed a possible fifth planet. Their data set consists of 226 HARPS measurements observed from October 2003 to April 2014.

This planet host constitutes a perfect test case for multisignal searches as was claimed to host up to seven planets in a wide range of periodicities. The published planetary signals are summarised in Table 3.2.

3.1.1.1 HD 40307 signal search and analysis

The FAP obtained via the methodology described in Section 2.5.1 provides us a threshold to decide whether the inclusion of a new signal in the search is significant. We have calculated 1500 classical bootstrap samples and we fitted both λ and M parameters of the Scargle's law, obtaining a value of $\text{FAP}_{.1\%} = 15.20$ ($\Delta \ln \mathcal{L}$).

The GLS search (Table 3.3, left columns) significantly detects up to five signals. These correspond to the announced planets in TA13 from b to f . Although an additional

Table 3.2: Six-planet solution of HD 40307 radial velocities according to TA13

Parameter	HD 40307 <i>b</i>	HD 40307 <i>c</i>	HD 40307 <i>d</i>
<i>P</i> (days)	$4.3123^{+0.0011}_{-0.0012}$	$9.6184^{+0.0050}_{-0.0049}$	$20.432^{+0.022}_{-0.024}$
<i>e</i>	$0.20^{+0.14}_{-0.16}$	$0.06^{+0.11}_{-0.06}$	$0.07^{+0.11}_{-0.07}$
<i>K</i> (ms ⁻¹)	$1.94^{+0.31}_{-0.27}$	$2.45^{+0.30}_{-0.28}$	$2.75^{+0.35}_{-0.30}$
	HD 40307 <i>e</i>	HD 40307 <i>f</i>	HD 40307 <i>g</i>
<i>P</i> (days)	$34.62^{+0.21}_{-0.20}$	$51.76^{+0.50}_{-0.46}$	$197.8^{+5.7}_{-9.0}$
<i>e</i>	$0.15^{+0.13}_{-0.15}$	$0.02^{+0.20}_{-0.02}$	$0.29^{+0.31}_{-0.29}$
<i>K</i> (ms ⁻¹)	$0.84^{+0.32}_{-0.31}$	$1.09^{+0.28}_{-0.32}$	$0.95^{+0.32}_{-0.30}$
<i>N_{meas}</i>	345		
σ_j (ms ⁻¹)	$0.73^{+0.14}_{-0.14}$		

significant component appears with a period of 4.5 days, the proximity to a previously extracted signal suggests that this is an artifact generated by the prewhitening process.

Regarding the procedure using MGLS, the code simultaneously detects a set of seven significant signals. Five of them correspond to the TA13 announced from *b* to *f*, and two other significant signals, a long-term one at 295 days and one at 15.6 days, also appear. The long-period signal is close to an already detected 320-day signal in TA13. However, this latter signal is strongly correlated to the chromospheric emission, and also shows chromaticity (TA13). Our MGLS search finds a significant tuple of seven frequencies. The seventh 15.6-day signal was not found in any previous report, and appears with a marginal $\Delta \ln \mathcal{L} = 15.3$ with respect to the six-component solution. However, its amplitude is only 0.44 m/s and the proximity to the 20-day signal suggests that it is likely to be an artifact. The full results of GLS/MGLS search are shown in Table 3.3.

The rotation period of this low-activity star was given by the analysis of the calcium index by M09 to be 48 days, and was subsequently refined by Lovis et al. (2011b) obtaining a value of 47.2 days. TA13 claims a signature of stellar activity at a period of 320 days. Suárez Mascareño et al. (2015a) found a rotational period of 32 days using photometric observations, whereas Díaz et al. (2016) suggest the existence of signals related to the rotational period from 37 to 47 days. None of these signals attributed to stellar rotation are found in the RV series using our approach, with the exception of the aforementioned 295-day signal, which may be related to the chromospheric emission at 320 days.

Figure 3.2 shows the detectability plot assuming white noise. The maximum detectability region is comprised between 2 and ~ 100 days for amplitudes larger than ~ 2 m/s. The number of observations (345) and their distribution make the dataset remarkably sensitive to low-amplitude signals (~ 20 cm/s) above the 0.1% threshold indicated with a dashed white line.

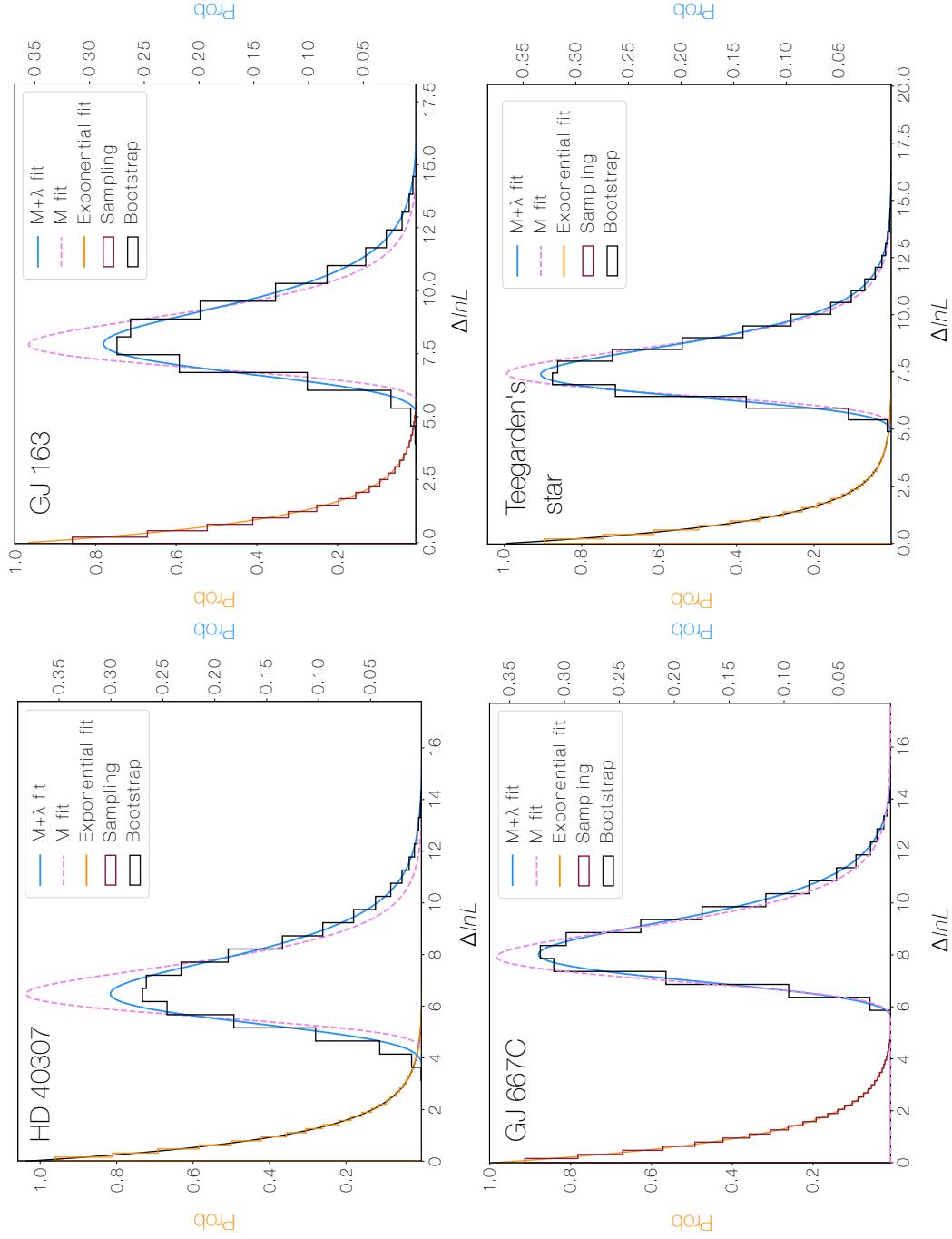


Figure 3.1: FAP determination following the recipe in Section 2.5.1. The plots show the fitted Scargle distribution (with a reduced set of bootstrap samples) by adjusting both M and λ parameters (in blue), and adjusting only the number of independent frequencies, M (in dashed pink). The background histograms are computed employing the classical bootstrap method and shown for comparison purposes.

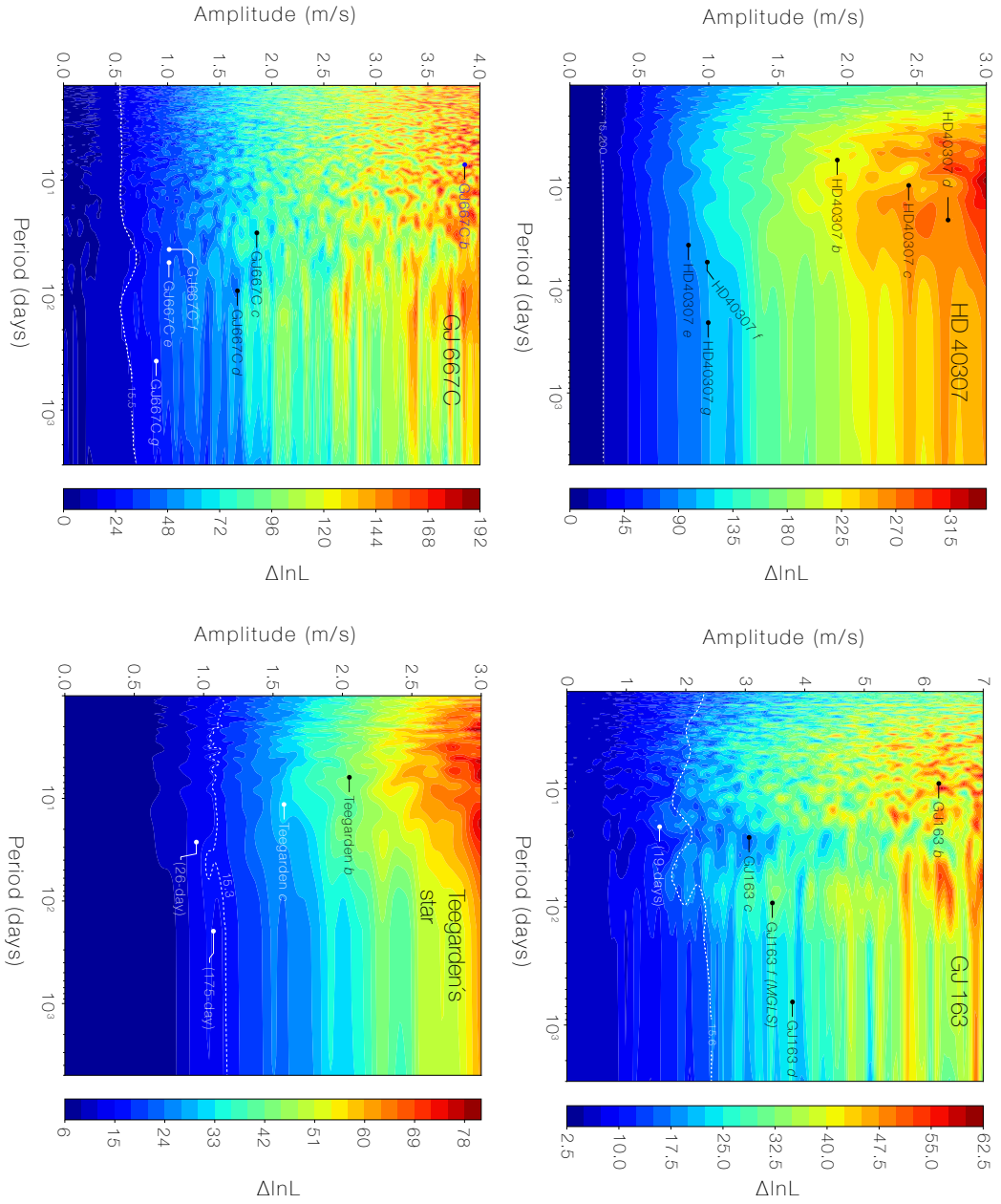


Figure 3.2: Detectability plots for simulated injected signals in tested stars. In each plot the 0.1% FAP is marked with a dashed white curve and the published planets are also located.

Table 3.3: Sequence of signals found in HD 40307; using GLS with prewhitening, and global analysis using MGLS.

GLS (prewhitening)						MGLS			
\mathcal{D} (dim.)	P_{opt} [days]	$K_{P_{opt}}$ [m/s]	$\ln \mathcal{L}$	Jitter [m/s]	$\{P_j\}$ [days]	$\{K_j\}$ [m/s]	$\ln \mathcal{L}$	Jitter [m/s]	
0	(Null model)		-865.58	$2.90^{+0.12}_{-0.11}$			-865.58	$2.90^{+0.12}_{-0.11}$	
1	$9.6209^{+0.0029}_{-0.0028}$	$2.77^{+0.23}_{-0.23}$	-771.16 ✓	$2.17^{+0.09}_{-0.09}$	$9.6209^{+0.0029}_{-0.0028}$	$2.77^{+0.23}_{-0.23}$	-771.16 ✓	$2.17^{+0.09}_{-0.09}$	
2	$20.423^{+0.013}_{-0.013}$	$2.13^{+0.17}_{-0.17}$	-662.94 ✓	$1.53^{+0.07}_{-0.06}$	$9.6164^{+0.0023}_{-0.0023}$ $20.420^{+0.012}_{-0.012}$	$2.58^{+0.13}_{-0.13}$ $2.24^{+0.17}_{-0.17}$	-657.83 ✓	$1.51^{+0.07}_{-0.06}$	
3	$4.3113^{+0.0005}_{-0.0005}$	$1.61^{+0.11}_{-0.11}$	-559.95 ✓	$1.07^{+0.05}_{-0.05}$	$4.3112^{+0.0004}_{-0.0019}$ $9.6188^{+0.0017}_{-0.0017}$ $20.441^{+0.007}_{-0.007}$	$1.8^{+0.1}_{-0.1}$ $2.50^{+0.09}_{-0.09}$ $2.55^{+0.09}_{-0.09}$	-528.61 ✓	$0.95^{+0.05}_{-0.04}$	
4	$51.17^{+0.18}_{-0.16}$	$0.66^{+0.13}_{-0.13}$	-537.69 ✓	$0.99^{+0.05}_{-0.05}$	$4.3109^{+0.0004}_{-0.0004}$ $9.6187^{+0.0016}_{-0.0016}$ $20.420^{+0.006}_{-0.006}$ $51.35^{+0.15}_{-0.14}$	$1.69^{+0.09}_{-0.09}$ $2.49^{+0.08}_{-0.08}$ $2.82^{+0.11}_{-0.11}$ $0.84^{+0.10}_{-0.10}$	-484.98 ✓	$0.80^{+0.04}_{-0.04}$	
5	$34.55^{+0.08}_{-0.08}$	$0.74^{+0.13}_{-0.13}$	-506.62 ✓	$0.88^{+0.05}_{-0.04}$	$4.3110^{+0.0004}_{-0.0004}$ $9.6184^{+0.0015}_{-0.0015}$ $20.424^{+0.006}_{-0.006}$ $34.53^{+0.08}_{-0.09}$ $51.33^{+0.13}_{-0.12}$	$1.71^{+0.08}_{-0.08}$ $2.56^{+0.07}_{-0.07}$ $1.56^{+0.17}_{-0.17}$ $0.61^{+0.10}_{-0.10}$ $0.89^{+0.09}_{-0.09}$	-458.59 ✓	$0.73^{+0.04}_{-0.04}$	

Notes:

[✓ / ✗] Signal accepted/rejected with FAP < 10^{-3}

Table 3.4: Continuation of Table 3.4 (HD 40307)

GLS (prewhitening)						MGLS			
\mathcal{D} (dim.)	P_{opt} [days]	$K_{P_{opt}}$ [m/s]	$\ln \mathcal{L}$	jitter [m/s]	$\{P_j\}$ [days]	$\{K_j\}$ [m/s]	$\ln \mathcal{L}$	jitter [m/s]	
6	$4.534^{+0.001}_{-0.001}$	$0.52^{+0.09}_{-0.09}$	-484.59 ✗ (**)	$0.81^{+0.04}_{-0.04}$	$4.3111^{+0.0004}_{-0.0004}$	$1.75^{+0.08}_{-0.08}$			
					$9.6181^{+0.0015}_{-0.0015}$	$2.54^{+0.07}_{-0.07}$			
					$20.422^{+0.006}_{-0.006}$	$2.69^{+0.10}_{-0.10}$			
					$34.28^{+0.13}_{-0.44}$	$0.54^{+0.11}_{-0.11}$			
					$51.26^{+0.14}_{-0.14}$	$0.99^{+0.09}_{-0.09}$			
					297^{+7}_{-6}	$0.67^{+0.10}_{-0.10}$			
						-437.26 ✓			
						$0.65^{+0.04}_{-0.03}$			
7					$4.3116^{+0.0003}_{-0.0004}$	$1.78^{+0.07}_{-0.07}$			
					$9.617^{+0.001}_{-0.001}$	$2.42^{+0.07}_{-0.07}$			
					$15.64^{+0.02}_{-0.03}$	$0.44^{+0.08}_{-0.08}$			
					$20.421^{+0.006}_{-0.006}$	$2.66^{+0.09}_{-0.09}$			
					$34.44^{+0.10}_{-0.10}$	$0.65^{+0.09}_{-0.09}$			
					$51.38^{+0.12}_{-0.12}$	$0.95^{+0.08}_{-0.08}$			
					295^{+5}_{-5}	$0.76^{+0.10}_{-0.10}$			
						-421.96 ✗)			
						$0.61^{+0.04}_{-0.04}$			

Notes:

[✓ / ✗] Signal accepted/rejected with $FAP < 10^{-3}$.(*) $FAP < 10^{-3}$ but not a planetary signal.(**) $FAP < 10^{-3}$ but wrong period found (very close to previous one).

Table 3.5: Signals found for GJ 163 in the bibliography.

B13					
	GJ 163 <i>b</i>	GJ 163 (<i>e</i>)	GJ 163 <i>c</i>	GJ 163 (<i>f</i>)	GJ 163 <i>d</i>
<i>P</i> (days)	8.631 ± 0.001	19.46 ± 0.02	25.60 ± 0.02	108.4 ± 0.5	603 ± 12
<i>e</i>	0.11 ± 0.04	0.32 ± 0.17	0.08 ± 0.08	0.41 ± 0.15	0.41 ± 0.07
<i>K</i> (m s ^{−1})	6.22 ± 0.26	1.69 ± 0.32	3.07 ± 0.27	1.94 ± 0.38	3.82 ± 0.38
<i>N_{meas.}</i>	153				

Tu13					
<i>P</i> (days)	8.6312 ^{+0.0038} _{−0.0045}		25.632 ^{+0.083} _{−0.075}	125.0 ^{+3.0} _{−1.7}	572 ⁺³¹ _{−41}
<i>e</i>	0.02 ^{+0.16} _{−0.02}		0.01 ^{+0.38} _{−0.01}	0.32 ^{+0.55} _{−0.32}	0.27 ^{+0.27}
<i>K</i> (m s ^{−1})	5.87 ^{+1.18} _{−1.32}		3.54 ^{+1.21} _{−1.34}	3.38 ^{+1.54} _{−2.35}	3.76 ^{+2.11} _{−1.88}
<i>N_{meas.}</i>	51				

Notes:

B13: Bonfils et al. (2013b)

Tu13: Tuomi et al. (2013a)

3.1.2 GJ 163

The young low-mass star GJ 163 (L 229-91, HIP19394) was observed on 51 epochs from October 2003 to August 2010 with the HARPS (High Accuracy Radial velocity Planet Searcher, Mayor et al. 2003) spectrograph, installed at the 3.6 m telescope at La Silla observatory (Chile). Using these data, Tuomi et al. (2013a), Tu13 hereafter, announced the presence of at least three orbiting planets. For their analysis, the authors used the Java-based TERRA pipeline and a Bayesian analysis code. They found up to four significant signals, but a solution composed by three Keplerian was deemed as the most robust one. This result was confirmed by Bonfils et al. (2013b), B13 hereafter, using a more extended dataset composed of 150 HARPS measurements spanning ~ 3000 days. In their analysis, the authors found the same three strong signals reported by Tu13 at periods 8.63, 25.6 and 604 days and made a claim for two additional tentative signals at 19.4 and 109 days, which could not be confirmed. All published signals found in GJ 163 are summarised in Table 3.5. In our analysis, as the main purpose is benchmarking the performance of the code, the dataset from Tu13 is used to further compare the MGLS results with those from B13. Besides, the low number of epochs (51) emphasises the improvement of employing the simultaneous approach.

3.1.2.1 GJ 163 signal search and analysis

Analogously as the previous section, to obtain the FAP value for this dataset we fitted the histogram with 1500 bootstrap samples, yielding a value of $\text{FAP}_{.1\%} = 16.74$ ($\Delta \ln \mathcal{L}$). Figure 3.1 shows the fit for GJ 163.

In Table 3.6, we show the results of GLS (left columns) and MGLS (right columns) searches. Similarly to TA13, we find conclusive evidence for a three-planet system, but

our analysis actually detects the yearly alias of the eccentric Keplerian orbit of the third signal,

$$\left(\frac{1}{600} + \frac{1}{365} \right)^{-1} \approx 227 \text{ days}$$

The GLS prewhitening recipe fails to find even a significant third signal, whereas the MGLS detects the third signal and detects confidently a fourth. When searching for 3 signals, MGLS gives a value of $\Delta \ln \mathcal{L} = 14.52$, which is slightly lower than the expected FAP at 0.1%. However, in the next step when fitting for four signals this metric increases to $\Delta \ln \mathcal{L} = 20.71$, enough to confidently confirm the simultaneous presence of four signals. When searching for five signals, the improvement of the metric is < 10 and the process is stopped.

Generally, the $\ln \mathcal{L}$ values of the simultaneous fits with MGLS are significantly higher and the amplitudes coincide more precisely with those presented by TA13 than those obtained with GLS. The search for sinusoidal variations fails for the eccentric orbit of the 570 day-signal, which might be the reason for the prewhitening process to stop after finding the other two signals.

The analysis of the GJ 163 dataset clearly shows the improvement in detection confidence of the simultaneous MGLS versus the prewhitening GLS in multiplanet datasets. GLS can only detect 2 significant signals, and one of them is probably the 232-day yearly alias of a longer period. Note the tuple of four signals does not contain the former detected alias signal, and this reveals as one of the great advantages of global optimisation strategy, in contrast to local optimisation in which previously detected signals are fixed for the next search.

GJ 163 displays a low activity level and has a rotational period of 61.0 ± 0.3 days found by calcium indicators (Suárez Mascareño et al. 2015b). In the dataset analysed here, we found a single, but significant period, of 65.3 days in the S-index ($\Delta \ln \mathcal{L} = 63.59$), which can be attributed to stellar rotation. However, no signs of rotation-induced periodicities were found in the analysed RV data.

Figure 3.2 shows the detectability plot. The low number of observations (51) and the long timespan (~ 2500 days) imply in a long average separation between measurements (~ 48 days) resulting in a good signal detectability from 3–4 days to the full observational time base for amplitudes above 5 m/s. The confirmed published (Bonfils et al. 2013b) planets *e* and *f* lie below our detection threshold. Our analysis with MGLS gives us an amplitude value for GJ 163 *f* of 3.51 m/s. These results are compatible with those given in Tuomi et al. (2013a) for the detection at 125 days with an amplitude of 3.38 m/s.

Table 3.6: Sequence of periodic signals found in GJ 163; with the GLS prewhitening (left columns), and with the MGLS code (right columns).

GLS (prewhitening)						MGLS			
\mathcal{D} (dim.)	P_{opt} [days]	$K_{P_{opt}}$ [m/s]	$\ln \mathcal{L}$	Jitter [m/s]	$\{P_j\}$ [days]	$\{K_j\}$ [m/s]	$\ln \mathcal{L}$	Jitter [m/s]	
0	(Null model)		-168.25	$6.54^{+0.74}_{-0.62}$			-168.25	$6.54^{+0.74}_{-0.62}$	
1	$8.630^{+0.003}_{-0.003}$	$6.03^{+1.31}_{-1.31}$	-152.17 ✓	$4.7^{+0.6}_{-0.5}$	$8.630^{+0.003}_{-0.003}$	$6.03^{+1.31}_{-1.31}$	-152.17 ✓	$4.7^{+0.6}_{-0.5}$	
2	$232.1^{+1.1}_{-1.3}$	$6.14^{+1.03}_{-1.03}$	-131.61 ✓	$3.0^{+0.4}_{-0.3}$	$8.632^{+0.002}_{-0.002}$ $230.9^{+1.4}_{-1.6}$	$6.08^{+0.65}_{-0.65}$ $6.58^{+1.06}_{-1.06}$	-129.64 ✓	$2.9^{+0.4}_{-0.3}$	
3	$25.60^{+0.03}_{-0.04}$	$2.94^{+0.68}_{-0.68}$	-117.76 ✗	$2.1^{+0.3}_{-0.3}$	$8.632^{+0.002}_{-0.002}$ $25.61^{+0.04}_{-0.04}$ $231.1^{+1.2}_{-1.4}$	$6.07^{+0.55}_{-0.55}$ $3.09^{+0.62}_{-0.62}$ $5.99^{+0.76}_{-0.76}$	-115.12 ✓	$2.0^{+0.3}_{-0.3}$	
4	$7.0^{+0.2}_{-0.5}$	$1.80^{+0.56}_{-0.56}$	-110.80 ✗	$1.9^{+0.3}_{-0.3}$	$8.6315^{+0.0012}_{-0.0012}$ $25.615^{+0.020}_{-0.020}$ $109.06^{+0.47}_{-0.43}$ 584^{+16}_{-14}	$5.71^{+0.37}_{-0.37}$ $3.75^{+0.40}_{-0.40}$ $3.51^{+0.43}_{-0.43}$ $3.96^{+0.68}_{-0.68}$	-94.41 ✓	$0.93^{+0.3}_{-0.3}$	
5					$8.632^{+0.001}_{-0.001}$ $25.60^{+0.02}_{-0.02}$ $32.9^{+0.1}_{-0.1}$ $109.3^{+0.8}_{-0.5}$ 595^{+16}_{-14}	$5.31^{+0.35}_{-0.35}$ $4.10^{+0.34}_{-0.34}$ $1.74^{+0.47}_{-0.47}$ $2.73^{+0.44}_{-0.44}$ $3.71^{+0.48}_{-0.48}$	-84.64 ✗	$0.48^{+0.3}_{-0.3}$	

Notes:
 [✓ / ✗] Signal accepted/rejected with FAP < 10^{-3}
 (**)

Table 3.7: Signals found for GJ667C in the bibliography.

TA13					
	GJ 667C <i>b</i>	GJ 667C <i>c</i>	GJ 667C <i>d</i>	GJ 667C <i>e</i> *	GJ 667C <i>f</i>
<i>P</i> (days)	7.2004 ^{+0.0017} _{−0.0017}	28.140 ^{+0.053} _{−0.065}	91.61 ^{+0.81} _{−0.89}	62.24 ^{+0.55} _{−0.55}	39.026 ^{+0.194} _{−0.211}
<i>e</i>	0.13 ^{+0.10} _{−0.11}	0.02 ^{+0.15} _{−0.02}	0.03 ^{+0.20} _{−0.03}	0.02 ^{+0.22} _{−0.02}	0.03 ^{+0.16} _{−0.03}
<i>K</i> (m s ^{−1})	3.93 ^{+0.42} _{−0.38}	1.71 ^{+0.47} _{−0.47}	1.52 ^{+0.43} _{−0.43}	0.92 ^{+0.48} _{−0.42}	1.08 ^{+0.47} _{−0.46}
	GJ 667C <i>g</i>	GJ 667C <i>h</i> **			
<i>P</i> (days)	256.2 ^{+13.8} _{−7.9}	16.946 ^{+0.051} _{−0.074}			
<i>e</i>	0.08 ^{+0.41} _{−0.08}	0.06 ^{+0.32} _{−0.06}			
<i>K</i> (m s ^{−1})	0.95 ^{+0.48} _{−0.44}	0.61 ^{+0.44} _{−0.49}			
<i>N</i> _{meas.}	171				
FH14					
	GJ 667C <i>b</i>	GJ 667C <i>c</i>			
<i>P</i> (days)	7.200 ^{+0.001} _{−0.001}	28.143 ^{+0.029} _{−0.029}			
<i>e</i>	0.122 ^{+0.078} _{−0.078}	0.133 ^{+0.098} _{−0.098}			
<i>K</i> (m s ^{−1})	3.977 ^{+0.193} _{−0.193}	1.663 ^{+0.291} _{−0.291}			
<i>N</i> _{meas.}	172				

Notes:

A13: Anglada-Escudé et al. (2013)

FH14: Feroz et al. (2014)

(*) Due to the presence of a strong alias, this period could be 53 days instead

(**) Only significantly detected considering circular orbits for all signals

3.1.3 GJ 667C

GJ 667 is a triple star system whose C component orbits the inner pair AB, separated 12.6 AU. The orbital distance of the C component ranges from 56 to 215 AU and reaches an apparent separation of up to 30 arcsec. GJ 667C is an M1.5V red dwarf located at a distance of 6.8 pc with a luminosity of $3 \cdot 10^{-3} L_{\odot}$ being the least massive component of the system with a mass of $0.33 \pm 0.03 M_{\odot}$ (Delfosse et al. 2013). Its proximity to the Solar System, and the fact that is a low mass star makes it an interesting target to search for low mass planets.

The story behind this planetary system around GJ667C is quite controversial due to the distinct results depending on the procedures employed in the analysis of RV datasets. An early work Bonfils et al. (2013a, B13 hereafter), reported the detection of a close planet (GJ667C b) using data from the HARPS spectrograph. Besides, the authors found evidence for two additional signals at 28 and 90 days, the nature of which they were not able to confidently determine. Using this HARPS dataset complemented with additional data gathered with HIRES@KECK and PFS@Magellan, and analysed with TERRA template matching pipeline, Anglada-Escudé et al. (2013, A13 hereafter), found evidence for six planets, or up to seven, if circular orbits were considered. The data

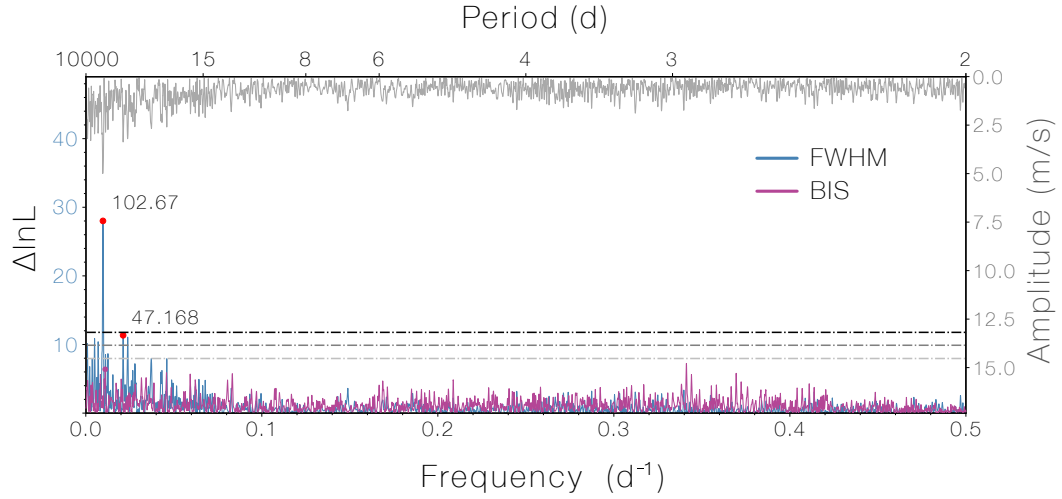


Figure 3.3: FWHM and BIS periodograms for GJ667C. The blue line shows the FWHM, reaching a strong peak at 102 days. The dashed lines are the FAP levels at 10%, 1% and 0.1% (corresponding to FWHM series). The purple line shows the bisector span periodogram, which peaks at 90 days but with insufficient statistical significance.

were re-analysed by means of a sophisticated Bayesian framework in presence of correlated noise of ~ 9 days timescale (Feroz et al. 2014, FH14 hereafter) concluding that the data only yields strong evidence for a planetary system with two components, GJ 667C b and GJ 667C c. Although the analysis in A13 included a correlated noise term following a moving average (MA) scheme, the global fit favored the white noise assumption. When assuming white noise scheme, FH14 obtained a set of five significant signals.

3.1.3.1 GJ 667C signal search and analysis

Analogously in the previous sections, the first step was estimating the FAP. In Figure 3.1 we show the fit using just 1500 random bootstrap samples in the same procedure described in the preceding chapter. The fitted Scargle law to GJ 667C data yields a value of 15.51 ($\Delta \ln \mathcal{L}$) at 0.1% FAP.

Tables 3.8 and 3.9 provide the results of both the sequential and simultaneous MGLS search for the GJ 667C dataset. The prewhitening GLS procedure is not capable of confidently detect more than four signals (the fifth 39-day signal appears with a $\Delta \ln \mathcal{L} = 13.35$). The simultaneous search is able to extract a set of five signals with strong significance, while the sixth signal at 271 days contributes with a $\Delta \ln \mathcal{L} = 13.95$ with respect to the 5-signal set, and therefore we consider statistically insignificant. Signals found in A13 analysis at ~ 256 days and 17 days appear in our search, but with marginally insufficient significance.

Figure 3.3 shows the periodogram for FWHM and BIS activity indices. In the case of the FWHM indicator, a clear and highly significant periodicity is found at 102 days, constituting convincing evidence for the stellar rotation period. The BIS index, does not show significant peaks even though the second highest peak is found at 90 days. A long (\sim hundreds of days) rotation period is expected for old stars (Barnes 2010) and the

102-day peak is in agreement with the 106-day found in Suárez Mascareño et al. (2017) using spectroscopic indicators and photometric monitoring.

Figure 3.2 shows the detectability plot. In the particular case of GJ 667C, the signal detectability is clearly high for periods $\lesssim 100$ days mainly due to the sampling distribution and observation timespan. Confirmed planets GJ 667C *b* and *c* are located clearly above the $\Delta \ln \mathcal{L} = 15.51$ detection threshold (FAP 0.1%) along with the signal at 91 days. Controversial signals found at 53, 39 and 271 days are hardly above the threshold limit. It is worth reminding that in case of considering some correlated noise scheme, the FAP threshold would be higher and these signals would become statistically insignificant.

3.1.4 Teegarden’s Star

Teegarden’s Star was discovered recently by Teegarden et al. (2003) and, located at 3.83 pc, is one of the closest late-type stars. Zechmeister et al. 2019b, ZE19 hereafter, announced the detection of two potentially habitable planets using CARMENES (Quirrenbach et al. 2018, installed at the 3.51 m telescope of the Calar Alto Observatory) spectra observed between January 2016 and March 2019 on 238 occasions in VIS channel and 244 in the NIR channel, with a total timespan of 1136 days. We give an overview of the planetary parameters found by the authors in Table 3.10.

3.1.4.1 Teegarden’s star signal search and analysis

Following the same steps as the other test cases, in Figure 3.1 we show the fit of 1500 bootstrap samples using the procedure described in the preceding chapter. The fitted Scargle law to Teegarden’s yields $\text{FAP}_{.1\%} = 15.03$ ($\Delta \ln \mathcal{L}$).

In this test, we explore a joint analysis of the VIS and NIR data. It is well known that activity signals are likely to be chromatic, i.e. the amplitude of these signals tend to decrease towards longer wavelengths while Keplerian signals are kept invariant both in amplitude and phase (Reiners et al. 2010; Barnes et al. 2011; Jeffers et al. 2014). By combining chromatic RV time-series in which the activity components have differential amplitudes results in a large decrease in statistical significance of the fitted signals. On the contrary, genuine planetary signals strengthen the fit metrics.

Table 3.11 shows a comparison between GLS sequential and MGLS simultaneous searches using only the VIS data. In our analysis, we find again parameter values, especially the amplitude, closer to the published ones using the MGLS rather than the GLS prewhitening (left columns). Differently from the previous two examples, we find quite similar $\ln \mathcal{L}$ values for the different multiplicities, as we go further until the four-planet model, which is not significant in neither method. Significant though is a third signal, which is equally described by the two methods with a 174 day period and a 1.12 m s^{-1} amplitude. This signal, due to the low-activity of Teegarden’s star, might well be related to the rotation period of the star, which the authors assume to be > 100 days, but could also be of Keplerian nature.

A joint VIS+NIR analysis is displayed in Figure 3.5 using a GLS periodogram. Red and blue solid lines show the periodograms considering only VIS and VIS+NIR data,

Table 3.8: Sequence of periodic signals found on GJ 667C; with the GLS prewhitening (middle columns), and with the MGLS code (right columns).

GLS (prewhitening)							MGLS			
\mathcal{D} (dim.)	P_{opt} [days]	$K_{P_{opt}}$ [m/s]	$\ln \mathcal{L}$	Jitter [m/s]	$\{P_j\}$ [days]	$\{K_j\}$ [m/s]	$\ln \mathcal{L}$	Jitter [m/s]		
0	(Null model)		-497.40	$4.34^{+0.26}_{-0.23}$			-497.40	$4.34^{+0.26}_{-0.23}$		
1	$7.2007^{+0.0012}_{-0.0012}$	$3.96^{+0.27}_{-0.27}$	-393.98 ✓	$2.26^{+0.15}_{-0.13}$	$7.2007^{+0.0012}_{-0.0012}$	$3.96^{+0.27}_{-0.27}$	-393.98 ✓	$2.26^{+0.15}_{-0.13}$		
2	$28.08^{+0.03}_{-0.03}$	$1.88^{+0.30}_{-0.30}$	-365.69 ✓	$1.85^{+0.13}_{-0.12}$	$7.199^{+0.001}_{-0.001}$ $28.07^{+0.03}_{-0.03}$	$3.90^{+0.24}_{-0.24}$ $1.96^{+0.30}_{-0.30}$	-364.42 ✓	$1.85^{+0.13}_{-0.12}$		
3	$92.09^{+0.38}_{-0.38}$	$1.49^{+0.20}_{-0.20}$	-334.91 ✓	$1.46^{+0.11}_{-0.10}$	$7.2000^{+0.0008}_{-0.0009}$ $28.098^{+0.024}_{-0.023}$ $92.10^{+0.38}_{-0.39}$	$3.95^{+0.19}_{-0.19}$ $1.83^{+0.20}_{-0.20}$ $1.52^{+0.21}_{-0.21}$	-333.18 ✓	$1.47^{+0.11}_{-0.10}$		
4	$62.30^{+0.25}_{-0.24}$	$0.93^{+0.22}_{-0.22}$	-320.08 ✓	$1.29^{+0.10}_{-0.10}$	$7.1994^{+0.0008}_{-0.0008}$ $28.098^{+0.023}_{-0.022}$ $53.23^{+0.15}_{-0.15}$ $92.10^{+0.32}_{-0.32}$	$3.90^{+0.18}_{-0.18}$ $2.06^{+0.18}_{-0.18}$ $1.07^{+0.20}_{-0.20}$ $1.69^{+0.19}_{-0.19}$	-313.63 ✓	$1.26^{+0.10}_{-0.09}$		
5	$38.87^{+0.09}_{-0.09}$	$0.765^{+0.165}_{-0.165}$	-306.73 ✗	$1.14^{+0.10}_{-0.09}$	$7.1997^{+0.0007}_{-0.0007}$ $28.11^{+0.02}_{-0.02}$ $35.17^{+0.08}_{-0.09}$ $53.26^{+0.12}_{-0.12}$ $91.6^{+0.4}_{-0.4}$	$3.96^{+0.16}_{-0.16}$ $2.08^{+0.18}_{-0.18}$ $0.98^{+0.19}_{-0.19}$ $1.20^{+0.15}_{-0.15}$ $1.56^{+0.17}_{-0.17}$	-296.09 ✓	$1.08^{+0.10}_{-0.09}$		

Notes:

[✓ / ✗] Signal accepted/rejected with $FAP < 10^{-3}$.

Table 3.9: Continuation of Table 3.8 (GJ67C)

GLS (prewhitening)						MGLS			
\mathcal{D} (dim.)	P_{opt} [days]	$K_{P_{opt}}$ [m/s]	$\ln \mathcal{L}$	jitter [m/s]	$\{P_j\}$ [days]	$\{K_j\}$ [m/s]	$\ln \mathcal{L}$	jitter [m/s]	
6	259^{+2}_{-2}	$0.76^{+0.17}_{-0.17}$	-294.40	$1.00^{+0.09}_{-0.09}$	$7.1998^{+0.0006}_{-0.0007}$	$3.89^{+0.14}_{-0.14}$			
					$28.105^{+0.021}_{-0.020}$	$1.88^{+0.16}_{-0.16}$			
					$39.02^{+0.07}_{-0.07}$	$1.08^{+0.18}_{-0.18}$			
					$53.27^{+0.14}_{-0.15}$	$1.03^{+0.14}_{-0.14}$			
					$91.66^{+0.31}_{-0.30}$	$1.66^{+0.13}_{-0.13}$			
					271^{+5}_{-5}	$0.83^{+0.19}_{-0.19}$			
7	$27.84^{+0.05}_{-0.05}$	$0.56^{+0.18}_{-0.18}$	-286.11	$0.93^{+0.09}_{-0.09}$	$7.1998^{+0.0006}_{-0.0007}$	$3.86^{+0.13}_{-0.13}$			
					$16.95^{+0.02}_{-0.02}$	$0.63^{+0.17}_{-0.17}$			
					$28.11^{+0.02}_{-0.02}$	$1.90^{+0.14}_{-0.14}$			
					$39.02^{+0.07}_{-0.07}$	$1.25^{+0.18}_{-0.18}$			
					$53.27^{+0.14}_{-0.15}$	$0.97^{+0.12}_{-0.12}$			
					$91.7^{+0.3}_{-0.3}$	$1.70^{+0.14}_{-0.14}$			
					271^{+5}_{-5}	$0.98^{+0.18}_{-0.18}$	-272.17	$0.82^{+0.09}_{-0.09}$	

Notes:
 [✓ / ✗] Signal accepted/rejected with FAP < 10⁻³.

Table 3.10: Significant periodic signals found for Teegarden’s star in ZE19. These were obtained by MCMC computation of 500,000 samples of the Keplerian parameters posterior distribution.

Parameter	Teegarden b	Teegarden c
P [days]	$4.9100^{+0.0014}_{-0.0014}$	$11.409^{+0.009}_{-0.009}$
K [m/s]	$2.02^{+0.19}_{-0.20}$	$1.61^{+0.19}_{-0.19}$
e	$0.00^{+0.16}$	$0.00^{+0.16}$
Jitter (m/s)	$1.21^{+0.16}_{-0.16}$	
$\ln \mathcal{L}$	-520.98	
$\Delta \ln \mathcal{L}$	69.16	

respectively. The top plot is the amplitude of the fitted sinusoid; the dashed lines are the computed FAP for 1%, 0.1% and 0.01% , in ascending order. Frequencies of the confirmed planet signals are clearly seen, while the two discussed signals appearing in the analysis of VIS channel (26 and ~ 174 days) are less significant (shaded yellow and green bands). Table 3.12 shows the joint VIS+NIR search, which in this case only two significant signals can be found.

The analysis of activity indicators does not shed light on the determination of the rotation period. Only in the periodogram of differential line-width (dWL, Zechmeister et al. 2018) a highly-significant peak is found at 121.9 days. Alternatively, the yearly alias of this period is

$$\left(\frac{1}{122} - \frac{1}{365} \right)^{-1} \approx 183 \text{ days}$$

which is slightly different from the 174-day found in the MGLS analysis of the RVs. The photometric time-series analysed by ZE19 does not yield any concluding information about the stellar rotation period. The origin of the 174-day signal may also be related to the time sampling.

Regarding the nature of the 26-day signal, it appears marginally insignificant when considering a 4-signal solution in the VIS dataset ($\Delta \ln \mathcal{L} = 14.37$). However, this suggesting signal shows an interesting proportionality with the 4.9 and 11.4 signals. Although the MGLS analysis cannot assure the significance of this signal, new observations of Teegarden’s star will contribute to shed light on its existence.

Figure 3.4 shows the two-dimensional periodogram produced with the `--bidim` option of our MGLS code. A zoom-in into the most significant region ($\Delta \ln \mathcal{L} \sim 70$) of the plot is shown. Since the images are mirrored on the diagonal, we show only one half. The colour code gives the $\ln \mathcal{L}$ differences with respect to the null model, showing the highest value at the location of the 4.9 and 11.4-day signals.

Figure 3.2 shows the detectability plot for the VIS channel dataset. In this particular case of Teegarden’s observations, signal detectability is clearly strong for periods $\lesssim 100$ days mainly due to the sampling distribution and observation timespan. Confirmed planets Teegarden b and c are located clearly above the $\Delta \ln \mathcal{L} = 15.3$ detection

Table 3.11: Sequence of signals found in Teggarden's star in VIS dataset, with the GLS prewhitening (left), and with the MGLS code (right).

GLS (prewhitening)							MGLS			
\mathcal{D} [dim.]	P_{opt} [days]	$K_{P_{opt}}$ [m/s]	$\ln \mathcal{L}$	jitter [m/s]	$\{P_j\}$ [days]	$\{K_j\}$ [m/s]	$\ln \mathcal{L}$	jitter [m/s]		
0	(Null model)		-590.13	$2.20^{+0.17}_{-0.16}$			-590.13	$2.20^{+0.17}_{-0.16}$		
1	$4.9094^{+0.0015}_{-0.0014}$	$2.10^{+0.25}_{-0.25}$	-554.76 ✓	$1.64^{+0.16}_{-0.16}$	$4.9094^{+0.0015}_{-0.0014}$	$2.10^{+0.25}_{-0.25}$	-554.76 ✓	$1.64^{+0.16}_{-0.16}$		
2	$11.409^{+0.009}_{-0.009}$	$1.62^{+0.20}_{-0.20}$	-521.69 ✓	$1.15^{+0.17}_{-0.16}$	$4.9096^{+0.0013}_{-0.0013}$	$2.06^{+0.22}_{-0.22}$	-522.19 ✓	$1.17^{+0.16}_{-0.16}$		
3	174^{+3}_{-3}	$1.10^{+0.26}_{-0.26}$	-506.26 ✓	$0.92^{+0.17}_{-0.19}$	$4.9097^{+0.0012}_{-0.0012}$	$2.02^{+0.20}_{-0.20}$	-506.67 ✓	$0.95^{+0.17}_{-0.18}$		
4	$25.99^{+0.07}_{-0.07}$	$0.82^{+0.21}_{-0.21}$	-493.76 ✓	$0.64^{+0.21}_{-0.27}$	$4.9088^{+0.0011}_{-0.0011}$	$2.01^{+0.18}_{-0.18}$	-492.3 ✓	$0.64^{+0.21}_{-0.27}$		
5	$79.3^{+0.7}_{-0.7}$	$0.77^{+0.20}_{-0.20}$	-483.64 ✗	$0.45^{+0.24}_{-0.27}$	$4.9086^{+0.0011}_{-0.0011}$	$1.97^{+0.16}_{-0.16}$	-481.3 ✗	$0.43^{+0.24}_{-0.27}$		
					$11.413^{+0.009}_{-0.009}$	$1.40^{+0.16}_{-0.16}$				
					$26.02^{+0.07}_{-0.07}$	$0.92^{+0.18}_{-0.18}$				
					177^{+2}_{-2}	$1.71^{+0.19}_{-0.19}$				
					378^{+10}_{-8}	$2.0^{+0.5}_{-0.5}$				

Notes:
 [✓/✗] Signal accepted/rejected with $FAP < 10^{-3}$
 In the null model only offset/s and jitter/s are fitted

Table 3.12: Sequence of signals found in Teegarden’s star in VIS+NIR datasets; with the GLS prewhitening (left), and with the MGLS code (right).

GLS (prewhitening)						MGLS			
\mathcal{D} [dim.]	P_{opt} [days]	$K_{P_{opt}}$ [m/s]	$\ln \mathcal{L}$	Jitter [m/s]	$\{P_j\}$ [days]	$\{K_j\}$ [m/s]	$\ln \mathcal{L}$	Jitter [m/s]	
0	(Null model)		-1402.98	$2.20^{+0.17}_{-0.16}$ $3.35^{+0.49}_{-0.49}$			-1402.98	$2.20^{+0.17}_{-0.16}$ $3.35^{+0.49}_{-0.49}$	
1	$4.9090^{+0.0013}_{-0.0013}$	$2.14^{+0.22}_{-0.22}$	-1359.33 ✓	$1.64^{+0.16}_{-0.16}$ $2.86^{+0.52}_{-0.55}$	$4.9090^{+0.0013}_{-0.0013}$	$2.14^{+0.22}_{-0.22}$	-1359.33 ✓	$1.64^{+0.16}_{-0.16}$ $2.86^{+0.52}_{-0.55}$	
2	$11.407^{+0.008}_{-0.008}$	$1.68^{+0.20}_{-0.20}$	-1318.47 ✓	$1.16^{+0.16}_{-0.17}$ $2.20^{+0.59}_{-0.72}$	$4.9093^{+0.0012}_{-0.0012}$ $11.407^{+0.008}_{-0.008}$	$2.09^{+0.19}_{-0.19}$ $1.66^{+0.20}_{-0.20}$	-1318.67 ✓	$1.17^{+0.16}_{-0.17}$ $2.22^{+0.60}_{-0.71}$	
3	$172.7^{+1.1}_{-1.1}$	$0.90^{+0.22}_{-0.22}$	-1307.02 ✗	$0.94^{+0.17}_{-0.18}$ $2.41^{+0.57}_{-0.66}$	$4.9095^{+0.0011}_{-0.0011}$ $11.409^{+0.008}_{-0.008}$ $172.6^{+1.2}_{-1.1}$	$2.05^{+0.19}_{-0.19}$ $1.62^{+0.19}_{-0.19}$ $0.90^{+0.22}_{-0.22}$	-1307.03 ✗	$0.95^{+0.17}_{-0.18}$ $2.45^{+0.56}_{-0.64}$	
4	$26.00^{+0.08}_{-0.07}$	$0.76^{+0.19}_{-0.19}$	-1296.59 ✗	$0.69^{+0.21}_{-0.26}$ $2.50^{+0.55}_{-0.61}$	$4.9093^{+0.0011}_{-0.0011}$ $11.415^{+0.008}_{-0.008}$ 379^{+3}_{-3} 175^{+1}_{-1}	$2.01^{+0.18}_{-0.18}$ $1.57^{+0.18}_{-0.18}$ $1.81^{+0.31}_{-0.31}$	-1295.23 ✗	$0.76^{+0.19}_{-0.23}$ $2.29^{+0.60}_{-0.68}$	

Notes:

[✓/✗] Signal accepted/rejected with $FAP < 10^{-3}$
 In the null model only offset/s and jitter/s are fitted

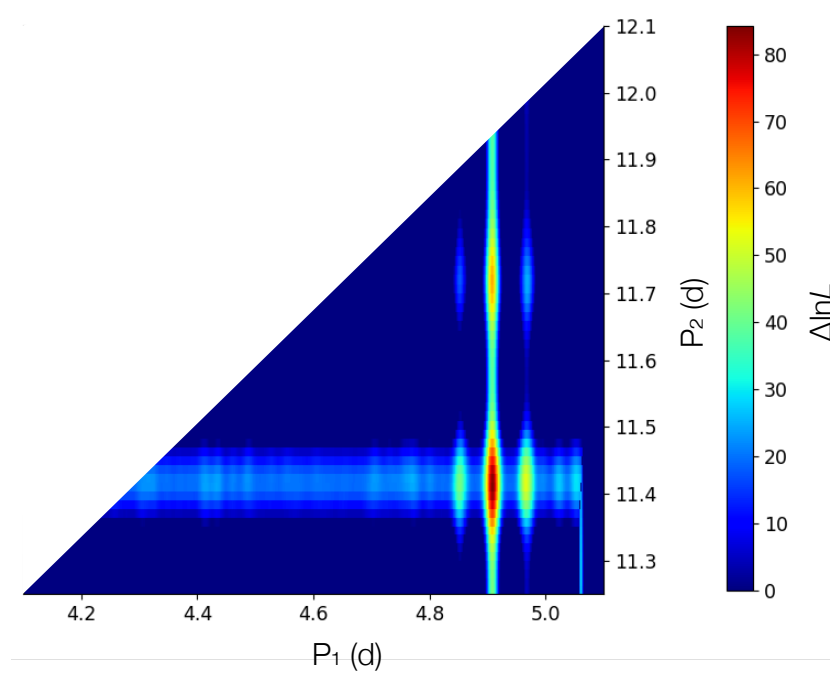


Figure 3.4: Teegarden's star 2-dimensional MGLS plot.

threshold (FAP 0.1%), while prospective signals at 26 and 174 days lie marginally below this threshold.

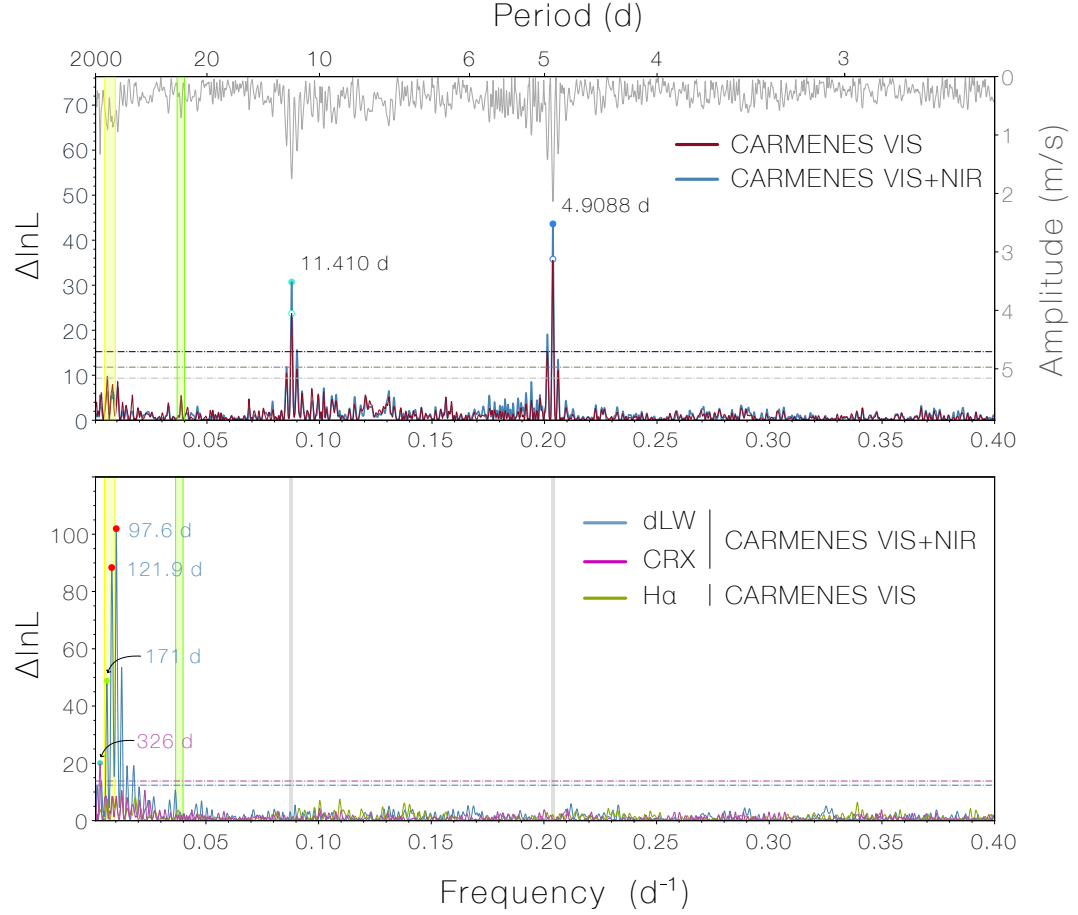


Figure 3.5: GLS periodogram for Teegarden's star using CARMENES-VIS measures (red) and VIS+NIR (blue). While the signals for the two candidate planets are clearly strengthened when the NIR observations are considered, the rest of the periodicities remain with roughly the same $\Delta \ln \mathcal{L}$. Yellow and green shaded bands mark the additional potentially significant signals. The top panel FAP lines are 10%, 1% and 0.1%; in bottom panel both FAP lines are 0.1%.

Table 3.13: Observational dataset of Barnard’s star.

Dataset	epochs	timespan (d)	$\langle \text{separation} \rangle^{(a)}$ (d)
1/ APF (Automated Planet Finder)	43	973.3	22.6
2/ CARMENES	201	642.5	3.2
3/ HARPS-N (1st run)	40	1196.8	29.9
4/ HARPS-N (2nd run)	69	92.8	1.3
5/ HARPS-N (3rd run)	118	2234.0	18.9
6/ HIRES	186	5904.7	31.7
7/ PFS (Planet Finder Spectrograph)	39	1824.0	46.8
8/ UVES	75	2357.8	31.4

Notes:

(a) Mean separation of measurements.

3.1.5 Barnard b: A super–Earth orbiting at the snow–line of Barnard’s star

Gl 699 (also known as Barnard’s star) has been observed historically for years using various methods to probe the presence of companions. Early astrometric ground observations of van de Kamp (1982) for 44 consecutive years, and using HST (Hubble Space Telescope) Benedict et al. (1999) found no evidence for any planet around Barnard’s star. Since then, the 4th star in proximity to the Sun did not show any significant sign of being orbited by an exoplanet despite high–precision RV measurements taken since the end of 90’s in HIRES@Keck and PSF spectrographs.

Recently, the analysis of a vast dataset of 771 Doppler measurements obtained with six different spectrographs in eight seasons for more than 20 years, allowed the announcement of a cold super–Earth candidate orbiting at the snow–line (Ribas et al. 2018).

3.1.5.1 Time-series analysis and discussion

The aim of this section is test the MGLS code with a large collection of data composed by eight independent datasets comprising a long timespan. A multiset analysis implies that each dataset has to be fitted globally with a specific offset and jitter. We analysed the eight datasets considering a white noise model for multiple periodicities simultaneously. One–dimensional log–likelihood periodograms are shown in Figure 3.6 which clearly shows a high significance peak at 233 days. Additionally, there are other significant peaks with longer periods as illustrated in the 2D–periodogram shown in Figure 3.7.

Tables 3.14 and 3.15 display the results of the sequential (left), and simultaneous (right) periodicity searches for Barnard’s star. The first 2–signal tuple shows no differences in both approaches. A strong significant signal at 233 days, and a second long–term signal at 1900 days, both of them close to 1 m/s amplitude. The difference in $\ln \mathcal{L}$ between GLS and MGLS is very small, -1698.28 vs. -1696.78 . The analysis is favourable to the inclusion of a linear term in the global fit, improving the metrics in $\Delta \ln \mathcal{L} \sim 1.8$

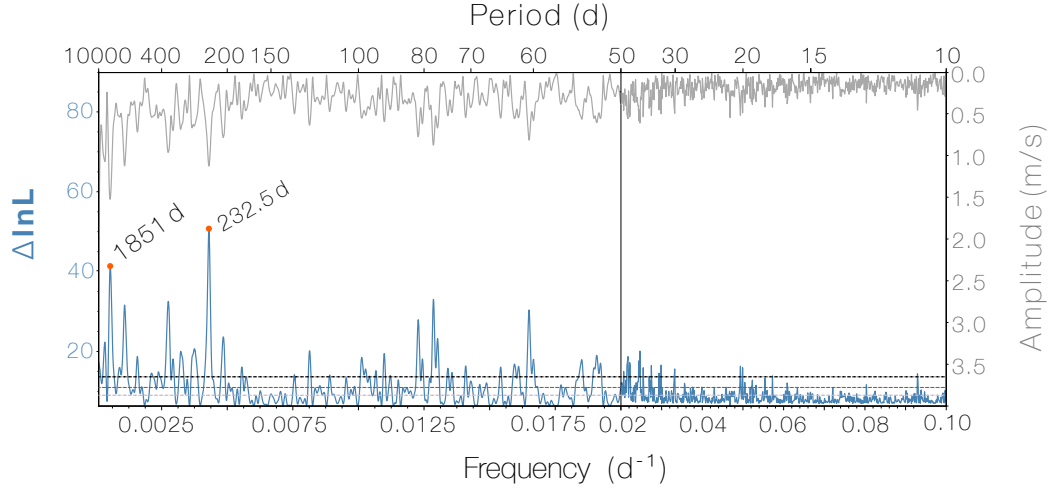


Figure 3.6: GLS periodogram for the Barnard's star data.

with respect to the model with no trend. By exploring higher dimensionalities, for $\mathcal{D} = 3$ a period of 122 days appears in both GLS/MGLS approaches despite its low amplitude of 0.62 m/s. Toledo-Adrón et al. (2019) (T19) found a rotation period of 145 ± 15 days through the analysis of spectroscopic activity indexes. This period is consistent with the expected rotation period according to the low activity level of the star and previous claims. Our detection of 122 days may be associated to stellar rotation as differential rotation may be responsible for the different peaks found in GLS periodogram between 130 to 180 days, as T19 mentioned. Moreover, the search with $\mathcal{D} = 3$ signals reveals a consistent long-term modulation at 1900 days that could arise from a magnetic activity cycle or from a more distant planetary object.

Figure 3.7 shows a MGLS bidimensional periodogram. The colour scale is the improvement of the $\Delta \ln \mathcal{L}$ with respect to the null model (no oscillatory components). Values of $\Delta \ln \mathcal{L} > 18.1$ correspond to a significant detection (FAP < 0.1%). The highest likelihood value corresponds to periods of 233 days and 1890 days ($\Delta \ln \mathcal{L} > 71$), but any combination of 233 days with periods longer than 2500 days yields $\Delta \ln \mathcal{L} > 65$ and thus are statistically equivalent. The ability to have a wide perspective across the plane is the great advantage of the presented bidimensional periodograms: although the optimal solutions are unique, the MGLS plot allows us to explore other quasi-optimal configurations for which their figure of merit is just marginally below the optimal one, or even statistically equivalent. In this particular analysis, R18 selected as the optimal configuration the 233 days signal with a long-term period of 6600 days by forcing the long term signal to be > 4000 days, from the use of an alternative method to determine the relative offsets taking advantage of the overlapped observations in different datasets.

Table 3.14: Sequence of periodic signals found in **Barnard's star**, with the GLS prewhitening (left columns), and with the MGLS code (right columns).

GLS (prewhitening)					MGLS			
\mathcal{D} (dim.)	P_{opt} (days)	$K_{P_{opt}}$ (m/s)	$\ln \mathcal{L}$	jitters (m/s)	$\{P_j\}$ (days)	$\{K_j\}$ (m/s)	$\ln \mathcal{L}$	jitters (m/s)
1	$232.39^{+0.45}_{-0.44}$	$1.11^{+0.12}_{-0.12}$	$-1725.16 \checkmark$	a $2.83^{+0.50}_{-0.44}$	$232.39^{+0.45}_{-0.44}$	$1.11^{+0.12}_{-0.12}$	$-1725.16 \checkmark$	a $2.83^{+0.50}_{-0.44}$
				b $1.83^{+0.15}_{-0.14}$				b $1.83^{+0.15}_{-0.14}$
				c $2.03^{+0.31}_{-0.26}$				c $2.03^{+0.31}_{-0.26}$
				d $1.18^{+0.19}_{-0.18}$				d $1.18^{+0.19}_{-0.18}$
				e $1.00^{+0.14}_{-0.14}$				e $1.00^{+0.14}_{-0.14}$
				f $2.52^{+0.19}_{-0.18}$				f $2.52^{+0.19}_{-0.18}$
				g $1.21^{+0.34}_{-0.34}$				g $1.21^{+0.34}_{-0.34}$
				h $2.33^{+0.24}_{-0.21}$				h $2.33^{+0.24}_{-0.21}$
2	1919^{+13}_{-13}	$1.22^{+0.17}_{-0.17}$	$-1698.28 \checkmark$	a $2.81^{+0.51}_{-0.44}$	$232.50^{+0.50}_{-0.50}$	$0.94^{+0.13}_{-0.13}$	$-1696.78 \checkmark$	a $2.80^{+0.50}_{-0.44}$
				b $1.74^{+0.15}_{-0.13}$				b $1.75^{+0.15}_{-0.14}$
				c $1.68^{+0.27}_{-0.23}$				c $1.62^{+0.27}_{-0.24}$
				d $1.18^{+0.19}_{-0.17}$				d $1.18^{+0.19}_{-0.17}$
				e $0.89^{+0.14}_{-0.14}$				e $0.88^{+0.14}_{-0.14}$
				f $2.41^{+0.19}_{-0.18}$				f $2.45^{+0.20}_{-0.19}$
				g $1.32^{+0.34}_{-0.34}$				g $1.29^{+0.35}_{-0.34}$
				h $2.16^{+0.23}_{-0.21}$				h $2.14^{+0.24}_{-0.21}$

Notes:

a) APF (Automated Planet Finder); b) CARMENES; c) HARPS-N (1st epoch); d) HARPS-N (2nd epoch); e) HARPS-N (3rd epoch); f) HIRES; g) PFS (Planet Finder Spectrograph); h) UVES.
 [✓ / ✗] Signal accepted/rejected with FAP < 10⁻³.

Table 3.15: Continuation of Table 3.14

GLS (prewhitening)					MGLS				
\mathcal{D} (dim.)	P_{opt} [days]	$K_{P_{opt}}$ [m/s]	$\ln \mathcal{L}$	Jitters [m/s]	$\{P_j\}$ [days]	$\{K_j\}$ [m/s]	$\ln \mathcal{L}$	Jitters [m/s]	
3	122.62 $^{+0.45}_{-0.44}$	0.55 $^{+0.15}_{-0.15}$	-1684.17 ✓	<i>a</i> 2.87 $^{+0.52}_{-0.46}$	233.28 $^{+0.52}_{-0.51}$	0.98 $^{+0.11}_{-0.11}$	-1681.06 ✓	<i>a</i> 2.86 $^{+0.50}_{-0.43}$	
				<i>b</i> 1.73 $^{+0.14}_{-0.14}$	122.62 $^{+0.19}_{-0.18}$	0.62 $^{+0.15}_{-0.15}$		<i>b</i> 1.72 $^{+0.15}_{-0.14}$	
				<i>c</i> 1.61 $^{+0.25}_{-0.22}$	1917 $^{+40}_{-38}$	1.20 $^{+0.17}_{-0.17}$		<i>c</i> 1.45 $^{+0.27}_{-0.23}$	
				<i>d</i> 1.19 $^{+0.20}_{-0.17}$				<i>d</i> 1.21 $^{+0.19}_{-0.17}$	
				<i>e</i> 0.79 $^{+0.14}_{-0.14}$				<i>e</i> 0.79 $^{+0.14}_{-0.14}$	
				<i>f</i> 2.32 $^{+0.20}_{-0.18}$				<i>f</i> 2.38 $^{+0.20}_{-0.18}$	
				<i>g</i> 1.43 $^{+0.34}_{-0.32}$				<i>g</i> 1.44 $^{+0.35}_{-0.32}$	
				<i>h</i> 2.05 $^{+0.22}_{-0.20}$				<i>h</i> 1.97 $^{+0.23}_{-0.20}$	
4	288.86 $^{+1.16}_{-1.13}$	0.48 $^{+0.15}_{-0.15}$	-1673.86 ✗	<i>a</i> 2.91 $^{+0.51}_{-0.45}$	233.19 $^{+0.47}_{-0.47}$	1.16 $^{+0.10}_{-0.10}$	-1669.64 ✗	<i>a</i> 2.70 $^{+0.50}_{-0.43}$	
				<i>b</i> 1.68 $^{+0.15}_{-0.13}$	179.23 $^{+0.56}_{-0.59}$	0.62 $^{+0.16}_{-0.16}$		<i>b</i> 1.67 $^{+0.15}_{-0.14}$	
				<i>c</i> 1.48 $^{+0.25}_{-0.22}$	2007 $^{+92}_{-83}$	1.21 $^{+0.19}_{-0.19}$		<i>c</i> 1.35 $^{+0.26}_{-0.22}$	
				<i>d</i> 1.21 $^{+0.18}_{-0.17}$	2709 $^{+180}_{-151}$	0.95 $^{+0.16}_{-0.16}$		<i>d</i> 1.15 $^{+0.19}_{-0.17}$	
				<i>e</i> 0.77 $^{+0.14}_{-0.14}$				<i>e</i> 0.75 $^{+0.14}_{-0.15}$	
				<i>f</i> 2.34 $^{+0.20}_{-0.18}$				<i>f</i> 2.34 $^{+0.19}_{-0.18}$	
				<i>g</i> 1.31 $^{+0.35}_{-0.33}$				<i>g</i> 1.18 $^{+0.35}_{-0.36}$	
				<i>h</i> 1.93 $^{+0.21}_{-0.19}$				<i>h</i> 2.12 $^{+0.24}_{-0.20}$	

Notes:

a) APF (Automated Planet Finder); b) CARMENES; c) HARPS-N (1st epoch); d) HARPS-N (2nd epoch); e) HARPS-N (3rd epoch); f) HIRES; g) PFS (Planet Finder Spectrograph); h) UVES.

[✓/✗] Signal accepted/rejected with FAP < 10⁻³.

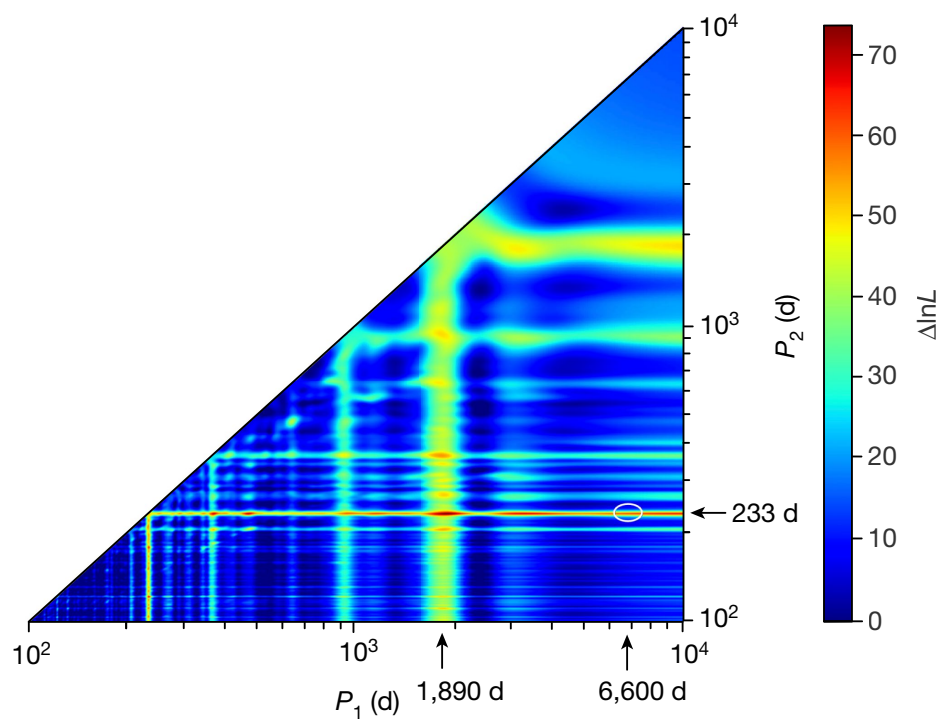


Figure 3.7: Application of a bidimensional MGLS code for multiset period search in Barnard's star data. The color bar shows the increase in log-likelihood of a 2-period solution versus a null model consisting of an offset and a jitter term. It is clearly shown a 233-day signal, that combined with a band of longer periods gives high-significant solutions. The maximum value of $\Delta \ln \mathcal{L} = 71.0$ is found in 233 and 1890 days.

3.2 Conclusions

In summary, the MGLS code tested in this chapter offers a number of improvements with respect to the unidimensional algorithm, namely

- Provides a generalisation of the GLS algorithm for systems where multisignal content is expected.
- The MGLS behaviour in presence of noise is particularly improved when recovering low-amplitude signals that may generate constructive interferences among them, and arise spurious signals by convolution with the sampling.
- The code can fit multidimensional components simultaneously, in contrast to most extended hierarchical designs. The need for global simultaneous fit in opposition to global and local fit is supported by the fact that there are many different configurations for which the difference in $\ln \mathcal{L}$ is small. In case a global fit is performed with frequencies found through the iterative process, the optimal configuration can be missed or misinterpreted.
- The improvement in detectability is demonstrated in the testbed examples. For all test cases, MGLS either detect more signals than the prewhitening approach, or the statistical significance is higher.
- The use of log-likelihood statistic permits the inclusion of a jitter parameter in the optimisation process. This accounts for the underestimation of data uncertainties and non-sinusoidal signals in data.
- Offers the functionality to produce 2D-periodogram plots, where a global view of the frequency plane may help to identify regions with multiple peaks of similar power.
- The algorithm is simple and robust since it is based on fitting a linear combination of sinusoidal functions. Therefore, the execution time of the code remains reasonable in most applications.

Part III

Modelling active stars

Chapter 4

The StarSim 2 project: Fast modelling code of heterogeneous photospheres of active stars

4.1 StarSim modelling code

StarSim is a code package for simulating rotational modulating effects caused by evolving dark spots and bright faculae on the stellar photosphere. The first version of the StarSim code was presented in Herrero et al. (2016) as a tool to investigate the effects of stellar activity arising from starspots on spectroscopic and spectrophotometric observations by simulating full spectra.

Atmosphere models are used to model the spectral distribution of the heterogeneous surface including quiet photosphere, spots, and faculae considering all significant effects such as limb-darkening for the photosphere and spots, and limb brightening for faculae. The integrated spectrum of the visible face, which is the quantity that compares with the observed flux of unresolved sources, is computed by summing the contributions of all small surface elements on a grid, each one with their projective elements according to the position on the sphere, the spectral distribution, which depends on the type of surface, and the position on the grid to compute those limb effects. Then, time-series spectra can be obtained covering a given time interval, and finally convolving it with a specific transfer function (filter) to account for photometric, atmospheric passbands or instrumental response profile. Hence, spectrophotometric or spectroscopic data can be simulated to study how activity works and what are its effects on planetary transits or radial velocities, and ultimately enhance the detection capabilities or small signals, optimise exoplanet observational campaigns or other stellar activity numerical experiments.

The StarSim software consists of three fundamental parts, *i*) The synthetic spectral energy distributions, which characterise the radiative surfaces, *ii*) The spot model,

which describes the distribution of active regions on the stellar surface and its restrictions (i.e. circular shape, non-overlapping condition, ...), and *iii*) A set of physical parameters that characterise the star and its active regions.

4.2 StarSim 2: A fast implementation of a parametric model

The original StarSim version (Herrero et al. 2016) performs the integration of all surface elements of a user-defined size. This implementation imposes an algorithmic complexity —proportional to time consumption— of $\sim \mathcal{O}^2$ to scan across the stellar surface, setting a limit on the grid resolution in order to ensure an affordable execution time.

One of the goals of the new StarSim¹2 version is to optimise the algorithm to dramatically decrease the running time by a factor ~ 100 to make the computation of the inverse problem, possible (developed in Section 4.4). Inverting light curves to derive spot maps is a hard computational task. Typically thousands of objective function evaluations are needed to find approximate solutions for the spot maps as they may contain dozens of spot elements. Each spot has five adjustable parameters (see Section 4.2.1.3) and the total number of variables to fit may reach hundreds. Furthermore, when performing a full dataset inversion, where both spot map and a stellar parameter set are obtained, \mathcal{S}_θ , the computational effort is almost prohibitive.

In StarSim 2, a number of approximations have been implemented to allow the computation of data inversion. Namely, the grid approach to surface integration is no longer used, and the spots are considered as a single surface element where their center position is used to calculate the physical properties, i.e. limb-darkening coefficient, geometric projection. Hence, the complexity of the spot loop is reduced to a linear order. As a counterpart, we have to work at the limit of small spots to ensure that the discontinuities arising from the appearance of spots from the limb, which lead to jumps on the series, are negligible.

4.2.1 Physical model

4.2.1.1 Atmosphere models

The code simulates a spot model of a heterogeneous surface by integration of the immaculate photosphere with circular active regions. Each of these active regions contains the spectrum of its type (quiet photosphere, ph, spots, sp or faculae, fc). The spectral distributions of each type of surface are given by BT-Settl synthetic spectra (Allard et al. 2013) generated from the Phoenix code (Hauschildt et al. 1999). The parameters that characterise the spectra for StarSim 2 are the effective temperature of the three types of surface, $T_{\text{sp}} < T_{\text{ph}} < T_{\text{fc}}$, the surface gravity, $\log g$ and several values for α enhancement and metallicity.

BT-Settl spectra are available for $2,600 \text{ K} < T_{\text{eff}} < 70,000 \text{ K}$ and $3.5 < \log g < 5.0$ and for several metallicities. These models include the effects induced by convection

1. The StarSim 2 code is publicly available at github.com/rosich/starsim-2

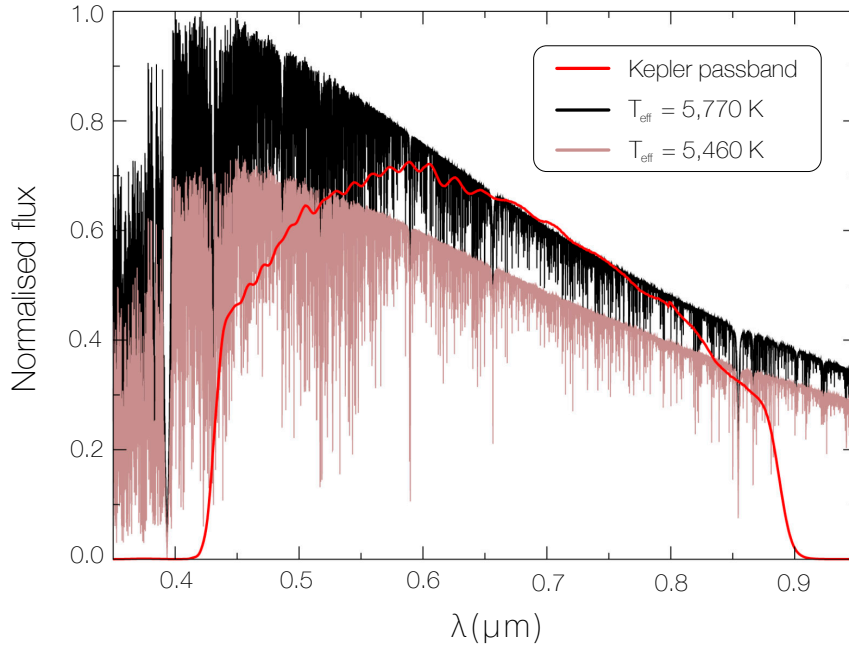


Figure 4.1: BT-Settl examples of synthetic spectra. In black is represented a Sun-like star, $T_{\text{eff}} = 5770$ K and $\log g = 4.5$ compared with a spectra $\Delta T = 310$ K lower, in brown. The transfer function of Kepler photometry system is drawn in red. Note the decreasing flux difference between both spectra having a temperature with wavelength. As moving towards longer wavelengths, the relative effect of spots in the flux is less pronounced.

as described in Vitense (1953) using 3D models which have a significant impact on line profile and therefore on the activity-induced radial velocity.

Even though BT-Settl spectra have a sufficiently high resolution to compute the flux distribution of the surface elements ($\gtrsim 0.05$ Å) this is not high enough to derive accurate RV values because the spectral lines are not well resolved. To simulate Doppler RVs, StarSim uses high-resolution spectra by Husser et al. (2013), also generated with the Phoenix code. These spectra have a resolution of $R = 500,000$ in the VIS/near IR and $R = 100,000$ in the IR, with effective temperature coverage of $2,300 \text{ K} < T_{\text{eff}} < 12,000 \text{ K}$; $0.0 < \log g < +6.0$, $-4.0 \leq [Fe/H] \leq +1.0$ and $0.2 \leq [\alpha/Fe] \leq +1.2$.

Figure 4.1 plots the BT-Settl spectral distributions of a solar-type star ($T_{\text{eff}} = 5,770$ K) and one 310 K cooler, which could correspond to the temperature of a spot. A large flux difference can be seen at short wavelengths, while towards the IR this difference narrows. As a consequence, the spot effects are less significant as the wavelength increases.

4.2.1.2 Parameter system

Simple physical systems are amenable to be modelled with parametric models. When the laws that govern the behaviour of the system are well known, the way to study, understand and make predictions is to fit a physical function that relates the input variables to the output ones. In parametric modelling, this functional dependence that

relates the input to the output is defined through a set of parameters, whose number is fixed and a-priori known. The distributions of the parameters are unknown and have to be estimated based on the available input–output observations (e.g. Theodoridis 2020).

StarSim is a parametric model in which those parameters are strongly motivated on the basis of a physical system of rotational spots on a simulated photosphere. In Table 4.1 lists the most important parameters contained in `starsim.conf` configuration file which serves as input.

In contrast to parametric modelling, there are the so-called nonparametric methods, a well-known example being machine learning techniques based on artificial neural networks. These models ignore the underlying laws that explain the relations between the input–output, but are data-driven, in the sense that they rely on a —usually large— training dataset to *train* the network on how to adjust their hyperparameters to correctly predict the output. If the trained network succeeds in predicting another testing dataset, the neural network model is well trained and will be able to predict new data initially unknown to the model. However, these approaches do not have, in general, any knowledge of the physics involved in the studied phenomena and, therefore, the calculated hyperparameters are not connected to any physical meaning.

The stellar activity problem is complex since many physical phenomena with multiscale effects are at play. Nevertheless, simple spot models have revealed effective in explaining a large fraction of the variance coming from rotational-induced variability (Lanza et al. 2011; Kipping 2012). Such rotational signals are often referred to as quasi-periodic because even though the star rotates at uniform speed, the evolutive properties of spots and the potential existence of differential rotation create sophisticated profiles in light curves or in RV signals. At this point, the parametrisation of a complete model of spotted stars may lead to include many parameters, some of them difficult to justify and many of them, totally unknown a-priori.

One of the most controversial parameters in physical spot modelling is the number of active regions to be adopted. The shape of spots, their uniform temperature inside them, and the number of them represent a strong regularisation method for this type of models. At first, one could think that the fit is better as more active regions are introduced. However, in the following chapters we demonstrate that if the spot size–temperature degeneracy is broken by the availability of color information in photometry, the absolute filling factor, i.e. the total coverage of spots, is constrained and a minimum number of spots is needed to assure an optimal filling factor. Hence, in case of considering a large number of spots, the optimiser will make them small and clustered together to mimic the behaviour of a larger active region.

4.2.1.3 *Spot model*

The original version of StarSim works by dividing the photosphere into a grid of surface elements, typically set with an angular size of 1° . Each of these elements is assigned to represent either photosphere, spot, or facula and has different spectral distribution and specific limb-darkening/brightening behaviour. Using this grid scheme allows StarSim to build spots of arbitrary shape, even though in the implemented version spots are assumed to be circular to optimise the numerical effort.

Table 4.1: Set of most relevant StarSim parameters

Parameter	Description
Spectral range [nm]	Wavelength range of the spectra
$\log g$	Stellar surface gravity
T_{eff}	Photosphere effective temperature [K]
ΔT_{sp}	Spot temperature contrast [K]
ΔT_{fc}	Facula, temperature contrast [K]
Q	facula-to-spot surface ratio
P_{rot}	Stellar rotation period [days]
R_*	Stellar radius [R_{\odot}] (only for RV)
i	Rotation axis inclination [°]
δ_0	Differential rotation constant (see Section 4.2.1.3)
B, C	Differential rotation parameters (Beck 2000)
Spot mean lifetime	mean life of spots [days]
σ lifetime	standard deviation of spot lifetime [days]
Spot growing factor	slope of the growing/decaying linear law [°/day]

In the first version of the code the stellar surface is divided into a grid and the code iterates over each cell. This resulting in a high time consumption that in StarSim 2 version we managed to solve by considering the spots as a single cell of a given area with those physical properties computed at the centre. This approach boosts the efficiency of the code by a factor of two orders of magnitude, but at the expense of some drawbacks. The most important is the sudden apparition and disappearance of the spots to the stellar limb when the central coordinates become visible. If the spots are large, this creates visible jump discontinuities on the observables that may be prevented by limiting the maximum spot size. In the simulations carried out in this Thesis, the spot size limit was set to 8° . With this restriction, the spot maps may contain a large number of small active elements that are able to cluster them to create arbitrary shapes to properly model the observables.

The fact that spots are imposed to be circular and have a uniform temperature, constitutes a strong restriction that breaks the degeneracy found in a lattice-type surface, in where a large number of small pixels can independently be tagged as photosphere or spots. In gridded surface models, a regularization prior must be implemented to avoid inconsistent solutions e.g. the maximum entropy (Lanza et al. 2011), which tend to favour surfaces with the lowest number of dark pixels (Luger et al. 2021b).

An active region consists of a central dark spot of radius R_{sp} with a circular ring of faculae having a total radius R_{fc} . We define the amount of faculae, Q , as the ratio of facula area over spot area and it is assumed to be a common parameter for all active regions. Hence,

$$Q = \left(\frac{R_{\text{fc}}}{R_{\text{sp}}} \right)^2 - 1$$

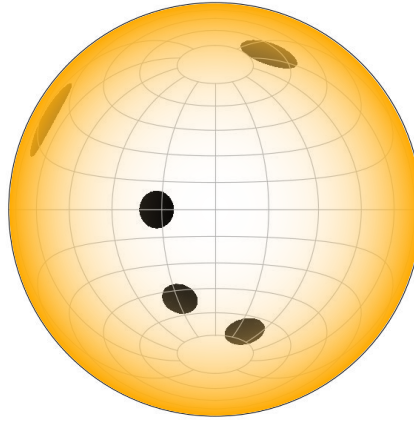


Figure 4.2: Representation of the StarSim spot model on a Lambert projection. The system is composed of five circular spots as defined in the model. The shaded yellow disk emulates the limb-darkening effect for spots near to the edge.

The choice of this definition of circular spots has as the main advantage the computational efficiency in comparison to the grid surface model. On the other hand, the fact that spots are defined as a single patch on the stellar surface, the spots appear and disappear abruptly from the limb once the central point is visible. To circumvent this issue, an approximation of small spots must be used, and an upper limit is set. In the simulations described in the next chapters, a limit of 8° is adopted.

The number of spots is a fixed parameter during the simulation. Furthermore, spots evolve with time with a linear growing phase, followed by a time interval of constant size, and finally a linear decrease until the spot vanishes. The mean life of spots and their standard deviation are parameters that affect all active elements. This scheme of appearing and disappearing spots at a certain rate produces surprisingly complex activity patterns even when considering only a few spots.

The StarSim model does not distinguish between umbra and penumbra in a star spot, but a mean temperature contrast is assumed for each active element and associated to the average of the different temperature components. Modelling star-spots as circular spots surrounded by a ring concentric faculae corona has been used in previous modelling codes (Lanza et al. 2003).

Figure 4.2 shows an example of a spot map plotted in a Lambert geographic projection having only dark spots. The spot map file contains a preselected number of spots, each one having five tunable parameters, namely initial time, lifetime, colatitude, longitude and radius. The spot file parameters can be assigned randomly, *ad-hoc* (set by hand), or can be the product of the inversion process of the observational time-series.

Differential rotation While stellar rotation periods have been measured for quite a number of stars, differential rotation is more complicated to assess because it cannot be easily measured. Reinhold et al. (2013) study more than 24 000 Kepler field active stars, and 18 000 were found to show some evidence of differential rotation. Therefore, high-precision observational data reveal that differential rotation is a very common

characteristic in spotted stars. This is in good agreement with the theory of magnetic field generation in the Sun that could be extended to other stars.

Looking at a well-sampled light curve of an active star, no simple rotation pattern occurs in the vast majority of stars, particularly when a long timespan is available. One of the most simple signposts of differential rotation is the appearance of light curve amplitude beating as a result of the combination of multiple periods close together.

The StarSim code implements a differential rotation model following Beck (2000), here each spot element rotates at a different angular speed according to its colatitude coordinate,

$$\omega = \omega_0 - \delta_{rot}[B \sin^2(\varphi) + C \sin^4(\varphi)] \quad (4.1)$$

where Ω_0 represents the fundamental rotation frequency, i.e. the equatorial frequency. φ is the latitude of the surface element and $B = 2.39 \text{ deg} \cdot \text{day}^{-1}$; $C = 1.78 \text{ deg} \cdot \text{day}^{-1}$ for a solar model (Snodgrass et al. 1990). δ_{rot} is a factor quantifying the degree of differential rotation. For $\delta_{rot} = 1$ the solar model is recovered, and $\delta_{rot} = 0$, corresponds to the solid sphere model.

4.3 StarSim Forward Problem

In this section, we describe how StarSim 2 computes photometry and spectroscopic time-series. The forward problem consists of calculating an observable quantity (photometry, spectra) from an input spot map and parameter set, \mathcal{S}, θ , in the form

$$\mathbf{X}(t) = F(\mathcal{S}, \theta) \quad (4.2)$$

where $\mathbf{X}(t)$ represents the time-series of some observable and F is the StarSim model.

4.3.1 Surface integration

The Kurucz (ATLAS9) (Kurucz 2017) models are computed to obtain the intensity quotients for different projection angles $\mu = \cos \theta$. The intensity profile across the disc is calculated according to the position of each element on the sphere,

$$\text{LD}_{\text{ph,sp}}(\lambda, \mu) = \frac{I_{\text{ph,sp}}(\lambda, \mu)}{I_{\text{ph,sp}}(\lambda, 0)} \quad (4.3)$$

where $I_{\text{ph,sp}}(\lambda, 0)$ is the intensity at the centre. In this way, the spectrum of each surface element is multiplied by this coefficient according to its relative position with respect to the line-of-sight,

$$f(\lambda, \mu) = f(\lambda) \cdot \text{LD}(\lambda, \mu) \quad (4.4)$$

In order to obtain the time-series observables, the flux of each element also projected to the line-of-sight to calculate its effective area. Thus, the element j flux becomes,

$$f_j(\lambda, \mu_j) = f(\lambda) \cdot \text{LD}(\lambda, \mu_j) \cdot a_j \cdot \max(0, \mu_j) \quad (4.5)$$

where a_j is the area of the active element. The expression $\max(0, \mu_j)$ guarantees the visibility condition of the surface element ($\mu \geq 0$). The area of a circular spot of angular radius r_j on the sphere (unit radius) is

$$a_j = 2\pi \left[1 - \cos(r_j) \right] \quad (4.6)$$

and

$$\mu_j = \sin i_* \sin \theta_j \cos(\phi_j + \omega t) + \cos i_* \cos \theta_j \quad (4.7)$$

is the cosine of the projection angle. i_* is the stellar axis inclination and θ, ϕ are co-latitude and longitude coordinates of the center of the active region, respectively. The quantity $\mu_j = \mu_j(t)$ is an explicit function of time as a result of considering a sphere's comobile (θ, ϕ) coordinate system rotating at angular speed ω .

Computing the full-disc integrated flux F , follows from summing over all surface elements,

$$F_{\text{tot}}(\lambda, t) = \sum_j f_{ph}^j + \Delta f_j \quad (4.8)$$

where the last sum is the flux variation induced by active regions,

$$\begin{aligned} \Delta f_j(\lambda) = & (f_{sp} \cdot J_{sp}(\lambda) - f_{ph}(\lambda) \cdot J_{ph}(\lambda) + \\ & + f_{fc} \cdot J_{fc}(\lambda) - f_{ph}(\lambda) \cdot J_{ph}(\lambda)) \end{aligned} \quad (4.9)$$

The first term is the flux deficit produced by spots, while the second accounts for the overflux given by faculae. The expressions J_{ph}, J_{sp}, J_{fc} account for the limb-darkening contribution and the projective effects over the flux, according to the position of each element.

$$\begin{aligned} J_{ph}^j(\lambda) &= \frac{I_{ph}(\lambda, \mu_j)}{I_{ph}(\lambda, 0)} \cdot a_j \cdot \max(0, \mu_j) \\ J_{sp}^j(\lambda) &= \frac{I_{sp}(\lambda, \mu_j)}{I_{sp}(\lambda, 0)} \cdot a_j \cdot \max(0, \mu_j) \\ J_{fc}^j(\lambda) &= c_{fc}(\mu_j) \cdot a_j \cdot \max(0, \mu_j) \end{aligned} \quad (4.10)$$

In the case of facula, it is known that they present a brightening at the limb (Frazier et al. 1978; Berger et al. 2007), instead of the classical limb darkening effect of the stellar photosphere. The observations show that facula dominate the radiation budget in solar-type stars where active regions show a brightness ~ 1.2 greater at the limb than at the center. In our code we use the formulation following Meunier et al. (2010a),

$$c_{fc}(\mu_j) = \left(\frac{T_{\text{eff}} + \Delta T_\mu(\mu_j)}{T_{\text{eff}} + \Delta T_{fc}} \right)^4, \quad (4.11)$$

where $\Delta T_\mu(\mu_j) = a_\mu + b_\mu \cdot \mu_j + c_\mu \cdot \mu_j^2$, with coefficients $a_\mu = 250.9$, $b_\mu = -407.4$ and $c_\mu = 190.9$; so that $c_{fc} \sim 1$ at the center of the disc, and $c_{fc} \sim 1.16$ near the limb for a Sun-type star.

4.3.2 Generating photometry observables

Once we derive the integrated flux over the stellar disc (Equation 4.8), the time-series is normalised with respect to the immaculate photosphere or its mean value, and it is convolved with the transmissivity function of a preselected filter, $T_f(\lambda)$

$$F_n(\lambda, t) = \frac{\sum_j f_{ph}^j + \Delta f_j}{\sum_j f_{ph}^j} \cdot T_f(\lambda) \quad (4.12)$$

and the integrated flux is

$$F_n(t) = \int_{\lambda} F_n(\lambda, t) d\lambda \quad (4.13)$$

in the case $T_f(\lambda) = 1$, we obtain a spectra time-series of bolometric flux. For an arbitrary T_f function and integrating over all wavelengths, the instrumental normalised flux is obtained in a given band, B , of the heterogeneous stellar surface.

4.3.2.1 Transiting planet model

We have added functionality to StarSim so that it is capable of simulating planetary transits on a spotted photosphere to investigate the effects induced by active regions on the profile and depth of transit events. The planet is modelled as a circular opaque disc, equivalent as a zero-temperature spot but assuming a circular projection throughout the stellar disc. The photometric signal is generated on the basis of relative planet size, r_p/R_* , the ephemeris and orbital orientation parameters (inclination, spin-orbit angle). The semi-major axis is computed from the parametrised orbital period of the planet and stellar mass through Kepler's third law, assuming $e = 0$. The current implementation does not consider the planet to have an atmosphere. When simulating transiting planets, particularly in high-precision chromatic transit studies, attention has to be paid to the size of the grid since a too coarse simulated photosphere may induce large errors in the transit curve of small planets. This has to be balanced against computing time, which increases quadratically with the number of surface elements.

Figure 4.3 shows the profile of three transits on a spotted star with $T_{\text{eff}} = 5000$ K, $\Delta T_{\text{sp}} = 700$ K and $P_{\text{rot}} = 15$ days. The planetary radius is 10% of the star's. It can be seen how spots alter the relative transit depths according to the observing wavelength. The bumps in the second and third transits are spot-crossing events: when the planet disc occults a spot instead of quiet photosphere, the flux difference makes the transit shallower. The upper spot models are pictured in Lambert projection to show more than a half of the sphere and only the spots inside the dashed circumference are visible. The transit path is marked with a grey band.

4.3.3 Generating spectroscopic observables

To simulate the observables derived from spectroscopic measures (radial velocity and activity indices, ie. bisector span, FWHM, ...), high-resolution spectra ($R > 100\,000$) is

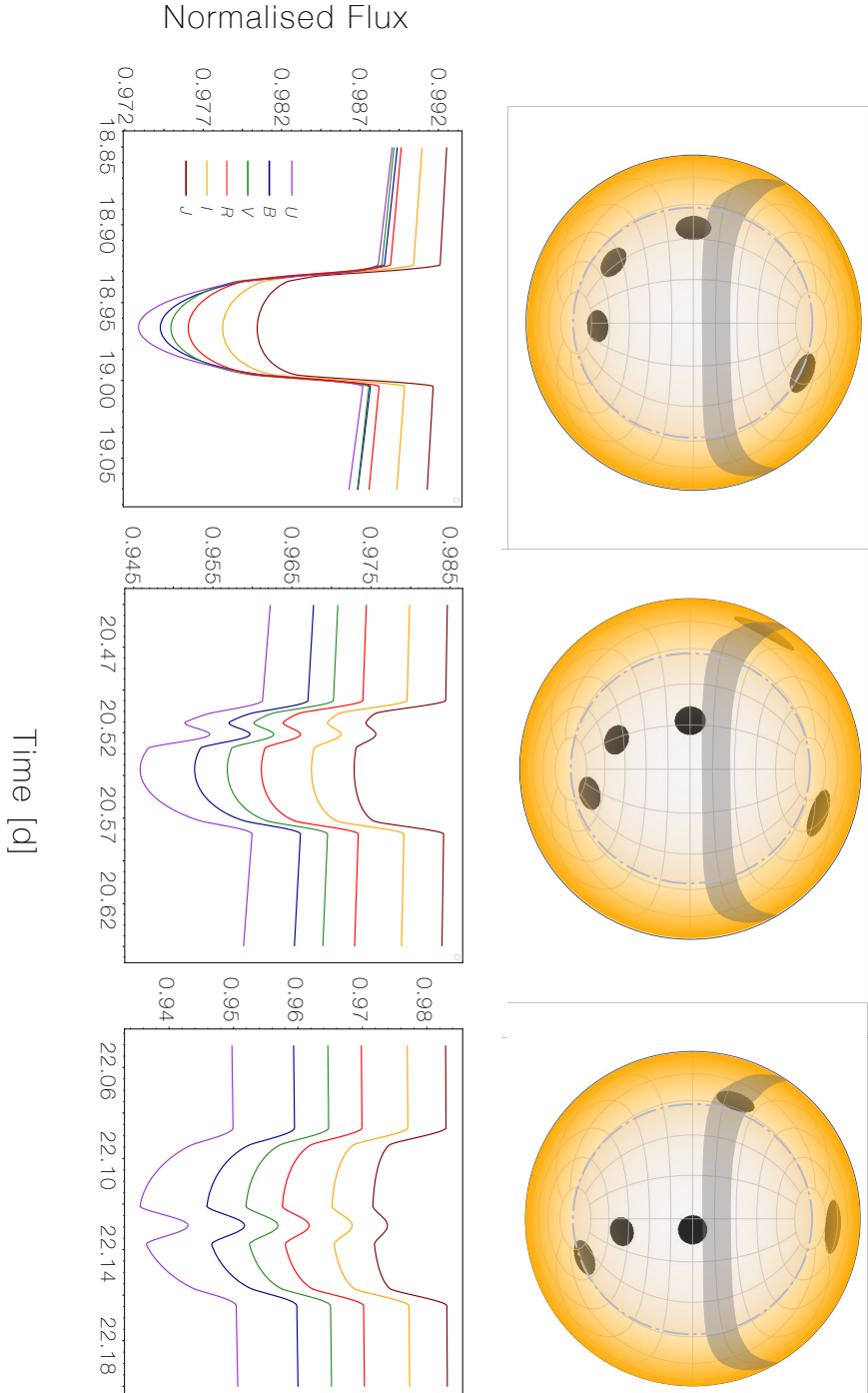


Figure 4.3: Simulated multi-band transits on a spotted surface. The synthetic map contains static (non-evolving) spots on a star $T_{\text{eff}} = 5000$ K, $\Delta T_{\text{sp}} = 700$ K and $P_{\text{rot}} = 15$ days. The gray band depicts the path of the transiting planet (r/R_* = 0.1, $b_{\text{in}} = 0.15$ and $P_{\text{pl}} = 1.58$ days) and appears curved due to the Lambert projection used to show the map, and therefore the most peripheral spots are not actually visible from the observer's line-of-sight. The innermost gray circle indicates the visible side of the sphere.

correlated with a specifically optimised mask of selected lines to minimise the sensitivity to stellar activity (Lafarga et al. 2020).

StarSim 2 uses high-resolution synthetic spectra generated with the Phoenix code (Husser et al. 2013) to compute the cross-correlation functions (CCF) between the spectra and the mask. To carry out determinations of precise radial velocities, the spectral resolution has to be very high, therefore requiring a significant computational effort. StarSim 2 operates in the CCF domain in spectroscopic mode instead of the natural wavelength space. The option of working on the space of CCFs offers the possibility of running the simulations in reasonable times, in comparison to the prohibitive computational expensiveness if we were dealing with high-resolution spectra. Nonetheless, the price to pay for this shortcut is the loss of information encoded in the spectral lines that collapses to an average CCF profile.

First, the spectra of all three types of surface are cross-correlated with an appropriate mask —HARPS, CARMENES, ...— and according to the stellar spectral type. Phoenix spectra include a modelisation of 1-D plane-parallel model (projection coefficient, $\mu = \cos \theta = 1$) convective effect. However, we initially remove it from the CCF profile by fitting a low-degree polynomial to its bisector. Subsequently, we incorporate a more sophisticated bisector using CFIST model spectra (Ludwig et al. 2009), which implements a more accurate 3D atmosphere model. CFIST models are available for Sun-like stars at several projection angles from the center to the stellar limb. The CFIST bisectors are computed by cross-correlation of the high-resolution spectra, and a 4-degree polynomial is fitted for 10 different cosine values of the projection angles. The new bisector at the position of the surface element is linearly interpolated using μ and finally introduced in the CCF profile of such element.

It is well established that convective motions are inhibited in surface magnetic features —spots and faculae— where strong magnetic fields prevent hotter plasma to emerge (Strassmeier 2009; Berdyugina 2005). In those active regions, high-resolution solar observations are used to model the bisector line to be implemented using a quadratic polynomial as follows,

$$\text{CCF}_{act}(v) \left[\frac{km}{s} \right] = -1.2908 + 4.5295 \cdot \text{CCF}_{sym}(v) - 3.0231 \cdot \text{CCF}_{sym}(v)^2 \quad (4.14)$$

where $\text{CCF}_{sym}(v)$ denotes the quiet photosphere CCF with Phoenix convective bisector subtracted. This model was obtained by high-resolution FTS (Fourier Transform Spectrograph) observations of sunspots, and originally used in the SOAP 2 tool (Dumusque et al. 2014).

Computation The way to proceed to integrate the surface in spectroscopic mode consists of determining a CCF for each of the surface elements, which is scaled according to its projection geometry and spectral flux with low-resolution BT-Settl spectra. Finally, the corresponding Doppler shift is added $\lambda_j = \lambda + \Delta\lambda_j$

$$\Delta\lambda_j \left[\frac{km}{s} \right] = 8.05 \cdot \lambda \cdot \frac{1}{c} \cdot R_{star} \cdot \omega \sin i_* \sin \theta_j \sin \phi_j \quad (4.15)$$

where θ_j i ϕ_j are colatitude and longitude coordinates of the surface element j , and ω is the angular rotation rate (Equation 4.1).

The next step is the substitution of the original CCF bisector from the Phoenix models by the more sophisticated and geometry-dependent CIFIST bisector,

$$\text{CCF}_j(v) = \frac{I_j}{I_o} \cdot \left[\text{CCF}(v) - \hat{\text{BS}}(\alpha_0, \dots, \alpha_6) + \text{BIS}(\text{CIFIST}) + \Delta\lambda_j \right] \quad (4.16)$$

where I_j is the flux intensity of BT-Settl spectra and the projection vector; I_o is the flux normalisation; $\hat{\text{BS}}(\alpha_0, \dots, \alpha_6)$ is the polynomial fit bisector model; and $\text{BIS}(\text{CIFIST})$ is the new CIFIST model bisector to be introduced.

The resulting CCF of the integrated surface is obtained by summing up all CCFs,

$$\text{CCF}(v) = \sum_j \text{CCF}_j(v) \quad (4.17)$$

Once the integrated spotted photosphere CCF is obtained, Gaussian fit is performed as this will allow to measure some key parameters, such as the position and momenta of the profile. The `curvefit` routine from `scipy` Python package is used to fit the function,

$$G(v) = c + A \exp \left[- \frac{(v - \langle v \rangle)^2}{2\sigma^2} \right] \quad (4.18)$$

where $\langle v \rangle$ is the net radial velocity of the spotted star. This constitutes the slowest step in the calculation as it takes up roughly 80% of the execution time of a full RV curve computation. To speed up the computation of the Gaussian fit, the initial hint parameters $A, \sigma, \langle v \rangle, c$ are set to those from the previous calculation in the RV series.

Once the optimal parameters from the Gaussian fit are obtained, a set of activity indicators can be built. The radial velocity is simply the average value $\langle v \rangle$ of G ; σ is related with Full-width-at-half-maximum (FWHM), and A is the contrast indicator. Additionally, an additional activity proxy can be computed by quantifying the shape of the integrated CCF by means of the bisector (BIS) index. It is computed by firstly calculating the curve that divides the CCF into two equal parts, and then subtract the mean value of the upper part between 60–90% from the lower part between 10–40% of the CCF height (Santos et al. 2001; Queloz et al. 2001).

Two main effects can be distinguished in the radial velocity modelling: the Doppler shift of each surface element weighted by its flux and the convective blueshift effect. In Figure 4.4 the RV signals of single spot model are shown with no faculae ($Q = 0$, in blue), and with faculae ($Q = 3$, in red) for a star rotating at $P_{\text{rot}} = 25$ days. Dotted lines show only the convective effect, which exhibits the symmetric nature with respect to the spot central passage time. At this moment the convective blueshift effect is maximum due to its radial geometry. In dashed lines, the isolated Doppler effect is shown, which has also a symmetric profile. The left half of the star is blueshifted ($v_r < 0$) while the right half is redshifted ($v_r > 0$). When in one of these halves is some active element, by being at different temperature, the integrated emissivity in both halves becomes different. When adding these two effects, the result is a non-symmetric profile because the convection is always blueshifted (shown in solid lines). Note that, although the

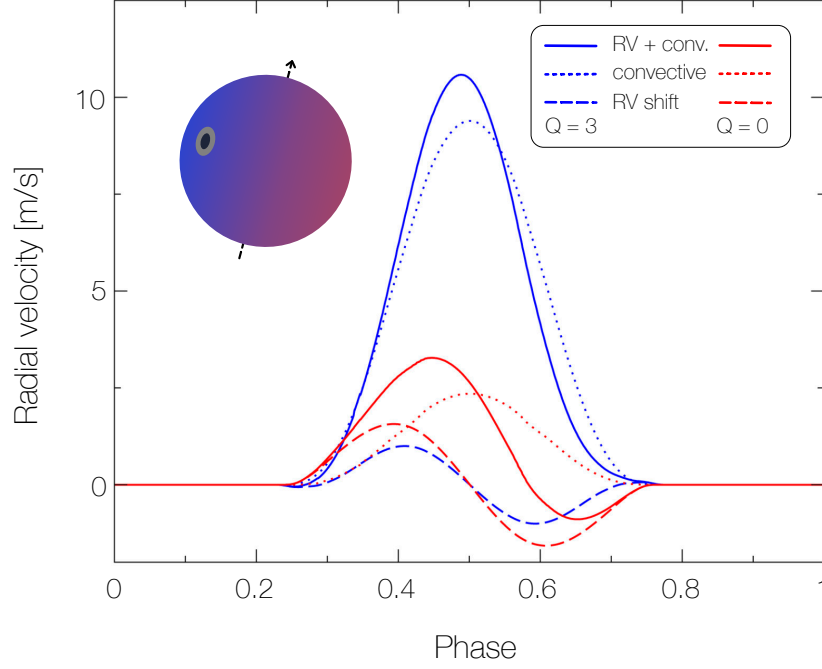


Figure 4.4: Composition of radial velocity signal budget for a solar-type star with an equatorial spot of $1.6 \cdot 10^{-3}$ fractional area coverage. In dashed lines, the Doppler contribution is plotted and the convective effect is shown in dotted lines for a no facula case (red) and for $Q = 3$ (blue). The total effect is represented in solid lines.

convective contribution is always blueshifted ($v < 0$), the effect shown in the plot is positive due to the subtraction of the systemic velocity.

Figure 4.5 shows an example of StarSim 2 generated observables using the parameters from Table 4.2.

4.4 StarSim Inverse Problem

The inversion functionality is the culmination of the StarSim project since allows the outstanding functionality of inferring starspot properties from observational data. Extracting information on the spot distribution and its properties has been a historical subject of interest and some studies have been published (e.g. Petrov et al. 1994; Kovari et al. 1997; Kipping 2012; Eker 1999a, 1999b; Luo et al. 2019; Luger et al. 2021b; Luger et al. 2021a).

The inverse problem consists of finding the parameters of a model that best explains a set of observables. In our case, finding \mathcal{S}_θ ,

$$\mathcal{S}_\theta = F^{-1}(\mathbf{X}, \theta) \quad (4.19)$$

where F is the activity model, \mathcal{S} is the spot map, and θ are the stellar parameters. To obtain a solution for the inverse problem means finding a spot map, but also a set of stellar

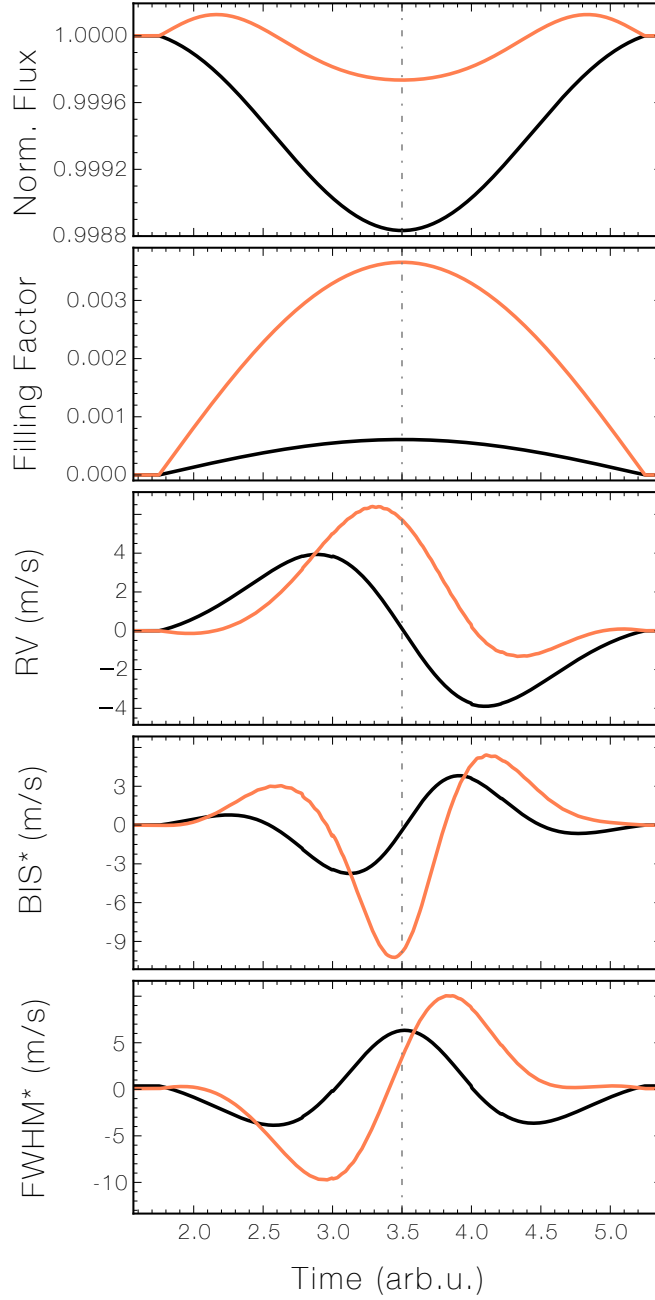


Figure 4.5: Examples of curves generated with StarSim 2 code for a K5V star and a map consisting of a single equatorial spot of 2° in radius.

Parameters: $T_{\text{eff}} = 4500 \text{ K}$, $\Delta T_{\text{sp}} = 800 \text{ K}$, $\Delta T_{\text{fc}} = 150 \text{ K}$, $P_{\text{rot}} = 7.0 \text{ d}$, $i = 90^\circ$ and without differential rotation ($\delta_{\text{diff}} = 0$). The solid black line shows the output for $Q = 0$ and the orange line is for $Q = 6$, keeping constant all other simulation parameters. Asterisks (*) in the bisector and FWHM indicate quantities normalised to the mean of the time-series, $\text{BIS}^* = \text{BIS} - \langle \text{BIS} \rangle$ i $\text{FWHM}^* = \text{FWHM} - \langle \text{FWHM} \rangle$. The vertical dashed line marks the passage of the spot's center across the stellar meridian.

Table 4.2: Stellar and simulation parameters of the example curves in Figure 4.5.

Parameter	Value
Spectral range [nm]	450-900
T_{eff} [K]	4500
$\log g$	4.5
Spot temperature contrast, ΔT_{sp} [K]	800
Facula temperature contrast, ΔT_{fc} [K]	150
facula–spot surface ratio, Q	0 / 6
Rotation period, P_0 [days]	7.5
Rotation axis inclination, i [°]	90

parameters, such as rotation period, photosphere temperature, spot temperature contrast, amount of facula, stellar axis inclination, spot growing factor; and observational-related parameters such as additional statistical jitter in data or passbands.

The inversion process of the observables is a computationally intensive task due to the fact that many —typically tens of thousands— evaluations of the forward StarSim problem are needed. In addition, some difficulties arise. Namely, *i*) The process of obtaining a spot map, keeping constant the stellar parameters, is inherently degenerate and only approximate estimates of the optima are attained; and *ii*) To solve the full inversion problem, i.e. spot maps with their associated parameter sets is only possible by performing a two-step approach in which the spot maps are inverted with randomised parameters. This first step is repeated until a statistically equivalent population of the top-ranked solutions can be selected by using some statistical criteria. In forthcoming sections (inverse problem in 5.3.2 and the application to WASP-52, 6.4), this methodology is described in detail.

4.4.1 Spot map optimisation

StarSim 2 finds the maps that best reproduce the observables by performing a Monte-carlo Simulated Annealing (MCSA) optimisation. This is a generic metaheuristic algorithm intended for global optimisation that allows to find reasonably good approximation to the optimum in large configuration-space problems.

MCSA is a probabilistic method intended to find a reasonably good solution of the global optimum of a cost function. This method excels in: *i*) finding an approximation to the solution since finding and assuring the global optimal is not attainable with a finite number of function evaluations; and *ii*) MCSA is designed to overcome getting trapped in local minima, which is one the most important drawbacks of gradient-based descent methods. As the functions to evaluate are commonly noisy, there are countless local minima where the algorithm could get stuck.

Simulated Annealing finds its inspiration in the likewise called metallurgy technique consisting of heating metals and cooling back several times to modify their crystallographic properties by moving and replacing the atoms of the lattice to minimize the energy of states. This procedure is top to bottom: first, the biggest defects are arranged, and as temperature decreases, the tiniest defects are corrected.

The first authors that linked the Simulated Annealing of metals crystallography with combinatorial minimization were Kirkpatrick et al. (1983). They replaced the energy with a cost function and the states of a physical system by a combinatorial solution that minimizes this cost function (merit function).

In the beginning, this technique was designed to solve combinatorial problems with discrete space objective functions. However, the mathematical formulation can be easily extended to continuous spaces (Dekkers et al. 1991) by an appropriate selection of minimal distance, $\delta\mathbf{x}$, between neighbour configurations to obtain perturbed configurations $\phi_j^p = \phi_j + \delta\mathbf{x}$.

The great advantage of MCSA is the ability to avoid local optimals given an adequate cooling schedule, which is the way the temperature is decreased. Hence, a favourable configuration after a perturbation is always accepted while a disfavourable perturbation can be accepted or not according to an exponential probability, making it possible to hillclimb to finally reach the global optimum.

4.4.1.1 Simulated Annealing

We show here a simple outline of the simulated annealing method implemented in StarSim 2 map optimisation routine.

1. Draw a random configuration (a list of random spots).
 - Set initial inverse temperature $\beta = \beta_0$, where $T = 1/\beta$ is defined as the cooling temperature of the algorithm.
 - Let $k = 0$ (number of temperature steps; max set to k_{max}).
 - Let $n_\beta = 0$ (number of steps for each temperature; maximum set to n_β^{MAX}).
2. Make the current configuration ϕ_j evolve to a perturbed configuration ϕ_j^p by slightly modifying one randomly selected parameter of one of the spots. $f(\phi)$ represents the forward problem.
3. If $f(\phi_j^p) > f(\phi_j)$ then
 accept ϕ_j^p
 else
 accept ϕ_j^p with probability $e^{-\beta\Delta f}$
 $n_\beta \leftarrow n_\beta + 1$. If $n_\beta \leq n_\beta^{MAX}$ go to 2.
4. Increase the inverse temperature $\beta \leftarrow \beta + \delta\beta$ according to a cooling schedule.
5. $k \leftarrow k + 1$. If $k < k_{max}$ go to 2, else go to 6.
6. End.

There is no consistent rule for choosing an optimal annealing schedule as it strongly depends on the specific function to be optimised. However, a fairly efficient scheme may be determined by trial and error (Kirkpatrick et al. 1983) to obtain a robust optimiser, and a faster convergence was found for a Kirkpatrick cooling schedule,

$$\beta_k = \beta_0 \alpha^{-k}, \quad (4.20)$$

in comparison to a linear scheme, both widely used (Nourani et al. 1998). Several trials were performed to adjust the parameters $\{T_0, \alpha, n_\beta, k\} = \{5000, 0.25, 500, 25\}$ until a reasonable behavior was reached.

4.4.2 Parameter inference

Solving the inverse problem means finding a spot map \mathcal{S} and the parameter tuple θ that reproduces best the observational data. Algorithmically, optimising a map and its associated parameters, \mathcal{S}_θ , cannot be performed at the same time due to the strong coupling between two data structures. Each of the parameters in the tuple affects the simulation as a whole. While moving a single spot produces typically small variations in the computed observable, a slight variation on rotation period or spot temperature will affect all spots and produce large variations on these observables. Spot maps have to satisfy the restriction of non-overlapping spots meaning that during the fitting process, each spot movement has to be checked to fulfill this condition, by perturbing the spot parameters one by one. In this sense, a spot map can be understood as a superparameter with internal structure. Therefore, fitting map and parameter set simultaneously is extremely difficult and a two-step solution is adopted.

In our approach to obtain the parameter distribution, the first step is to fit the map with parameters kept constant during all the process. Hence, an optimal map with a given set of parameters is obtained, denoted as $\mathcal{S}_{\theta_0}^0$. By repeating this last process repetitively with randomised parameters, each instance will result in a different fitting metric. Those variations are motivated by two facts: *i*) Parametric randomisation, and *ii*) The inherent uncertainty in map fitting. This last uncertainty is the most concerning since it cannot be easily estimated. Subsequently, statistical criteria may be used to group a set of equivalent solutions, conforming an ensemble to extract the true parameter distributions. The criteria used in StarSim analysis of WASP-52 is developed in the following Chapter 6.

The process of fitting a map is numerically hard mainly due to the total number of parameters involved and the fact that all spots have the same weight on the fit, which implies that only approximate optimals can be reached. This class of optimisation problems falls into the category of noisy optimisation which in our case the evaluation of the objective function is always lower than the real optimal. If the estimated map given constant θ yields $\hat{\mathcal{S}}_\theta = F^{-1}(\mathbf{X}, \theta)$, the returned forward problem is given by $F(\hat{\mathcal{S}}_\theta) = F(\mathcal{S}_\theta^{opt}) - |\zeta|$ where ζ is a random noise variable produced by the SA optimisation algorithm.

This situation appears frequently in fields such as machine learning, robotics, sensor networks, physics simulations or any other optimisation application domain in

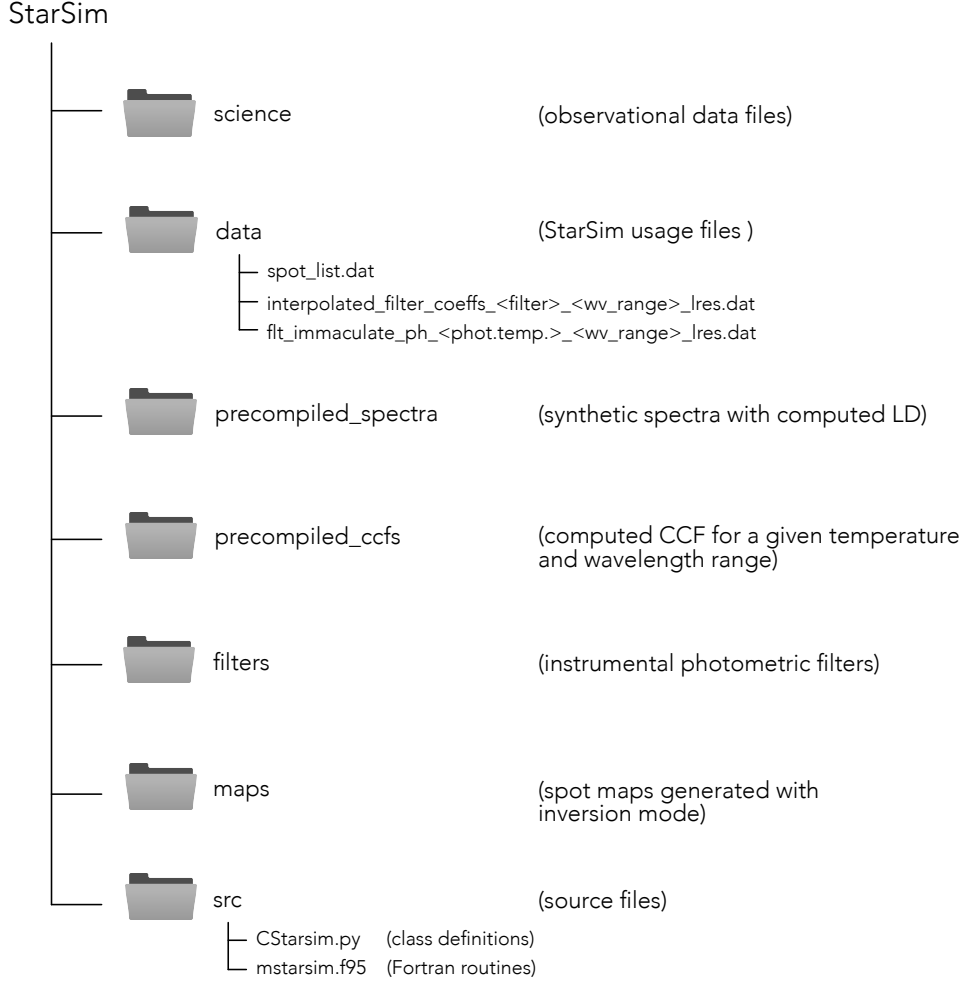


Figure 4.6: StarSim application file tree with most important folders and files.

which the computational cost of evaluating the objective function is high, or we just can access to approximations of them.

For those optimisation problems several modern techniques have been implemented. One of the most successful is Bayesian optimisation consisting of fitting a surrogate hypersurface to a small group noisy evaluations. Based on this first approximation to the objective space, new points are sampled according to conditional probability densities until convergence is reached. Then the fitted surface is sampled through a standard MCMC routine taking advantage of its high computing speed. An example of this methodology is the Python implemented **approxposterior** (Fleming et al. 2018), allowing speed boosting up to a factor of $\times 1000$ in sampling heavy functions. Unfortunately, the computing times (see Table 4.3) of the photometric inverse problem is still too high, and the associated uncertainties are too large.

Table 4.3: StarSim 2 running times for a typical dataset.

(HD 25825)	computational time		
	Photometry ^a	Radial velocity ^b	Spot map ^d
Forward problem	~ 3.5 ms	~ 0.1 s	Evolving 35-spot
	~ 0.003 s	~ 0.076 s	Unevolving 3-spot
Inverse problem	68 s	25 min	Evolving 35-spot
	13 s	8 min	Unevolving 3-spot
Inversion (Ph+RV)		2.2 h ^c	Evolving 35-spot
		8 min ^c	Unevolving 3-spot

^a 191 points (original data binned).

^b 59 measurements; HARPS (24) + TRES (23) + SOPHIE (12) (see Section 8.2.1.1).

^c The inversion of two observables is not the sum of the times for photometry and RV inverse problems because of the increase in spot-placement difficulty, since much more spot movements have to be made to reach an acceptable fitting statistic. All these values do not consider the loading times needed to compute ph/rv curves such as reading spectras, precompiled CCFs... in such a case, 1-2 seconds shall be added.

^d Evolving spotmap means that not all the spots are active at a given time but they can appear at any moment and have an individual lifetime. Unevolving pattern means all the spots are present since the beginning of the simulation and no spot can appear during the simulation.

4.5 Code architecture, usage and benchmarking

The new version of StarSim is coded in Python (2.7+/3.x) but the most numerically-intensive routines are written in Fortran 90 to optimise the performance. The code is intended for fast computation of the forward problem, but also offers a flexible environment to solve many different configurations of the observables, e.g. computing a single inverse problem by inverting a set of multiband photometric observations; solving N inverse problems with randomised parameters in an arbitrary configuration of observables (photometry, RV, BIS,...); and computing the forward problem of a collection of maps and parameter sets. Object-oriented programming (OOP) is the most adequate paradigm to write the code as each of the observables has specific parameters and data only affecting themselves, and, all the observables need to access common methods like spectra or IO operations taking advantage of encapsulation and inheritance, two of the fundamental features of the OOP.

The architecture relies on the design of a container class `StarsimSimulation`, which reads from a configuration file where the observational data time-series can be selected. Then, an object for each type of observable is instantiated and the methods enclosed all across the inherited classes allow the computation of the different data products. Figure 4.7 shows the UML class diagram where the inheritance relations and composition among the classes can be traced. The notation `1..*` in `StarsimSimulation` means one or more instances can be created, while `0..*` indicates an arbitrary number and combination of observable objects can be created by the container class.

The core idea of the design relies on the OOP encapsulation concept. Each of the observables has specific data and parameters, e.g. a photometry series dataset has a specific filter, spectral range, time sampling, quadrature added jitter, ...and, needs specific methods (functions) to compute merit function metrics ($\ln \mathcal{L}$). All these features

are kept encapsulated into an object and cannot be altered nor interfered with by any other objects. However, there are other attributes (variables and parameters) that are common throughout the simulation process, namely those related to stellar parameters and the spot model. Hence, the class container `StarsimSimulation` inherits the classes `Spots` (which in turn inherits `Star`), `Planet` and `SimulatedAnnealing` to allow the access to their attributes and methods.

This architecture allows building complex data structures such as creating a tuple of `StarsimSimulation` objects, each one with different simulation parameters, to perform parameter inference. This option can be run in parallel in multicore machines and is truly computationally expensive.

Some examples of the computing time for representative StarSim 2 forward and inverse problems can be found in Table 4.3. Times were benchmarked on an Intel i5-8259U@2.3GHz using a single core for typical space-based photometry and spectroscopic observations.

Package dependences Starsim is written in Python 2.x with precompiled libraries in Fortran 90 which are linked through `f2py` interfacier (Peterson 2009). This wrapper facilitates the integration of fast compiled Fortran code into Python environments resulting in a dramatic speed-up in those routines in where Fortran excels in speed, i.e. loops, conditional statements or basic algebra. The Fortran subroutines must be compiled before to be callable by the main Python code using this built-in bash script,

```
$ bash compile_fmodule.sh
```

`f2py` is a part of Python `Numpy` package and can also be installed as a standalone command line tool. In addition to this, StarSim 2 also needs `scipy 0.18.1+` for optimisation routines and Lapack library for fast linear algebra routines.

File tree and important files Figure 4.6 sketches the file structure of the StarSim 2 showing the most important files. In `/science` folder are found the observational data files; `/data` folder contains those files used in all simulations, e.g. the spot map list (`spot_list.dat`), the interpolated filters (`interpolated_filter_coeffs_*`), low resolution spectra of integrated immaculate photosphere (`flt_immaculate_ph_*`), among other utility files. In `/precompiled_spectra` and `/precompiled_ccfs` folders are found the already generated spectra and CCFs for a given set of parameters. In case of StarSim 2 does not find these files, they are generated at the expense of more time consumption. `/src` folder contains Fortran and Python source code files.

StarSim 2 basic usage Next we provide a list of the commands implemented in the code. Following the structure,

```
./starsim.py --mode=[available_modes] --other-options
```

a number of modes are available:

1. `--mode=ph`: perform photometric forward problem
2. `--mode=rv`: perform RV forward problem
3. `--mode=bis`: perform bisector forward problem
4. `--mode=contrast`: perform contrast forward problem
5. `--mode=fwhm`: perform full width at half maximum (fwhm) forward problem
6. `--mode=inversion`: perform inverse problem using the observables checked at `multifit.conf` configuration file
7. `--mode=N-inversion`: perform N equivalent instances of the inverse problem with the same stellar parameters
8. `--mode=param_fit`: perform full inverse problem with randomly varying parameter space
9. `--mode=forward`: perform forward problem for all spot maps found in `/maps` folder and a parameter list for each map
10. `--inhibit_msg`: do not show informative screen messages
11. `--ncpus=`: number of computing cores (parallelism implemented on inverse problem)

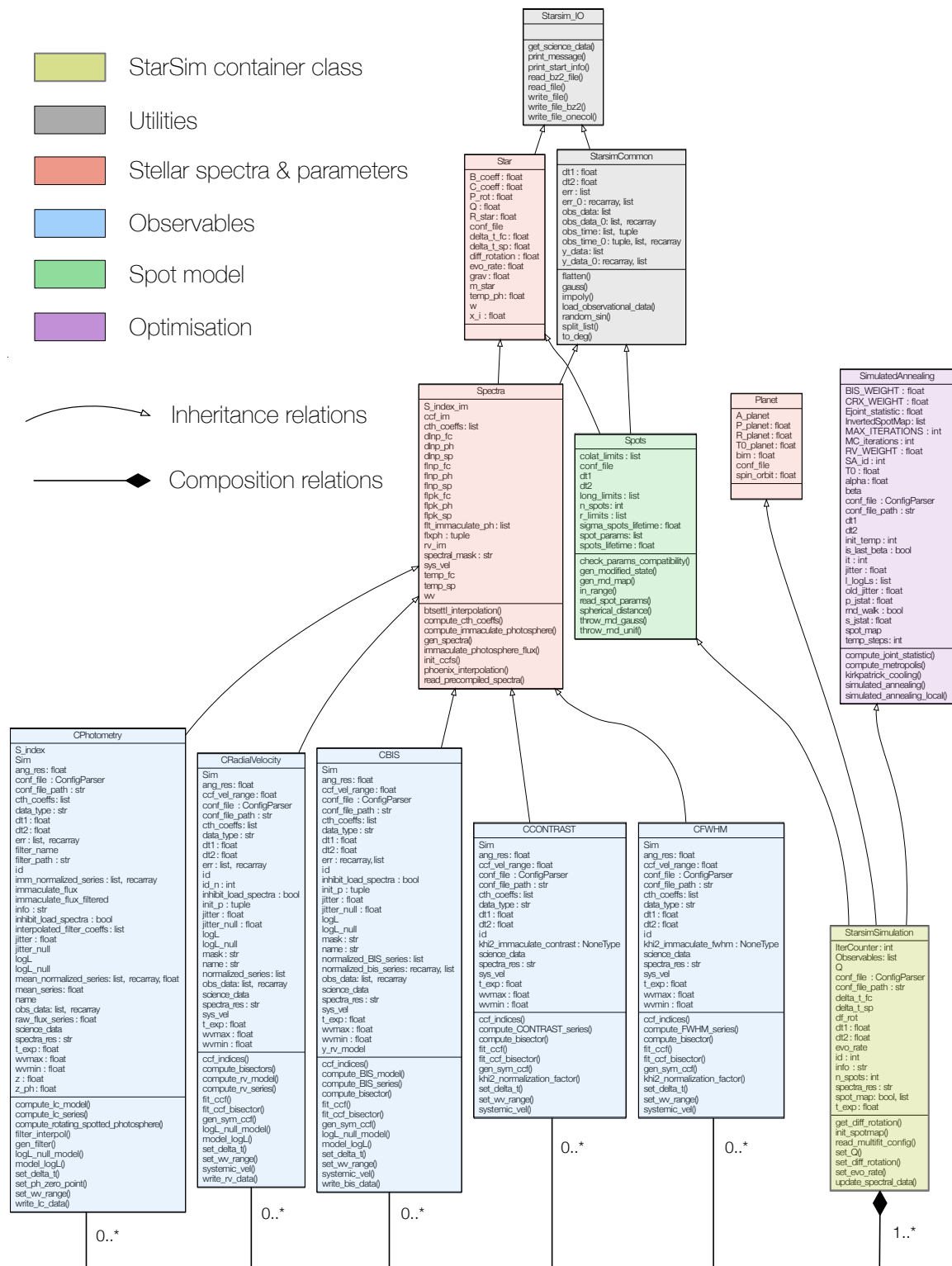


Figure 4.7: StarSim UML class diagram showing the relations among the classes.

Chapter 5

The Photometric Inverse Problem

The findings presented in this chapter are published in Rosich et al. (2020).

5.1 Introduction

In this chapter, the analytical formulation and numerical simulations of the multiband photometric inverse problem are presented. The inverse problem consists of finding the model parameters that satisfy some observational data. Unlike the forward problem which has—in deterministic physics—a unique solution, the inverse problems are typically *ill-posed* meaning there is no uniqueness or stability of solutions, i.e. arbitrarily small errors in the measurement data may lead to indefinitely large errors in the solutions (Tarantola 2005; Kabanikhin et al. 2008).

Inverse problem theory has wide applications in fields such as optics, acoustics, imaging, geophysics, remote sensing, and astrophysics. In the stellar activity context, it means finding a set of physical stellar parameters (i.e. rotation period, spot temperature, ...) and a spot map, containing the distribution of the active elements and their evolution. In general, the photometric inverse problem is computationally expensive as it requires a large number of *forward* evaluations, the solution is not unique, and often, only approximate estimations can be attained.

The implemented spot model is also a key constrain to tackle the large multiplicity of equivalent maps retrieved. In discrete spot maps, like the StarSim model developed in the preceding chapter, the degeneracy-breaking prior is the assumption of circular spots and the fact that they have uniform contrast temperature and sharp boundaries, on a perfectly uniform photosphere (Luger et al. 2021b). In gridded stellar surface models (like the old StarSim version) a regularisation metric is often assumed to assure the convergence to favor the simplest solutions containing the fewest dark pixels solutions (e.g. Vogt et al. 1987; Lanza et al. 2011).

The following sections aim to describe the computational inverse problem to retrieve the stellar properties and the active maps in the framework of StarSim spot model. We explore a variety of degeneracies found in the problem, concluding that by using

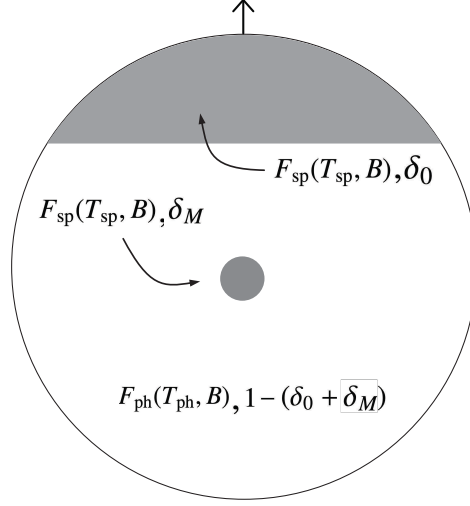


Figure 5.1: Example sketch of a stellar photosphere including modulating and non-modulating spots. The model is composed of a polar cap, which does not contribute to light curve modulation, with a projected filling factor δ_0 , and an equatorial small spot with a maximum projected filling factor δ_M . $F(\cdot)$ is the brightness of each surface element.

multiband photometry is possible to break the spot size—temperature degeneracy, and therefore, reconstruct the time-evolving projected filling factor allowing to constrain its chromatic effect on planet transit events.

5.2 Analytical foundations of the inverse problem

Before moving on to discuss how StarSim 2 handles the inversion of photometric light curves, we shall present the analytical foundations of the model and the relevant variables.

We consider a stellar surface with active regions covering a total projected filling factor δ_{sp} , which is made of two components: one that produces flux variations over time due to rotational modulation, δ_M , and another one that stays constant as the star rotates (e.g., a polar cap or a uniformly spotted latitudinal band), δ_0 .

To demonstrate analytically that light curves can carry information of both spottedness levels we assume a simple model consisting of a star with a circular spot on its equator, with a surface δ_M , and a polar cap covering a surface δ_0 , as illustrated in Figure 5.1.

The amplitude of the photometric modulation can be estimated as the ratio of the flux of the star when the modulating spot is out of view,

$$f_1 = F_{\text{sp}}(\mathcal{B}, T_{\text{sp}}) \delta_0 + F_{\text{ph}}(\mathcal{B}, T_{\text{ph}}) (1 - \delta_0), \quad (5.1)$$

with respect to the phase when the spot is at the center of the visible stellar disc,

$$f_2 = F_{\text{sp}}(\mathcal{B}, T_{\text{sp}}) (\delta_0 + \delta_M) + F_{\text{ph}}(\mathcal{B}, T_{\text{ph}}) (1 - \delta_0 - \delta_M), \quad (5.2)$$

where F_{ph} and F_{sp} are the surface fluxes of the spotted and immaculate surface of the star, respectively, which depend on the temperature of each surface element (T_{ph} and T_{sp}) and the spectral band (\mathcal{B}). The relative amplitude of the photometric variability can be computed as $(f_1 - f_2)/f_1$, which is equivalent to the measurement provided by observational light curves. Rearranging Equations 5.1 and 5.2, the photometric amplitude can be written as,

$$A_{\text{obs}}(\mathcal{B}, \Delta T_{\text{sp}}, \delta_M, \delta_0) = \frac{\delta_M}{\Phi(\mathcal{B}, \Delta T_{\text{sp}}) - \delta_0} \quad (5.3)$$

where

$$\Phi(\mathcal{B}, \Delta T_{\text{sp}}) = \frac{1}{1 - \left(\frac{F_{\text{sp}}}{F_{\text{ph}}} \right)_{\mathcal{B}, \Delta T_{\text{sp}}}}, \quad (5.4)$$

with ΔT_{sp} being the difference between the photospheric and spot temperature ($\Delta T_{\text{sp}} = T_{\text{ph}} - T_{\text{sp}}$). We assume that the effective temperature of the star can be estimated independently.

These equations illustrate that the photometric amplitude increases linearly with the modulating filling factor, δ_M , and decreases with a combination of the non modulating filling factor, δ_0 , and the brightness contrast between the photosphere and the spots Φ , which is, in turn, a function of the photometric band (\mathcal{B}) and the temperature contrast (ΔT_{sp}). We remark here that these quantities satisfy some constraints, namely, $0 \leq \delta_0 \leq 1$, $0 \leq \delta_M \leq 1$, $0 \leq \delta_0 + \delta_M \leq 1$ and $\Phi > 1$, for cool spots.

5.2.1 Implications on attainable information

Figure 5.2 illustrates Equation 5.3 and its dependence on the three independent parameters. If only information from a single band is available (e.g., typical data from exoplanet surveys), the linear dependence of the light curve amplitude with δ_M permits the determination of the modulating filling factor provided the non-modulating filling factor is neglected and the spot temperature contrast is adopted as an external constraint. This has been a common practice in the literature.

In case two or more bands are available, the possibility of determining another variable such as ΔT_{sp} or δ_0 arises. For typical spot temperature contrasts and visible bands, it can be shown that Φ is significantly greater than δ_0 and therefore the former parameter dominates. Thus, from two or more photometric bands (preferably covering a large interval in wavelength, i.e., large Φ variation), the simultaneous determination of δ_M and ΔT_{sp} becomes possible. A practical application can be found in Mallonn et al. (2018).

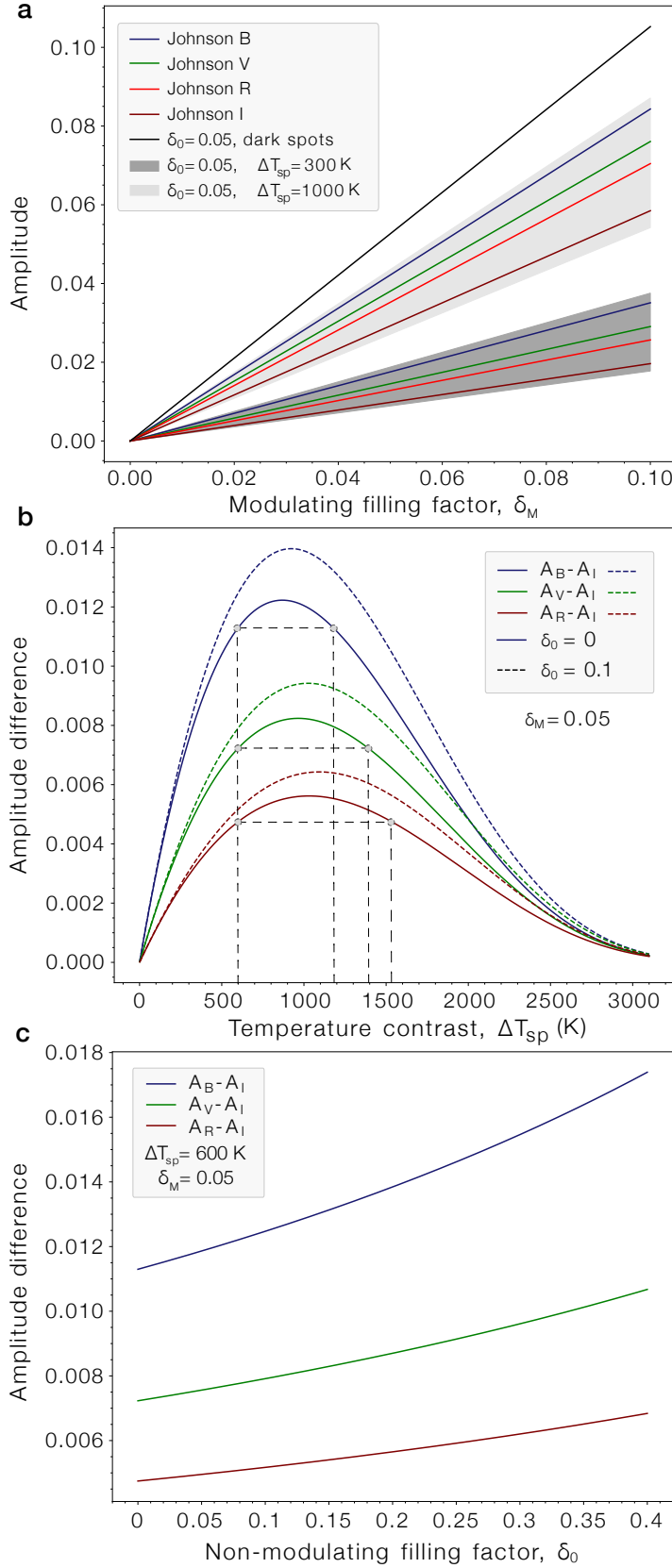


Figure 5.2: Graphical representation of Equation 5.3 for the parameters of interest assuming a stellar photospheric temperature of $T_{ph} = 5000$ K and a black-body for the spectral energy distribution of each surface element. **(a)** Photometric amplitude as a function of the modulating filling factor for different spot temperature contrasts. A non-modulating filling factor $\delta_0 = 0.05$ is employed, but adopting other values has negligible effect at this scale. The gray area covers the spectral range from 400 nm to 1 μ m. **(b)** Difference in activity-induced amplitude for photometric bands *BVR* with respect to the *I* band as a function of spot temperature contrast. A value of 0.05 is adopted as modulating filling factor (δ_M) and examples for non-modulating filling factors of $\delta_0 = 0$ and 0.1 are shown. The gray lines illustrate how the bi-valuate nature of the temperature contrast effect can be resolved if more than two bands are used. **(c)** Difference in activity-induced amplitude for photometric bands *BVR* with respect to the *I* band as a function of the non-modulating filling factor. A modulating filling factor $\delta_M = 0.05$ and a spot temperature contrast $\Delta T_{sp} = 600$ K are adopted.

In the particular case of two bands, the solution becomes bi-valuate, with two possible ΔT_{sp} reproducing the amplitude difference, as shown in Figure 5.2b. However, if a third band is available, such degeneracy can be broken, as illustrated by the gray lines in the figure.

Equation 5.3 further shows that, from three bands or more, one can theoretically determine at the same time the three relevant variables. For that to be possible, the photometric information needs to be of sufficient precision to discriminate the changes induced by each variable. The curvature in Figure 5.2 (i.e., the variation in the amplitude difference) is what makes it possible to determine the non-modulating filling factor from multi-colour photometry. Nevertheless, the scale of the variations makes reliable estimates of δ_0 very challenging for typical ground-based photometric precisions. We remark that we have presented a simplified version of the problem, defined by only two epochs (maximum and minimum light). However, photometric time-series also carry information on the relevant parameters because of the correlations present among the different measurement epochs, thus adding additional constraints to the spot properties.

The formulation discussed here proves that it is possible to simultaneously determine the total spot filling factor and spot temperature contrast as long as good multi-band photometric data are available, therefore providing theoretical foundation to the inverse problem.

5.3 The inverse problem

The most recent StarSim version can perform the inverse problem. The goal is to obtain the underlying properties, i.e. a stellar activity model as described by the parameters of the star and its surface map, that reproduce the observed time-series data \mathbf{X} . Such nonlinear problem can be expressed as

$$\mathbf{X} = F(\mathcal{S}, \theta) + \epsilon, \quad (5.5)$$

where \mathbf{X} is the time-series data, F is the activity model, and θ is a set of stellar parameters. The surface map \mathcal{S} is the set of parameters that describe the surface distribution, sizes, and lifetimes of all the active elements considered, each of them defined as a small circular spot surrounded by a bright facular region. Finally, ϵ is an additional noise term, or jitter, that we adopt as uncorrelated and following a Gaussian distribution (white noise).

The inverse problem consists in finding the surface map that best reproduces the data \mathbf{X} for a given θ , expressed as

$$\hat{\mathcal{S}}_\theta = F^{-1}(\mathbf{X}, \theta), \quad (5.6)$$

where $\hat{\mathcal{S}}_\theta$ means the optimal surface map linked to a specific set of stellar parameters. Subsequently, this map and its associated parameters provide an optimal fit to the photometric and spectroscopic time-series data when applying the forward model (Equation 5.5).

5.3.1 Objective function

The statistical function to optimise, or figure of merit, is a linear combination of the logarithmic likelihood function of all the time-series data defined as

$$\ln \mathcal{L}(\mathcal{S}, \theta) = \sum_j^{\text{Obs.}} a_j \ln \mathcal{L}_j(\mathbf{X}_j | \mathcal{M}_j(\mathcal{S}, \theta)), \quad (5.7)$$

where $\ln \mathcal{L}_j$ is the log-likelihood of the fitted model \mathcal{M}_j according to the observational data \mathbf{X}_j for the j -th observable. The quantities a_j are the weights associated to each set of observables (here we assume $a_j = 1$). As shown in Equation 5.5, we consider the simplest case of non-correlated Gaussian uncertainties (white noise). The likelihood function is then written as

$$\mathcal{L}_j = \prod_i^N \frac{1}{\sqrt{2\pi}(\sigma_i^2 + s_j^2)} \exp \left[-\frac{(y_i - \mathcal{M}_{ij})^2}{2(\sigma_i^2 + s_j^2)} \right], \quad (5.8)$$

where \mathcal{M}_j is the StarSim-generated model of the j -th time-series observable with N measurements, and y_i are the observational data points. σ_i is the nominal error of the measurement i , and s_j is a quadrature-added jitter, that accounts for a possible incompleteness of the model, underestimated uncertainties or traces of correlated noise.

5.3.2 Optimising the surface distribution of active regions

Surface maps describing the distribution, size and evolution of the active elements are obtained through maximisation of the figure of merit (Equation 5.7) given a fixed set of stellar parameters, θ , typically including the rotation period, P_{rot} , the spot temperature contrast, ΔT_{sp} , and the facula-to-spot area ratio, Q . Thus, the optimal surface map is given by

$$\hat{\mathcal{S}}_\theta = \arg \max_{\mathcal{S}_j \in \mathcal{S}} \ln \mathcal{L}_\theta(\mathcal{S}), \quad (5.9)$$

where $\hat{\mathcal{S}}_\theta$ is the optimal surface map constrained to a specific set of parameters and \mathcal{S}_j is any surface map among all possible maps, \mathcal{S} .

5.3.2.1 Simulated annealing optimization

In order to optimise the expression in Equation 5.9, we implemented a Monte Carlo Simulated Annealing optimisation algorithm (hereafter MCSA, Kirkpatrick et al. 1983).

Surface map The surface map optimisation starts with a random distribution of a fixed number of active elements, each one characterised by five adjustable parameters: time of appearance, lifetime, latitude, longitude, and angular radius. Active elements do not appear or disappear abruptly. They are assumed to grow or shrink linearly in radius at a rate of 0.5 deg/day (see Herrero et al. 2016, for additional details). The value of Q is assumed to be the same for all spot elements. The total number of active regions

is a parameter of StarSim and can be defined by the user. However, it is advisable to limit the maximum number of active elements thus forcing the model to retrieve simpler surface maps, avoiding the appearance of a large number of very small spots, especially when modelling noisy time-series data.

Optimisation The optimisation process follows successive iterations where the algorithm randomly selects one of the adjustable parameters from a randomly-selected active element and modifies it slightly. Then, with this new map, the forward problem is re-computed and $\ln \mathcal{L}$ is evaluated for each time-series dataset. If such perturbed map improves the quality of the fit as given by Equation 5.7, the change is accepted. Otherwise, it is accepted with a probability $e^{-\beta \Delta \ln \mathcal{L}}$, where β is a parameter that grows in each step. With this strategy, the optimiser avoids getting trapped in local maxima. A more detailed explanation of the implementation of MCSA in StarSim can be found in Herrero et al. (2016).

Due to the large number of parameters describing the stellar surface map and the intrinsic randomness of Monte Carlo algorithms, several maps retrieved with the same θ may not be identical in spite of producing very similar time-series datasets with the forward model. This effect can be mitigated by performing a number of solutions with different initial spot maps and subsequently exploring the variance of the likelihood statistic. For each target and set of observations, these tests can help the user to define parameters such as the number of iterations of the MCSA algorithm per β step and the number of active regions on the stellar surface. The execution time (which increases linearly with both parameters) and some regularisation criteria (selecting the minimum number of spots to avoid unnecessarily complex surface structures) need to be factored in when deciding on the optimal procedure. Once these parameters are adopted and a large number of inversions with varying initial spot conditions are run, the final product of the inversion for a fixed set of parameters, θ , is a spot map calculated by averaging the total sample of optimal maps. The result is a smooth time-evolving surface map that contains valuable information about the surface distribution of the active regions, their typical lifetimes and possible differential rotation.

5.3.3 Optimising stellar parameters

Equation 5.9 describes the optimisation of the surface maps \mathcal{S}_θ when fixing a set of stellar parameters θ . However, tests show that \mathcal{S} has a dependence on θ . Due to the non-linear and multimodal nature of the problem, \mathcal{S}_θ and θ are strongly coupled, and small variations on the parameters potentially imply large changes in \mathcal{S} . Therefore, fitting both sets of parameters simultaneously is computationally challenging. Therefore, we choose a two-step approach.

The retrieval of optimal parameters in this context belong to the class of noisy optimisation problems (e.g. Grill et al. 2015), which are characterised by long evaluation times and noisy outputs for the objective function. In our particular case, we consider a large number of randomly-generated parameter sets drawn according to a prior distribution (generally uniform) and calculate the inverse problem to produce an optimal surface map starting from random initial conditions for each of the parameter sets.

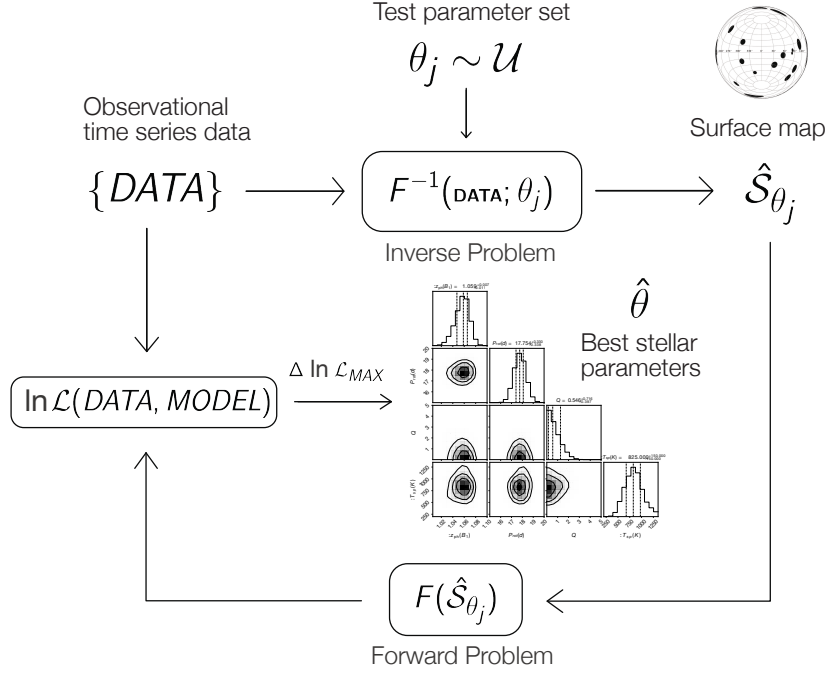


Figure 5.3: Flow chart of the parameter retrieval approach as presented in Section 5.3. The inverse problem block fits the observational data to generate a surface map constrained to a fixed set of stellar parameters, which are drawn uniformly distributed and serve as *test* values. Once the optimal map is obtained, a forward model is computed using this map and parameter set. Finally, the $\ln \mathcal{L}$ is computed and all iterations are sorted to select the N -best solutions according to a maximum $\Delta \ln \mathcal{L}$ in order to build the parameter distributions.

Using their $\ln \mathcal{L}$ values, a certain interval enclosing the best statistically-equivalent solutions can be defined. From the selected best solutions, the optimal parameter set and the corresponding uncertainties can be determined. The procedure can be expressed as

$$\begin{aligned} \{\theta\}^* &= \{\theta_{\ln \mathcal{L}_0}, \theta_{\ln \mathcal{L}_1}, \dots, \theta_{\ln \mathcal{L}_N} | \ln \mathcal{L}_j \geq \ln \mathcal{L}_{j+1}, \Delta \ln \mathcal{L}_{MAX} = \\ &= \ln \mathcal{L}_0 - \ln \mathcal{L}_N \sim \Delta \ln \mathcal{L}_{lim}\}, \quad (5.10) \end{aligned}$$

where $\Delta \ln \mathcal{L}_{lim}$ is the threshold defining equivalent solutions. Then, the median and 68% confidence interval of $\{\theta\}^*$ are used as the best estimates of stellar parameters and their uncertainties. A flow chart of the full StarSim inversion problem is displayed in Figure 5.3.

5.4 Description and handling the degeneracies on the inverse problem

The methodology of StarSim to calculate the inverse problem implies accounting for a large number of parameters and potential correlations among them, which may produce degenerate solutions. This is especially important when only a particular type of data is considered (i.e. photometry or spectroscopy). As degenerate solutions, we understand a set of spot maps and parameters that results in equivalent surface integrated observables. The casuistics can be summarised into the next three categories.

5.4.1 Size–latitude degeneracy

A number of different active region configurations can yield very similar simulated datasets, with figures of merit that are not significantly different. This can be easily illustrated by the example of considering a system with a single spot on an equator-on star ($i_* = 90^\circ$). In such case, the location in either hemisphere produces identical results for all datasets (hemispherical degeneracy). Also, spots with different sizes at different latitudes can produce similar effects, as they may yield the same projected area towards the observer. This results in size–latitude correlations. Such type of degeneracies could in principle be solved through accurate modelling of limb darkening. However, the uncertainty of photometric and spectroscopic measurements usually make limb darkening variations indistinguishable. This is why, especially for stars that are nearly equator-on, we cannot retrieve the latitude distribution of the active regions but only the stellar longitudes occupied by spots. Therefore, retrieved surface maps have to be understood as the filling factor of active regions projected on the longitude axis (see Section 6.4.3.1 for the case study presented here).

At least from a theoretical point of view the size–latitude degeneracy can be solved with photometry of sufficiently high precision. That may capture the tiny variations in light curves due to different limb darkening profiles drawn by the chords followed by spots at different latitudes. Therefore, we could refer to those class of degeneracies as *weak*, in contrast to *strong* degeneracy, such as hemispherical symmetry, which cannot be solved.

5.4.2 Size–temperature degeneracy

Another important potential degeneracy is caused by the correlation between the size of the active regions and their temperature contrast. However, there is potential relief to such degeneracy because the temperature contrast of spots and faculae introduces a chromatic effect, as we have shown above. This can be measurable when multiband time–series photometry is available. When cool spots are present on the stellar surface, the amplitude of the photometric variability is larger at bluer wavelengths, and lower in the red and infrared bands. Such chromatic signature can be studied by performing the inverse problem with StarSim on multiband photometric measurements, thus allowing to derive the temperature contrast of the active regions with respect to the photosphere and thus break the temperature–size correlation.

5.4.3 Spot–facula degeneracy

Different combinations of the filling factor of spots and the amount of faculae, Q , can also result in similar solutions for the modeled observables. This is because bright regions can partially compensate the effects of dark spots. However, the presence of faculae can also be identified when analysing multiband photometric time-series, through proper modelling of limb brightening (as opposed to limb darkening for dark spots) and the corresponding chromatic signature, which differs from the one produced by cool spots. Also, due to the blocking of convection by active regions and the subsequent decrease of the net convective blueshift in radial velocities, the spot–facula degeneracy can also be broken when contemporaneous photometry and high-precision spectroscopy are available.

5.4.4 Latitude–inclination degeneracy

In case of considering the stellar axis inclination (i_*) as a free parameter, another relative orientation degeneracy may occur with those systems with the latitude and inclination parameters being swapped. Hence, it becomes observationally indistinguishable a system with a spot at a latitude θ in a star with an inclination axis i_* , from a system with inclination $i_* = \theta$ and $\theta = i_*$.

5.4.5 Zero point degeneracy

When polar or circumpolar active regions are present, these give rise to a non-modulating effect that produces a constant offset of the photometric flux. As we have mentioned before, such offset is only mildly chromatic and therefore difficult to determine. Performing the inverse problem for these stars if they are not equator on is particularly challenging when only single-band photometry is available. This degeneracy, also referred to as normalisation degeneracy (Luger et al. 2021b) is related to the fact that in the disc-integrated flux of a star the level that corresponds to the immaculate photosphere is a priori unknown, and therefore it is not possible to normalise the simulated light curves with respect to the unspotted surface. If the light curves are normalised with respect to their mean value, degeneracy appears in case of considering surface maps made of combinations of modulating (δ_M) and non-modulating (δ_0) active regions.

As an example, we may consider the comparison of two systems as shown in Figure 5.4. One, is composed of a small equatorial spot, and the other has the same equatorial spot plus a polar cap. Obviously, these two systems have a different spot maps but if the light curves are normalised to the average flux or to the maximum flux level, they are indistinguishable in single-band observations. This is because the immaculate photosphere flux level is in principle unknown.

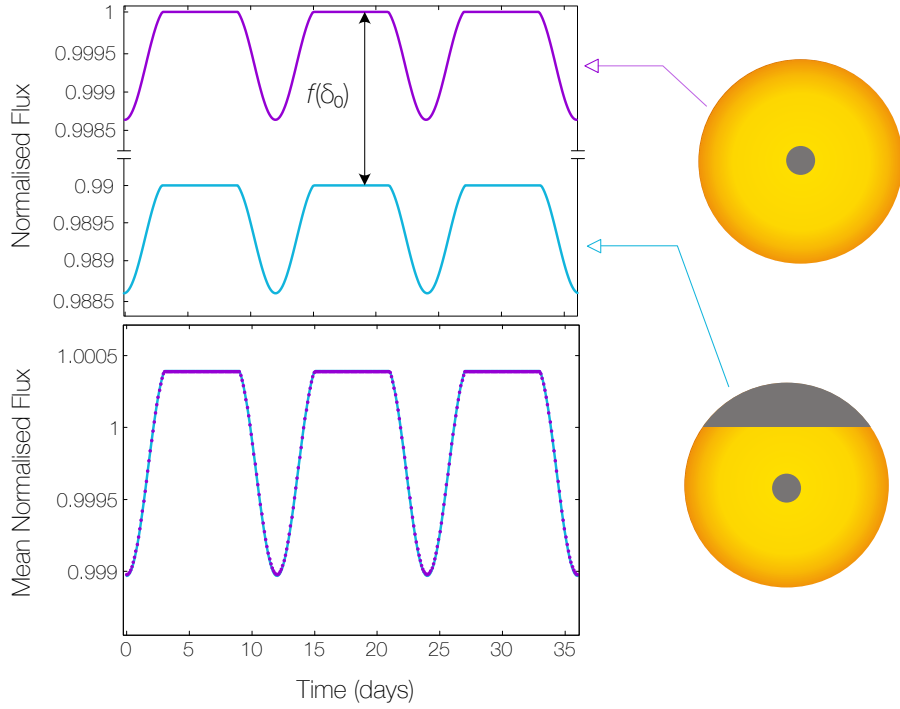


Figure 5.4: StarSim 2 single-band simulation of two examples of heterogeneous surfaces to illustrate their photometric signal dependence on the normalisation used. If the flux is normalised over the immaculate photosphere (top panel) the effect of the polar cap modifies the integrated flux. At practice, the immaculate photosphere flux is unknown. Thus, normalising the light curves over their mean leads to an observational degeneracy impossible to break without color information (lower panel).

5.5 Toy model inversion

As a further test of the validity of our approach, we conducted an inversion exercise using synthetic data generated with the forward functionality of StarSim to check the inversion feasibility.

5.5.1 Model

We considered a star with $T_{ph} = 5000$ K, $i_* = 90$ deg, $P_{rot} = 15$ days and with its surface covered by 5 spots having a $\Delta T_{sp} = 700$ K and no faculae ($Q = 0$). We considered spots of various sizes, placing them at different latitudes and longitudes in such a way that a photometric modulation is present and also ensuring that spotted regions are visible at all times (i.e., projected spot filling factor never being zero), to produce a nonzero non-modulating filling factor.

We assumed the spots to be of constant size and static in the frame of reference of the rotating star. We generated synthetic light curves in the *UBVRIJ* bands covering a

Table 5.1: Results of inversion tests using synthetic data with StarSim. δ_{sp} is the total projected filling factor, $\delta_0 + \delta_M$.

Parameter	Input	Model		
		5 spots	10 spots	20 spots
$P_{rot}(\text{d})$	15	$15.001^{+0.011}_{-0.011}$	$15.002^{+0.011}_{-0.013}$	$14.995^{+0.015}_{-0.012}$
$\Delta T_{sp}(\text{K})$	700	650^{+100}_{-125}	700^{+50}_{-75}	675^{+50}_{-50}
Q	0	< 0.23	< 0.22	< 0.13
$\delta_{sp}(\text{max})$	0.049	$0.048^{+0.004}_{-0.004}$	$0.051^{+0.003}_{-0.003}$	$0.053^{+0.004}_{-0.004}$
$\delta_{sp}(\text{min})$	0.012	$0.012^{+0.002}_{-0.002}$	$0.013^{+0.001}_{-0.001}$	$0.015^{+0.002}_{-0.002}$

timespan of 90 days (6 rotations) and with a uniform observational cadence of one measurement every 0.5 days. Noise was added to the photometric measurements following a heteroskedastic scheme, consisting of non-correlated Gaussian noise $\sim \mathcal{N}(0, \sigma)$ with $\sigma \sim \mathcal{U}(0.0085, 0.00115)$ (i.e., ≈ 1000 ppm).

5.5.2 Analysis

The resulting 6 light curves were inverted following the scheme sketched in Figure 5.3. The fitted parameters were P_{rot} , Q , and ΔT_{sp} . We ran several thousand inversions exploring the relevant parameter space, all starting from a random map. We considered three different assumptions regarding the number of active regions in the model, namely 5, 10, and 20, to evaluate the impact on the quality of the solution and on the resulting filling factor. We furthermore considered that spots have a finite lifetime and that their growth or decrease rate is $0.5 \text{ deg}\cdot\text{day}^{-1}$. Each spot was characterised by six parameters (appearance time, lifetime, longitude, latitude, radius) that were allowed to vary during the optimisation process.

We established a likelihood criterion to select the best maps and solutions based on the inherent scatter of the solution when performing a large number of inversions from random maps and assuming the correct parameters. This yielded a number of 100–150 good solutions for all three cases. The statistical results of such solutions are given in Table 5.1, and a graphical illustration for the 10-spot case is shown in Figure 5.5.

The results of the inversion tests allow assessing the retrieval performance of the StarSim model. As can be seen in Table 5.1, all fitted parameters are retrieved within the uncertainties regardless of the number of spots assumed. No significant bias is observed except for a slight tendency to underestimate the spot temperature contrast. However, additional tests showed that this may be caused by the relatively low number of solutions used and therefore it should not be reason for concern. It is interesting to note that the spot filling factor is also accurately obtained, including the non-modulating fraction. This indicates that the algorithm is able to reduce the size of the spots and place them appropriately to reproduce the correct spottedness of the star. We do not see a dependence on the number of active regions except for a slight tendency to overestimate the filling factor when 20 spots are used.

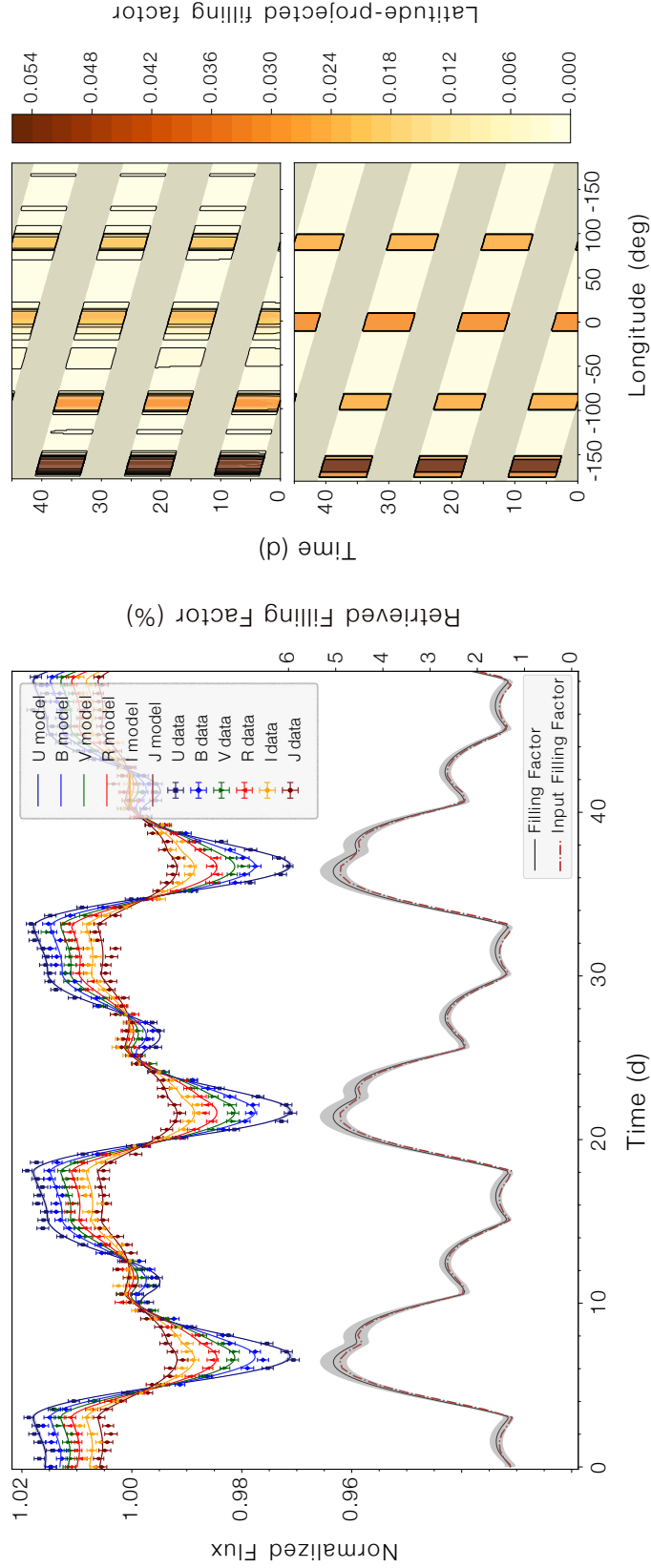


Figure 5.5: The left panel shows the StarSim model fits to multiband synthetic light curves assuming the parameters in Table 5.1, for a subset of 50 days. Solid curves represent the mean of ~ 20 optimal solutions of the inverse problem. The shaded bands indicate the $1-\sigma$ ranges. The gray line at the bottom is the projected spot filling factor of the maps, also showing the mean and $1-\sigma$ ranges. The right panels plot the longitudinal spot filling factor projected on the stellar equator as a function of time, for both the input (bottom) and the retrieved (top) spot maps. Only stellar longitudes visible at each time are shown (hence the band structure). The colour scale indicates the fractional latitude-projected spot coverage for each 1° -degree longitude bin.

The graphics in Figure 5.5 show the high quality of the multiband fits and the low dispersion of the solutions. Also, the comparison of the input latitude-projected filling factor with the inversion results on the right panels of the figure indicates a very precise retrieval of both the longitudes and sizes of the active regions. We additionally ran tests using only the *BVRI* passbands, thus restricting the wavelength baseline. The solutions converged well only with slightly larger uncertainties, as expected. In all cases, the input parameters were well within the error bars.

Chapter 6

An activity model for WASP-52

The findings presented in this chapter are published in Rosich et al. (2020).

6.1 Introduction

As an example of the use of StarSim for the inverse problem, we use multiband photometric time-series data of the active planet host star WASP-52 to retrieve an optimal set of stellar parameters and reconstruct a time-dependent map of the filling factor of active regions.

6.2 The WASP-52 exoplanetary system

WASP-52 is a young and active K2 V star hosting an inflated Jupiter-sized exoplanet with an orbital period of 1.75 days (Hébrard et al. 2013). The inclination of the planetary orbit i is such that results in a transit of the planet. Hébrard et al. (2013), Loudén et al. (2017), Mancini et al. (2017), May et al. (2018), Oshagh et al. (2018), and Öztürk et al. (2019) described and refined the different planetary parameters using light curves and the Rossiter–McLaughlin effect (see Table 6.1), whereas Kirk et al. (2016), Loudén et al. (2017), Chen et al. (2017), and Alam et al. (2018a) described its cloudy sodium-bearing atmosphere. In many of those studies, in-transit anomalies appearing on the transit light curves and the effects of active regions on the stellar photosphere are discussed. May et al. (2018) and Bruno et al. (2018) quantify the differences of the chromatic effect on transit depths from a spotted and unspotted photosphere. The different values found for the stellar rotation period P_{rot} and the spot temperature differences ΔT_{sp} , which are fitted parameters of our inverse problem, are given in Table 6.2.

6.2.1 Transmission spectroscopy early studies

In Alam et al. (2018b), a transmission spectroscopy study was done using three transits observed in the visible and NIR wavelengths with HST/STIS, combined with Spitzer/IRAC

photometry. A stellar activity correction was applied by fitting the baseline flux from ground-based photometry using a quasi-periodic Gaussian process. The effective temperature of the starspots was assumed to be 4750 K, i.e., 250 K cooler than the photosphere. A recent atmosphere study of WASP-52 b by Bruno et al. (2019) combine the STIS and IRAC data from Alam et al. (2018b) and HST/WFC3 from Bruno et al. (2018). The spot effects were corrected using the approach described in Rackham et al. (2017) for an heterogeneous photosphere. In both cases, the stellar photosphere was assumed to be dominated by cool spots. It is worth emphasizing that Bruno et al. (2019) present a joint fit of the atmospheric model and a stellar contamination correction, parameterised by the fraction of stellar surface occupied by activity features and their temperature. Such approach may lead to a biased solution since genuine planetary features could be modelled as stellar contamination.

6.3 Photometric observations of WASP-52

Photometric light curves obtained from two different ground-based observatories, the 1.2-m STELLA telescope at Izaña Observatory in Tenerife and the 0.8-m Joan Oró telescope (TJO) at the Montsec Astronomical Observatory in Catalonia, were used in this work. WASP-52 was observed as part of our monitoring survey VARIability MONitoring of exoplanet host Stars (VAMOS, Mallonn et al. 2018). Measurements cover a time interval of more than 500 days in two different observing seasons (2016 and 2017), and were obtained contemporaneously with both telescopes using four different passbands (*BVRI*).

The STELLA telescope and its wide-field imager WiFSIP (Strassmeier et al. 2004; Strassmeier et al. 2010) obtained nightly blocks of three individual exposures in Johnson B and three in Johnson V from May 2016 until January 2018, completed by three exposure in Cousins I since September 2017. The detector is a single $4k \times 4k$ back-illuminated thinned CCD with $15 \mu\text{m}$ pixels, providing a field of view of 22×22 arcmin. The telescope was slightly defocused to minimise scintillation noise and improve the quality of the photometry (Mallonn et al. 2016).

The data reduction employed the same ESO-MIDAS routines used for previous monitoring programs of exoplanet host stars with STELLA/WiFSIP (Mallonn et al. 2015; Mallonn et al. 2018). Bias and flat-field correction was done by the STELLA pipeline. We used SExtractor (Bertin et al. 1996) for aperture photometry and extracted the flux in elliptical apertures (SExtractor aperture option MAG_AUTO). Data points of low quality due to non-photometric conditions were discarded. Weighted nightly averages were considered for each filter, resulting finally in a total of 112 measurements in the *B* filter, 110 in *V*, and 16 in *I*. The use of nightly averages is justified because the noise floor for each night is typically limited by systematic errors in the measurement procedure and calibration of ground-based photometric observations. We avoid giving excessive weight to nights with larger number of measurements by computing the nightly averages and adding a jitter term to account for random errors not included in the model of the observations such as night-to-night calibration errors. This strategy is appropriate because we are interested in rotational modulation timescales, with the few-hour time domain being irrelevant.

The TJO telescope provided photometric data using its main imager MEIA2. The instrument has a single $2k \times 2k$ back-illuminated thinned CCD yielding a field of view of 12.3×12.3 arcmin and a resolution of 0.36 arcsec/pixel. The images were calibrated with darks, bias and flat fields using the ICAT pipeline (Colome et al. 2006) and differential photometry was extracted using AstroImageJ (Collins et al. 2017). A total of 66 and 77 epochs, i.e., weighted nightly averages, were obtained with the Johnson–Cousins B and R filters, respectively.

6.4 Photometric inverse problem for WASP-52

6.4.1 Fixed StarSim parameters

The basic stellar parameters (relevant to select the appropriate atmosphere models) were adopted to be $T_{\text{eff}} = 5000$ K and $\log g = 4.5$, which are close to the literature values (see Table 6.1). The stellar inclination was chosen to be 90° , which is a reasonable assumption given the measured values of the planet’s orbital inclination and spin-orbit angle. The lifetime of the active elements was assumed to be Gaussian-distributed around 250 ± 100 days. The choice of this value is done considering the full timespan of the series and the fact that there are two separate epochs covered by our datasets, each of ~ 200 days in duration. This allows for the presence of different active groups in each epoch, as well as some common regions in both, creating a flexible evolving map of the active regions. Our selection also matches the range of spot lifetimes from 70 to 350 days described for WASP-52 by Mancini et al. (2017). Several tests showed that those values are not critical in terms of fitting quality and do not produce any bias in other parameters.

Furthermore, we did not consider fitting for differential rotation (which StarSim could) as this would add even more complexity to the parameter space. If significant differential rotation was present, this would be naturally accounted for by the fitting algorithm through the appearance and disappearance of active regions at slowly varying longitudes. The temperature contrast between faculae and the photosphere was fixed to $\Delta T_{\text{fc}} = 50$ K. This is consistent with the results from Solanki (1993) (see Herrero et al. 2016, for further discussion).

As will be shown in Section 6.4.2, our inversion process yields solutions with insignificant facular coverage independently of the temperature contrast ΔT_{fc} . Finally, we set a restriction to the active element colatitudes, namely that we only allowed their presence in one of the stellar hemispheres. This is possible in our case because of latitudinal symmetry. By doing so, we guarantee the absence of spot crossing events when we study planetary transits later in our analysis.

Table 6.1: Important parameters of WASP-52 (top) and WASP-52 b (bottom).

parameter	unit	value	ref.
α	J2000	23:13:58.75	Ga18
δ	J2000	+08:45:39.9	Ga18
Sp.type	-	K2 V	He13
G	mag	$11.95^{+0.13}_{-0.36}$	Ga18
μ_α	mas a^{-1}	-6.914 ± 0.079	Ga18
μ_δ	mas a^{-1}	-44.248 ± 0.054	Ga18
Distance	pc	175.7 ± 1.3	Ga18
M_*	M_\odot	0.81 ± 0.05	Ma17
R_*	R_\odot	$0.860^{+0.021}_{-0.027}$	Ga18
L_*	L_\odot	0.4189 ± 0.0046	Ga18
T_{eff}^a	K	5010^{+80}_{-60}	Ga18
$\log g^a$	cgs	4.582 ± 0.014	He13
age	Ga	$0.4^{+0.3}_{-0.2}$	He13
$\log R'_{HK}$	cgs	-4.4 ± 0.2	He13
λ	deg	$5.4^{+4.6}_{-4.2}$	Öz18
M	M_J	0.46 ± 0.02	He13
r	R_J	1.223 ± 0.062	Öz19
P^b	days	1.7497828 ± 0.0000006	Öz19
T_0^b	BJD	$2,405,793.68128 \pm 0.00049$	Öz19
r/R_*^b	-	0.159 ± 0.004	Öz19
b/R_*^b	-	0.60 ± 0.02	He13
Orbit inc., i^b	deg	5.24 ± 0.84	Öz19
a	au	0.0272 ± 0.0003	He13

(a) fixed parameters in Section 6.4

(b) fixed parameters in Section 7

References:

He13: Hébrard et al. (2013)

Kil6: Kirk et al. (2016)

Ma17: Mancini et al. (2017)

Os18: Oshagh et al. (2018)

Ga18: Gaia Collaboration et al. (2018)

Öz19 (Öztürk et al. 2019)

Optimiser configuration As discussed in Section 5.3.2, there are some other parameters that need to be set beforehand. One is the number of iterations of the MCSA algorithm. Ideally, this should be sufficiently large to ensure consistency in the resulting surface map but at the same time be computationally affordable. The number of iterations is also linked to the quantity of active elements considered on the stellar surface, since each element contributes 5 potential parameters. This number has to be sufficient large to generate the inhomogeneities of the stellar surface causing the photometric variability, yet not as large as to overfit instrumental noise. It partially depends on the timespan of the dataset and the lifetime of spots, and the maximum radius of the active elements, which we set to 10 degrees to be consistent with the small spot approximation.

We ran a battery of tests considering different number of iterations and surface active elements. An initial exploration of the parameter space was used to obtain valid guesses of the optimal parameters P_{rot} , ΔT_{sp} , Q , and s (jitter). We systematically explored combinations of number of iterations, from 500 to 10 000, and number of spots from 60 to 150 and evaluated the resulting $\ln \mathcal{L}$ values. For each of the explored pairs, we ran 112 realisations starting with different random spot maps, except for the runs with 10 000 iterations, for which we considered 56 realisations.

Results The results of the tests are shown in Figure 6.1. It is not surprising to see that the quality of the fits improves with the number of iterations, as it also does with the number of spots. The criterion to select an adequate number of iterations is mostly related to computational cost. For the present problem, numbers above 3 000 iterations per MCSA step are prohibitive. Nevertheless, 3 000 iterations already delivers very consistent solutions from random starting points. The final variance of the statistic $\ln \mathcal{L}$ from the sample converged solutions is about 5, which is sufficient to guarantee stable solutions. For lower numbers of iterations, this number increases to 7 (1500) and 17 (500). Note that for 10 000 iterations ($3\times$ longer computational time) the final variance is approximately 3. Regarding the number of active elements we see that less than 80 spots is clearly insufficient to fit our data (precision & timespan); and we also see that the improvement in likelihood flattens out quite apparently beyond 120 spots. For simplicity arguments, and factoring in again computational costs, we find that using 100 spots delivers sufficiently reliable fits, and note as well that the average $\ln \mathcal{L}$ values for 150 and 100 spots overlap at the 1-sigma level. Thus, our adopted values regarding the number of MCSA iterations and spots was 3 000 and 100, respectively, and this produces a optimal solutions with a variance of $\sigma_{\ln \mathcal{L}} = 4.8$.

6.4.2 Photometric fits results

The available datasets consist of multiband photometry and we are especially interested in exploring the parameters related to the chromatic properties of active regions, which play a major role on the characterisation of the effects of activity on transit spectroscopy measurements studied in Chapter. 7. The parameters describing stellar properties must be fitted simultaneously to provide consistent solutions, as explained in Section 5.3.3. In our case, besides the 100-spot map, those parameters are the rotation period, P_{rot} ,

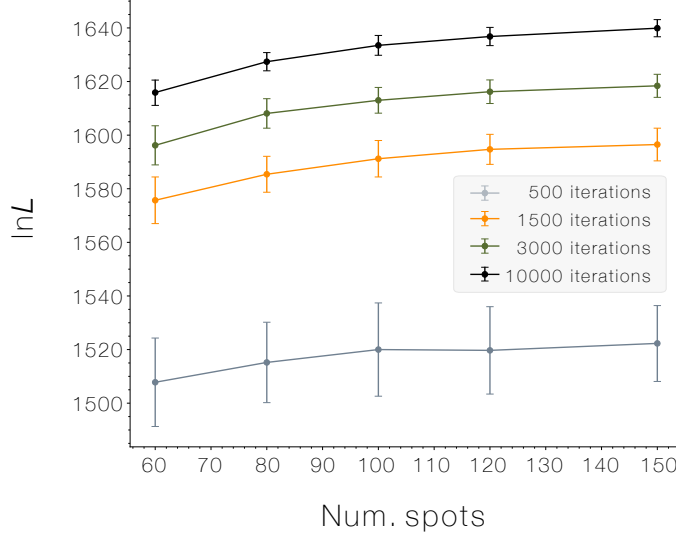


Figure 6.1: Trial tests showing the statistic $\ln \mathcal{L}$ as a function of the number of spots and for several values of the number of iterations of the MCSA algorithm. The size of the error bars corresponds to the 1-sigma intervals around the optimal solutions found from 112 random initial spot maps, except for the case with 10 000 iterations, where we employed 56.

which drives the timescale of the variability, the facula-to-spot area ratio, Q , which determines the fractional coverage of bright active elements, the temperature contrast of the spots, $\Delta T_{\text{sp}} = T_{\text{ph}} - T_{\text{sp}}$, and the additional noise term or jitter, s .

As discussed in Section 5.3.2 and shown in Figure 5.3, the inversion procedure with StarSim consists of firstly selecting a parameter set drawn from a prior distribution covering the relevant parameter space, and subsequently optimizing random spot maps for each realisation. The resulting statistic (Equation 5.7) is then used to evaluate the relative quality of each set of parameters. We have shown before that the intrinsic variance of the statistic for our problem is $\sigma_{\ln \mathcal{L}} = 4.8$, which indicates that differences of such order for different parameter sets are statistically indistinguishable. Therefore, the recipe that we adopt is taking the best (larger $\ln \mathcal{L}$) solution from the batch and then considering as statistically equivalent those that are within 3-sigma ($\Delta \ln \mathcal{L}_{\text{lim}} = 14.4$). We are aware that this is not completely satisfactory because the interval may include some solutions with mildly sub-optimal parameter sets, but we adopt this criterion to ensure a statistically significant collection of *good* solutions at the expense of some increase in the parameter uncertainties.

Increasing the number of MCSA iterations or speeding up convergence could make this criterion more stringent. Admittedly, this is a limitation of the current algorithm imposed by the complexity of the optimisation method that should be improved in subsequent releases of StarSim.

After initial tests, for the case of WASP-52 we adopted uniform priors on the parameter sets within the ranges given in Table 6.2. The priors are quite narrow to avoid

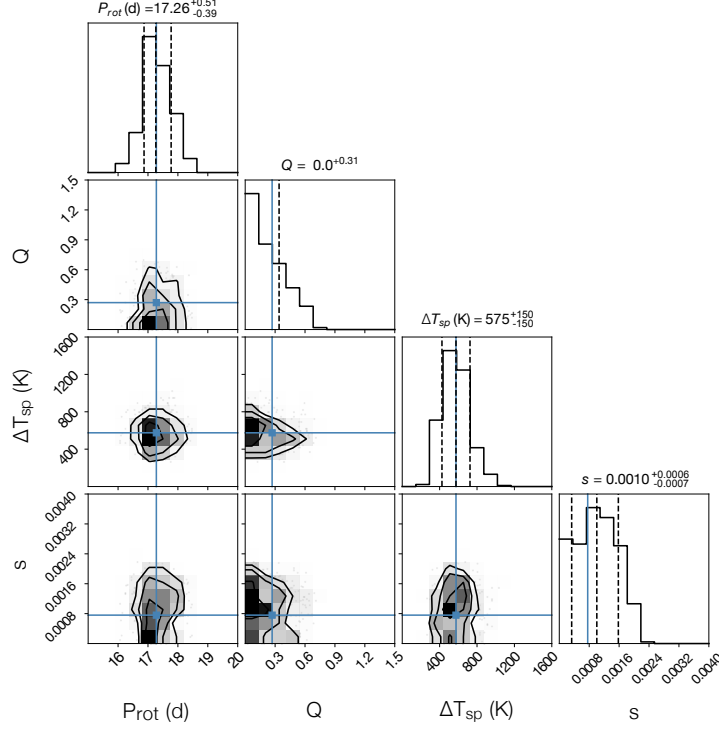


Figure 6.2: Histograms and correlations of 365 statistically-equivalent solutions for the fitted StarSim parameters: the rotational period, P_{rot} , the facula-to-spot area ratio, Q , the spot-photosphere temperature contrast, ΔT_{sp} , and the photometric jitter, s . The dotted lines mark the median and 68% percentiles of the parameter distributions and the blue lines indicate the parameters of the solution with the best likelihood value found. Plot created with `corner.py` (Foreman-Mackey 2016)

unnecessarily exploring irrelevant parameter space but, as seen below, the intervals are sampling regions beyond 3-sigma from the optimal parameters. A total of 21 296 realisations were performed and 108 satisfied the likelihood criterion $\Delta \ln \mathcal{L} < 14.4$ from the highest likelihood one. In a second step, we further narrowed down the parameter priors to include only the ranges defined by the 108 good solutions. An additional 2576 realisations yielded another 257 good solutions, leading to a grand total of 365.

These are the θ -configurations that were used to build the parameter distributions. The distributions are plotted in Figure 6.2, and the relevant optimal parameters and their uncertainties are listed in Table 6.2, compared with other values from the literature.

Table 6.2: Priors and results from the StarSim inversion of WASP-52 photometry. The values given correspond to the optimal solution and uncertainties are estimated from a sample of 365 equivalent solutions. Literature values for the same parameters are provided when available.

Parameter	Prior	Value	Ref.
P_{rot} (days)	$\mathcal{U}(16.0, 18.5)$	$17.26^{+0.51}_{-0.39}$	This work
		$11.8 \pm 3.3^{(a)}$	He13
		16.40 ± 0.04	He13
		15.53 ± 1.96	Ma17
		17.79 ± 0.05	Lo17
ΔT_{sp} (K)	$\mathcal{U}(50, 2000)$	18.06 ± 0.2	Br19
		575 ± 150	This work
		$1250 - 1500$	Ki16
		~ 270	Ma17
		~ 950	Br18
$Q^{(b)}$	$\mathcal{U}(0, 3)$	250	Al18
		2230	Br19
		< 0.31 (1- σ)	This work
s (jitter)	$\mathcal{U}(0, 0.0045)$	$0.0010^{+0.0006}_{-0.0007}$	This work

(a) Obtained from spectral line broadening.

(b) Q is unilaterally distributed.

References:

He13: Hébrard et al. (2013);

Ki16: Kirk et al. (2016)

Ma17: Mancini et al. (2017)

Lo17: Louden et al. (2017)

Al18: Alam et al. (2018b)

Br18: Bruno et al. (2018)

Br19: Bruno et al. (2019)

Stellar parameters results Our model favours a heterogeneous surface dominated by dark spots with a temperature contrast of 575 ± 150 K with respect to the surrounding photosphere. This value is intermediate to those found in the literature, which can be separated into two groups, one with low-temperature contrasts of ~ 250 K and another one with contrasts $\gtrsim 1000$ K. The temperature contrast we find is generally lower than found by Andersen et al. (2015) for K dwarfs and also lower than the prediction by the relationship of Berdyugina (2005), which would give ~ 1300 K for a K2 V star. This is, however, not surprising since the strong degeneracy between spot temperature contrast and filling factor can severely bias some determinations. We believe that our value, based on multi-colour photometry, provides a robust estimate. As already mentioned, the results from our analysis do not support a significant presence of bright faculae as expected for rapidly-rotating young K dwarfs (Radick et al. 1983; Lockwood et al. 2007). The 365 best solutions indicate a unilateral distribution consistent with $Q = 0$. We furthermore find a rotation period of $17.26^{+0.51}_{-0.39}$ days, which is in marginal agreement with the values found by Loudén et al. (2017) and Bruno et al. (2019).

6.4.3 Multiband fits

The best fits to the observational multiband photometry using the modeled parameters of the star and their associated spot maps are shown in Figure 6.3, together with the multiband photometric measurements. The solid lines show the average models resulting from the optimal 21 surface maps, and the shaded bands show the $1-\sigma$ range. The two epochs as shown in the top and bottom panel are separated by a ~ 140 -day gap. The gray curves at the bottom of each panel show the projected spot filling factor, also showing the average model and the $1-\sigma$ range. During our observational campaign, total filling factor values covered a range $\sim 3\%$ - 14% along an observational timespan of ~ 600 days in 2016–2017. This implies flux modulations of $\sim 4\%$ - 7% due to time-varying spot coverage.

6.4.3.1 The evolving surface of WASP-52

In addition to the optimal parameters $\hat{\theta}$ and their uncertainties, our inversion realisations also provide a picture of the stellar surface as a function of time. From the 365 accepted solutions we selected only those satisfying $\Delta \ln \mathcal{L} < 4.8$ with respect to the best likelihood value. This yields 21 maps, which should be a good representation of the optimal model considering both the intrinsic randomness of the map inversion from the MSCA algorithm and the statistical variance of the stellar parameters.

Longitudinal projected filling factor Figure 6.4 shows the longitudinal spot filling factor projected on the stellar equator, averaged for the 21 optimal maps, as a function of time. We perform a latitudinal projection because of the degeneracy present (see Section 5.4). The longitudes in the map are plotted in the reference frame of the rotational period found in our analysis (see Table 6.2). The map suggests that there is a clear dominant active region at a longitude of ~ 30 deg at the beginning of the first epoch, which changes into a more complex longitudinal pattern later in the season,

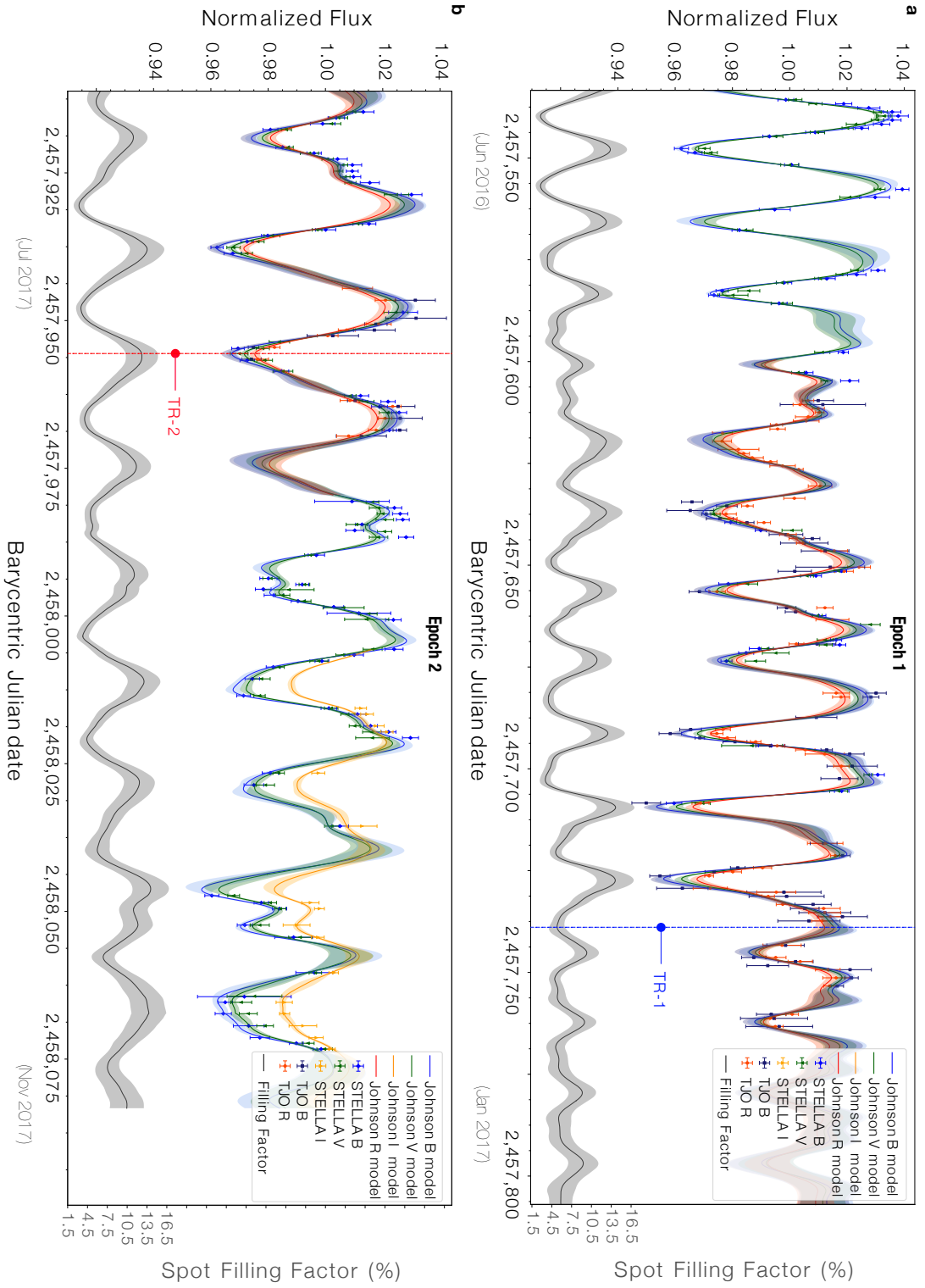


Figure 6.3: Starsim model fits to multiband photometric ground-based observational data as described. Solid curves represent the mean of 21 optimal solutions of the inverse problem. The shaded bands indicate the 1- σ ranges. The gray line at the bottom is the projected spot filling factor of the maps, also showing the mean and 1- σ ranges. TR1 and TR2 indicate transit events as discussed in Chapter 7

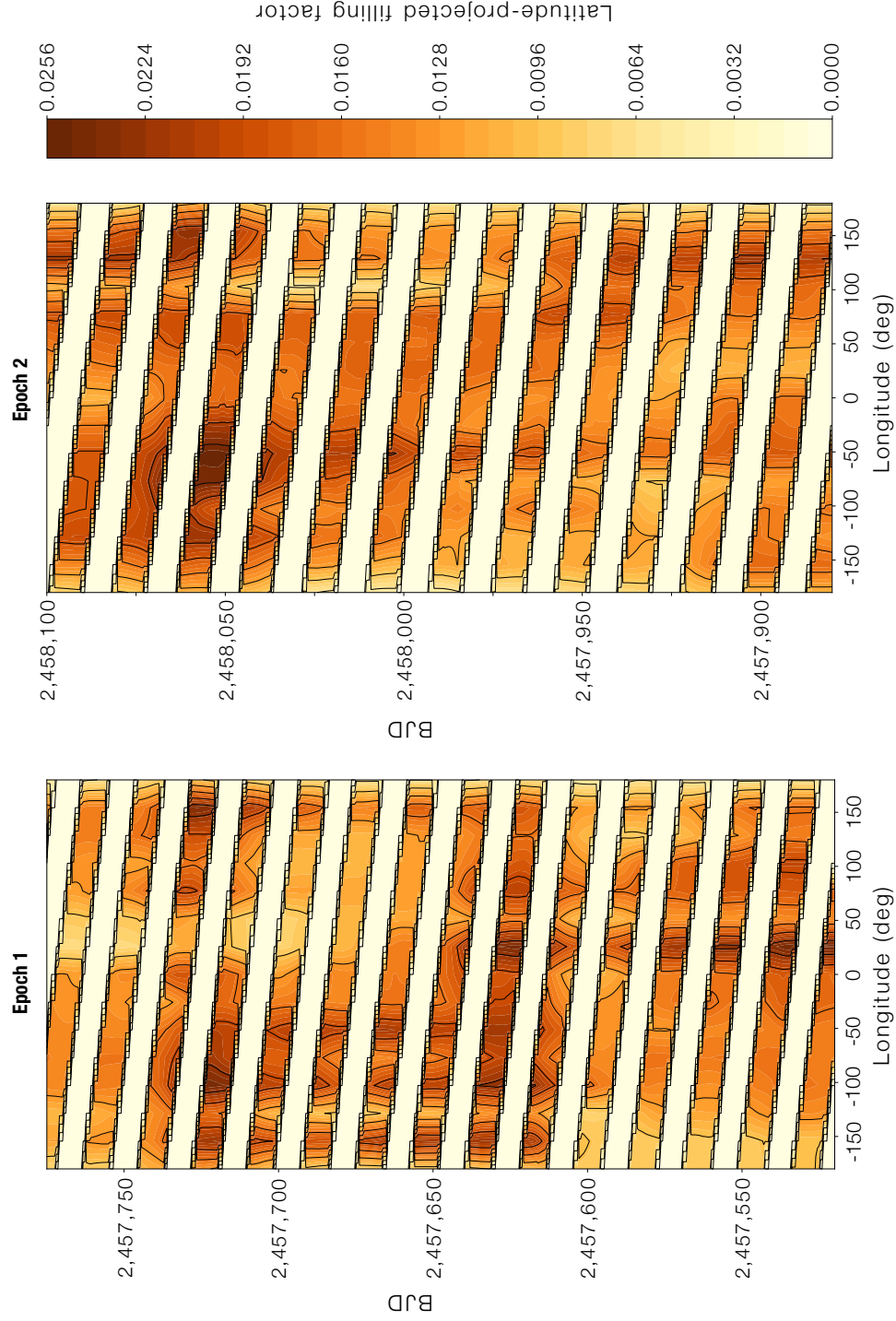


Figure 6.4: Longitudinal spot filling factor projected on the stellar equator of WASP-52 as a function of time, covering photometric epochs in 2016 and 2017 and corresponding to the average of 21 optimal solutions. Only stellar longitudes visible at each time are shown (hence the band structure). The colour scale indicates the fractional latitude-projected spot coverage for each 15-degree longitude bin.

and finally the spottedness level suffers a sudden decrease. Active regions typically last for ~ 8 rotations or ~ 140 days. The second epoch appears to show a better-defined dominance of spot regions, with two alternating active longitudes at about ~ 130 deg and ~ -50 deg, i.e., in almost perfect opposition.

Differential rotation The distribution does not indicate signs of differential rotation along the sampled period of time. The existence of measurable differential rotation would be seen in Figure 6.4 through the presence of two or more active regions with drifting longitudes. We do not see such effect in the time-series of WASP-52 thus probably indicating that either differential rotation is negligible or that dominant active regions appear at similar latitudes. Such dominant regions are used by the fitting algorithm to define the rotation period.

Chapter 7

Chromatic spot effects on simulated transits

The findings presented in this chapter are published in Rosich et al. (2020).

7.1 Introduction

Surface magnetic features have been identified as a source of systematic effects for transmission spectroscopy. Studying the imprint of active regions is of crucial importance for future instruments and projects aiming at the study of exoatmospheres such as, e.g., the JWST (James Webb Space Telescope; Gardner et al. 2006) and Ariel (Atmospheric Remote-sensing Infrared Exoplanet Large-survey; Tinetti et al. 2018; Encrenaz et al. 2018) space missions.

In this chapter, we discuss the chromatic effect of rotational-induced photometric variability on planetary transits. We explore the behaviour of the transit depth with wavelength by means of the surface modelling simulation techniques developed in previous sections, and we propose a methodology to estimate and correct them out.

Transiting events offer a unique opportunity to study planetary atmospheres by transmission spectroscopy technique. The central idea behind this technique relies on the chromatic dependence of the planetary transit depth since potential atmospheres absorb selectively certain wavelengths. At high opacity, the planetary radius appears larger than at wavelengths of weaker opacity (Sing et al. 2016; Mallonn et al. 2017). However, brightness inhomogeneities caused by magnetic activity, starspots or faculae (Strassmeier 2009), modify the derived transit parameters and can mimic spectral features of the planetary atmosphere (e.g. Oshagh et al. 2013b; Herrero et al. 2016; Mallonn et al. 2018). For example, a trend of increasing planetary radius toward blue wavelengths can be caused by scattering properties in the atmosphere of the planet, but also by the presence of starspots on the visible hemisphere of the host star (e.g. Jiang et al. 2021). Moreover, simple spot models compatible with observed photometry, may

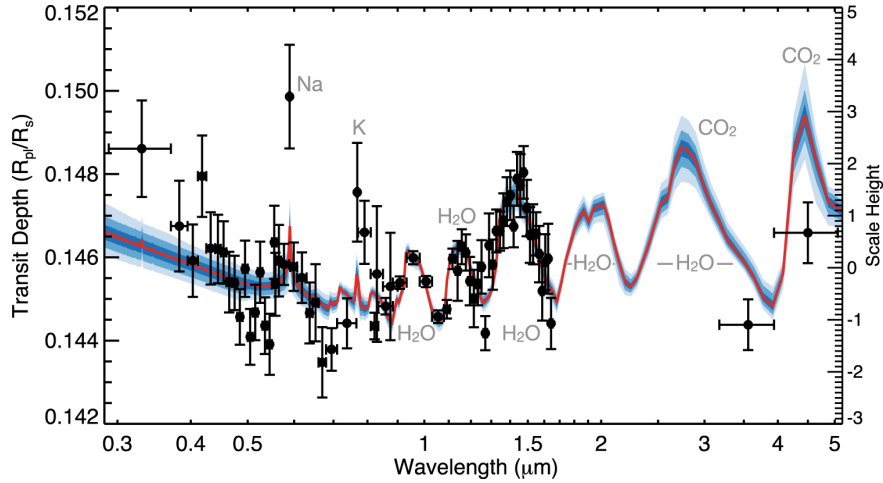


Figure 7.1: Transmission spectra of WASP-39 b using observations from HST STIS and WFC3, Spitzer IRAC, and VLT FORS2 instruments from 0.3 to 5.0 μm . Using the ATMO retrieval code, a best-fit atmospheric model is determined (red line). The 1, 2, and 3 σ confidence regions are displayed in dark to light blue. The small depth variability induced by the atmosphere ($\sim 6 \cdot 10^{-3}$) highlights the importance of a precise activity correction to confidently extract the atmospheric features. (From Wakeford et al. 2017)

reproduce depth profiles leading to the detection of singular molecules in transmission spectroscopy studies (Barclay et al. 2021).

The variability induced by exoatmospheres in the chromatic transit depth is tiny. Moreover, stellar contamination signal can be more than an order of magnitude larger than the transit depth changes expected for atmospheric features in rocky exoplanets. In Figure 7.1 shows an example of atmosphere model retrieval for WASP-39 b. The overall amplitude of those variations is smaller than $6 \cdot 10^{-3}$, therefore, precise modelling of spot-driven activity is needed to disentangle the atmospheric features from those stellar effects.

7.2 Transit depth variability due to spots

For magnetically-active stars showing spots on their surface, the transit depth caused by a planet passing in front of the host star depends both on the planetary effective radius and the potential inhomogeneities on the stellar photosphere. The effect of unocculted spots induces a negative slope of the depth profile, meaning that the depth of transit is enhanced due to the flux balance as the planet blocks photosphere instead of spots —with lower flux—. On the contrary, unocculted bright regions —with higher flux— the balance makes the transit to appear shallower. The effect of spot crossing events appears on transit curves as bumps with a positive or negative contribution depending on the relative brightness of the occulted surface. In Figure 4.3 some spot-crossing simulations are displayed.

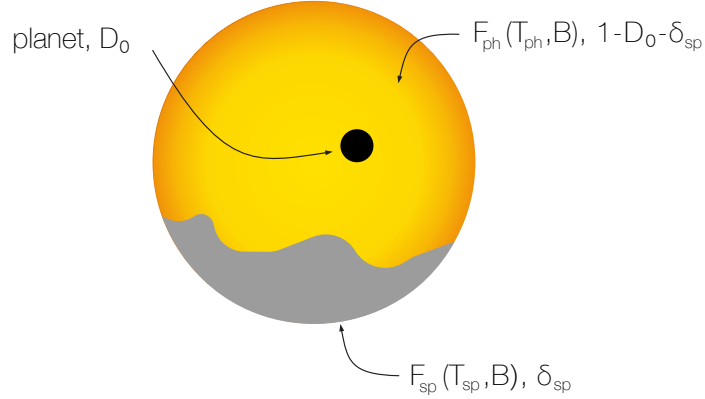


Figure 7.2: Spotted transit sketch. The grey area represents a cold active region composed both modulating and non-modulating components ($\delta_{sp} = \delta_M + \delta_0$).

7.2.1 Analytical formalism

Using the same nomenclature used in Section 5.2, it is straightforward to show that a planet producing a depth \mathcal{D}_0 when transiting an immaculate star, produces a transit depth \mathcal{D}_{sp} when crossing a spotted star (see sketch in Figure 7.2) given by,

$$\mathcal{D}_{sp}(\lambda) = \frac{\Delta F}{F_0} = \frac{F_{ph} \mathcal{D}_0}{F_{ph}(1 - \delta_{sp}) + F_{sp} \delta_{sp}} = \frac{\mathcal{D}_0}{1 - \delta_{sp} \left(1 - \frac{F_{sp}}{F_{ph}}\right)} \quad (7.1)$$

redefining $\Phi^{-1} = 1 - \frac{F_{sp}}{F_{ph}}$, the spotted depth is,

$$\mathcal{D}_{sp}(\lambda) = \frac{\mathcal{D}_0}{1 - \delta_{sp} \Phi^{-1}(\mathcal{B}, \Delta T_{sp})}, \quad (7.2)$$

with δ_{sp} being the total spot filling factor, and \mathcal{D}_0 the non-chromatic transit depth.

Such very simple model illustrates the chromaticity of the problem, since the function Φ is wavelength dependent. However, the latter expression only accounts for the total filling factor, ignoring the spatial distribution of the heterogeneities across the disc and those effects when limb darkening is accounted for.

We can employ StarSim 2 to estimate the impact of spots on the transit depth of WASP-52, by simulating transits of the planet. According to the ephemeris of WASP-52, a total of 112 and 104 transits occurred during the timespan covered by the first and second seasons of our light curves, respectively. We used the StarSim model of stellar activity obtained in Section 6.4 for WASP-52. It is important to emphasise that we solely focus on the effects of unocculted starspots, whose properties cannot be constrained from the transit itself. Therefore, in our simulations we avoided planet-spot crossings by positioning active regions outside the swath covered by the planet during the transit.

Occulted spots during transits may also have a significant impact on the planet properties. However, if photometry of sufficient precision and time cadence is available (such as expected for JWST and Ariel), spot-crossing events could be identified from

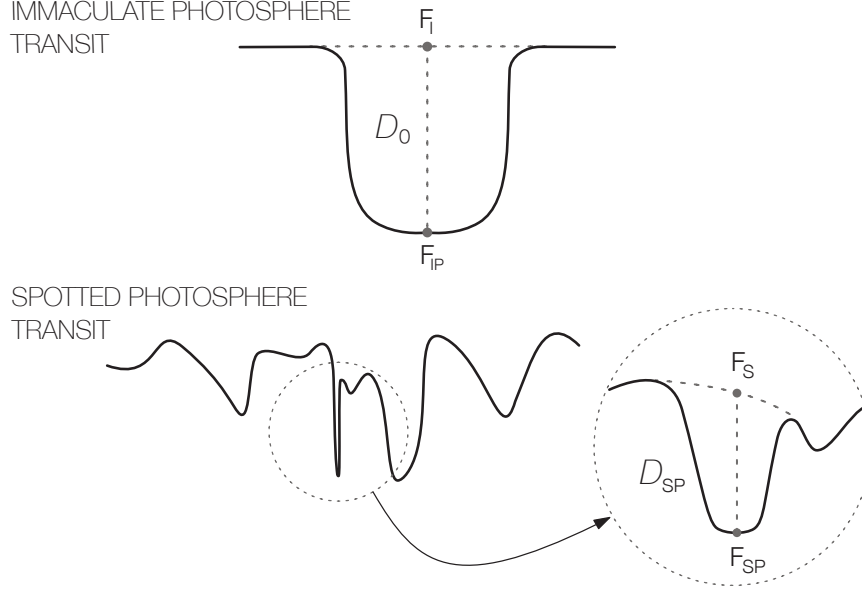


Figure 7.3: Definition of transit depth with immaculate and spotted photosphere. F_I , F_{IP} and F_S , F_{SP} refers to flux out/in transit for immaculate and spotted photosphere, respectively.

the transit observations themselves and potentially modelled. Unocculted spots, on the contrary, leave no visible signature.

We define the transit depth as the difference between the out-of-transit and the in-transit (at transit mid-time) observed flux. The out-of-transit is the flux of the model without transit calculated at the mid-transit time (Figure 7.3). Hence, both stellar spots and limb darkening effects cause variations that depend on wavelength. Note that we take this approach for computational simplicity, because we only need to evaluate the flux at the transit mid point and not the full transit event. The transit depth can then be written as,

$$\mathcal{D}'(\lambda) = \mathcal{D}_0 + \text{LD}(\lambda) + \text{SP}(\lambda), \quad (7.3)$$

where, $\text{SP}(\lambda)$ and $\text{LD}(\lambda)$ are the additive chromatic effects on depths induced by spots and by stellar limb darkening, respectively, and \mathcal{D}_0 is the non-chromatic (asymptotic) transit depth, which we assume to be constant, i.e. we do not consider here the planet atmosphere. This can be written as

$$\mathcal{D}_0 = \left(\frac{r}{R_*} \right)^2, \quad (7.4)$$

where r and R_* are the planetary and stellar radii, respectively. Here, we assume $\mathcal{D}_0 = 0.0253$ for WASP-52 b (Öztürk et al. 2019). The transit depth including chromaticity effects can then be expressed as

$$\mathcal{D}'(\lambda) = \left(\frac{r + \Delta r}{R_*} \right)^2, \quad (7.5)$$

where $\Delta r = \Delta r(\lambda)$ accounts for the effect of spots and limb darkening on the planetary radius as a function of wavelength. Expanding Equation 7.3, and neglecting second order terms, we derive

$$\mathcal{D}'(\lambda) - \mathcal{D}_0 = 2 \left(\frac{r}{R_*} \right) \left(\frac{\Delta r}{R_*} \right) \quad (7.6)$$

then,

$$\frac{\Delta r}{R_*} = \frac{1}{2} \frac{1}{\sqrt{\mathcal{D}_0}} [\text{LD}(\lambda) + \text{SP}(\lambda)]. \quad (7.7)$$

Therefore, we can separate the chromatic contributions due to limb darkening (Δr_{LD}) and spots (Δr_{SP}) as,

$$\frac{\Delta r_{\text{LD}}}{R_*} = \frac{1}{2} \frac{1}{\sqrt{\mathcal{D}_0}} \cdot \text{LD}(\lambda), \quad (7.8)$$

and

$$\frac{\Delta r_{\text{SP}}}{R_*} = \frac{1}{2} \frac{1}{\sqrt{\mathcal{D}_0}} \cdot \text{SP}(\lambda). \quad (7.9)$$

7.2.2 Transit simulations

Differently than the case of the inverse problem seen in the previous chapter, in transit simulation, StarSim employs the grid integration technique to compute the observables. To evaluate the effect of spottedness in transit depth, we simulated a series of transits using the parameters displayed in Table 7.1. The retrieved optimal spot maps in the fitted activity model, along with the stellar parameters for WASP-52 are used in these simulations. Since the effect of spots on the depth is usually small, the size of the grid is set to 0.05° to make negligible the numerical noise from the tessellation of the surface. To avoid spot crossings, the activity model was forced to reject allocating spots in the path swept by the planet during the transits, by defining an exclusion latitude range.

Results Figure 7.4 shows the results of the transit depth simulations. The shaded gray region in panel (a) displays the range of transit depth values as a function of wavelength for the simulated transits comprised within the timespan of our datasets. Such transits differ in the amount and distribution of spots in the projected stellar surface. As a reference, the red line shows the transit depth in the case of an immaculate star. In this latter case, $\mathcal{D}'_{\text{imm}} = \mathcal{D}_0 + \text{LD}(\lambda)$, the wavelength dependence is solely due to the limb darkening effect, which for WASP-52 produces transit depths from 0.0285 to 0.0256 in flux ratio units from the visual to the infrared bands, respectively. As reference, the dot-dashed line indicates the expected transit depth for a uniform planetary disk corresponding to $\mathcal{D}_0 = (r/R_*)^2 = 0.0253$. The shaded grey region in panel (b) shows the same results when removing the contribution of limb darkening, $\Delta \mathcal{D}' = \mathcal{D}'_{\text{sp}} - \mathcal{D}'_{\text{imm}} = \text{SP}(\lambda)$, where \mathcal{D}'_{sp} and $\mathcal{D}'_{\text{imm}}$ denote the depth in a spotted and immaculate photosphere, respectively.

The results indicate that transit depth variations of WASP-52 b due to spots vary from $\sim 2 \cdot 10^{-4}$ to $\sim 3 \cdot 10^{-3}$ in the visual region, and from $\sim 10^{-4}$ to $\sim 7 \cdot 10^{-4}$ in the NIR region. The range in the variations is caused by both different spot filling factors

Table 7.1: StarSim simulation parameters for WASP-52 transits. Those parameters tagged as *model distribution* refers to those statistically equivalent solutions found in the photometric model developed in Section 6.4.2.

Parameter	Value
Spectral range [nm]	300-8000
T_{eff} [K]	5000
$\log g$	4.5
Spot temperature contrast, ΔT_{sp} [K]	<i>model distribution</i>
Facula temperature contrast, ΔT_{fc} [K]	150
facula–spot surface ratio, Q	<i>model distribution</i>
Rotation period, P_0 [days]	<i>model distribution</i>
Rotation axis inclination, i [°]	90
Differential rotation constant, δ	0
Spot growing factor [°/day]	1.5
Grid size [°]	0.05
Planet radius [star units]	0.159
Planet parameter of impact	0.6
Planet orbital period [days]	1.7497828
T_0 [BJD]	2 456 862.79776

during each transit, and the uncertainty on the exact distribution of spots. In other words, our results show that the transit depth of WASP-52 b can be overestimated by up to $\sim 12\%$ and $\sim 3\%$ in the visual and NIR bands, respectively, thus potentially affecting the retrieval of planetary atmosphere parameters, which could be one or two orders of magnitude smaller (Kreidberg 2018).

To understand the impact of different spot configurations we selected two transit events representing low- and high-activity levels at well-sampled epochs of the light curves. These are labeled in Figure 6.3 as TR-1 at BJD = 2,457,739.475 (16 Dec 2016, low-activity case), when the projected filling factor of spots is estimated to be close to its minimum value around 5%, and as TR-2 at BJD = 2,457,954.698 (20 Jul 2017, high-activity case), when spot projected filling factor reaches $\sim 12\%$. Panel (b) of Figure 7.4 shows the transit depth variation of TR-1 and TR-2 in blue and red, respectively, for the 21 optimal maps, with vertical error bars indicating the $1\text{-}\sigma$ variance. These test cases clearly show that the strength of the chromatic effect is ~ 3 times larger when the projected spot filling factor is larger by about the same factor. This implies that the effect will also be correspondingly larger for stars showing higher levels of stellar activity.

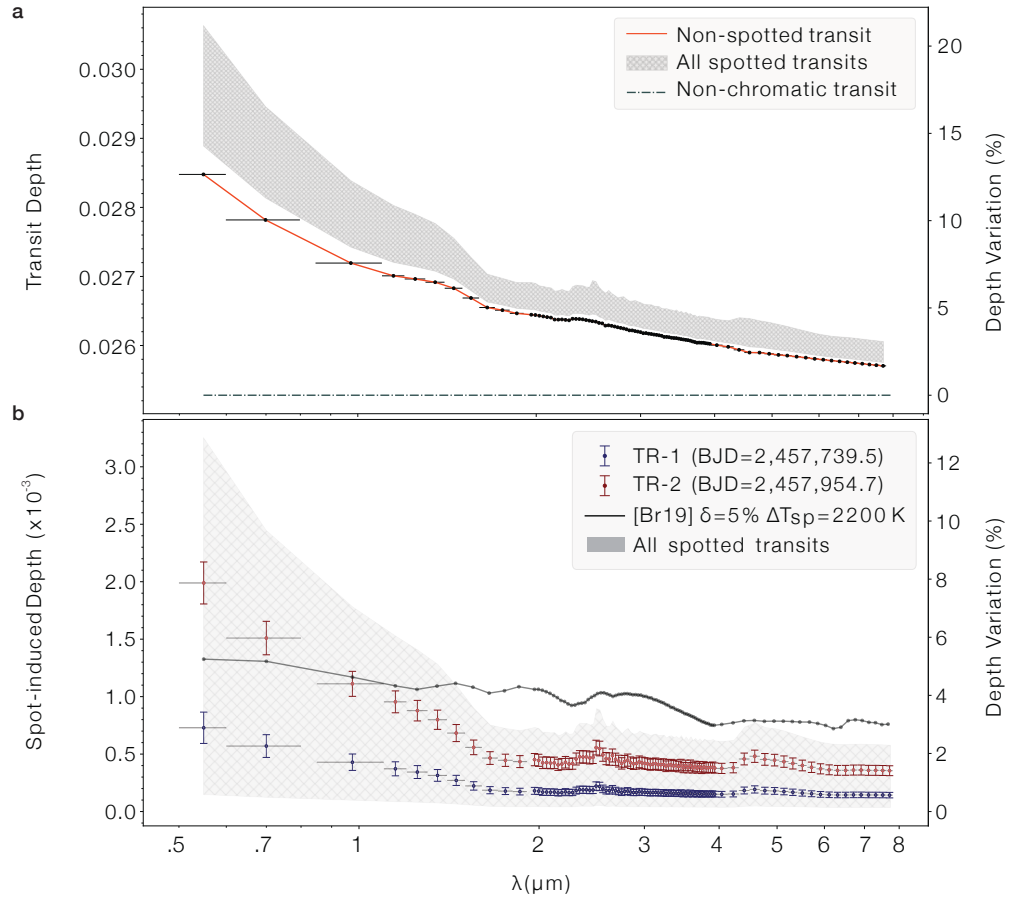


Figure 7.4: (a) Transit depth as a function of wavelength for the WASP-52 system simulated assuming an immaculate surface (red) and including spots in the photosphere (gray band), avoiding transit spot crossings. The gray area illustrates the total coverage from all simulated transits for the 21 optimal spot configurations, clearly showing the transit depth chromatic bias produced by non-occulted spots. The dotted blue line is the non-chromatic transit depth assuming a uniform photosphere ($(r/R_*)^2$) according to the parameters in Öztürk et al. (2019).

(b) Chromatic signature on transit depths due to spots after subtracting limb darkening effect (in red in top panel). TR-1 and TR-2 show two fiducial transits corresponding to low- and high-activity —marked with vertical lines in Figure 6.3— respectively, and the gray band represents the area covered by all simulated transits. The black solid line with symbols shows the chromatic effect estimated using stellar spot parameters derived by Bruno et al. (2019).

Table 7.2: Mean values of transit depth variations due to spots, $\text{SP}(\lambda) = \Delta\mathcal{D}'$, on WASP-52 simulated transits. The corresponding planet relative radius variations, $\frac{\Delta r_{SE}}{R_*}$, are computed following Equation 7.9. Values for the low-activity transit TR-1 and the high-activity transit TR-2 cases are provided. λ_{eff} refers to the central wavelength of the corresponding channel.

Instrument	$\lambda_{\text{eff}} (\mu\text{m})$	$((\text{SP}(\lambda)) \pm \sigma) \times 10^{-3}$		$((\frac{\Delta r_{SE}}{R_*}) \pm \sigma) \times 10^{-3}$	
		low-activity TR-1	high-activity TR-2	low-activity TR-1	high-activity TR-2
VIS-Phot	0.55	0.73 ± 0.13	1.99 ± 0.18	2.30 ± 0.41	6.26 ± 0.57
FGS-1	0.7	0.57 ± 0.10	1.51 ± 0.15	1.79 ± 0.31	4.75 ± 0.47
FGS-2	0.975	0.43 ± 0.07	1.11 ± 0.11	1.35 ± 0.22	3.49 ± 0.35
NIR-Spec	1.525	0.22 ± 0.04	0.56 ± 0.06	0.69 ± 0.13	1.76 ± 0.19
AIRS-Ch0	2.95	0.17 ± 0.03	0.43 ± 0.05	0.53 ± 0.10	1.35 ± 0.16
AIRS-Ch1	5.875	0.15 ± 0.02	0.37 ± 0.04	0.47 ± 0.10	1.16 ± 0.13

Table 7.2 provides the statistics of the solutions for the TR-1 and TR-2 transit cases for different wavelengths, encompassing the visual channels of the Ariel mission (~ 550 nm and ~ 705 nm), the infrared (~ 975 nm) and the low-resolution spectrographs (NIR, AIRS). We computed the mean transit depth variation induced by spots, $\langle \text{SP}(\lambda) \rangle$, and its corresponding planetary radius variation $\langle \Delta r / R_* \rangle$, along with the standard deviation σ induced by the uncertainty in the determination of the spot map. The statistics reveal that, even for the relatively low stellar activity level of WASP-52, with a spot filling factor of $\sim 5\%$, the signature of stellar activity on the transit depth can be close to 10^{-3} of the flux in the visual bands and $\sim 2 \cdot 10^{-4}$ in the NIR bands. This translates into relative planetary radius changes from $\sim 2 \cdot 10^{-3}$ to $\sim 5 \cdot 10^{-4}$, depending on the wavelength considered, which are of the same order as the typical signature of exoplanet atmospheres on transmission spectra. Furthermore, at higher activity level such as that of TR-2, the effect of the starspots is about twice as large. This illustrates the complication of studying atmospheres of exoplanets orbiting relatively active host stars.

7.3 Correcting transit depth for stellar activity

Various of methodologies to correct the impact of stellar activity on transmission spectroscopy of transiting planets have been suggested (see e.g. Knutson et al. 2012; Pont et al. 2013; Alam et al. 2018b). Typically, they are based on measuring the photometric variability of the host star, from which the spot filling factor and the temperature contrast or a combination of both are estimated. Furthermore, Rackham et al. (2018) claim that the methodologies applied to calculate these corrections could still be biased as they are usually computed assuming a single spot on the surface of the star.

The problem is that photometric variations do not generally provide a realistic representation of the spot filling factor throughout the rotation cycle. For example, a uniformly densely spotted star could produce a bias on the transit depth while not producing any observable photometric variability. Using semi-empirical relations between the filling factor of spots and the photometric variability of stars, Rackham et al. (2018) conclude that the chromatic effect on the transit depth due to spots could be several times larger than the signal of atmospheric features, thus greatly difficulting the general use of transmission spectroscopy for stars with some level of magnetic variability.

We demonstrate here that using multiband photometric light curves contemporaneous to transit observations, we can overcome this problem, at least partly. The advantage of using several photometric bands is that, as discussed in the previous section (see also Mallonn et al. 2018), it is possible to independently estimate the zero point with respect to the unspotted photosphere (z), and thus the absolute spot filling factor (including any persistent level of spottedness during the rotation cycle), and the temperature contrast. This provides extremely valuable information to correct transit depth variations due to spots. We can actually use our StarSim 2 simulations of WASP-52 to investigate at which level of accuracy we can correct the effects of starspots on transit data.

7.3.1 Simulations

The StarSim simulations of the WASP-52 planetary system discussed in the previous section reveal that, without additional information allowing us to infer the distribution and filling factor of spots during the transit, the planet-to-star radius cannot be determined with an accuracy better than a few percent. This is because in that case, we do not know at which rotation phase of the star the transit is observed, and therefore, we can only estimate a range of possible filling factors based on the amplitude of the photometric modulation. However, if we know the rotation phase at the time of transit, and we can estimate the spot properties and distribution from light curves, we can infer the correction that needs to be applied to the relative radius of the exoplanet (or transit depth) with an accuracy of order 10^{-4} , as shown in Table 7.2.

Results Figure 7.5 illustrates the sequence of spot corrections on the transit depth as a function of time. We plot the planet radius variation due to unocculted spots, $\Delta r_{\text{SP}}/R_*$, following our StarSim model describing the photometric light curve in Figure 6.3. Each panel corresponds to a different wavelength band of interest for the Ariel mission, as an example. To test all possible filling factors, we consider transits occurring at any time, but actual WASP-52 b transits are indicated as filled dots. The average value of all the simulations performed is displayed as a solid line. It is clear from this figure that the variation of the apparent planet radius depends on the projected spot filling factor and follows the stellar rotation period.

The relative effect on the radius diminishes towards longer wavelengths due to the lower flux contrast of spots, from an average of $\sim 4 \cdot 10^{-3}$ to $\sim 8 \cdot 10^{-4}$ at 550 nm (VIS-Phot) and $6 \mu\text{m}$ (AIRS-Ch1), respectively. In all cases it is possible to estimate the unocculted spot correction to the planetary relative radius with a precision close to 10%, estimated as the standard deviation (σ) of the sample of 21 optimal fits to the light curves. As explained above, this residual uncertainty is due to the variance of the stellar parameters and the randomness of the determination of the active region map.

7.3.2 Activity attenuation factor

It is interesting to quantify the improvement on a realistic transit depth determination that we can reach using this approach. We measure this improvement through what we call *activity attenuation factor*, which we compute as the ratio between the effect of spots on the transit depth and the uncertainty of the correction resulting from the StarSim model ($\langle \text{SP}(\lambda) \rangle / \sigma_{\text{SP}}$).

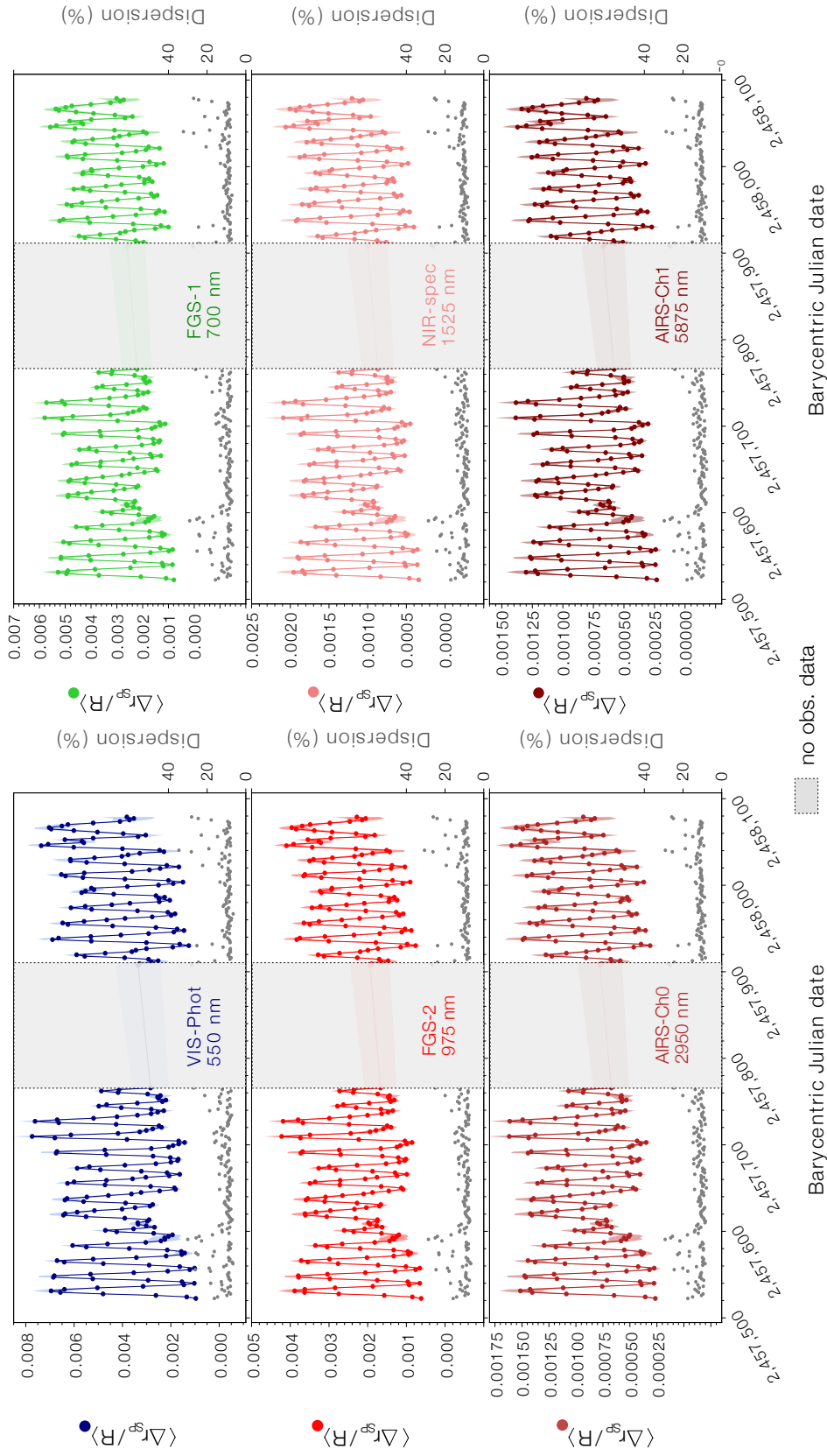


Figure 7.5: Effect of unocculted dark spots on the planetary radius $\Delta r_{sp}/R_*$ over time for different wavelength bands of the Ariel mission. The solid lines correspond to the average value of 21 optimal fits, and the 1- σ confidence level is shown as a light-shaded band. Solid dots indicate the timestamps of WASP-52 b transits according to the available ephemeris. Note the different vertical scale for each wavelength panel. Gray dots at the bottom of the panels represent the variance (σ , as a percentage value) of the results calculated from the different maps, as labeled in the right axis. The gap in the middle of each plot corresponds to the period between observational seasons without available photometric data (see Fig 6.3).

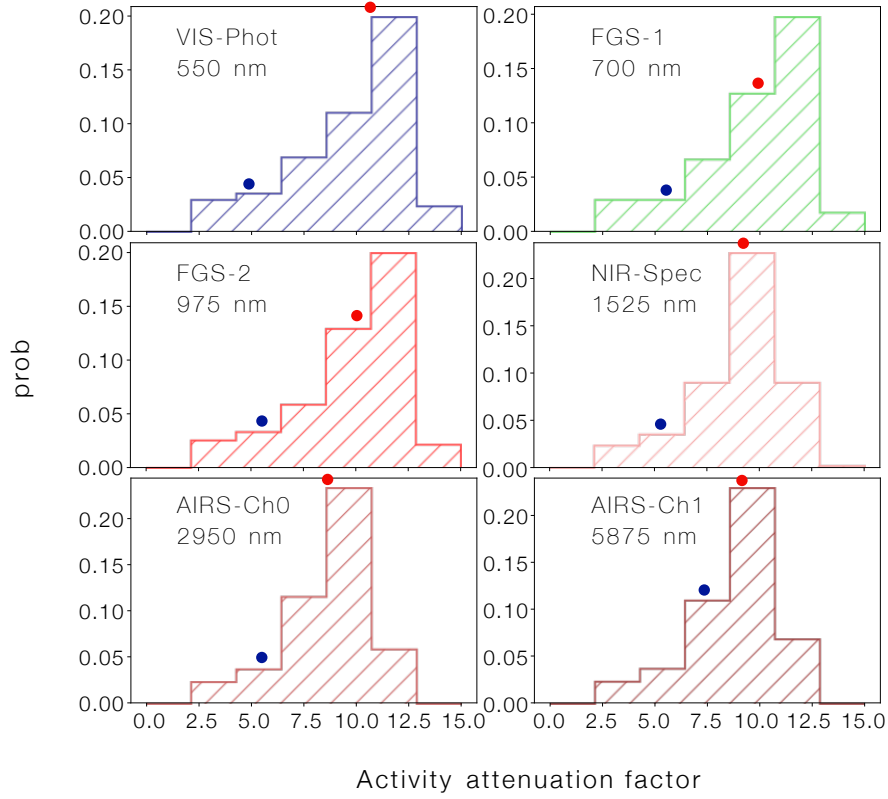


Figure 7.6: Distribution of activity attenuation factors from StarSim modelling considering all transit simulations. This factor is calculated as the ratio between the mean effect of spots on the transit depth and the uncertainty of the correction given by StarSim model ($\langle SP(\lambda) \rangle / \sigma_{SP}$). Blue and red dots indicate the position in the histograms of the fiducial transits TR-1 and TR-2, respectively.

The distribution of attenuation factors for the different simulated transits is shown in Figure 7.6, and this is a measure of the ability of the StarSim model to correct starspot effects. The plots indicate that using simultaneous photometry, we can correct the effects of spots on the relative radius of WASP-52 b by factors between 2 and 15 depending on the spot filling factor and the accuracy and coverage of the contemporaneous photometric monitoring. Lower attenuation factors mainly correspond to transits occurring when the projected filling factor of spots is smaller and having poor photometric coverage. On the other hand, larger factors are mainly due to transits with greater spottedness at times of well-sampled photometric light curves. On average, we can expect an attenuation factor around 10 on the relative radius measured for WASP-52 b.

The simulations of this planetary system show that, using contemporaneous photometric data, we are able to estimate transit depth corrections due to unocculted spots with uncertainties of around few times 10^{-5} in the NIR, regardless of the fraction of photosphere covered by spots (see error bars in the first two columns of Table 7.2). And we recall that this corresponds to a star displaying a modulation of about 7% in flux in the visible band, and a spot filling factor of about 3–14%.

7.3.3 Comparison with alternative activity correction methods for WASP-52 b

Several approaches have been presented to account for and correct out the chromatic effects of activity in transit observations. Instead of fitting photometric light curves, Rackham et al. (2019) estimate the filling factor of FGK spectral type stars from the peak-to-valley variability measured from Kepler photometry using the simple analytical form in Equation 7.2. The temperature contrast of the active regions and the projected filling factor of spots are strongly degenerate when only a single passband is used to measure photometric variability. To overcome this problem, Rackham et al. (2019) assume a temperature for the spots consistent with empirical values reported by Berdyugina (2005). The authors also assume a uniform distribution of dark spots over the stellar photosphere. If we apply this approach to WASP-52, the temperature contrast of spots would be 1290 K. However, in spite of being significantly larger than the value we find in our fits, with this value it is not possible to reproduce the $\sim 7\%$ peak-to-peak photometric variability of our light curves (see Figure 2 in Rackham et al. 2019) when assuming uniform spot coverage. Unphysically high spot coverage levels, a much larger spot temperature contrast or non-uniform distributions would be needed to explain the large observed amplitude. In addition to the different temperature contrast we obtain, the surface distribution of active regions derived with StarSim is distinctly non-uniform (see Figure 6.4).

Bruno et al. (2019) also estimate the properties of starspots for WASP-52, following the procedure of Huitson et al. (2013) and the light curve normalisation factor suggested by Aigrain et al. (2012). Based on data in Alam et al. (2018b), the authors fit both parameters from ASAS-SN and AIT photometry and transit spectroscopy from HST/STIS, WFC3 IR, and Spitzer/IRAC. They obtain temperatures < 3000 K for the starspots and a $\sim 5\%$ filling factor. Although the filling factor is consistent with our results, the temperature of spots is much larger than the value we measure from multiband photometry.

The chromatic signature produced by the parameters of Bruno et al. (2019) is plotted in panel (b) of Figure 7.4 (black solid line) using Equation 7.2. At wavelengths below $\sim 1.5 \mu\text{m}$, the resulting values are within the region allowed by our StarSim runs, and below $1 \mu\text{m}$ the predicted corrections are not far from the TR-2 ($\sim 12\%$ spot filling factor) case. The chromatic dependence, however, is significantly different. It is therefore possible that the transmission spectrum of Bruno et al. (2019) is somewhat affected by the different chromatic slope, although the average correction value should not be strongly biased. Note, however, that if we were to use the spot parameters adopted by Bruno et al. (2019) ($\Delta T_{\text{sp}} = 2200 \text{ K}$, compared with 575 K in our work) at NIR wavelengths, very strong deviations would occur, with spot corrections overestimated by factors of 2 to 5.

The differences found highlight the importance of extracting the chromatic signature from multiband photometric datasets so that the degeneracy between spot temperature and filling factor can be broken, and to reproduce realistic spot distributions. The determination of physical properties of active regions is a key point for the study of their chromatic signature on transit observations. Besides, using independent data, instead of the same transmission spectroscopy used for exoplanet characterisation, prevents misinterpreting atmospheric features.

7.3.4 Higher order effects: Limb darkening

The procedure employed in the literature above corrects the chromatic effects of active regions in transits by estimating the spot filling factor and temperature. However, the spot distribution on the stellar surface is still relevant because of limb darkening effects. Figure 7.7 shows a comparison between two *ad-hoc* spot maps (MAP-1 and MAP-2), a fitted map of WASP-52 at epoch BJD = 2,457,645 and the theoretical depth variation with homogeneous distribution (Equation 7.2). All these models have been calculated with the same projected spot coverage of 7% , $T_{\text{eff}} = 5000 \text{ K}$, $\Delta T_{\text{sp}} = 575 \text{ K}$. MAP-1 has a single spot at the center of the disk while MAP-2 has spot group close to the limb. The red line is the $\text{SP}(\lambda)$ profile calculated through Equation 7.2 and using low-resolution BT-Settl synthetic spectra. The differences found between the two extremal maps are considerably larger than the fiducial example of WASP-52 (green line), even considering the error bars. The sampling and bandwidths correspond to the visual channels of the Ariel mission ($\sim 550 \text{ nm}$ and $\sim 705 \text{ nm}$), the infrared ($\sim 975 \text{ nm}$) and the low-resolution spectrographs (NIR, AIRS).

The depth differences induced by these maps are quite significant, especially in the visible. From around $\sim 2 \mu\text{m}$, the effect of limb darkening dilutes as the photosphere tends to be more transparent to longer wavelengths. Hence, the chromatic depth contribution due to the spot distribution is roughly flat and barely exceeds 1% in depth variation. This simple exercise illustrates the need to consider the full stellar surface at the time of transit to properly correct for spot effects. The quality of the multiband photometric data is a crucial factor to invert the light curves and obtain precise surface maps to account for the distribution of spots.

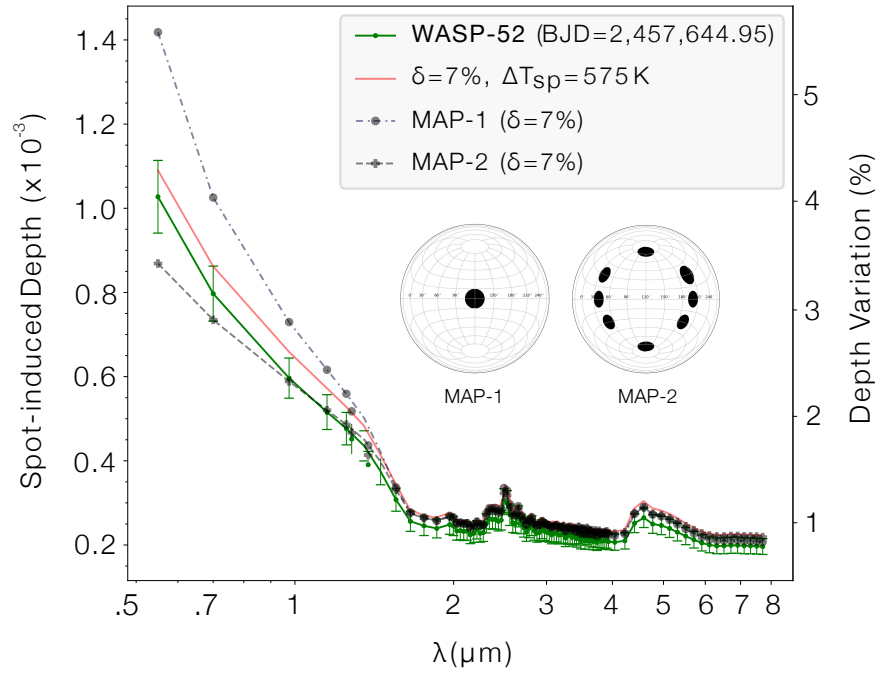


Figure 7.7: Chromatic effects due to spot distributions with identical filling factors. MAP-1 and MAP-2 are two *ad-hoc* generated spot maps to show the differences in their depth profiles and the comparison with that provided with Equation 7.2, shown in red. The green curve is a fitted map for WASP-52 at epoch BJD= 2,457,644.95 with the parametric distribution error bands.

Part IV

Conclusions and Future Work

Chapter 8

Discussion and Future Work

8.1 On time-series analysis

The first part of the Thesis presents the development of an extension of the well-known in the community, Generalised Lomb-Scargle (GLS) code. GLS is an algorithm for searching periodicities in time-series data. Among the advantages in using Lomb-Scargle approaches in radial velocity time-series are the ubiquity in treating unequally spaced measurements by least-squares fitting sine waves, in contrast to classical methods such as Fourier transform.

Taking as the starting point the mathematical foundations of the GLS algorithm, we developed a multidimensional extension (Multidimensional Generalised Lomb-Scargle, MGLS) to allow the simultaneous fitting of n frequency components, that is, full sine waves represented as a linear combination of a sin and a cos. The core idea is to produce a ready-to-use code package, written mostly in Python and, plus a few numerical routines written in Fortran 95, with the aim of achieving optimal efficiency and high execution speed.

A common approach in analysing multiplanetary systems using GLS has been searching for a single periodicity and iteratively re-analyse the residuals at each step with the same procedure. The loop is continued until the highest peak found in residuals of the n -th step is not statistically significant by considering some *ad-hoc* measurement criterion (e.g. 0.1% FAP).

The main shortcomings of the hereby procedure —also known as prewhitening— are the biases carried over through each one-dimensional fits, which may induce significant fake detections (both false positives and false negatives). An illustrative example is shown in Section 2.2.1.

The algorithmic structure of MGLS consists of fitting a model composed of linear and non-linear parameters (see Section 2.3.1). The linear part is solved using LAPACK linear algebra Fortran libraries, while the non-linear parameters, the tuple of frequencies and jitters are optimised by means of a Monte Carlo Simulated Annealing (MCSA) approach, in spite of the fact that this type of search is considerably numerically hard since the hypervolume of the solution is tiny in comparison to the huge search space.

In addition, the multimodality of the problem makes the optimiser particularly prone to be trapped in local optima and, therefore, faster and more efficient optimisation strategies may easily fail.

Hence, the strategy consists of setting an initial tuple of non-linear parameters, θ_{nl} , and as the algorithm tries perturbed configurations following the cooling schedule, the optimal linear parameter solution is computed at each step. To improve the performance and assure the optimal, this last MCSA computation is repeated in parallel for different initial conditions, and finally, a descendent gradient algorithm computes the final optimum of each thread. The best optimal of this list is the optimal MGLS solution.

We carried out benchmarking tests of MGLS in real multiplanetary systems to assess the capabilities of this simple, but useful approach. Furthermore, we developed a robust and fast methodology to determine the significance of the signals and the dimensionality of the optimal solution. The general performance is good and we proved to detect confidently signals in multiplanetary systems, attaining comparable results to those obtained through sophisticated Bayesian inference techniques. In addition, the code is flexible to be modified and adapted to other specific tasks.

The presented MGLS employs an approximation of circular orbits, which has some pros and cons. On the one hand, the assumption $e = 0$ permits a fast and stable solutions, while on the other hand, for in systems with high eccentricity orbits, the results could suffer from aliases and may lead to misdetections. We note, however, that from orbit stability criteria, the orbital eccentricity values of multiplanet systems are not expected to reach extreme values, therefore validating the circular orbit assumption.

The MGLS code is publicly available at github.com/rosich/mgls

8.1.1 Future work

Taking advantage of the algorithmic logic of the code in separating linear and non-linear parameters, the inclusion of Keplerian fitting functions will be quite straightforward in forthcoming upgrades.

In addition, there are plenty of archival radial velocity data with the potential to be re-analysed with MGLS. Many of the oldest measurements were analysed by means of LS or Fourier methods, which would be interesting to revisit.

8.2 On active surface modelling techniques

The second part of the Thesis addresses the deterministic physical modelling of active stars from varying and rotating surface inhomogeneities. In particular, we investigate the ubiquitous and important problem of precisely understanding how heterogeneous photospheres affect the surface-integrated observables. From this, we provide a methodology to correct the observations, these being spectroscopic measurements of radial velocity or the chromatic effects in transiting planets for atmospheric structure determination.

We developed a software package StarSim 2 that upgrades the previous version — only intended for the computation of the observables—, and among other capabilities, allows the computation of the inverse problem. By refactoring and optimising the older version of StarSim, we gained a speed-up factor of ~ 1000 in the forward problem evaluation by imposing a set of approximations such as considering the spots as single surface elements, in which their physical properties, geometric and spectral characteristics are calculated at the center of the active element. This allows a dramatic increase in numerical performance in comparison to the grid integration scheme, at the expense of working on the small-size spot limit due to the discontinuities arising in large spots when appearing and disappearing from the limb. As an advantage, ensembles of small spots have the interesting ability to form arbitrary complex shapes by setting up clusters.

The inversion functionality is the newly-developed part consisting of inferring spot maps and stellar parameters that best describe the observational data. The approach we have developed to perform the inverse problem consists of finding a spot map along with a set of stellar physical parameters. The key point in this formulation is that both the map and physical properties are, in fact, a single set of parameters with the difference that the map can be understood as a parameter with some internal structure since it is made of a selected number of spots, each of them having five distinct properties (appearance time, lifetime, colatitude, longitude, and radius). Due to the enormous computational cost, and the fact that this problem belongs to the class of *noisy optimisation* problems, characterised by approximate and long function evaluation times, StarSim 2 implements a two-step approach to perform the inversion. First, by employing a Simulated Annealing (SA) optimisation algorithm generates a spot map with stellar parameters kept constant; and second, repeat this process (typically thousands of times) for randomly selected parameters to finally obtain the distributions of these parameters. In addition, the forward simulation routines have been migrated from Fortran 77 to Python (framework) and Fortran 95 (numerical routines) and parallelised to be executed in multicore servers.

The StarSim 2 code is publicly available at github.com/rosich/starsim-2

The effects induced by dark spots and bright faculae are responsible for the observed stellar signals in the timescale of few days to hundreds of days, which is related to the stellar rotation period. In the case of RV measurements, the presence of these structured signals hampers the procedure to extract small signals related to true planetary companions. Even in the case where the periods of the star and the planets are sufficiently different, correlated noise arising from the basal activity imposes great difficulties in those stellar systems with signals at the limit of their detectability.

It is important to remark that our aim in this part is not aimed at inferring precise active region maps, but our goal is to predict the effects produced and establish a process to correct observational data.

In our work we discuss the analytical foundations of the inverse problem if multi-band photometry is available. We show that color information allows breaking the degeneracy between the spot size and its contrast temperature. Otherwise, without differential fluxes through color filters, the optimiser would be unable to determine a unique

ΔT_{sp} since it is possible to obtain a very similar flux modulation by properly balancing the interplay with the size of the spots and temperature contrast.

To further understand the effects of the various parameters, we developed a toy model as a simplified version of the stellar surface. We define a heterogeneous surface with a polar cap—a large spot centered in the pole—and an equatorial small spot. We understand the contributions of these two types of spots as modulating and non-modulating, and we define their filling factors as δ_M and δ_0 , respectively. The relevant variables in our problem are δ_M , δ_0 , and ΔT_{sp} . We study analytically the impact on the photometry and we reach the following conclusions depending on the availability of multiband photometry:

- If a single photometric band is available, the inverse problem can only yield a value for the modulating filling factor. Neither the non-modulating component (which may be neglected) nor the temperature contrast can be univocal determined. A common practice up to now has been assuming a typical temperature contrast from the bibliography.
- In case of two-band photometric monitoring, it is possible to determine, in addition to δ_M , ΔT_{sp} although the solution becomes bivariate regarding the contrast temperature.
- If having three or more photometric bands, the degeneracy involved in spot map generation is broken and it is possible to determine all relevant variables (δ_M , δ_0 , and ΔT_{sp}). However, the determination of the non-modulating component reveals to be very challenging due to its weak chromatic effect. Nevertheless, for the same reason, its contamination effect on the observables will be, in consequence, rather small.

We tested the aforementioned principles of the photometric inverse problem in ground-based multiband observations of the exoplanet-host WASP-52 covering a total timespan of ~ 600 days that implies capturing about 30 full stellar rotations.

The results show the feasibility of light curve inversion and relevant parameter retrieval with enough photometric accuracy and time coverage. The most important parameters fitted are listed next,

- We found a temperature contrast $\Delta T_{sp} = 575 \pm 150$ K below that of the surrounding photosphere.
- An absolute projected filling factor function, $FF(t)$, is obtained, showing a spottedness ground value of 3% and a peak value of 14%. Permanent spot coverage on WASP-52, is therefore, kept for a long period of time, i.e. no immaculate photosphere existed for the entire photometric monitoring period.
- In addition to ΔT_{sp} and $FF(t)$, the inversion process presented is also capable of yielding latitude-projected filling factor maps.

The last chapter of the part dedicated to stellar activity modelling addresses the chromatic effects on simulated transits. The transit depth is the flux decrease at the

time of transit measured with respect to the flux level immediately before and after the event. Its value is equal to the ratio between the radii of the planet and the host star, respectively. In the presence of a stellar spotted surface, the depth presents a dependence as a function of wavelength. As a consequence, transit depths can be intrinsically chromatic, therefore hampering the determination of the transmission spectrum of an exoplanet, whose measurement is based on exactly the same principle.

The main advantage of the approach we present to compute the chromatic effect on the depth of exoplanet transits is that it is based on an independent and consistent deterministic method allowing to accurately determine the stellar parameters, the filling factor and the distribution of spots. Thus, the effect of spots and their different positions on the stellar disk is also taken into account. We have only considered here the effect of unocculted spots, which increase the transit depth. On the other hand, occulted spots produce the opposite effect.

We used the WASP-52 multiband photometric model to simulate transits and study the chromatic effects on the transit depth obtained at different epochs corresponding to different stellar spot distributions. In WASP-52, which has peak-to-peak flux variations of $\sim 7\%$, the chromatic effect of the spot map results in uncertainties of $\sim 10\%$ in transit depth ($SP(\lambda)/\mathcal{D}_0$) and of up to $\sim 5\%$ in the planetary radius ($\Delta r_{SP}/r$), at VIS. After the correction, we are able to reduce residual depth uncertainties down to $\sim 4 \cdot 10^{-5}$ at 550 nm (ARIEL/VIS-Phot) and $\sim 10^{-5}$ at $6 \mu\text{m}$ (ARIEL/AIRS-Ch1). The remaining uncertainty of the correction, encloses several effects, most importantly: *i*) the photometric precision and phase coverage of the monitoring along time, *ii*) the incompleteness of the physical model and, *iii*) the uncertainties in retrieving optimal surface maps and stellar parameters. We expect that higher activity attenuation factors could be achieved by optimizing the strategy of the photometric observations, i.e., making sure that multi-color measurements are obtained before and after the transits to better constrain the model. Photometric observations from space would even boost the accuracy in light curve modelling, and hence, the correction capabilities.

8.2.1 Future Work

The potential shown by simulation techniques based on physically-motivated models is enormous and especially adequate for facing the challenges in exoplanet research for the next decade with regard to stellar activity.

The most important bottleneck of the StarSim code is the inversion module to generate spot maps by using observational data. Table 4.3 summarises typical running times for both forward and inverse problems in a test case studied in this Thesis. These performance numbers are referred to the parameter-constant inversion, i.e. keeping the stellar parameters fixed, finding the optimal map. When performing a full inversion consisting of a collection of maps and their associated parameters all of them statistically equivalent in probability, means the need of executing thousands of inversion replicates with a substantial amount of time even in a dedicated multicore server. As explained in Chapter 6, a full inversion of BVRI light curves with ~ 300 measurements required about 3 weeks in a 56-core server. An amount of time that if we wanted to use in massive data processing, would be prohibitive.

In addition, the forward computation of RV series takes an order of magnitude more time than photometry series generation. This imposes a barrier when it comes to processing just a few target stars of a survey campaign.

Therefore, with the purpose of making the code usable when, in the immediate future, a large amount of high-precision data of contemporaneous spectroscopic and space-based photometric observations will be available, StarSim requires a number of improvements we separate into two subgroups:

On the physical model In the current version, faculae are always associated with a spot as a concentric ring. A new version should allow the independent modelling of spots and faculae as two types of different active elements. Improved physical modelling of faculae and a new generation of synthetic spectra in addition to the incorporation of chromospheric activity indices would allow enhancing the precision and functionality of StarSim.

On the code implementation To speed up StarSim's productivity in parameter inference, the exploration of Deep Generative modelling will allow generating samples of the maps by efficiently sampling a probabilistic model previously trained with calculated StarSim inverted data. A generative model is a neural network that describes the way a dataset is created in terms of a trained probabilistic model. By sampling from this model, new data are generated, not belonging to the training set, but created just from the probabilistic distributions intrinsically encoded in the neural network.

The use of generative deep learning techniques has yielded an astonishing performance in image processing (Goodfellow et al. 2014; Kingma et al. 2014) or music composition (Dhariwal et al. 2020). In scientific fields have shown promising results, i.e in geophysical inverse problems (Lopez-Alvis et al. 2021); astrophysics (e.g. Martínez-Palomera et al. 2020; Gabbard et al. 2019). Machine learning generative modelling provides a way of creating new datasets not just interpolating/extrapolating the existent data on the training sets but entirely creating new data by the capability of these architectures in capturing the essential structure by fitting probabilistic models.

Particularly interesting for this goal, are the Generative Adversarial Networks (GAN) and the Variational Autoencoders (VAE), which are amongst the most promising deep learning techniques. GANs are essentially two competitive neural networks, the first generating data from a random input, and a second one trying to discriminate if the data generated is real or not on the basis of a training set. Hence, by training both networks in best performing their duty, it ends up in a zero-sum game in which the generator learns how to create new data to fool the discriminator. VAEs are based on the autoencoder architecture of neural networks, in which two copies of the same data are fed in the input and output of the network. In the middle, a reduced number of neurons called latent space encodes the data by training the network's weights to encode-decode the data. However, the encoded information in the latent layer does not have any particular structure and if perturbed, the result of the decoding would surely be nonsense. Instead, if the latent space has learned to build the probabilistic distributions of the input data, just by sampling from these distributions we are able to generate new data. This is the purpose of VAEs and its variant cVAE (conditional-VAE) in

which additional metadata are included acting as a conditional probability. The stellar parameter tuple (θ) can play this role. Both methods are able to learn the underlying structure of data, and then acquire the ability to generate new observations.

GAN and cVAE are potentially useful in the StarSim parameter inference routine as defined in Section 5.3.3. The evaluation of the metrics of the photometric model—that is, compute $F(\mathcal{S}, \theta)$ to generate observables—takes $\sim 1\text{--}5$ ms depending on the data extension. Therefore, if a GAN/cVAE is reasonably efficient in producing likelihood samples of \mathcal{S}_θ , we can expect a good ratio of acceptance to build an ensemble of statistically equivalent models or, alternatively, run a MCMC routine to make parameter inference. However, we do not expect a realistic improvement in code optimisation that will be sufficient to change the defined parameter inference retrieval paradigm. At best, code refactoring and algorithm optimisation may lead to 5–10% performance gains, totally insufficient to overturn our current strategy.

On the surface mapping The current StarSim 2 version maps the stellar surface as a flat and uniform photosphere with circular spots of constant temperature contrast, which in turn can have a concentric corona to model faculae. This representation imposes a constraint on the possible surfaces that can be reproduced. If the mapping was a grid of cells where each one can represent photosphere or spots in a free-forming structure approach, we would be able to reproduce any sophisticated surface, but the degree of degeneracy would be unaffordable. In fact, a free lattice model has many different configurations to represent a light curve, and regularisation techniques have to be implemented (Vogt et al. 1987; Lanza et al. 2011; Luo et al. 2019).

An alternative approach to map stellar heterogeneous surfaces is through spherical harmonics (SH) decomposition (e.g. Russell 1906; Luger et al. 2021b; Luger et al. 2021a) that is equivalent to Fourier series on the sphere. The SH expansion can represent any continuous function as a SH series on the sphere, providing the ability to represent any type of spotted surface with arbitrarily complex shapes. This may be implemented in the StarSim scheme, particularly for the solution of the inverse problem taking advantage of the semi-analytical formulation this approach offers.

8.2.1.1 Prospects of joint photometry and spectroscopic modelling

The aim of simultaneous modelling is to obtain a solution of the inverse problem, \mathcal{S}_θ , through fitting several observables at the same time. One of the applications of joint modelling is the disentanglement of true Keplerian signals from those arising from stellar activity, by employing observables with no planetary signals such as photometry or spectroscopic indices.

Here we present preliminary results of an application example of the StarSim 2 modelling methodology with contemporary photometry and spectroscopic observations using observations of the Hyades member star HD 25825. This solar-type star ($T_{\text{eff}} = 6100$ K) is a part of a sample that was observed contemporaneously with K2 space telescope and ground-based spectrographs. These were: HARPS-N (Pepe et al. 2005) mounted on the Galileo telescope in La Palma island (Tenerife); SOPHIE (Perruchod et al. 2008) mounted on the 1.93-m telescope at the Observatoire de Haute-Provence and

TRES mounted on the 1.5-m Tillinghast telescope at Fred L. Whipple Observatory on Mt. Hopkins (Arizona).

RV inference using photometry By using high-precision space photometry, an inversion process is performed to obtain a large number of optimal maps, \mathcal{S}_θ , and the best models are selected to create a population of statistically equivalent solutions. Once obtained, this set of optimal maps is used as input parameters for the computation of the predicted RV signal.

Photometry time-series does not contain enough information to univocally infer the RV counterpart of spotted stars. In accordance with the conclusions reached in Chapter 5 using K2 single-band photometry, it is only possible to retrieve the modulating filling factor of the maps and no information on spot temperature contrast is attained. Therefore, in these types of inversion problems, the spot temperature contrast has to be fixed initially, either arbitrarily or using a semiempirical $T_{\text{eff}} - \Delta T_{\text{sp}}$ law (e.g. Berdyugina 2005). Otherwise, if included as a parameter, the spot temperature distribution obtained will be flat.

Figure 8.1 shows a prototype fit to generate the RV signal by using K2 photometry. A pool of 20,000 light curve inversions were performed by randomising two parameters, the rotation period and the photometric statistical jitter, and a total of 20 of the best models were picked to generate the average RV model and its uncertainty. The spot maps used were set to have 35 active regions and the facula Q value was set to zero as no chromatic information is available. The results of these prototype RV inference is listed in Table 8.1. \mathcal{L}_0 is obtained by fitting the RV data with a constant, and the \mathcal{L} metrics of the generated RV model for each instrument dataset is shown. Once a set of statistically equivalent solutions of the inverse problem is obtained, $\mathcal{S}_{\theta_0}, \mathcal{S}_{\theta_1}, \dots, \mathcal{S}_{\theta_n}$, a RV forward problem is computed using these parameters. An individual offset term is fitted for each RV dataset to reach the best fit metric.

RV inference using photometry and BIS index Figure 8.2 shows the result of K2+BIS simultaneous fits to predict RV. The improvement in the prediction variates depending on the dataset. HARPS data show a very significant improvement with respect to the results in photometry fits. In these fits, the Q value was included as an adjustable parameter, with the rotation period and photometric jitter. A total of 10,000 inversions were performed to finally select the best 50 models to compute the RV signal. Table 8.1 show the results of the RV generated models.

The integration of spectroscopic data in the joint fits still holds a number of degeneracies in the inversion model such as the convective parameters. To obtain information on the stellar convection, Baroch et al. (2020) modelled CARMENES chromatic radial velocities to determine the properties of active regions such as spot-filling factor, temperature contrast, and convective shift. In the StarSim 2 code, the convective model is set to a solar analog.

One of the most important difficulties in spectroscopic joint inversions is related to the computational effort required. StarSim 2 takes few milliseconds to invert a typical light curve, while in the case of a RV or BIS series a tenth of a second is needed

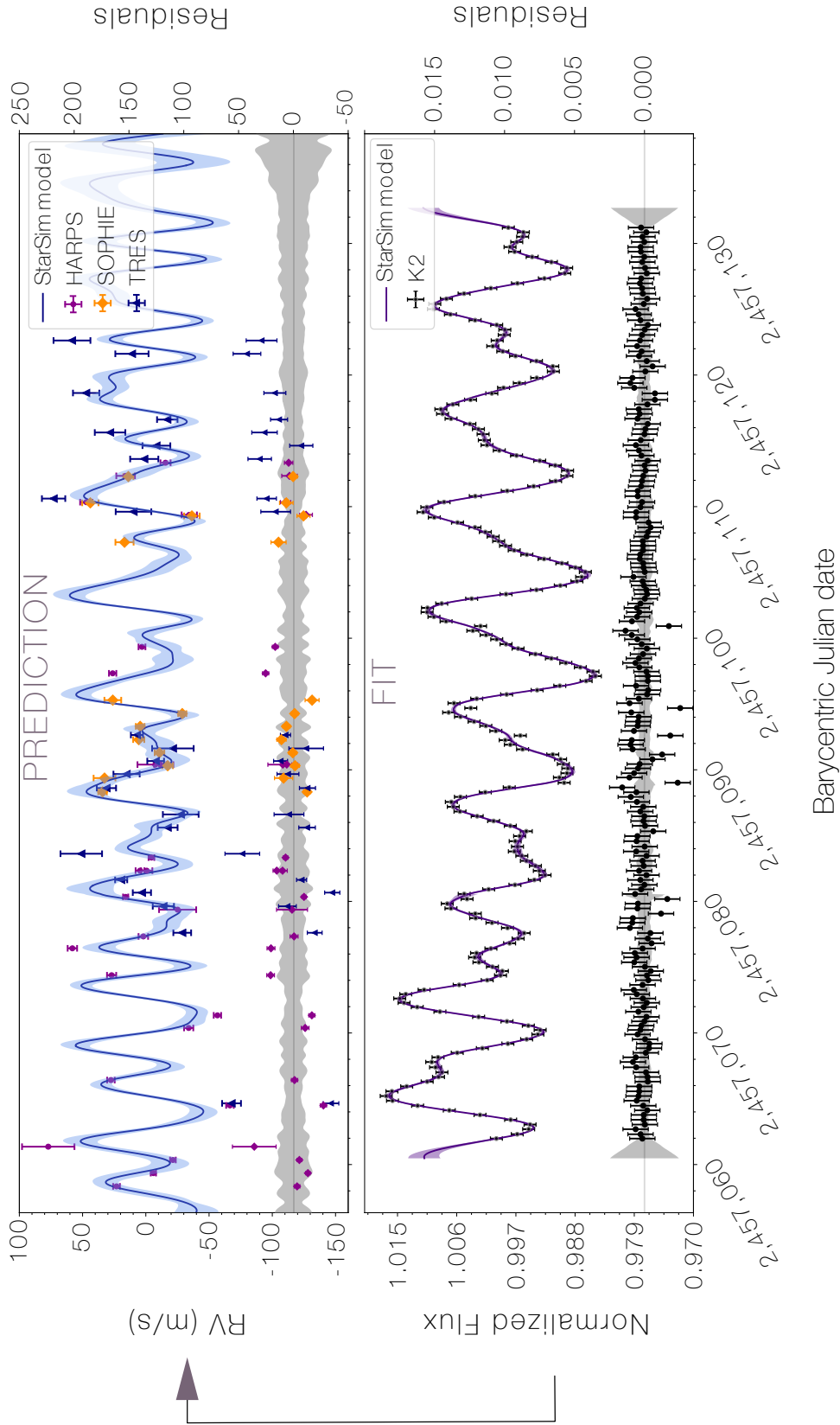


Figure 8.1: Photometry fit of HD 25825 and resulting RV prediction. The solid curves and their shaded area represent the mean and $1\text{-}\sigma$ deviation of the best 20 independently inverted curves out of a pool of $\sim 20,000$ random-parameter inversions. The spot map has a fixed 35 fittable active elements and we determined the best values of stellar parameters for P_{tot} and jitter for the photometry measurements. Q factor was set to zero. Using the products of the inversion (maps and parameters) the radial velocity curves were generated for each of the 3 independent datasets from HARPS, SOPHIE and TRES spectrographs. The gray shaded area shows the $1\text{-}\sigma$ uncertainty of the model.

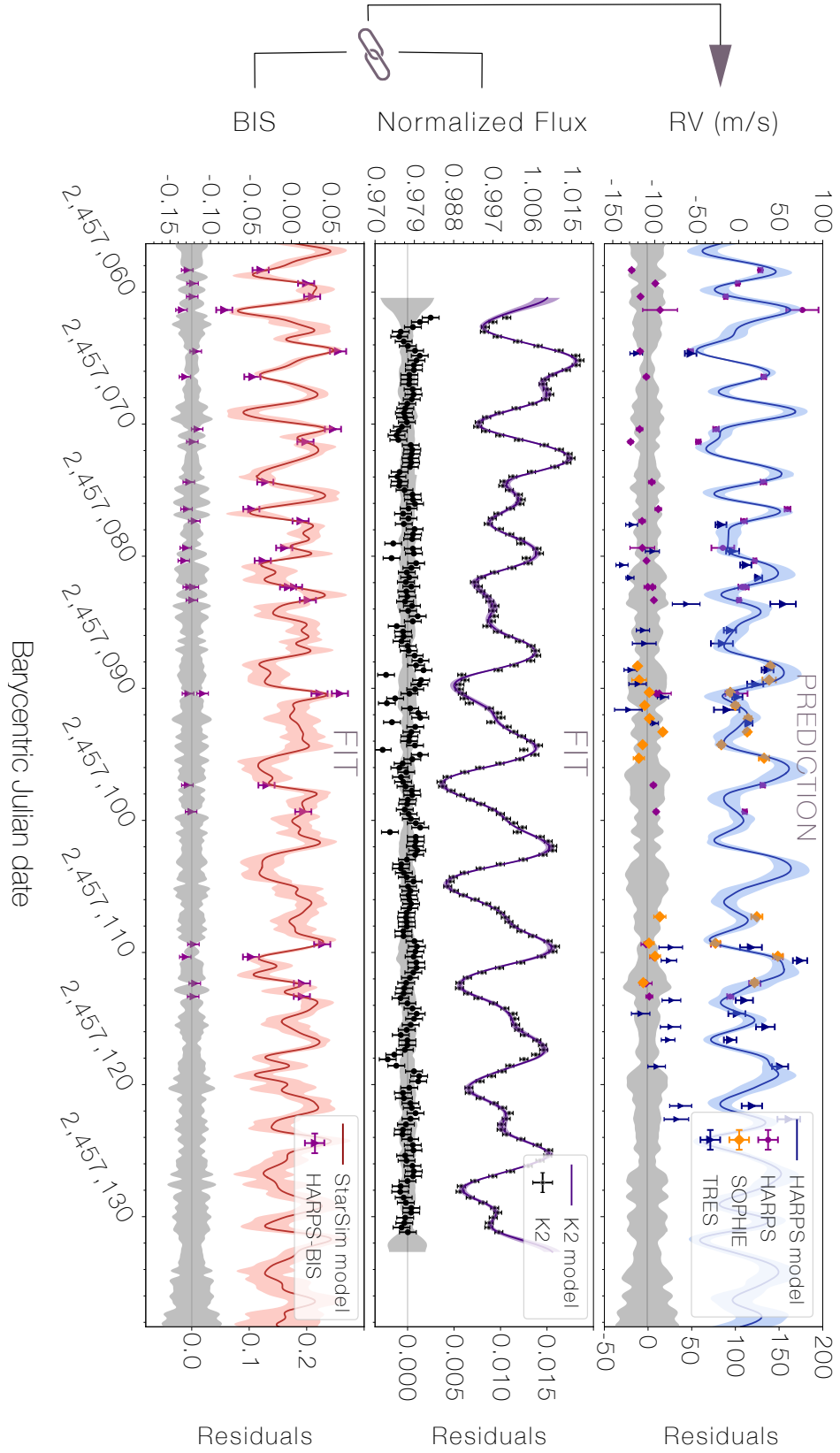


Figure 8.2: Simultaneous photometry and BIS fits for HD 25825. The solid curves and their shaded area represent the mean and 1- σ deviation of the best 50 independently inverted curves. The spot map has a fixed 35 fittable active elements and we determined the best values of stellar parameters for P_{rot} , Q and jitter for the photometry measurements. We only considered for the simultaneous fits the HARPS velocities because of their better precision compared with the other datasets.

Table 8.1: Overview of prediction results through K2 and K2+HARPS BIS simultaneous fits for each instrument. The only fitted parameter in RV data is an offset for each instrument by maximising the $\ln \mathcal{L}$ with respect to the mean StarSim 2 model. No additional jitter is applied to RV data errors.

<i>Parameter</i>	HD 25825		
	HARPS	SOPHIE	TRES
Measurements	24	12	22
Timespan (days)	55	24	58
Mean separation (days)	2.29	1.99	2.63
$\ln \mathcal{L}_0^{(a)}$	−977.40	−204.25	−205.35
<i>K2 photometry fit</i>			
Offset (m/s)	3.64	−0.09	9.33
$\ln \mathcal{L}$	−280.46	−50.02	−131.75
<i>Joint K2 + BIS (HARPS) fit</i>			
Offset (m/s)	14.73	12.68	21.65
$\ln \mathcal{L}^{(b)}$	−158.40	−57.79	−131.77

(a) The null model likelihood, $\ln \mathcal{L}_0$, is calculated with respect to data mean.

(b) Computed with respect to residuals and mean StarSim model.

(see Table 4.3). This implies a major challenge to the calculation of the inverse problem, which takes a long time —often unaffordable— to obtain large enough ensembles of statistically-equivalent models to construct the parameter distributions. In order to tackle this issue, new techniques have to be explored to efficiently compute the spectroscopic models, such as reversible neural networks as suggested in previous sections.

Colophon The disentanglement of the Keplerian signals from the stellar activity is both a formidable and timely challenge in exoplanet science. To achieve the detection and characterisation of Earth-like planets, including their potential atmospheres is one of the most exciting and relevant scientific endeavours of the next decade. The presented preliminary results of joint photometry and spectroscopic indices are very promising and encourage us to think they have great potential. The combination of high-precision photometry from space missions like TESS or PLATO, with spectroscopic ground observations with the cutting-edge instruments such as ESPRESSO, capable of reaching long-term accuracies of ~ 0.5 m/s, will play a fundamental role in addressing this problem. The StarSim project may be a key tool in making progress in this field. With the proven capability to model spot-modulated stellar activity and correct planetary transits employing multiband photometry, besides the promising potential of precisely predicting RVs, the StarSim tool has a role to play.

List of Figures

1	Descobriments d'exoplanetes pel mètode de la velocitat radial, trànsits i altres mètodes	xvi
2	Mapa de color de l'escombrat MGLS bidimensional per dades de l'estel de Barnard	xviii
3	Exemples d'espectres sintètics BT-Settl	xix
4	Diferents contribucions de la velocitat radial per un sistema d'una taca equatorial	xxiii
5	Exemples de corbes generades per StarSim 2 i un mapa compost per una sola taca equatorial	xxiv
6	Esquema de fotosfera activa definint part modulant i no-modulant	xxvi
7	(a) Profunditat dels trànsits simulats en funció de λ . Assumint una fotosfera immaculada (línia vermella) i tacada (banda gris) (b) Contribució de les taques a la profunditat dels trànsits per tota la simulació de WASP-52 (gris), i per dos trànsits que exemplifiquen casos d'alta i baixa activitat. .	xxviii
1.1	Exoplanet discoveries by radial velocity, transits, and other methods . . .	5
1.2	High-resolution image of the Sun's photosphere showing the convective cells.	9
1.3	SDO images of the quiet and spotted Sun	10
1.4	Simulated multiband transits on a spotted surface.	13
2.1	Comparison between LS and GLS periodogram	22
2.2	Simulated two-period system solved using pre-whitening technique and joint multifrequency fit (MGLS).	26
2.3	MCMC for the period and jitter distribution of Teegarden star observations (see Section 3.1.4) searching for 4 signals.	35
2.4	Example of bootstrap histogram fit	40
3.1	FAP determination following the recipe in 2.5.1	51
3.2	Detectability plots for simulated injected signals in tested stars	52
3.3	GJ667C FWHM and BIS activity indices	59
3.4	Teegarden's star 2-dimensional MGLS plot	66
3.5	Teegarden's star GLS periodogram VIS+NIR	67
3.6	GLS periodogram for the Barnard's star data	69
3.7	Application of a bidimensional MGLS code for multiset period search in Barnard data	72
4.1	BT-Settl example spectra	79
4.2	StarSim spot model	82

4.3	StarSim 2 simulated transits	86
4.4	Composition of radial velocity signal budget for a Solar type star with an equatorial spot of $1.6 \cdot 10^{-3}$ coverage	89
4.5	Examples of curves generated with the StarSim 2 code for a K5V star and a map consisting of a single equatorial spot of 2° in radius	90
4.6	StarSim application file tree	94
4.7	StarSim UML class diagram	98
5.1	Sketch of a stellar photosphere including modulating and non-modulating spots	100
5.2	Graphical representation of inverse problem Equation 5.3 for different configurations of δ_0 , δ_M and ΔT_{sp}	102
5.3	Flow chart of the parameter retrieval approach	106
5.4	StarSim 2 single-band simulation of two examples of heterogeneous surfaces to illustrate their photometric signal dependence on the normalisation used	109
5.5	Photometric inverse problem toy model	111
6.1	Trial tests showing the statistic $\ln \mathcal{L}$ as a function of the number of spots and for several values of the number of iterations of the MCSA algorithm	118
6.2	Histograms and correlations of 365 statistically-equivalent solutions for the fitted StarSim parameters	119
6.3	StarSim model fits to multiband photometric ground-based observational data	122
6.4	Longitudinal spot filling factor projected on the stellar equator of WASP-52	123
7.1	Transmission spectra of WASP-39 b	126
7.2	Spotted transit sketch	127
7.3	Definition of transit depth with immaculate and spotted photosphere	128
7.4	Chromatic transit depth for the WASP-52 transiting simulated system.	131
7.5	Effect of unocculted dark spots on the planetary radius $\Delta r_{SP}/R_*$ over time for different wavelength bands	135
7.6	Distribution of activity attenuation factors from StarSim modelling	136
7.7	Chromatic effects due to spot distributions with identical filling factors	139
8.1	Photometry fit of HD 25825 and resulting RV prediction	151
8.2	Simultaneous photometry and BIS fits for HD 25825	152

List of Tables

1	Paràmetres estel·lars i de simulació de les corbes d’exemple de la Figura 5.	xxv
2.1	Significant periodic signals found on ad-hoc toy model to demonstrate the potential flaws of time-series iterative processing.	27
2.2	MGLS running times	43
3.1	Overview of the stellar parameters for the analysed stars	48
3.2	Six-planet solution of the radial velocities of HD 40307 in bibliography	50
3.3	Sequence of signals found in HD 40307 using GLS with prewhitening, and global analysis using MGLS	53
3.4	Continuation of Table 3.3 (HD 40307)	54
3.5	Signals found for GJ 163 in the bibliography	55
3.6	Sequence of signals found in GJ 163	57
3.7	Signals found for GJ667C in the bibliography	58
3.8	Sequence of periodic signals found on GJ 667C with the GLS prewhitening, and with the MGLS code	61
3.9	Continuation of Table 3.8 (GJ667C)	62
3.10	Significant periodic signals found for Teegarden’s star in ZE19	63
3.11	Sequence of signals found in Teegarden’s star in VIS dataset; with the GLS prewhitening (left), and with the MGLS code (right).	64
3.12	Sequence of signals found in Teegarden’s star in VIS+NIR datasets; with the GLS prewhitening (left), and with the MGLS code (right).	65
3.13	Observational dataset of Barnard’s star.	68
3.14	Sequence of periodic signals found in Barnard’s star; with the GLS prewhitening (middle columns), and with the MGLS code (right columns)	70
3.15	Continuation of Table 3.14	71
4.1	Set of most relevant StarSim parameters	81
4.2	Stellar and simulation parameters of the example curves in Figure 4.5.	91
4.3	starsimdos running times	95
5.1	Results of inversion tests using synthetic data with StarSim	110
6.1	Important parameters of WASP-52 (top) and WASP-52 b (bottom).	116
6.2	Priors and results from the StarSim inversion of WASP-52 photometry	120
7.1	StarSim simulation parameters for WASP-52 transits. Those parameters tagged as <i>model distribution</i> refers to those statistically equivalent solutions found in the photometric model developed in Section 6.4.2.	130

7.2	Mean values of transit depth variations due to spots, $SP(\lambda) = \Delta\mathcal{D}'$, on WASP-52 simulated transits	132
8.1	Overview of prediction results through K2 and K2+HARPS-BIS fits	153

Abbreviations

APS	Automated Planet Searcher
BIS	Bisector Inverse Slope
BJD	Barycentric Julian Date
CCF	Cross-Correlation Function
CARMENES	Calar Alto high-Resolution for M dwarfs with Exoearths with Near-infrared and optical Échelle Spectrographs
FAP	False Alarm Probability
FWHM	Full Width at Half Maximum
GEV	Generalised Extreme Value distribution
GLS	Generalised Lomb-Scargle
MCSA	Monte Carlo Simulated Annealing
MGLS	Multidimensional Generalised Lomb-Scargle
NIR	Near Infrared
HARPS	High Accuracy Radial velocity Planet Searcher
HIRES	High Resolution Echelle Spectrometer
PFS	Planet Finder Spectrograph
RV	Radial Velocity
SOPHIE	Spectrographe pour l'Observation des Phénomènes des Intérieurs stellaires et des Exoplanètes
TJO	Telescopi Joan Oró
TRES	Tillinghast Reflector Echelle Spectrograph
WASP	Wide Angle Search for Planets

Publications

Directly related to the Thesis

★ Papers with surface modelling StarSim contribution.

≈ Papers with MGLS time-series analysis contribution.

1. ★ **Rosich, A.**, Herrero, E., Mallonn, M., Ribas, I., Morales, J. C., Perger, M., Anglada-Escudé, G., and Granzer, T., 2020, *Correcting for chromatic stellar activity effects in transits with multiband photometric monitoring: application to WASP-52*, A&A, 641, A82, <https://doi.org/10.1051/0004-6361/202037586>
2. ≈ Ribas, I., Tuomi, M., Reiners, A., Butler, R. P., Morales, J. C., Perger, M., Dreizler, S., Rodríguez-López, C., González Hernández, J. I., **Rosich, A.**, Feng, F., Trifonov, T., Vogt, S. S., Caballero, J. A., Hatzes, A., Herrero, E., Jeffers, S. V., Lafarga, M., Murgas, F., Nelson, R. P., Rodríguez, E., Strachan, J. B. P., Tal-Or, L., Teske, J., Toledo-Adroán, B., Zechmeister, M., Quirrenbach, A., Amado, P. J., Azzaro, M., Béjar, V. J. S., Barnes, J. R., Berdiñas, Z. M., Burt, J., Coleman, G., Cortés-Contreras, M., Crane, J., Engle, S. G., Guinan, E. F., Haswell, C. A., Henning, T., Holden, B., Jenkins, J., Jones, H. R. A., Kaminski, A., Kiraga, M., Kürster, M., Lee, M. H., López-González, M. J., Montes, D., Morin, J., Ofir, A., Pallé, E., Rebolo, R., Reffert, S., Schweitzer, A., Seifert, W., Shectman, S. A., Staab, D., Street, R. A., Suárez Mascareño, A., Tsapras, Y., Wang, S. X., Anglada-Escudé, G., 2018, *A candidate super-Earth planet orbiting near the snow line of Barnard's star*, Nature, 563, 365-368, <https://ui.adsabs.harvard.edu/abs/2018Natur.563..365R>
3. ★ Mallonn, M., Herrero, E., Juvan, I. G., von Essen, C., **Rosich, A.**, Ribas, I., Granzer, T., Alexoudi, X., Strassmeier, K. G., 2018, *GJ 1214: Rotation period, starspots, and uncertainty on the optical slope of the transmission spectrum*, A&A, 614, A35, <https://ui.adsabs.harvard.edu/abs/2018A&A...614A...35M>
4. ★ Perger, M., G. Anglada-Escudé, I. Ribas, **A. Rosich**, E. Herrero, J. C. Morales, 2021, *Auto-correlation functions of astrophysical processes, and their relation to Gaussian processes: Application to radial velocities of different starspot configurations*, A&A, <https://doi.org/10.1051/0004-6361/202039594>
5. ★ Baroch, D.; Morales, J. C.; Ribas, I.; Herrero, E.; **Rosich, A.**; Perger, M.; Anglada-Escudé, G.; Reiners, A.; Caballero, J. A.; Quirrenbach, A.; Amado, P. J.; Jeffers, S. V.; Cifuentes, C.; Passegger, V. M.; Schweitzer, A.; Lafarga, M.; Bauer, F. F.; Béjar, V. J. S.; Colomé, J.; Cortés-Contreras, M. Dreizler, S.; Galadí-Enríquez, D.; Hatzes, A. P.; Henning, Th.; Kaminski, A.; Kürster, M.; Montes, D.; Rodríguez-López, C.; Zechmeister, M., 2020, *The CARMENES search for exoplanets around M dwarfs. Convective shift and starspot constraints from chromatic radial velocities*, A&A, 641, A69, <https://ui.adsabs.harvard.edu/abs/2020A&A...641A...69B>
6. ★ Herrero, E., I. Ribas, C. Jordi, J. C. Morales, M. Perger, **A. Rosich**, 2016, *Modelling the photosphere of active stars for planet detection and characterization*, A&A, 586, A131, <https://doi.org/10.1051/0004-6361/201425369>

Non-Directly related to the Thesis / other contributions

1. Lafarga, M. et al., 2020, *The CARMENES search for exoplanets around M dwarfs. Radial velocities and activity indicators from cross-correlation functions with weighted binary masks*, A&A, 636, A36, <https://ui.adsabs.harvard.edu/abs/2020A&A...636A..36L>
2. Morales, J. C. et al., 2019, *A giant exoplanet orbiting a very-low-mass star challenges planet formation models*, Science, 6460, 365, 1441-1445, <https://ui.adsabs.harvard.edu/abs/2019Sci...365.1441M>
3. Zechmeister, M. et al., 2019, *The CARMENES search for exoplanets around M dwarfs. Two temperate Earth-mass planet candidates around Teegarden's Star*, A&A, 627, A49, <https://ui.adsabs.harvard.edu/abs/2019A&A...627A..49Z>
4. Lalitha, S. et al., 2019, *The CARMENES search for exoplanets around M dwarfs. Detection of a mini-Neptune around LSPM J2116+0234 and refinement of orbital parameters of a super-Earth around GJ 686 (BD+18 3421)*, A&A, 627, A116, <https://ui.adsabs.harvard.edu/abs/2019A&A...627A.116L>
5. Reiners, A. et al., 2018, *The CARMENES search for exoplanets around M dwarfs. HD147379 b: A nearby Neptune in the temperate zone of an early-M dwarf*, A&A, 609, L5, <https://ui.adsabs.harvard.edu/abs/2018A&A...609L...5R>
6. Tinetti, G. et al., 2018, *A chemical survey of exoplanets with ARIEL*, Experimental Astronomy, 46, 1, 135-209, <https://ui.adsabs.harvard.edu/abs/2018ExA...46..135T>
7. Reiners, A. et al., 2018, *The CARMENES search for exoplanets around M dwarfs. High-resolution optical and near-infrared spectroscopy of 324 survey stars*, A&A, 612, A49, <https://ui.adsabs.harvard.edu/abs/2018A&A...612A..49R>
8. Trifonov, T. et al., 2018, *The CARMENES search for exoplanets around M dwarfs. First visual-channel radial-velocity measurements and orbital parameter updates of seven M-dwarf planetary systems*, A&A, 609, A117, <https://ui.adsabs.harvard.edu/abs/2018A&A...609A.117T>
9. Baroch, D. et al., 2018, *The CARMENES search for exoplanets around M dwarfs. Nine new double-line spectroscopic binary stars*, A&A, 619, A32, <https://ui.adsabs.harvard.edu/abs/2018A&A...619A..32B>
10. Quirrenbach, A. et al., 2018, *CARMENES: high-resolution spectra and precise radial velocities in the red and infrared*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, <https://ui.adsabs.harvard.edu/abs/2018SPIE10702E..0WQ>
11. Suárez Mascareño, A. et al., 2017, *HADES RV Programme with HARPS-N at TNG. V. A super-Earth on the inner edge of the habitable zone of the nearby M dwarf GJ 625*, A&A, 605, A92, <https://ui.adsabs.harvard.edu/abs/2017A&A...605A..92S>

12. Scandariato, G. et al., 2017, *HADES RV Programme with HARPS-N at TNG. IV. Time resolved analysis of the Ca II H&K and H α chromospheric emission of low-activity early-type M dwarfs*, A&A, 598, A28, <https://ui.adsabs.harvard.edu/abs/2017A&A...598A..28S>
13. Perger, M. et al., 2017, *HADES RV Programme with HARPS-N at TNG. VI. GJ 3942 b behind dominant activity signals*, A&A, 608, A63, <https://ui.adsabs.harvard.edu/abs/2017A&A...608A..63P>
14. Affer, L. et al., 2016, *The Hades RV Programme With Harps-N@TNG GJ 3998: An Early M-Dwarf Hosting a System of Super-Earths*, A&A, 593, A117, <https://ui.adsabs.harvard.edu/abs/2016A&A...593A.117A>
15. Perger, M. et al., 2016, *HADES RV Programme with HARPS-N at TNG. II. Data treatment and simulations*, A&A, 598, A26, <https://ui.adsabs.harvard.edu/abs/2017A&A...598A..26P>

Conferences / Talks / Posters

1. ★ Rosich, A., 2018, *Talk: Modeling the photosphere of active stars and the inverse problem: Starsim/2 project*, Workshop: Observing the Sun as a Star: Would we find the solar system if we saw it?, September 10–13 2018; Göttingen, <http://sun-as-a-star.astro.physik.uni-goettingen.de/abstracts.php>
2. ★ Rosich, A., 2016, *Talk: Modeling the photosphere of active stars for exoplanet search*, XII Reunión científica de la Sociedad Española de Astronomía, July 18-22 2016; Bilbao, <http://www.ajax.ehu.es/SEA2016/pdf/ID234.pdf>
3. ≈ Rosich, A., Ribas, I., Perger, M., Herrero, E., Morales, J.C., 2015, *Poster: Multidimensional Generalized Lomb-Scargle Periodogram (MGLS)*, Workshop: Extreme Precision Radial Velocities, July 5-8 2015; Yale (New Haven, USA), <https://docs.google.com/a/yale.edu/viewer?a=v&pid=sites&srcid=\eWFsZS5lZHV8ZXBydi1wb3N0ZXJzGd40jdi0ThmM2EyY2IzZmY2ODg>
4. del Ser, D., O. Fors, J. Núñez, H. Voss, A. Rosich, and V. Kouprianov, 2015, *Detrending Algorithms in Large Time Series: Application to TFRM-PSES Data*, Astronomical Society of the Pacific Conference Series, Vol. 496, p.301, <https://arxiv.org/abs/1411.5320>
5. ★ Rosich, A., Herrero, E., Ribas, I., Perger, M., Jordi, C., 2014, *Poster: Deriving photospheric spot patterns by fitting high-precision lightcurves*, Confence: Towards other Earths II, September 15-19, 2014, Porto, <https://www.astro.up.pt/investigacao/conferencias/toe2014/index.php?opt=posters>

Bibliography

- Aigrain, S., F. Pont, and S. Zucker. 2012. “A simple method to estimate radial velocity variations due to stellar activity using photometry.” *MNRAS* 419 (February): 3147–3158. doi:[10.1111/j.1365-2966.2011.19960.x](https://doi.org/10.1111/j.1365-2966.2011.19960.x). arXiv: [1110.1034](https://arxiv.org/abs/1110.1034) [[astro-ph.SR](#)].
- Alam, Munazza K., Nikolay Nikolov, Mercedes López-Morales, David K. Sing, Jayesh M. Goyal, Gregory W. Henry, Jorge Sanz-Forcada, et al. 2018a. “The HST PanCET Program: Hints of Na I and Evidence of a Cloudy Atmosphere for the Inflated Hot Jupiter WASP-52b.” *AJ* 156, no. 6, 298 (December): 298. doi:[10.3847/1538-3881/aaee89](https://doi.org/10.3847/1538-3881/aaee89). arXiv: [1811.00935](https://arxiv.org/abs/1811.00935) [[astro-ph.EP](#)].
- . 2018b. “The HST PanCET Program: Hints of Na I and Evidence of a Cloudy Atmosphere for the Inflated Hot Jupiter WASP-52b.” *AJ* 156, no. 6, 298 (December): 298. doi:[10.3847/1538-3881/aaee89](https://doi.org/10.3847/1538-3881/aaee89). arXiv: [1811.00935](https://arxiv.org/abs/1811.00935) [[astro-ph.EP](#)].
- Allard, F., D. Homeier, and B. Freytag. 2013. “Atmospheres from very low-mass stars to extrasolar planets.” *memsai* 84:1053.
- Alonso-Floriano, F. J., J. C. Morales, J. A. Caballero, D. Montes, A. Klutsch, R. Mundt, M. Cortés-Contreras, I. Ribas, A. Reiners, and P. J. Amado. 2015. “CARMENES input catalogue of M dwarfs. I. Low-resolution spectroscopy with CAFOS.” *A&A* 577, A128 (May): A128. doi:[10.1051/0004-6361/201525803](https://doi.org/10.1051/0004-6361/201525803). arXiv: [1502.07580](https://arxiv.org/abs/1502.07580) [[astro-ph.SR](#)].
- Andersen, J. M., and H. Korhonen. 2015. “Stellar activity as noise in exoplanet detection - II. Application to M dwarfs.” *MNRAS* 448 (April): 3053–3069. doi:[10.1093/mnras/stu2731](https://doi.org/10.1093/mnras/stu2731). arXiv: [1501.01302](https://arxiv.org/abs/1501.01302) [[astro-ph.SR](#)].
- Anderson, E., Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, et al. 1999. *LAPACK Users ’ Guide*. Third. Philadelphia, PA: Society for Industrial / Applied Mathematics. isbn: 0-89871-447-8 (paperback).
- Anglada-Escudé, P. J. Amado, J. Barnes, Z. M. Berdiñas, R. P. Butler, G. A. L. Coleman, I. de La Cueva, et al. 2016a. “A terrestrial planet candidate in a temperate orbit around Proxima Centauri.” *Nature* 536 (August): 437–440. doi:[10.1038/nature19106](https://doi.org/10.1038/nature19106). arXiv: [1609.03449](https://arxiv.org/abs/1609.03449) [[astro-ph.EP](#)].

- Anglada-Escudé, Pedro J. Amado, John Barnes, Zaira M. Berdiñas, R. Paul Butler, Gavin A. L. Coleman, Ignacio de La Cueva, Stefan Dreizler, Michael Endl, and Benjamin Giesers. 2016b. “A terrestrial planet candidate in a temperate orbit around Proxima Centauri.” *Nature* 536, no. 7617 (August): 437–440. doi:[10.1038/nature19106](https://doi.org/10.1038/nature19106). arXiv: [1609.03449](https://arxiv.org/abs/1609.03449) [[astro-ph.EP](#)].
- Anglada-Escudé and Mikko Tuomi. 2015. “Comment on “Stellar activity masquerading as planets in the habitable zone of the M dwarf Gliese 581” .” *Science* 347 (6226): 1080. doi:[10.1126/science.1260796](https://doi.org/10.1126/science.1260796). eprint: <http://www.sciencemag.org/content/347/6226/1080.2.full.pdf>. <http://www.sciencemag.org/content/347/6226/1080.2.abstract>.
- Anglada-Escudé, Mikko Tuomi, Enrico Gerlach, Rory Barnes, René Heller, James S. Jenkins, Sebastian Wende, et al. 2013. “A dynamically-packed planetary system around GJ 667C with three super-Earths in its habitable zone.” *A&A* 556, A126 (August): A126. doi:[10.1051/0004-6361/201321331](https://doi.org/10.1051/0004-6361/201321331). arXiv: [1306.6074](https://arxiv.org/abs/1306.6074) [[astro-ph.EP](#)].
- Baluev. 2008. “Assessing the statistical significance of periodogram peaks.” *MNRAS* 385, no. 3 (April): 1279–1285. doi:[10.1111/j.1365-2966.2008.12689.x](https://doi.org/10.1111/j.1365-2966.2008.12689.x). arXiv: [0711.0330](https://arxiv.org/abs/0711.0330) [[astro-ph](#)].
- . 2013a. “Detecting multiple periodicities in observational data with the multi-frequency periodogram. I. Analytic assessment of the statistical significance.” *Monthly Notices of the Royal Astronomical Society* 436 (August). doi:[10.1093/mnras/stt1617](https://doi.org/10.1093/mnras/stt1617).
- . 2013b. “Detecting multiple periodicities in observational data with the multi-frequency periodogram II. Frequency Decomposer, a parallelized time-series analysis algorithm.” *Astronomy and Computing* 3 (November): 50–57. doi:[10.1016/j.ascom.2013.11.003](https://doi.org/10.1016/j.ascom.2013.11.003). arXiv: [1309.0100](https://arxiv.org/abs/1309.0100) [[astro-ph.IM](#)].
- . 2013c. “The impact of red noise in radial velocity planet searches: only three planets orbiting GJ 581?” *MNRAS* 429 (March): 2052–2068. doi:[10.1093/mnras/sts476](https://doi.org/10.1093/mnras/sts476). arXiv: [1209.3154](https://arxiv.org/abs/1209.3154) [[astro-ph.EP](#)].
- Barclay, Thomas, Veselin B. Kostov, Knicole D. Colón, Elisa V. Quintana, Joshua E. Schlieder, Dana R. Louie, Emily A. Gilbert, and Susan E. Mullally. 2021. “Stellar surface inhomogeneities as a potential source of the atmospheric signal detected in the K2-18 b transmission spectrum.” *arXiv e-prints*, arXiv:2109.14608 (September): arXiv:2109.14608. arXiv: [2109.14608](https://arxiv.org/abs/2109.14608) [[astro-ph.EP](#)].
- Barnes, J. R., S. V. Jeffers, and H. R. A. Jones. 2011. “The effect of M dwarf starspot activity on low-mass planet detection thresholds.” *MNRAS* 412, no. 3 (April): 1599–1610. doi:[10.1111/j.1365-2966.2010.17979.x](https://doi.org/10.1111/j.1365-2966.2010.17979.x). arXiv: [1011.1125](https://arxiv.org/abs/1011.1125) [[astro-ph.SR](#)].
- Barnes, Sydney A. 2010. “A Simple Nonlinear Model for the Rotation of Main-sequence Cool Stars. I. Introduction, Implications for Gyrochronology, and Color-Period Diagrams.” *ApJ* 722, no. 1 (October): 222–234. doi:[10.1088/0004-637X/722/1/222](https://doi.org/10.1088/0004-637X/722/1/222).

- Baroch, D., J. C. Morales, I. Ribas, E. Herrero, A. Rosich, M. Perger, G. Anglada-Escudé, et al. 2020. “The CARMENES search for exoplanets around M dwarfs. Convective shift and starspot constraints from chromatic radial velocities.” *A&A* 641, A69 (September): A69. doi:[10.1051/0004-6361/202038213](https://doi.org/10.1051/0004-6361/202038213). arXiv: [2006.16608](https://arxiv.org/abs/2006.16608) [[astro-ph.SR](#)].
- Barros, S. C. C., G. Boué, N. P. Gibson, D. L. Pollacco, A. Santerne, F. P. Keenan, I. Skillen, and R. A. Street. 2013. “Transit timing variations in WASP-10b induced by stellar activity.” *MNRAS* 430 (April): 3032–3047. doi:[10.1093/mnras/stt111](https://doi.org/10.1093/mnras/stt111). arXiv: [1301.3760](https://arxiv.org/abs/1301.3760) [[astro-ph.EP](#)].
- Beck, John G. 2000. “A comparison of differential rotation measurements - (Invited Review).” *solphys* 191, no. 1 (January): 47–70. doi:[10.1023/A:1005226402796](https://doi.org/10.1023/A:1005226402796).
- Benedict, G. Fritz, Barbara McArthur, D. W. Chappell, E. Nelan, W. H. Jefferys, W. van Altena, J. Lee, et al. 1999. “Interferometric Astrometry of Proxima Centauri and Barnard’s Star Using HUBBLE SPACE TELESCOPE Fine Guidance Sensor 3: Detection Limits for Substellar Companions.” *The Astronomical Journal* 118 (2): 1086. <http://stacks.iop.org/1538-3881/118/i=2/a=1086>.
- Berdyugina, Svetlana V. 2005. “Starspots: A Key to the Stellar Dynamo.” *Living Reviews in Solar Physics* 2 (8). doi:[10.1007/lrsp-2005-8](https://doi.org/10.1007/lrsp-2005-8). <http://www.livingreviews.org/lrsp-2005-8>.
- Berger, T. E., L. Rouppe van der Voort, and M. Löfdahl. 2007. “Contrast Analysis of Solar Faculae and Magnetic Bright Points.” *ApJ* 661, no. 2 (June): 1272–1288. doi:[10.1086/517502](https://doi.org/10.1086/517502).
- Bertin, E., and S. Arnouts. 1996. “SExtractor: Software for source extraction.” *A&ASS* 117 (June): 393–404. doi:[10.1051/aas:1996164](https://doi.org/10.1051/aas:1996164).
- Blackford, L Susan, Antoine Petitet, Roldan Pozo, Karin Remington, R Clint Whaley, James Demmel, Jack Dongarra, Iain Duff, Sven Hammarling, Greg Henry, et al. 2002. “An updated set of basic linear algebra subprograms (BLAS).” *ACM Transactions on Mathematical Software* 28 (2): 135–151.
- Boisse, I., X. Bonfils, and N. C. Santos. 2012. “SOAP. A tool for the fast computation of photometry and radial velocity induced by stellar spots.” *A&A* 545, A109 (September): A109. doi:[10.1051/0004-6361/201219115](https://doi.org/10.1051/0004-6361/201219115). arXiv: [1206.5493](https://arxiv.org/abs/1206.5493) [[astro-ph.IM](#)].
- Boisse, I., C. Moutou, A. Vidal-Madjar, F. Bouchy, F. Pont, G. Hébrard, X. Bonfils, et al. 2009. “Stellar activity of planetary host star HD 189 733.” *A&A* 495, no. 3 (March): 959–966. doi:[10.1051/0004-6361:200810648](https://doi.org/10.1051/0004-6361:200810648). arXiv: [0811.3923](https://arxiv.org/abs/0811.3923) [[astro-ph](#)].
- Boldt, S., M. Oshagh, S. Dreizler, M. Mallonn, N. C. Santos, A. Claret, A. Reiners, and E. Sedaghati. 2020. “Stellar activity consequence on the retrieved transmission spectra through chromatic Rossiter-McLaughlin observations.” *A&A* 635, A123 (March): A123. doi:[10.1051/0004-6361/201937419](https://doi.org/10.1051/0004-6361/201937419). arXiv: [2002.06373](https://arxiv.org/abs/2002.06373) [[astro-ph.EP](#)].

- Bonfils, X., X. Delfosse, S. Udry, T. Forveille, M. Mayor, C. Perrier, F. Bouchy, et al. 2013a. “The HARPS search for southern extra-solar planets. XXXI. The M-dwarf sample.” *A&A* 549, A109 (January): A109. doi:[10.1051/0004-6361/201014704](https://doi.org/10.1051/0004-6361/201014704). arXiv: [1111.5019](https://arxiv.org/abs/1111.5019) [[astro-ph.EP](#)].
- Bonfils, X., G. Lo Curto, A. C. M. Correia, J. Laskar, S. Udry, X. Delfosse, T. Forveille, et al. 2013b. “The HARPS search for southern extra-solar planets. XXXIV. A planetary system around the nearby M dwarf GJ 163, with a super-Earth possibly in the habitable zone.” *A&A* 556, A110 (August): A110. doi:[10.1051/0004-6361/201220237](https://doi.org/10.1051/0004-6361/201220237). arXiv: [1306.0904](https://arxiv.org/abs/1306.0904) [[astro-ph.EP](#)].
- Bruno, Giovanni, Nikole K. Lewis, Munazza K. Alam, Mercedes López-Morales, Joanna K. Barstow, Hannah R. Wakeford, David Sing, et al. 2019. “WASP-52b. The effect of starspot correction on atmospheric retrievals.” *arXiv e-prints*, arXiv:1911.05179 (November): arXiv:1911.05179. arXiv: [1911.05179](https://arxiv.org/abs/1911.05179) [[astro-ph.EP](#)].
- Bruno, Giovanni, Nikole K. Lewis, Kevin B. Stevenson, Joseph Filippazzo, Matthew Hill, Jonathan D. Fraine, Hannah R. Wakeford, Drake Deming, Mercedes López-Morales, and Munazza K. Alam. 2018. “Starspot Occultations in Infrared Transit Spectroscopy: The Case of WASP-52b.” *AJ* 156, no. 3, 124 (September): 124. doi:[10.3847/1538-3881/aac6db](https://doi.org/10.3847/1538-3881/aac6db). arXiv: [1808.09514](https://arxiv.org/abs/1808.09514) [[astro-ph.EP](#)].
- Budding, E. 1977. *A&SS* 48,207.
- Burrows, A. S. 2014. “Highlights in the study of exoplanet atmospheres.” *Nature* 513 (September): 345–352. doi:[10.1038/nature13782](https://doi.org/10.1038/nature13782). arXiv: [1409.7320](https://arxiv.org/abs/1409.7320) [[astro-ph.EP](#)].
- Buse, A. 1982. “The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note.” *The American Statistician* 36 (3): 153–157. issn: 00031305. <http://www.jstor.org/stable/2683166>.
- Cegla, C. Lovis, V. Bourrier, B. Beeck, C. A. Watson, and F. Pepe. 2016. “The Rossiter-McLaughlin effect reloaded: Probing the 3D spin-orbit geometry, differential stellar rotation, and the spatially-resolved stellar spectrum of star-planet systems.” *A&A* 588, A127 (April): A127. doi:[10.1051/0004-6361/201527794](https://doi.org/10.1051/0004-6361/201527794). arXiv: [1602.00322](https://arxiv.org/abs/1602.00322) [[astro-ph.EP](#)].
- Cegla, C. A. Watson, S. Shelyag, W. J. Chaplin, G. R. Davies, M. Mathioudakis, III Palumbo M. L., S. H. Saar, and R. D. Haywood. 2018. “Stellar Surface Magneto-convection as a Source of Astrophysical Noise. II. Center-to-limb Parameterization of Absorption Line Profiles and Comparison to Observations.” *ApJ* 866, no. 1, 55 (October): 55. doi:[10.3847/1538-4357/aaddfc](https://doi.org/10.3847/1538-4357/aaddfc). arXiv: [1807.11423](https://arxiv.org/abs/1807.11423) [[astro-ph.SR](#)].
- Charbonneau, David, Timothy M. Brown, David W. Latham, and Michel Mayor. 2000. “Detection of Planetary Transits Across a Sun-like Star.” *ApJ Letter* 529, no. 1 (January): L45–L48. doi:[10.1086/312457](https://doi.org/10.1086/312457). arXiv: [astro-ph/9911436](https://arxiv.org/abs/astro-ph/9911436) [[astro-ph](#)].

- Charbonneau, David, Timothy M. Brown, Robert W. Noyes, and Ronald L. Gilliland. 2002. "Detection of an Extrasolar Planet Atmosphere." *The Astrophysical Journal* 568, no. 1 (March): 377–384. issn: 1538-4357. doi:[10.1086/338770](https://doi.org/10.1086/338770). <http://dx.doi.org/10.1086/338770>.
- Chen, G., E. Pallé, L. Nortmann, F. Murgas, H. Parviainen, and G. Nowak. 2017. "The GTC exoplanet transit spectroscopy survey. VI. Detection of sodium in WASP-52b's cloudy atmosphere." *A&A* 600, L11 (April): L11. doi:[10.1051/0004-6361/201730736](https://doi.org/10.1051/0004-6361/201730736). arXiv: [1703.06716](https://arxiv.org/abs/1703.06716) [[astro-ph.EP](#)].
- Collins, Karen A., John F. Kielkopf, Keivan G. Stassun, and Frederic V. Hessman. 2017. "AstroImageJ: Image Processing and Photometric Extraction for Ultra-precise Astronomical Light Curves." *AJ* 153, no. 2, 77 (February): 77. doi:[10.3847/1538-3881/153/2/77](https://doi.org/10.3847/1538-3881/153/2/77). arXiv: [1701.04817](https://arxiv.org/abs/1701.04817) [[astro-ph.IM](#)].
- Colome, J., and I. Ribas. 2006. "ICAT: a General Purpose Image Reduction and Analysis Tool for Robotic Observatories." *IAU Special Session 6*, 11 (August): 11.
- Cramér. 2016. *Mathematical Methods of Statistics (PMS-9)*. Princeton University Press. isbn: 9781400883868. doi:[doi:10.1515/9781400883868](https://doi.org/10.1515/9781400883868). <https://doi.org/10.1515/9781400883868>.
- Cumming, A. 2004. "Detectability of extrasolar planets in radial velocity surveys." *MNRAS* 354 (November): 1165–1176. doi:[10.1111/j.1365-2966.2004.08275.x](https://doi.org/10.1111/j.1365-2966.2004.08275.x). eprint: [astro-ph/0408470](https://arxiv.org/abs/astro-ph/0408470).
- Dekkers, Anton, and Emile Aarts. 1991. "Global optimization and simulated annealing." *Mathematical Programming* 50 (1): 367–393. issn: 1436-4646. doi:[10.1007/BF01594945](https://doi.org/10.1007/BF01594945). <http://dx.doi.org/10.1007/BF01594945>.
- Delfosse, X., X. Bonfils, T. Forveille, S. Udry, M. Mayor, F. Bouchy, M. Gillon, et al. 2013. "The HARPS search for southern extra-solar planets. XXXIII. Super-Earths around the M-dwarf neighbors Gl 433 and Gl 667C." *A&A* 553, A8 (May): A8. doi:[10.1051/0004-6361/201219013](https://doi.org/10.1051/0004-6361/201219013). arXiv: [1202.2467](https://arxiv.org/abs/1202.2467) [[astro-ph.EP](#)].
- Dhariwal, Prafulla, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. 2020. *Jukebox: A Generative Model for Music*. arXiv: [2005.00341](https://arxiv.org/abs/2005.00341) [[eess.AS](#)].
- Díaz, R. F., D. Ségransan, S. Udry, C. Lovis, F. Pepe, X. Dumusque, M. Marmier, R. Alonso, W. Benz, and F. Bouchy. 2016. "The HARPS search for southern extra-solar planets. XXXVIII. Bayesian re-analysis of three systems. New super-Earths, unconfirmed signals, and magnetic cycles." *A&A* 585, A134 (January): A134. doi:[10.1051/0004-6361/201526729](https://doi.org/10.1051/0004-6361/201526729). arXiv: [1510.06446](https://arxiv.org/abs/1510.06446) [[astro-ph.EP](#)].
- Donati, Jean-François, Thierry Forveille, Andrew Collier Cameron, John R. Barnes, Xavier Delfosse, Moira M. Jardine, and Jeff A. Valenti. 2006. "The Large-Scale Axisymmetric Magnetic Topology of a Very-Low-Mass Fully Convective Star." *Science* 311, no.

- 5761 (February): 633–635. doi:[10.1126/science.1121102](https://doi.org/10.1126/science.1121102). arXiv: [astro-ph/0602069](https://arxiv.org/abs/astro-ph/0602069) [[astro-ph](#)].
- Dorren, J.D. 1987. *ApJ* 320,756.
- Dumusque, X., I. Boisse, and N. C. Santos. 2014. “SOAP 2.0: A Tool to Estimate the Photometric and Radial Velocity Variations Induced by Stellar Spots and Plages.” *ApJ* 796, 132 (December): 132. doi:[10.1088/0004-637X/796/2/132](https://doi.org/10.1088/0004-637X/796/2/132). arXiv: [1409.3594](https://arxiv.org/abs/1409.3594) [[astro-ph.SR](#)].
- Dumusque, X., N. C. Santos, S. Udry, C. Lovis, and X. Bonfils. 2011a. “Planetary detection limits taking into account stellar noise. II. Effect of stellar spot groups on radial-velocities.” *A&A* 527, A82 (March): A82. doi:[10.1051/0004-6361/201015877](https://doi.org/10.1051/0004-6361/201015877). arXiv: [1101.0954](https://arxiv.org/abs/1101.0954) [[astro-ph.EP](#)].
- Dumusque, X., S. Udry, C. Lovis, N. C. Santos, and M. J. P. F. G. Monteiro. 2011b. “Planetary detection limits taking into account stellar noise. I. Observational strategies to reduce stellar oscillation and granulation effects.” *A&A* 525, A140 (January): A140. doi:[10.1051/0004-6361/201014097](https://doi.org/10.1051/0004-6361/201014097). arXiv: [1010.2616](https://arxiv.org/abs/1010.2616) [[astro-ph.EP](#)].
- Duric, Neb. 2003. “Orbital mechanics.” In *Advanced Astrophysics*, 3–25. Cambridge University Press. doi:[10.1017/CBO9780511800177.003](https://doi.org/10.1017/CBO9780511800177.003).
- Efron. 1979. “Bootstrap Methods: Another Look at the Jackknife.” *Ann. Statist.* 7, no. 1 (January): 1–26. doi:[10.1214/aos/1176344552](https://doi.org/10.1214/aos/1176344552). <https://doi.org/10.1214/aos/1176344552>.
- . 1982. *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial / Applied Mathematics. doi:[10.1137/1.9781611970319](https://doi.org/10.1137/1.9781611970319). eprint: <https://epubs.siam.org/doi/pdf/10.1137/1.9781611970319>. <https://epubs.siam.org/doi/abs/10.1137/1.9781611970319>.
- Eker, Z. 1999a. “Photometric imaging of starspots, techniques and reliability.” *Turkish Journal of Physics* 23, no. 2 (February): 357–377.
- . 1999b. “The uniqueness problem in starspot models; a critical review.” *na* 4, no. 5 (August): 365–376. doi:[10.1016/S1384-1076\(99\)00028-7](https://doi.org/10.1016/S1384-1076(99)00028-7).
- Encrenaz, Thérèse, G. Tinetti, and A. Coustenis. 2018. “Transit spectroscopy of temperate Jupiters with ARIEL: a feasibility study.” *Experimental Astronomy* 46, no. 1 (November): 31–44. doi:[10.1007/s10686-017-9561-2](https://doi.org/10.1007/s10686-017-9561-2).
- Endl, M., M. Kürster, S. Els, A. P. Hatzes, and W. D. Cochran. 2001. “The planet search program at the ESO Coudé Echelle spectrometer. II. The alpha Centauri system: Limits for planetary companions.” *A&A* 374 (August): 675–681. doi:[10.1051/0004-6361:20010723](https://doi.org/10.1051/0004-6361:20010723).

- Feroz, F., and M. P. Hobson. 2014. “Bayesian analysis of radial velocity data of GJ667C with correlated noise: evidence for only two planets.” *MNRAS* 437 (February): 3540–3549. doi:[10.1093/mnras/stt2148](https://doi.org/10.1093/mnras/stt2148). arXiv: [1307.6984](https://arxiv.org/abs/1307.6984) [[astro-ph.EP](#)].
- Fleming, D. P., and J. VanderPlas. 2018. “approxposterior: Approximate Posterior Distributions in Python.” *The Journal of Open Source Software* 3 (September): 781. doi:[10.21105/joss.00781](https://doi.org/10.21105/joss.00781).
- Foreman-Mackey, Daniel. 2016. “corner.py: Scatterplot matrices in Python.” *The Journal of Open Source Software* 1, no. 2 (June): 24. doi:[10.21105/joss.00024](https://doi.org/10.21105/joss.00024). <https://doi.org/10.21105/joss.00024>.
- Foreman-Mackey, Daniel, David W. Hogg, Dustin Lang, and Jonathan Goodman. 2013. “emcee: The MCMC Hammer.” *Publications of the Astronomical Society of the Pacific* 125, no. 925 (March): 306–312. issn: 1538-3873. doi:[10.1086/670067](https://doi.org/10.1086/670067). <http://dx.doi.org/10.1086/670067>.
- Fortney, J. J., M. Shabram, A. P. Showman, Y. Lian, R. S. Freedman, M. S. Marley, and N. K. Lewis. 2010. “Transmission Spectra of Three-Dimensional Hot Jupiter Model Atmospheres.” *ApJ* 709, no. 2 (February): 1396–1406. doi:[10.1088/0004-637X/709/2/1396](https://doi.org/10.1088/0004-637X/709/2/1396). arXiv: [0912.2350](https://arxiv.org/abs/0912.2350) [[astro-ph.EP](#)].
- Foster, Grant. 1995. “The Cleanest Fourier Spectrum.” *AJ* 109 (April): 1889. doi:[10.1086/117416](https://doi.org/10.1086/117416).
- Frazier, E. N., and J. O. Stenflo. 1978. “Magnetic, velocity and brightness structure of solar faculae.” *A&A* 70, no. 6 (December): 789–799.
- Frescura, F. A. M., C. A. Engelbrecht, and B. S. Frank. 2008. “Significance of periodogram peaks and a pulsation mode analysis of the Beta Cephei star V403 Car.” *Monthly Notices of the Royal Astronomical Society* 388, no. 4 (August): 1693–1707. issn: 0035-8711. doi:[10.1111/j.1365-2966.2008.13499.x](https://doi.org/10.1111/j.1365-2966.2008.13499.x). eprint: <https://academic.oup.com/mnras/article-pdf/388/4/1693/3040898/mnras0388-1693.pdf>. <https://doi.org/10.1111/j.1365-2966.2008.13499.x>.
- Gabbard, Hunter, Chris Messenger, Ik Siong Heng, Francesco Tonolini, and Roderick Murray-Smith. 2019. “Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy.” *arXiv e-prints*, arXiv:1909.06296 (September): arXiv:1909.06296. arXiv: [1909.06296](https://arxiv.org/abs/1909.06296) [[astro-ph.IM](#)].
- Gaia Collaboration, A. G. A. Brown, A. Vallenari, T. Prusti, J. H. J. de Bruijne, C. Babusiaux, C. A. L. Bailer-Jones, M. Biermann, D. W. Evans, and L. Eyer. 2018. “Gaia Data Release 2. Summary of the contents and survey properties.” *A&A*, A1 (August): A1. doi:[10.1051/0004-6361/201833051](https://doi.org/10.1051/0004-6361/201833051). arXiv: [1804.09365](https://arxiv.org/abs/1804.09365) [[astro-ph.GA](#)].

- Gardner, Jonathan P., John C. Mather, Mark Clampin, Rene Doyon, Matthew A. Greenhouse, Heidi B. Hammel, John B. Hutchings, et al. 2006. "The James Webb Space Telescope." *Space Science Reviews* 123, no. 4 (November): 485–606. issn: 1572-9672. doi:[10.1007/s11214-006-8315-7](https://doi.org/10.1007/s11214-006-8315-7). arXiv: [astro-ph/0606175](https://arxiv.org/abs/astro-ph/0606175) [[astro-ph](#)].
- Geballe, T. R., G. R. Knapp, S. K. Leggett, X. Fan, D. A. Golimowski, S. Anderson, J. Brinkmann, et al. 2002. "Toward Spectral Classification of L and T Dwarfs: Infrared and Optical Spectroscopy and Analysis." *ApJ* 564, no. 1 (January): 466–481. doi:[10.1086/324078](https://doi.org/10.1086/324078). arXiv: [astro-ph/0108443](https://arxiv.org/abs/astro-ph/0108443) [[astro-ph](#)].
- Giampaoli, Iacopo, Wing Lon Ng, and Nick Constantinou. 2009. "Analysis of ultra-high-frequency financial data using advanced Fourier transforms." *Finance Research Letters* 6 (1): 47–53. issn: 1544-6123. doi:<https://doi.org/10.1016/j.frl.2008.11.002>. <https://www.sciencedirect.com/science/article/pii/S1544612308000603>.
- Gillon, M., A. H. M. J. Triaud, B.-O. Demory, E. Jehin, E. Agol, K. M. Deck, S. M. Lederer, et al. 2017. "Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1." *ArXiv e-prints* (March). arXiv: [1703.01424](https://arxiv.org/abs/1703.01424) [[astro-ph.EP](#)].
- Glynn, Earl F., Jie Chen, and Arcady R. Mushegian. 2005. "Detecting periodic patterns in unevenly spaced gene expression time series using Lomb–Scargle periodograms." *Bioinformatics* 22, no. 3 (November): 310–316. issn: 1367-4803. doi:[10.1093/bioinformatics/bti789](https://doi.org/10.1093/bioinformatics/bti789). eprint: <https://academic.oup.com/bioinformatics/article-pdf/22/3/310/18529532/bti789.pdf>. <https://doi.org/10.1093/bioinformatics/bti789>.
- Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. *Generative Adversarial Networks*. arXiv: [1406.2661](https://arxiv.org/abs/1406.2661) [[stat.ML](#)].
- Gray, R. O., C. J. Corbally, R. F. Garrison, M. T. McFadden, E. J. Bubar, C. E. McGahee, A. A. O’Donoghue, and E. R. Knox. 2006. "Contributions to the Nearby Stars (NStars) Project: Spectroscopy of Stars Earlier than M0 within 40 pc-The Southern Sample." *AJ* 132, no. 1 (July): 161–170. doi:[10.1086/504637](https://doi.org/10.1086/504637). arXiv: [astro-ph/0603770](https://arxiv.org/abs/astro-ph/0603770) [[astro-ph](#)].
- Grill, Jean-Bastien, Michal Valko, and Rémi Munos. 2015. "Black-box optimization of noisy functions with unknown smoothness." In *Neural Information Processing Systems*. Montréal, Canada. <https://hal.inria.fr/hal-01222915>.
- Günther, Maximilian N., and Tansu Daylan. 2020. "Allesfitter: Flexible Star and Exoplanet Inference From Photometry and Radial Velocity." *arXiv e-prints*, arXiv:2003.14371 (March): arXiv:2003.14371. arXiv: [2003.14371](https://arxiv.org/abs/2003.14371) [[astro-ph.EP](#)].
- Gururajan, Gautham, and Shantanu Desai. 2020. "Generalized Lomb–Scargle analysis of ^{123}I and $^{99\text{m}}\text{Tc}$ decay rate measurements." *The European Physical Journal C* 80

- (11): 1071. doi:[10.1140/epjc/s10052-020-08663-8](https://doi.org/10.1140/epjc/s10052-020-08663-8). <https://doi.org/10.1140/epjc/s10052-020-08663-8>.
- Haddad, Caroline N. 2009. “Cholesky factorizationCholesky Factorization.” In *Encyclopedia of Optimization*, edited by Christodoulos A. Floudas and Panos M. Pardalos, 374–377. Boston, MA: Springer US. isbn: 978-0-387-74759-0. doi:[10.1007/978-0-387-74759-0_67](https://doi.org/10.1007/978-0-387-74759-0_67). https://doi.org/10.1007/978-0-387-74759-0_67.
- Hardegree-Ullman, Kevin K., Michael C. Cushing, Philip S. Muirhead, and Jessie L. Christiansen. 2019. “Kepler Planet Occurrence Rates for Mid-type M Dwarfs as a Function of Spectral Type.” *AJ* 158, no. 2, 75 (August): 75. doi:[10.3847/1538-3881/ab21d2](https://doi.org/10.3847/1538-3881/ab21d2). arXiv: [1905.05900](https://arxiv.org/abs/1905.05900) [astro-ph.EP].
- Hauschildt, P. H., and E. Baron. 1999. “Numerical solution of the expanding stellar atmosphere problem.” *Journal of Computational and Applied Mathematics* 109, no. 1 (September): 41–63. arXiv: [astro-ph/9808182](https://arxiv.org/abs/astro-ph/9808182) [astro-ph].
- Hébrard, G., A. Collier Cameron, D. J. A. Brown, R. F. Delfino, F. Faedi, B. Smalley, D. R. Anderson, D. Armstrong, S. C. C. Barros, and J. Bento. 2013. “WASP-52b, WASP-58b, WASP-59b, and WASP-60b: Four new transiting close-in giant planets.” *A&A* 549, A134 (January): A134. doi:[10.1051/0004-6361/201220363](https://doi.org/10.1051/0004-6361/201220363). arXiv: [1211.0810](https://arxiv.org/abs/1211.0810) [astro-ph.EP].
- Herrero, E., I. Ribas, C. Jordi, J. C. Morales, M. Perger, and A. Rosich. 2016. “Modelling the photosphere of active stars for planet detection and characterization.” *A&A* 586 (February): A131. doi:[10.1051/0004-6361/201425369](https://doi.org/10.1051/0004-6361/201425369). arXiv: [1511.06717](https://arxiv.org/abs/1511.06717) [astro-ph.EP].
- Horne, J. H., and S. L. Baliunas. 1986. “A prescription for period analysis of unevenly sampled time series.” *ApJ* 302 (March): 757–763. doi:[10.1086/164037](https://doi.org/10.1086/164037).
- Huitson, C. M., D. K. Sing, F. Pont, J. J. Fortney, A. S. Burrows, P. A. Wilson, G. E. Ballester, et al. 2013. “An HST optical-to-near-IR transmission spectrum of the hot Jupiter WASP-19b: detection of atmospheric water and likely absence of TiO.” *MNRAS* 434, no. 4 (October): 3252–3274. doi:[10.1093/mnras/stt1243](https://doi.org/10.1093/mnras/stt1243). arXiv: [1307.2083](https://arxiv.org/abs/1307.2083) [astro-ph.EP].
- Husser, T.-O., S. Wende-von Berg, S. Dreizler, D. Homeier, A. Reiners, T. Barman, and P. H. Hauschildt. 2013. “A new extensive library of PHOENIX stellar atmospheres and synthetic spectra.” *aap* 553, A6 (May): A6. doi:[10.1051/0004-6361/201219058](https://doi.org/10.1051/0004-6361/201219058). arXiv: [1303.5632](https://arxiv.org/abs/1303.5632) [astro-ph.SR].
- Jeffers, S. V., J. R. Barnes, H. R. A. Jones, A. Reiners, D. J. Pinfield, and S. C. Marsden. 2014. “Is it possible to detect planets around young active G and K dwarfs?” *MNRAS* 438, no. 4 (March): 2717–2731. doi:[10.1093/mnras/stt1950](https://doi.org/10.1093/mnras/stt1950). arXiv: [1311.3617](https://arxiv.org/abs/1311.3617) [astro-ph.SR].

- Jiang, C., G. Chen, E. Palle, F. Murgas, H. Parviainen, F. Yan, and Y. Ma. 2021. “Evidence for stellar contamination in the transmission spectra of HAT-P-12b.” *arXiv e-prints*, arXiv:2109.11235 (September): arXiv:2109.11235. arXiv: [2109.11235 \[astro-ph.EP\]](#).
- Juvan, Ines G., M. Lendl, P. E. Cubillos, L. Fossati, J. Tregloan-Reed, H. Lammer, E. W. Guenther, and A. Hanslmeier. 2018. “PyTranSpot: A tool for multiband light curve modeling of planetary transits and stellar spots.” *A&A* 610, A15 (February): A15. doi:[10.1051/0004-6361/201731345](#). arXiv: [1710.11209 \[astro-ph.EP\]](#).
- Kabanikhin, S., N Tikhonov, V Ivanov, and M Lavrentiev. 2008. “Definitions and examples of inverse and ill-posed problems.” *Journal of Inverse and Ill-posed Problems - J INVERSE ILL-POSED PROBL* 16 (January): 317–357.
- Kingma, Diederik P, and Max Welling. 2014. *Auto-Encoding Variational Bayes*. arXiv: [1312.6114 \[stat.ML\]](#).
- Kipping, David M. 2012. “An analytic model for rotational modulations in the photometry of spotted stars.” *MNRAS* 427, no. 3 (December): 2487–2511. doi:[10.1111/j.1365-2966.2012.22124.x](#). arXiv: [1209.2985 \[astro-ph.SR\]](#).
- Kirk, J., P. J. Wheatley, T. Louden, S. P. Littlefair, C. M. Copperwheat, D. J. Armstrong, T. R. Marsh, and V. S. Dhillon. 2016. “Transmission spectroscopy of the inflated exoplanet WASP-52b, and evidence for a bright region on the stellar surface.” *MNRAS* 463, no. 3 (December): 2922–2931. doi:[10.1093/mnras/stw2205](#). arXiv: [1608.08993 \[astro-ph.EP\]](#).
- Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi. 1983. “Optimization by simulated annealing.” *SCIENCE* 220 (4598): 671–680.
- Knutson, Heather A., Nikole Lewis, Jonathan J. Fortney, Adam Burrows, Adam P. Showman, Nicolas B. Cowan, Eric Agol, et al. 2012. “3.6 and 4.5 μm Phase Curves and Evidence for Non-equilibrium Chemistry in the Atmosphere of Extrasolar Planet HD 189733b.” *ApJ* 754, no. 1, 22 (July): 22. doi:[10.1088/0004-637X/754/1/22](#). arXiv: [1206.6887 \[astro-ph.EP\]](#).
- Koen. 1990. “Significance Testing of Periodogram Ordinates.” *ApJ* 348 (January): 700. doi:[10.1086/168277](#).
- Koen, C., D. Kilkeny, F. van Wyk, and F. Marang. 2010. “UBV(RI)_C JHK observations of Hipparcos-selected nearby stars.” *MNRAS* 403, no. 4 (April): 1949–1968. doi:[10.1111/j.1365-2966.2009.16182.x](#).
- Kopparapu, R. K. 2013. “A Revised Estimate of the Occurrence Rate of Terrestrial Planets in the Habitable Zones around Kepler M-dwarfs.” *ApJ Letter* 767, L8 (April): L8. doi:[10.1088/2041-8205/767/1/L8](#). arXiv: [1303.2649 \[astro-ph.EP\]](#).
- Kovari, Zs., and J. Bartus. 1997. “Testing the stability and reliability of starspot modelling.” *A&A* 323 (July): 801–808.

- Kraft, D. 1988. *A software package for sequential quadratic programming*. Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt Köln: Forschungsbericht. Wiss. Berichtswesen d. DFVLR. <https://books.google.es/books?id=4rKaGwAACAAJ>.
- Kreidberg, Laura. 2018. “Exoplanet Atmosphere Measurements from Transmission Spectroscopy and Other Planet Star Combined Light Observations.” In *Handbook of Exoplanets*, ISBN 978-3-319-55332-0. 100. Springer International Publishing AG, part of Springer Nature, 2018, id.100. doi:[10.1007/978-3-319-55333-7_100](https://doi.org/10.1007/978-3-319-55333-7_100).
- Kulesa, Anthony, Martin Krzywinski, Paul Blainey, and Naomi Altman. 2015. “Sampling distributions and the bootstrap.” *Nature Methods* 12 (6): 477–478. doi:[10.1038/nmeth.3414](https://doi.org/10.1038/nmeth.3414). <https://doi.org/10.1038/nmeth.3414>.
- Kunimoto, Michelle, and Jaymie M. Matthews. 2020. “Searching the Entirety of Kepler Data. II. Occurrence Rate Estimates for FGK Stars.” *AJ* 159, no. 6, 248 (June): 248. doi:[10.3847/1538-3881/ab88b0](https://doi.org/10.3847/1538-3881/ab88b0). arXiv: [2004.05296](https://arxiv.org/abs/2004.05296) [astro-ph.EP].
- Kurucz, Robert L. 2017. *ATLAS9: Model atmosphere program with opacity distribution functions*, October. ascl: [1710.017](https://ascl.net/1710.017).
- Lafarga, M., I. Ribas, C. Lovis, M. Perger, M. Zechmeister, F. F. Bauer, M. Kürster, et al. 2020. “The CARMENES search for exoplanets around M dwarfs. Radial velocities and activity indicators from cross-correlation functions with weighted binary masks.” *A&A* 636, A36 (April): A36. doi:[10.1051/0004-6361/201937222](https://doi.org/10.1051/0004-6361/201937222). arXiv: [2003.07471](https://arxiv.org/abs/2003.07471) [astro-ph.IM].
- Lagrange, A.-M., M. Desort, and N. Meunier. 2010. “Using the Sun to estimate Earth-like planets detection capabilities . I. Impact of cold spots.” *A&A* 512, A38 (March): A38. doi:[10.1051/0004-6361/200913071](https://doi.org/10.1051/0004-6361/200913071). arXiv: [1001.1449](https://arxiv.org/abs/1001.1449) [astro-ph.EP].
- Lanza, A. F., I. Boisse, F. Bouchy, A. S. Bonomo, and C. Moutou. 2011. “Deriving the radial-velocity variations induced by stellar activity from high-precision photometry. Test on HD 189733 with simultaneous MOST/SOPHIE data.” *A&A* 533, A44 (September): A44. doi:[10.1051/0004-6361/201117270](https://doi.org/10.1051/0004-6361/201117270). arXiv: [1107.4864](https://arxiv.org/abs/1107.4864) [astro-ph.SR].
- Lanza, A. F., M. Rodonò, I. Pagano, P. Barge, and A. Llebaria. 2003. “Modelling the rotational modulation of the Sun as a star.” *A&A* 403 (June): 1135–1149. doi:[10.1051/0004-6361:20030401](https://doi.org/10.1051/0004-6361:20030401).
- Lockwood, G. W., B. A. Skiff, G. W. Henry, S. Henry, R. R. Radick, S. L. Baliunas, R. A. Donahue, and W. Soon. 2007. “Patterns of Photometric and Chromospheric Variation among Sun-like Stars: A 20 Year Perspective.” *ApJSS* 171 (July): 260–303. doi:[10.1086/516752](https://doi.org/10.1086/516752). eprint: [astro-ph/0703408](https://arxiv.org/abs/astro-ph/0703408).
- Lomb, N. R. 1976. “Least-squares frequency analysis of unequally spaced data.” *A&SS* 39 (February): 447–462. doi:[10.1007/BF00648343](https://doi.org/10.1007/BF00648343).

- Lopez-Alvis, Jorge, Eric Laloy, Frédéric Nguyen, and Thomas Hermans. 2021. “Deep generative models in inversion: The impact of the generator’s nonlinearity and development of a new approach based on a variational autoencoder.” *Computers and Geosciences* 152, 104762 (July): 104762. doi:[10.1016/j.cageo.2021.104762](https://doi.org/10.1016/j.cageo.2021.104762). arXiv: [2008.12056](https://arxiv.org/abs/2008.12056) [[physics.geo-ph](#)].
- Louden, Tom, Peter J. Wheatley, Patrick G. J. Irwin, James Kirk, and Ian Skillen. 2017. “A precise optical transmission spectrum of the inflated exoplanet WASP-52b.” *MNRAS* 470, no. 1 (September): 742–754. doi:[10.1093/mnras/stx984](https://doi.org/10.1093/mnras/stx984). arXiv: [1703.09285](https://arxiv.org/abs/1703.09285) [[astro-ph.EP](#)].
- Lovis, C., X. Dumusque, N. C. Santos, F. Bouchy, M. Mayor, F. Pepe, D. Queloz, D. Ségransan, and S. Udry. 2011a. “The HARPS search for southern extra-solar planets. XXXI. Magnetic activity cycles in solar-type stars: statistics and impact on precise radial velocities.” *arXiv e-prints*, arXiv:1107.5325 (July): arXiv:1107.5325. arXiv: [1107.5325](https://arxiv.org/abs/1107.5325) [[astro-ph.SR](#)].
- . 2011b. “The HARPS search for southern extra-solar planets. XXXI. Magnetic activity cycles in solar-type stars: statistics and impact on precise radial velocities.” *arXiv e-prints*, arXiv:1107.5325 (July): arXiv:1107.5325. arXiv: [1107.5325](https://arxiv.org/abs/1107.5325) [[astro-ph.SR](#)].
- Lovis, C., D. Ségransan, M. Mayor, S. Udry, W. Benz, J.-L. Bertaux, F. Bouchy, et al. 2011c. “The HARPS search for southern extra-solar planets. XXVIII. Up to seven planets orbiting HD 10180: probing the architecture of low-mass planetary systems.” *A&A* 528, A112 (April): A112. doi:[10.1051/0004-6361/201015577](https://doi.org/10.1051/0004-6361/201015577). arXiv: [1011.4994](https://arxiv.org/abs/1011.4994) [[astro-ph.EP](#)].
- Ludwig, H. -G., E. Caffau, M. Steffen, B. Freytag, P. Bonifacio, and A. Kuvcinskas. 2009. “The CIFIST 3D model atmosphere grid.” *memsai* 80 (January): 711. arXiv: [0908.4496](https://arxiv.org/abs/0908.4496) [[astro-ph.SR](#)].
- Luger, Rodrigo, Daniel Foreman-Mackey, and Christina Hedges. 2021a. “Mapping Stellar Surfaces. II. An Interpretable Gaussian Process Model for Light Curves.” *AJ* 162, no. 3, 124 (September): 124. doi:[10.3847/1538-3881/abfdb9](https://doi.org/10.3847/1538-3881/abfdb9). arXiv: [2102.01697](https://arxiv.org/abs/2102.01697) [[astro-ph.SR](#)].
- Luger, Rodrigo, Daniel Foreman-Mackey, Christina Hedges, and David W. Hogg. 2021b. “Mapping stellar surfaces I: Degeneracies in the rotational light curve problem.” *arXiv e-prints*, arXiv:2102.00007 (January): arXiv:2102.00007. arXiv: [2102.00007](https://arxiv.org/abs/2102.00007) [[astro-ph.SR](#)].
- Luo, Tao, Yanyan Liang, and Wing-Huen Ip. 2019. “Efficient Starspot Reconstruction via Light-curve Inversion with Bipartite Regularization.” *AJ* 157, no. 6, 238 (June): 238. doi:[10.3847/1538-3881/ab1b46](https://doi.org/10.3847/1538-3881/ab1b46).
- Mallonn, M., I. Bernt, E. Herrero, S. Hoyer, J. Kirk, P. J. Wheatley, M. Seeliger, et al. 2016. “Broad-band spectrophotometry of HAT-P-32 b: search for a scattering signature

- in the planetary spectrum.” *MNRAS* 463, no. 1 (November): 604–614. doi:[10.1093/mnras/stw1999](https://doi.org/10.1093/mnras/stw1999). arXiv: [1608.02346](https://arxiv.org/abs/1608.02346) [[astro-ph.EP](#)].
- Mallon, M., E. Herrero, I. G. Juvan, C. von Essen, A. Rosich, I. Ribas, T. Granzer, X. Alexoudi, and K. G. Strassmeier. 2018. “GJ 1214: Rotation period, starspots, and uncertainty on the optical slope of the transmission spectrum.” *A&A* 614, A35 (June): A35. doi:[10.1051/0004-6361/201732300](https://doi.org/10.1051/0004-6361/201732300). arXiv: [1803.05677](https://arxiv.org/abs/1803.05677) [[astro-ph.SR](#)].
- Mallon, M., C. von Essen, J. Weingrill, K. G. Strassmeier, I. Ribas, T. A. Carroll, E. Herrero, T. Granzer, A. Claret, and A. Schwöpe. 2015. “Transmission spectroscopy of the inflated exo-Saturn HAT-P-19b.” *A&A* 580, A60 (August): A60. doi:[10.1051/0004-6361/201423778](https://doi.org/10.1051/0004-6361/201423778). arXiv: [1506.05685](https://arxiv.org/abs/1506.05685) [[astro-ph.EP](#)].
- Mallon, M., and Hannah R. Wakeford. 2017. “Near-ultraviolet transit photometry of HAT-P-32 b with the Large Binocular Telescope: Silicate aerosols in the planetary atmosphere.” *Astronomische Nachrichten* 338, no. 7 (August): 773–780. doi:[10.1002/asna.201713376](https://doi.org/10.1002/asna.201713376). arXiv: [1707.08328](https://arxiv.org/abs/1707.08328) [[astro-ph.EP](#)].
- Mancini, L., J. Southworth, G. Raia, J. Tregloan-Reed, P. Mollière, V. Bozza, M. Bretton, et al. 2017. “Orbital alignment and star-spot properties in the WASP-52 planetary system.” *MNRAS* 465 (February): 843–857. doi:[10.1093/mnras/stw1987](https://doi.org/10.1093/mnras/stw1987). arXiv: [1608.02001](https://arxiv.org/abs/1608.02001) [[astro-ph.EP](#)].
- Marcy, G. W., and R. P. Butler. 1996. “A Planetary Companion to 70 Virginis.” *ApJ Letter* 464 (June): L147. doi:[10.1086/310096](https://doi.org/10.1086/310096).
- Martínez-Palomera, Jorge, Joshua S. Bloom, and Ellianna S. Abrahams. 2020. “Deep Generative Modeling of Periodic Variable Stars Using Physical Parameters.” *arXiv e-prints*, arXiv:2005.07773 (May): arXiv:2005.07773. arXiv: [2005.07773](https://arxiv.org/abs/2005.07773) [[astro-ph.IM](#)].
- May, E. M., M. Zhao, M. Haidar, E. Rauscher, and J. D. Monnier. 2018. “MOPSS. I. Flat Optical Spectra for the Hot Jupiters WASP-4 b and WASP-52b.” *AJ* 156, no. 3, 122 (September): 122. doi:[10.3847/1538-3881/aad4a8](https://doi.org/10.3847/1538-3881/aad4a8). arXiv: [1807.06561](https://arxiv.org/abs/1807.06561) [[astro-ph.EP](#)].
- Mayor, M., F. Pepe, D. Queloz, F. Bouchy, G. Rupprecht, G. Lo Curto, G. Avila, W. Benz, J. -L. Bertaux, and X. Bonfils. 2003. “Setting New Standards with HARPS.” *The Messenger* 114 (December): 20–24.
- Mayor, M., and D. Queloz. 1995. “A Jupiter-mass companion to a solar-type star.” *Nat* 378 (November): 355–359.
- Mayor, M., S. Udry, C. Lovis, F. Pepe, D. Queloz, W. Benz, J.-L. Bertaux, F. Bouchy, C. Mordasini, and D. Segransan. 2009. “The HARPS search for southern extra-solar planets. XIII. A planetary system with 3 super-Earths.” *A&A* 493 (January): 639–644. doi:[10.1051/0004-6361:200810451](https://doi.org/10.1051/0004-6361:200810451). arXiv: [0806.4587](https://arxiv.org/abs/0806.4587).

- McCullough, P. R., N. Crouzet, D. Deming, and N. Madhusudhan. 2014. “Water Vapor in the Spectrum of the Extrasolar Planet HD 189733b. I. The Transit.” *ApJ* 791, no. 1, 55 (August): 55. doi:[10.1088/0004-637X/791/1/55](https://doi.org/10.1088/0004-637X/791/1/55). arXiv: [1407.2462](https://arxiv.org/abs/1407.2462) [[astro-ph.SR](#)].
- McLaughlin. 1924. “Some results of a spectrographic study of the Algol system.” *ApJ* 60 (July): 22–31. doi:[10.1086/142826](https://doi.org/10.1086/142826).
- Meunier, N., M. Desort, and A.-M. Lagrange. 2010a. “Using the Sun to estimate Earth-like planets detection capabilities . II. Impact of plages.” *A&A* 512, A39 (March): A39. doi:[10.1051/0004-6361/200913551](https://doi.org/10.1051/0004-6361/200913551). arXiv: [1001.1638](https://arxiv.org/abs/1001.1638) [[astro-ph.EP](#)].
- . 2010b. “Using the Sun to estimate Earth-like planets detection capabilities . II. Impact of plages.” *A&A* 512, A39 (March): A39. doi:[10.1051/0004-6361/200913551](https://doi.org/10.1051/0004-6361/200913551). arXiv: [1001.1638](https://arxiv.org/abs/1001.1638) [[astro-ph.EP](#)].
- Meunier, N., A. -M. Lagrange, S. Borgniet, and M. Rieutord. 2015. “Using the Sun to estimate Earth-like planet detection capabilities. VI. Simulation of granulation and supergranulation radial velocity and photometric time series.” *A&A* 583, A118 (November): A118. doi:[10.1051/0004-6361/201525721](https://doi.org/10.1051/0004-6361/201525721).
- Montalto, M., G. Boué, M. Oshagh, I. Boisse, G. Bruno, and N. C. Santos. 2014. “Improvements on analytic modelling of stellar spots.” *MNRAS* 444, no. 2 (October): 1721–1728. doi:[10.1093/mnras/stu1530](https://doi.org/10.1093/mnras/stu1530). arXiv: [1407.2155](https://arxiv.org/abs/1407.2155) [[astro-ph.EP](#)].
- Morris, Brett. 2020. “fleck: Fast approximate light curves for starspot rotational modulation.” *Journal of Open Source Software* 5 (March): 2103. doi:[10.21105/joss.02103](https://doi.org/10.21105/joss.02103).
- Mortier, A., J. P. Faria, C. M. Correia, A. Santerne, and N. C. Santos. 2015. “BGLS: A Bayesian formalism for the generalised Lomb-Scargle periodogram.” *A&A* 573, A101 (January): A101. doi:[10.1051/0004-6361/201424908](https://doi.org/10.1051/0004-6361/201424908). arXiv: [1412.0467](https://arxiv.org/abs/1412.0467) [[astro-ph.IM](#)].
- Murakami, Yukei S., Connor Jennings, Andrew M. Hoffman, James Sunseri, Raphael Baer-Way, Benjamin E. Stahl, Arjun B. Savel, Ivan Altunin, Nachiket Girish, and Alexei V. Filippenko. 2021. “PIPS, an advanced platform for period detection in time series – I. Fourier-likelihood periodogram and application to RR Lyrae Stars.” *arXiv e-prints*, arXiv:2107.14223 (July): arXiv:2107.14223. arXiv: [2107.14223](https://arxiv.org/abs/2107.14223) [[astro-ph.IM](#)].
- Murdoch, K. A., J. B. Hearnshaw, and M. Clark. 1993. “A search for substellar companions to southern solar-type stars.” *ApJ* 413 (August): 349–363. doi:[10.1086/173003](https://doi.org/10.1086/173003).
- Nourani, Yaghout, and Bjarne Andresen. 1998. “A comparison of simulated annealing cooling strategies.” *Journal of Physics A: Mathematical and General* 31 (41): 8373. <http://stacks.iop.org/0305-4470/31/i=41/a=011>.
- Oláh, K., Zs. Kovári, K. Petrovay, W. Soon, S. Baliunas, Z. Kolláth, and K. Vida. 2016. “Magnetic cycles at different ages of stars.” *A&A* 590, A133 (June): A133. doi:[10.1051/0004-6361/201628479](https://doi.org/10.1051/0004-6361/201628479). arXiv: [1604.06701](https://arxiv.org/abs/1604.06701) [[astro-ph.SR](#)].

- Oshagh, M., I. Boisse, G. Boué, M. Montalto, N. C. Santos, X. Bonfils, and N. Haghighipour. 2013a. “SOAP-T: a tool to study the light curve and radial velocity of a system with a transiting planet and a rotating spotted star.” *A&A* 549, A35 (January): A35. doi:[10.1051/0004-6361/201220173](https://doi.org/10.1051/0004-6361/201220173). arXiv: [1211.1311](https://arxiv.org/abs/1211.1311) [[astro-ph.EP](#)].
- Oshagh, M., N. C. Santos, I. Boisse, G. Boué, M. Montalto, X. Dumusque, and N. Haghighipour. 2013b. “Effect of stellar spots on high-precision transit light-curve.” *A&A* 556, A19 (August): A19. doi:[10.1051/0004-6361/201321309](https://doi.org/10.1051/0004-6361/201321309). arXiv: [1306.0739](https://arxiv.org/abs/1306.0739) [[astro-ph.EP](#)].
- Oshagh, M., N. C. Santos, D. Ehrenreich, N. Haghighipour, P. Figueira, A. Santerne, and M. Montalto. 2014a. “Impact of occultations of stellar active regions on transmission spectra. Can occultation of a plage mimic the signature of a blue sky?” *A&A* 568, A99 (August): A99. doi:[10.1051/0004-6361/201424059](https://doi.org/10.1051/0004-6361/201424059). arXiv: [1407.2066](https://arxiv.org/abs/1407.2066) [[astro-ph.EP](#)].
- . 2014b. “Impact of occultations of stellar active regions on transmission spectra. Can occultation of a plage mimic the signature of a blue sky?” *A&A* 568, A99 (August): A99. doi:[10.1051/0004-6361/201424059](https://doi.org/10.1051/0004-6361/201424059). arXiv: [1407.2066](https://arxiv.org/abs/1407.2066) [[astro-ph.EP](#)].
- Oshagh, M., A. H. M. J. Triaud, A. Burdanov, P. Figueira, A. Reiners, N. C. Santos, J. Faria, G. Boue, R. F. Diaz, and S. Dreizler. 2018. “Activity induced variation in spin-orbit angles as derived from Rossiter-McLaughlin measurements.” *A&A* 619, A150 (November): A150. doi:[10.1051/0004-6361/201833709](https://doi.org/10.1051/0004-6361/201833709). arXiv: [1809.01027](https://arxiv.org/abs/1809.01027) [[astro-ph.EP](#)].
- Öztürk, Oguz, and Ahmet Erdem. 2019. “New photometric analysis of five exoplanets: CoRoT-2b, HAT-P-12b, TrES-2b, WASP-12b, and WASP-52b.” *MNRAS* 486, no. 2 (June): 2290–2307. doi:[10.1093/mnras/stz747](https://doi.org/10.1093/mnras/stz747).
- Pepe, S. Cristiani, R. Rebolo, N. C. Santos, H. Dekker, A. Cabral, P. Di Marcantonio, et al. 2021. “ESPRESSO at VLT.” *A* 645 (January): A96. issn: 1432-0746. doi:[10.1051/0004-6361/202038306](https://doi.org/10.1051/0004-6361/202038306). <http://dx.doi.org/10.1051/0004-6361/202038306>.
- Pepe, M. Mayor, D. Queloz, W. Benz, J.-L. Bertaux, F. Bouchy, C. Lovis, et al. 2005. “On the Track of Very Low-mass Planets with HARPS.” *The Messenger* 120 (June): 22–25.
- Perger, M., I. Ribas, M. Damasso, J. C. Morales, L. Affer, A. Suárez Mascareño, G. Micela, J. Maldonado, J. I. González Hernández, and R. Rebolo. 2017. “HADES RV Programme with HARPS-N at TNG. VI. GJ 3942 b behind dominant activity signals.” *A&A* 608, A63 (December): A63. doi:[10.1051/0004-6361/201731307](https://doi.org/10.1051/0004-6361/201731307). arXiv: [1709.06851](https://arxiv.org/abs/1709.06851) [[astro-ph.SR](#)].
- Perger, M., G. Scandariato, I. Ribas, J. C. Morales, L. Affer, M. Azzaro, P. J. Amado, G. Anglada-Escudé, D. Baroch, and D. Barrado. 2019. “Gliese 49: activity evolution and detection of a super-Earth. A HADES and CARMENES collaboration.” *A&A* 624, A123 (April): A123. doi:[10.1051/0004-6361/201935192](https://doi.org/10.1051/0004-6361/201935192). arXiv: [1903.04808](https://arxiv.org/abs/1903.04808) [[astro-ph.EP](#)].

- Perruchod, S., D. Kohler, F. Bouchy, Y. Richaud, Pierre Richaud, G. Moreaux, Mourad Merzougui, et al. 2008. “The SOPHIE spectrograph: design and technical key-points for high throughput and high stability - art. no. 70140J,” vol. 7014. July. doi:[10.1117/12.787379](https://doi.org/10.1117/12.787379).
- Peterson, Pearu. 2009. “F2PY: a tool for connecting Fortran and Python programs.” *IJSE* 4:296–305.
- Petrov, P. P., V. A. Shcherbakov, S. V. Berdyugina, V. S. Shevchenko, K. N. Grankin, and S. Y. Melnikov. 1994. “Photometric and spectroscopic observations of the spotted T Tauri star V410 Tauri.” *A&ASS* 107 (October): 9–22.
- Pinhas, A., and N. Madhusudhan. 2017. “On signatures of clouds in exoplanetary transit spectra.” *MNRAS* 471 (November): 4355–4373. doi:[10.1093/mnras/stx1849](https://doi.org/10.1093/mnras/stx1849). arXiv: [1705.08893](https://arxiv.org/abs/1705.08893) [[astro-ph.EP](#)].
- Pont, F., D. K. Sing, N. P. Gibson, S. Aigrain, G. Henry, and N. Husnoo. 2013. “The prevalence of dust on the exoplanet HD 189733b from Hubble and Spitzer observations.” *MNRAS* 432, no. 4 (July): 2917–2944. doi:[10.1093/mnras/stt651](https://doi.org/10.1093/mnras/stt651). arXiv: [1210.4163](https://arxiv.org/abs/1210.4163) [[astro-ph.EP](#)].
- Pont, F., S. Zucker, and D. Queloz. 2006. “The effect of red noise on planetary transit detection.” *MNRAS* 373 (November): 231–242. doi:[10.1111/j.1365-2966.2006.11012.x](https://doi.org/10.1111/j.1365-2966.2006.11012.x). eprint: [astro-ph/0608597](https://arxiv.org/abs/astro-ph/0608597).
- Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. 1993. *Numerical Recipes in FORTRAN; The Art of Scientific Computing*. 2nd. USA: Cambridge University Press. isbn: 0521437164.
- Priestley, M.B. 1981. *Spectral Analysis and Time Series*. Probability and mathematical statistics : A series of monographs and textbooks v. 1. Academic Press. <https://books.google.es/books?id=RVTYvwEACAAJ>.
- Queloz, D., G. W. Henry, J. P. Sivan, S. L. Baliunas, J. L. Beuzit, R. A. Donahue, M. Mayor, D. Naef, C. Perrier, and S. Udry. 2001. “No planet for HD 166435.” *A&A* 379 (November): 279–287. doi:[10.1051/0004-6361:20011308](https://doi.org/10.1051/0004-6361:20011308). eprint: [astro-ph/0109491](https://arxiv.org/abs/astro-ph/0109491).
- Quirrenbach, A., P. J. Amado, J. A. Caballero, R. Mundt, A. Reiners, I. Ribas, W. Seifert, et al. 2014. “CARMENES instrument overview.” In *Ground-based and Airborne Instrumentation for Astronomy V*, vol. 9147, 91471F. procspie. July. doi:[10.1117/12.2056453](https://doi.org/10.1117/12.2056453).
- Quirrenbach, A., P. J. Amado, I. Ribas, A. Reiners, J. A. Caballero, W. Seifert, J. Aceituno, M. Azzaro, D. Baroch, and D. Barrado. 2018. “CARMENES: high-resolution spectra and precise radial velocities in the red and infrared.” In *procspie*, vol. 10702, 107020W. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. July. doi:[10.1117/12.2313689](https://doi.org/10.1117/12.2313689).

- Rabus, M., R. Alonso, J. A. Belmonte, H. J. Deeg, R. L. Gilliland, J. M. Almenara, T. M. Brown, D. Charbonneau, and G. Mandushev. 2009. “A cool starspot or a second transiting planet in the TrES-1 system?” *A&A* 494, no. 1 (January): 391–397. doi:[10.1051/0004-6361/200811110](https://doi.org/10.1051/0004-6361/200811110). arXiv: [0812.1799](https://arxiv.org/abs/0812.1799) [astro-ph].
- Rackham, Dániel Apai, and Mark S. Giampapa. 2018. “The Transit Light Source Effect: False Spectral Features and Incorrect Densities for M-dwarf Transiting Planets.” *ApJ* 853, no. 2, 122 (February): 122. doi:[10.3847/1538-4357/aaa08c](https://doi.org/10.3847/1538-4357/aaa08c). arXiv: [1711.05691](https://arxiv.org/abs/1711.05691) [astro-ph.EP].
- . 2019. “The Transit Light Source Effect. II. The Impact of Stellar Heterogeneity on Transmission Spectra of Planets Orbiting Broadly Sun-like Stars.” *AJ* 157, no. 3, 96 (March): 96. doi:[10.3847/1538-3881/aaf892](https://doi.org/10.3847/1538-3881/aaf892). arXiv: [1812.06184](https://arxiv.org/abs/1812.06184) [astro-ph.EP].
- Rackham, Néstor Espinoza, Dániel Apai, Mercedes López-Morales, Andrés Jordán, David J. Osip, Nikole K. Lewis, et al. 2017. “ACCESS I: An Optical Transmission Spectrum of GJ 1214b Reveals a Heterogeneous Stellar Photosphere.” *ApJ* 834, no. 2, 151 (January): 151. doi:[10.3847/1538-4357/aa4f6c](https://doi.org/10.3847/1538-4357/aa4f6c). arXiv: [1612.00228](https://arxiv.org/abs/1612.00228) [astro-ph.EP].
- Radick, R. R., D. Mihalas, G. W. Lockwood, D. T. Thompson, A. Warnock III, L. W. Hartmann, S. P. Worden, G. W. Henry, and J. M. Sherlin. 1983. “The photometric variability of solar-type stars. III - Results from 1981-82, including parallel observations of thirty-six Hyades stars.” *PASP* 95 (September): 621–634. doi:[10.1086/131229](https://doi.org/10.1086/131229).
- Rajpaul, V., S. Aigrain, and S. Roberts. 2016. “Ghost in the time series: no planet for Alpha Cen B.” *MNRAS* 456, no. 1 (February): L6–L10. doi:[10.1093/mnrasl/slv164](https://doi.org/10.1093/mnrasl/slv164). arXiv: [1510.05598](https://arxiv.org/abs/1510.05598) [astro-ph.EP].
- Reiners, A., J. L. Bean, K. F. Huber, S. Dreizler, A. Seifahrt, and S. Czesla. 2010. “Detecting Planets Around Very Low Mass Stars with the Radial Velocity Method.” *ApJ* 710, no. 1 (February): 432–443. doi:[10.1088/0004-637X/710/1/432](https://doi.org/10.1088/0004-637X/710/1/432). arXiv: [0909.0002](https://arxiv.org/abs/0909.0002) [astro-ph.SR].
- Reiners, A., M. Zechmeister, J. A. Caballero, I. Ribas, J. C. Morales, S. V. Jeffers, P. Schöfer, L. Tal-Or, A. Quirrenbach, and P. J. Amado. 2018. “The CARMENES search for exoplanets around M dwarfs. High-resolution optical and near-infrared spectroscopy of 324 survey stars.” *aap* 612, A49 (April): A49. doi:[10.1051/0004-6361/201732054](https://doi.org/10.1051/0004-6361/201732054). arXiv: [1711.06576](https://arxiv.org/abs/1711.06576) [astro-ph.SR].
- Reinhold, Timo, Ansgar Reiners, and Gibor Basri. 2013. “Rotation and differential rotation of activeKeplerstars.” *Astronomy and Astrophysics* 560 (November): A4. issn: 1432-0746. doi:[10.1051/0004-6361/201321970](https://doi.org/10.1051/0004-6361/201321970). <http://dx.doi.org/10.1051/0004-6361/201321970>.
- Ribas, I., M. Tuomi, A. Reiners, R. P. Butler, J. C. Morales, M. Perger, S. Dreizler, et al. 2018. “A candidate super-Earth planet orbiting near the snow line of Barnard’s

- star.” *Nature* 563 (November): 365–368. doi:[10.1038/s41586-018-0677-y](https://doi.org/10.1038/s41586-018-0677-y). arXiv: [1811.05955](https://arxiv.org/abs/1811.05955) [astro-ph.EP].
- Robertson, M. Endl, G. W. Henry, W. D. Cochran, P. J. MacQueen, and M. H. Williamson. 2015a. “Stellar Activity and its Implications for Exoplanet Detection on GJ 176.” *ApJ* 801, 79 (March): 79. doi:[10.1088/0004-637X/801/2/79](https://doi.org/10.1088/0004-637X/801/2/79). arXiv: [1501.02807](https://arxiv.org/abs/1501.02807) [astro-ph.SR].
- Robertson, Suvrath Mahadevan, Michael Endl, and Arpita Roy. 2014. “Stellar activity masquerading as planets in the habitable zone of the M dwarf Gliese 581.” *Science* 345 (6195): 440–444. doi:[10.1126/science.1253253](https://doi.org/10.1126/science.1253253). eprint: <https://www.science.org/doi/pdf/10.1126/science.1253253>. <https://www.science.org/doi/abs/10.1126/science.1253253>.
- Robertson, A. Roy, and S. Mahadevan. 2015b. “Stellar Activity Mimics a Habitable-zone Planet around Kapteyn’s Star.” *ApJ Letter* 805, L22 (June): L22. doi:[10.1088/2041-8205/805/2/L22](https://doi.org/10.1088/2041-8205/805/2/L22).
- Rosich, A., Herrero, E., Mallonn, M., Ribas, I., Morales, J. C., Perger, M., Anglada-Escudé, G., and Granzer, T. 2020. “Correcting for chromatic stellar activity effects in transits with multiband photometric monitoring: application to WASP-52.” *A&A* 641:A82. doi:[10.1051/0004-6361/202037586](https://doi.org/10.1051/0004-6361/202037586). <https://doi.org/10.1051/0004-6361/202037586>.
- Rossiter. 1924. “On the detection of an effect of rotation during eclipse in the velocity of the brighter component of beta Lyrae, and on the constancy of velocity of this system.” *ApJ* 60 (July): 15–21. doi:[10.1086/142825](https://doi.org/10.1086/142825).
- Russell, H. N. 1906. “On the light variations of asteroids and satellites.” *ApJ* 24 (July): 1–18. doi:[10.1086/141361](https://doi.org/10.1086/141361).
- Santos, N. C., J. Gomes da Silva, C. Lovis, and C. Melo. 2010. “Do stellar magnetic cycles influence the measurement of precise radial velocities?” *A&A* 511, A54 (February): A54. doi:[10.1051/0004-6361/200913433](https://doi.org/10.1051/0004-6361/200913433). arXiv: [0912.2901](https://arxiv.org/abs/0912.2901) [astro-ph.EP].
- Santos, N. C., M. Mayor, D. Naef, D. Queloz, and S. Udry. 2001. “Bisector analysis as a diagnostic of intrinsic radial-velocity variations.” *ArXiv Astrophysics e-prints* (January). eprint: [astro-ph/0101377](https://arxiv.org/abs/astro-ph/0101377).
- Scargle, J. D. 1982. “Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data.” *ApJ* 263 (December): 835–853. doi:[10.1086/160554](https://doi.org/10.1086/160554).
- Schwarzenberg-Czerny, A. 1998. “The distribution of empirical periodograms: Lomb-Scargle and PDM spectra.” *MNRAS* 301, no. 3 (December): 831–840. doi:[10.1046/j.1365-8711.1998.02086.x](https://doi.org/10.1046/j.1365-8711.1998.02086.x).

- Schweitzer, A., V. M. Passegger, C. Cifuentes, V. J. S. Béjar, M. Cortés-Contreras, J. A. Caballero, C. del Burgo, S. Czesla, M. Kürster, and D. Montes. 2019. “The CARMENES search for exoplanets around M dwarfs. Different roads to radii and masses of the target stars.” *aap* 625, A68 (May): A68. doi:[10.1051/0004-6361/201834965](https://doi.org/10.1051/0004-6361/201834965). arXiv: [1904.03231](https://arxiv.org/abs/1904.03231) [[astro-ph.SR](#)].
- Seager, S., and D. D. Sasselov. 2000. “Theoretical Transmission Spectra during Extra-solar Giant Planet Transits.” *ApJ* 537 (July): 916–921. doi:[10.1086/309088](https://doi.org/10.1086/309088). eprint: [astro-ph/9912241](https://arxiv.org/abs/astro-ph/9912241).
- Simon, Theodore, and Jr. Fekel Francis C. 1987. “The Dependence of Ultraviolet Chromospheric Emission upon Rotation among Late-Type Stars.” *ApJ* 316 (May): 434. doi:[10.1086/165212](https://doi.org/10.1086/165212).
- Sing, David K., Jonathan J. Fortney, Nikolay Nikolov, Hannah R. Wakeford, Tiffany Kataria, Thomas M. Evans, Suzanne Aigrain, et al. 2015. “A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion.” *Nature* 529, no. 7584 (December): 59–62. issn: 1476-4687. doi:[10.1038/nature16068](https://doi.org/10.1038/nature16068). <http://dx.doi.org/10.1038/nature16068>.
- Sing, David K., Jonathan J. Fortney, Nikolay Nikolov, Hannah R. Wakeford, Tiffany Kataria, Thomas M. Evans, Suzanne Aigrain, et al. 2016. “A continuum from clear to cloudy hot-Jupiter exoplanets without primordial water depletion.” *Nature* 529, no. 7584 (January): 59–62. doi:[10.1038/nature16068](https://doi.org/10.1038/nature16068). arXiv: [1512.04341](https://arxiv.org/abs/1512.04341) [[astro-ph.EP](#)].
- Snellen, I. A. G. 2004. “A new method for probing the atmospheres of transiting exoplanets.” *Monthly Notices of the Royal Astronomical Society* 353, no. 1 (September): L1–L6. issn: 0035-8711. doi:[10.1111/j.1365-2966.2004.08169.x](https://doi.org/10.1111/j.1365-2966.2004.08169.x). eprint: <https://academic.oup.com/mnras/article-pdf/353/1/L1/18662172/353-1-L1.pdf>. <https://doi.org/10.1111/j.1365-2966.2004.08169.x>.
- Snodgrass, H. B., and R. K. Ulrich. 1990. “Rotation of Doppler features in the solar photosphere.” *ApJ* 351 (March): 309–316. doi:[10.1086/168467](https://doi.org/10.1086/168467).
- Solanki, Sami K. 1993. “Smallscale Solar Magnetic Fields - an Overview.” 63, nos. 1-2 (March): 1–188. doi:[10.1007/BF00749277](https://doi.org/10.1007/BF00749277).
- Stevenson, Kevin B. 2016. “Quantifying and Predicting the Presence of Clouds in Exoplanet Atmospheres.” *ApJ Letter* 817, no. 2, L16 (February): L16. doi:[10.3847/2041-8205/817/2/L16](https://doi.org/10.3847/2041-8205/817/2/L16). arXiv: [1601.03492](https://arxiv.org/abs/1601.03492) [[astro-ph.EP](#)].
- Strassmeier. 2009. “Starspots.” *The Astronomy and Astrophysics Review* 17 (3): 251–308. issn: 1432-0754. doi:[10.1007/s00159-009-0020-6](https://doi.org/10.1007/s00159-009-0020-6). <http://dx.doi.org/10.1007/s00159-009-0020-6>.

- Strassmeier, K. G., T. Granzer, M. Weber, M. Woche, M. I. Andersen, J. Bartus, S. -M. Bauer, F. Dionies, E. Popow, and T. Fechner. 2004. “The STELLA robotic observatory.” *Astronomische Nachrichten* 325, no. 6 (October): 527–532. doi:[10.1002/asna.200410273](https://doi.org/10.1002/asna.200410273).
- Strassmeier, K. G., Thomas Granzer, Michael Weber, Manfred Woche, Emil Popow, Arto Järvinen, Janos Bartus, Svend-Marian Bauer, Frank Dionies, and Thomas Fechner. 2010. “The STELLA Robotic Observatory on Tenerife.” *Advances in Astronomy* 2010, 970306 (January): 970306. doi:[10.1155/2010/970306](https://doi.org/10.1155/2010/970306).
- Suárez Mascareño, A., R. Rebolo, J. I. González Hernández, and M. Esposito. 2015a. “Rotation periods of late-type dwarf stars from time series high-resolution spectroscopy of chromospheric indicators.” *MNRAS* 452 (September): 2745–2756. doi:[10.1093/mnras/stv1441](https://doi.org/10.1093/mnras/stv1441). arXiv: [1506.08039](https://arxiv.org/abs/1506.08039) [[astro-ph.EP](#)].
- . 2015b. “Rotation periods of late-type dwarf stars from time series high-resolution spectroscopy of chromospheric indicators.” *MNRAS* 452 (September): 2745–2756. doi:[10.1093/mnras/stv1441](https://doi.org/10.1093/mnras/stv1441). arXiv: [1506.08039](https://arxiv.org/abs/1506.08039) [[astro-ph.EP](#)].
- Suárez Mascareño, A., R. Rebolo, J. I. González Hernández, and M. Esposito. 2017. “Characterization of the radial velocity signal induced by rotation in late-type dwarfs.” *Monthly Notices of the Royal Astronomical Society* 468, no. 4 (March): 4772–4781. issn: 1365-2966. doi:[10.1093/mnras/stx771](https://doi.org/10.1093/mnras/stx771). <http://dx.doi.org/10.1093/mnras/stx771>.
- Süveges, M., L. P. Guy, L. Eyer, J. Cuypers, B. Holl, I. Lecoœur-Taïbi, N. Mowlavi, et al. 2015. “A comparative study of four significance measures for periodicity detection in astronomical surveys.” *MNRAS* 450 (June): 2052–2066. doi:[10.1093/mnras/stv719](https://doi.org/10.1093/mnras/stv719). arXiv: [1504.00782](https://arxiv.org/abs/1504.00782) [[astro-ph.IM](#)].
- Süveges, Maria. 2012. “Extreme-value modelling for the significance assessment of periodogram peaks.” *Monthly Notices of the Royal Astronomical Society* 440 (December). doi:[10.1093/mnras/stu372](https://doi.org/10.1093/mnras/stu372).
- Swain, Mark R., Gautam Vasisht, and Giovanna Tinetti. 2008. “The presence of methane in the atmosphere of an extrasolar planet.” *Nature* 452, no. 7185 (March): 329–331. doi:[10.1038/nature06823](https://doi.org/10.1038/nature06823).
- Tarantola, Albert. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial / Applied Mathematics. doi:[10.1137/1.9780898717921](https://doi.org/10.1137/1.9780898717921). eprint: <https://epubs.siam.org/doi/pdf/10.1137/1.9780898717921>. <https://epubs.siam.org/doi/abs/10.1137/1.9780898717921>.
- Teegarden, B. J., S. H. Pravdo, M. Hicks, K. Lawrence, S. B. Shaklan, K. Covey, O. Fraser, S. L. Hawley, T. McGlynn, and I. N. Reid. 2003. “Discovery of a New Nearby Star.” *ApJ Letter* 589 (May): L51–L53. doi:[10.1086/375803](https://doi.org/10.1086/375803). eprint: [astro-ph/0302206](https://arxiv.org/abs/astro-ph/0302206).

- Theodoridis, Sergios. 2020. "Chapter 3 - Learning in Parametric Modeling: Basic Concepts and Directions." In *Machine Learning (Second Edition)*, Second Edition, edited by Sergios Theodoridis, 67–120. Academic Press. isbn: 978-0-12-818803-3. doi:<https://doi.org/10.1016/B978-0-12-818803-3.00012-X>. <https://www.sciencedirect.com/science/article/pii/B978012818803300012X>.
- Tinetti, Giovanna, Pierre Drossart, Paul Eccleston, Paul Hartogh, Astrid Heske, J  r  my Leconte, Giusi Micela, et al. 2018. "A chemical survey of exoplanets with ARIEL." *Experimental Astronomy* 46, no. 1 (November): 135–209. doi:[10.1007/s10686-018-9598-X](https://doi.org/10.1007/s10686-018-9598-X).
- Tinetti, Giovanna, Alfred Vidal-Madjar, Mao-Chang Liang, Jean-Philippe Beaulieu, Yuk Yung, Sean Carey, Robert J. Barber, et al. 2007. "Water vapour in the atmosphere of a transiting extrasolar planet." *Nature* 448, no. 7150 (July): 169–171. doi:[10.1038/nature06002](https://doi.org/10.1038/nature06002).
- Toledo-Padr  n, B., J. I. Gonz  lez Hern  ndez, C. Rodr  guez-L  pez, A. Su  rez Mascare  o, R. Rebolo, R. P. Butler, I. Ribas, et al. 2019. "Stellar activity analysis of Barnard’s Star: very slow rotation and evidence for long-term activity cycle." *MNRAS* 488, no. 4 (October): 5145–5161. doi:[10.1093/mnras/stz1975](https://doi.org/10.1093/mnras/stz1975). arXiv: [1812.06712](https://arxiv.org/abs/1812.06712) [astro-ph.SR].
- Triaud, Amaury H. M. J. 2017. "The Rossiter–McLaughlin Effect in Exoplanet Research." In *Handbook of Exoplanets*, edited by Hans J. Deeg and Juan Antonio Belmonte, 1–27. Cham: Springer International Publishing. isbn: 978-3-319-30648-3. doi:[10.1007/978-3-319-30648-3_2-1](https://doi.org/10.1007/978-3-319-30648-3_2-1). https://doi.org/10.1007/978-3-319-30648-3_2-1.
- Triaud, D. Queloz, F. Bouchy, C. Moutou, A. Collier Cameron, A. Claret, P. Barge, et al. 2009. "The Rossiter-McLaughlin effect of CoRoT-3b and HD 189733b." *A&A* 506, no. 1 (October): 377–384. doi:[10.1051/0004-6361/200911897](https://doi.org/10.1051/0004-6361/200911897). arXiv: [0907.2956](https://arxiv.org/abs/0907.2956) [astro-ph.EP].
- Tsiaras, Angelos, Ingo P. Waldmann, Giovanna Tinetti, Jonathan Tennyson, and Sergey N. Yurchenko. 2019. "Water vapour in the atmosphere of the habitable-zone eight-Earth-mass planet K2-18 b." *Nature Astronomy* 3 (September): 1086–1091. doi:[10.1038/s41550-019-0878-9](https://doi.org/10.1038/s41550-019-0878-9). arXiv: [1909.05218](https://arxiv.org/abs/1909.05218) [astro-ph.EP].
- Tuomi, M., and G. Anglada-Escud  . 2013a. "Up to four planets around the M dwarf GJ 163. Sensitivity of Bayesian planet detection criteria to prior choice." *A&A* 556, A111 (August): A111. doi:[10.1051/0004-6361/201321174](https://doi.org/10.1051/0004-6361/201321174). arXiv: [1306.1717](https://arxiv.org/abs/1306.1717) [astro-ph.EP].
- Tuomi, M., G. Anglada-Escud  , E. Gerlach, H. R. A. Jones, A. Reiners, E. J. Rivera, S. S. Vogt, and R. P. Butler. 2013b. "Habitable-zone super-Earth candidate in a six-planet system around the K2.5V star HD 40307." *A&A* 549, A48 (January): A48. doi:[10.1051/0004-6361/201220268](https://doi.org/10.1051/0004-6361/201220268). arXiv: [1211.1617](https://arxiv.org/abs/1211.1617) [astro-ph.EP].
- van de Kamp, P. 1982. "The planetary system of Barnard’s star." *Vistas in Astronomy* 26:141–157. doi:[10.1016/0083-6656\(82\)90004-6](https://doi.org/10.1016/0083-6656(82)90004-6).

- van Leeuwen, F. 2007. “Validation of the new Hipparcos reduction.” *A&A* 474, no. 2 (November): 653–664. doi:[10.1051/0004-6361:20078357](https://doi.org/10.1051/0004-6361:20078357). arXiv: [0708.1752](https://arxiv.org/abs/0708.1752) [[astro-ph](#)].
- VanderPlas, Jacob T. 2018. “Understanding the Lomb-Scargle Periodogram.” *apjs* 236, no. 1, 16 (May): 16. doi:[10.3847/1538-4365/aab766](https://doi.org/10.3847/1538-4365/aab766). arXiv: [1703.09824](https://arxiv.org/abs/1703.09824) [[astro-ph.IM](#)].
- VanderPlas, Andrew J. Connolly, Zeljko Ivezic, and Alex Gray. 2012. “Introduction to astroML: Machine learning for astrophysics.” *2012 Conference on Intelligent Data Understanding* (October). doi:[10.1109/cidu.2012.6382200](https://doi.org/10.1109/cidu.2012.6382200). <http://dx.doi.org/10.1109/CIDU.2012.6382200>.
- Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, et al. 2020. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.” *Nature Methods* 17:261–272. doi:[10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).
- Vitense, E. 1953. “Die Wasserstoffkonvektionszone der Sonne. Mit 11 Textabbildungen.” *zap* 32 (January): 135.
- Vogt, Steven S., G. Donald Penrod, and Artie P. Hatzes. 1987. “Doppler Images of Rotating Stars Using Maximum Entropy Image Reconstruction.” *ApJ* 321 (October): 496. doi:[10.1086/165647](https://doi.org/10.1086/165647).
- Wakeford, H. R., D. K. Sing, D. Deming, N. K. Lewis, J. Goyal, T. J. Wilson, J. Barstow, et al. 2017. “The Complete Transmission Spectrum of WASP-39b with a Precise Water Constraint.” *The Astronomical Journal* 155, no. 1 (December): 29. doi:[10.3847/1538-3881/aa9e4e](https://doi.org/10.3847/1538-3881/aa9e4e). <https://doi.org/10.3847/1538-3881/aa9e4e>.
- Wittenmyer, R. A., M. Tuomi, R. P. Butler, H. R. A. Jones, G. Anglada-Escudé, J. Horner, C. G. Tinney, et al. 2014. “GJ 832c: A Super-Earth in the Habitable Zone.” *ApJ* 791, 114 (August): 114. doi:[10.1088/0004-637X/791/2/114](https://doi.org/10.1088/0004-637X/791/2/114). arXiv: [1406.5587](https://arxiv.org/abs/1406.5587) [[astro-ph.EP](#)].
- Zechmeister, M., S. Dreizler, I. Ribas, A. Reiners, J. A. Caballero, F. F. Bauer, V. J. S. Béjar, L. González-Cuesta, E. Herrero, and S. Lalitha. 2019a. “The CARMENES search for exoplanets around M dwarfs. Two temperate Earth-mass planet candidates around Teegarden’s Star.” *arXiv e-prints*, arXiv:1906.07196 (June): arXiv:1906.07196. arXiv: [1906.07196](https://arxiv.org/abs/1906.07196) [[astro-ph.EP](#)].
- . 2019b. “The CARMENES search for exoplanets around M dwarfs. Two temperate Earth-mass planet candidates around Teegarden’s Star.” *arXiv e-prints*, arXiv:1906.07196 (June): arXiv:1906.07196. arXiv: [1906.07196](https://arxiv.org/abs/1906.07196) [[astro-ph.EP](#)].
- Zechmeister, M., and M. Kürster. 2009. “The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms.” *A&A* 496 (March): 577–584. doi:[10.1051/0004-6361:200811296](https://doi.org/10.1051/0004-6361:200811296). arXiv: [0901.2573](https://arxiv.org/abs/0901.2573) [[astro-ph.IM](#)].

Zechmeister, M., A. Reiners, P. J. Amado, M. Azzaro, F. F. Bauer, V. J. S. Béjar, J. A. Caballero, et al. 2018. “Spectrum radial velocity analyser (SERVAL). High-precision radial velocities and two alternative spectral indicators.” *A&A* 609, A12 (January): A12. doi:[10.1051/0004-6361/201731483](https://doi.org/10.1051/0004-6361/201731483). arXiv: [1710.10114](https://arxiv.org/abs/1710.10114) [[astro-ph.IM](#)].

