# Gaussian process tools for modelling stellar signals and studying exoplanets


Vinesh Maguire Rajpaul

Merton College

University of Oxford

A thesis submitted for the degree of

*Doctor of Philosophy*

Trinity 2017

To my parents and grandparents,
who have been a lifelong source
of love, support, and inspiration.

# Abstract

The discovery of exoplanets represents one of the greatest scientific revolutions in history, and exoplanetary science has rapidly become uniquely positioned to address profound questions about the origins of life, and about humanity's place (and future) in the cosmos. Since the discovery of the first exoplanet over two decades ago, the radial velocity (RV) method has been one of the most productive techniques for discovering new planets. It has also become indispensable for characterising exoplanets detected via other techniques, notably transit photometry. Unfortunately, signals intrinsic to stars themselves – especially magnetic activity signals – can induce RV variations that can drown out or even mimic planetary signals. Modelling and thus mitigating these signals is notoriously difficult, which represents a major obstacle to using next-generation instruments to detect lower mass planets, planets with longer periods, and planets around more magnetically-active stars. Enter Gaussian processes (GPs), which have a number of features that make them very well suited to the joint modelling of stochastic activity processes and dynamical (e.g. planetary) signals. In this thesis, I leverage GPs to enable the study of smaller planets around a wider variety of stars than has previously been possible. In particular, I develop a principled and sophisticated Bayesian framework, based on GPs, for modelling RV time series jointly with ancillary activity-sensitive proxies, thus allowing activity signals to be constrained and disentangled from genuine planetary signals. I show that my framework succeeds even in cases where existing techniques would fail to detect planets, e.g. the case of a weak planetary signal with period identical to its host star's rotation period. In a first application of the framework, I demonstrate that Alpha Centauri Bb – until 2016, thought to be the closest exoplanet to Earth, and also the lowest minimum-mass exoplanet around a Sun-like star – was, in fact, an astrophysical false positive. Next, I use the framework to re-characterise the well-studied Kepler-10 system, thereby resolving a mystery surrounding the mass of planet Kepler-10c. I also use the framework to help discover or characterise various exoplanets. Finally, the activity modelling framework aside, I also present in outline form a few promising applications of GPs in the context of modelling stellar signals and studying exoplanets, *viz.* GPs for (i) enhanced characterisation of stellar rotation; (ii) generating realistic synthetic observations, and modelling in a systematic way the effects of an observing window function; and (iii) ultra-precise extraction of RV shifts directly from observed spectra, without requiring template cross-correlation.

# Acknowledgements

Being a doctoral research student at the University of Oxford has been a profoundly rewarding and transformative experience. I've met fascinating, talented people from all over the world; I've made the most wonderful friends; and, of course, I've learnt a great deal (or so I hope).

I am enormously indebted to the Rhodes Trust for the financial support that allowed me to come to Oxford in the first place, and to read for a DPhil in a city that was as expensive as it was inspiring. I hope that this thesis befits the trust they placed in me, and that in the years to come, I shall always esteem their charge to 'fight the world's fight'. I also wish to record my gratitude to the National Research Foundation (NRF) of South Africa, Merton College, and the Royal Astronomical Society, for helping to fund various parts of my research, including significant travel abroad.

On the academic front, my thanks are due first and foremost to Prof. Suzanne Aigrain. A glance at the acknowledgements noted in the theses of her former students reveals only the loftiest of praise for Suzanne, and I am glad to be able to add my voice to that hearty chorus. Suzanne's deep knowledge and fierce committment to academic rigour make her an inspiring scientist. Yet she is also a truly humble and caring supervisor. Thank you, Suzanne, for allowing me to work with you on such an interesting research project, for teaching me a great deal, and for rescuing me whenever I felt I was floundering. Thank you also for giving me the freedom to procrastinate, travel, and err as much as I did. I could not have hoped for a better supervisor, and I very much look forward to learning from and collaborating with you in future.

I also wish to thank my co-supervisor, Prof. Stephen Roberts, without whose formidable technical insights and expertise some of my research might never have got off the ground. I hope that we, too, will be able to keep working together and bringing powerful Bayesian tools to bear on interesting exoplanetary problems.

Finally, while still on the academic front, I am grateful to Prof. Didier Queloz and Prof. Pat Roche for kindly agreeing to examine my thesis, for reading it all, and for giving me a good deal of constructive feedback during my viva.

On an institutional level, I am very glad to have been hosted and supported by the Department of Physics at the University of Oxford. Merton College and

Rhodes House were both also integral to my time in Oxford, and provided vibrant, diverse, and even eclectic academic and social communities that truly made Oxford feel like a second home.

Further afield, I enjoyed and benefitted greatly from research visits to a number of other institutions, including the School of Physics and Astronomy at the University of St Andrews; the Department of Astronomy at Yale University; the Harvard–Smithsonian Center for Astrophysics; the Cavendish Laboratory at the University of Cambridge; and the Centre for Star and Planet Formation at the University of Copenhagen. I am most obliged to my collaborators at these institutions for their kindness and hospitality.

On a more personal note, I am deeply grateful to my family for their unconditional love and support throughout my life. In particular, I wish to thank my parents and grandparents for being excellent role models, and for the many sacrifices they made to ensure that I would grow up in a stable, nurturing environment, and that I could enjoy a relatively carefree childhood. I am also indebted to them for instilling in me the belief that 'education is the most powerful weapon which you can use to change the world' (as Nelson Mandela put it) – a belief I hold more strongly today than ever before.

And to my dear friends, in Oxford and further afield, old and new: thank you for all the happy memories, and for your precious companionship and unstinting support. Your friendship is my wealth.

Finally, to Victoria, who inspires me, challenges me, and suffuses my life with more meaning and joy and love than I ever dreamed possible. Of all the worlds around all the stars in the innumerable galaxies scattered across the Universe, I could not possibly be gladder that I was born in this world, and that I am alive at the same time as you.

— 31 July 2017

# Declaration

The work in this thesis is based on research I carried out at the University of Oxford between Michaelmas Term 2013 and Trinity Term 2017. This research led to the following peer-reviewed publications (of which I was the principal author).

1. Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., and Roberts, S. (2015). A Gaussian process framework for modelling stellar activity signals in radial velocity data. *MNRAS*, 452(3):2269–2291
2. Rajpaul, V., Aigrain, S., and Roberts, S. (2016). Ghost in the time series: no planet for Alpha Cen B. *MNRAS*, 456:L6–L10
3. Rajpaul, V., Buchhave, L. A., and Aigrain, S. (2017). Pinning down the mass of Kepler-10c: the importance of sampling and model comparison. *MNRAS*, 471:L125–L130

Chapters 3, 4, and 5 of this thesis are based heavily on papers 1, 2 and 3, respectively, as enumerated above. I carried out all of the analyses contained in these papers, and wrote almost all (more than 95%) of the text. My coauthors advised on technical aspects of the work, recommended relevant literature, and provided editorial feedback on the papers.

My doctoral research also led to contributions to the following peer-reviewed publications.

4. Mortier, A., Faria, J. P., Santos, N. C., Rajpaul, V., et al. (2016). The HARPS search for southern extra-solar planets. XXXIX. HD 175607, the most metal-poor G dwarf with an orbiting sub-Neptune. *A&A*, 585:A135
5. Dumusque, X., Borsa, F., Damasso, M., Díaz, R. F., Gregory, P. C., Hara, N. C., Hatzes, A., Rajpaul, V., et al. (2017). Radial-velocity fitting challenge. II. First results of the analysis of the data set. *A&A*, 598:A133
6. Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., and Rajpaul, V. (2018). Inferring probabilistic stellar rotation periods using gaussian processes. *MNRAS*, 474(2):2094–2108

The first section of Chapter 6 of this thesis is based on extracts from paper 4, above. The focus is on my own contributions to the paper, although a few passages written by co-authors are included for context. The remainder of Chapter 6 is my own work, and has not yet been published, although some of this work will form the basis for a paper I am currently preparing.

The first section of Chapter 7 of this thesis is based on extracts from paper 6, above. Again, the focus is on my own contributions to the paper, although passages written by co-authors are included for context, supplemented by some of my own results that have not yet been published. The remainder of Chapter 7 is my own work, and has not yet been published.

# Contents

# List of Abbreviations

**ACF** . . . . . Autocorrelation function.

**AWGN** . . . . Additive white Gaussian noise.

**BIS** . . . . . Bisector inverse slope.

**BJD** . . . . . Barycentric Julian Date

**DRS** . . . . . (HARPS/HARPS-N) Data Reduction Software.

**FAP** . . . . . False alarm probability.

**FWHM** . . . . Full width at half maximum.

**GP** . . . . . Gaussian process.

**HARPS** . . . . High Accuracy Radial Velocity Planet Searcher.

**HIRES** . . . . High Resolution Echelle Spectrometer.

**HJD** . . . . . Heliocentric Julian Date

**K-...** . . . . . Kepler-..., e.g. K-10b = Kepler-10b

**(G)LS** . . . . . (Generalised) Lomb-Scargle.

**MAP** . . . . . Maximum *a posteriori*.

**QP** . . . . . Quasi-periodic.

**RMS** . . . . . Root mean square.

**RV** . . . . . Radial velocity.

**SE** . . . . . Squared exponential.

**SNR** . . . . . Signal-to-noise ratio.

# List of Figures

# List of Tables

*A time will come when men will stretch out their eyes. They should see planets like our Earth.*

— Sir Christopher Wren, in his inaugural lecture as Professor of Astronomy at Gresham College (1657).

# 1 Introduction

## Contents



## 1.1 Context and motivation

In September 1977, NASA launched *Voyager 1*, a robotic spacecraft on a mission to study the Outer Solar System through flybys of Jupiter, Saturn, and Saturn's largest moon, Titan. In February 1990, having crossed the orbits of Neptune and Pluto, and a decade after the successful completion of its primary mission, *Voyager 1* was rotated to look back at Earth, and instructed to capture a final series of images before its cameras were permanently powered down. The 60 frames thus acquired formed a so-called 'Family Portrait' of the Solar System, and included Venus, Earth, Jupiter, Saturn, Uranus, Neptune, and the Sun (itself just a single pixel!), all set against a vast expanse of space.[1]

[1] Mercury was too close to the Sun to be seen. Mars was obscured by scattered sunlight, and Pluto was too faint to be imaged.

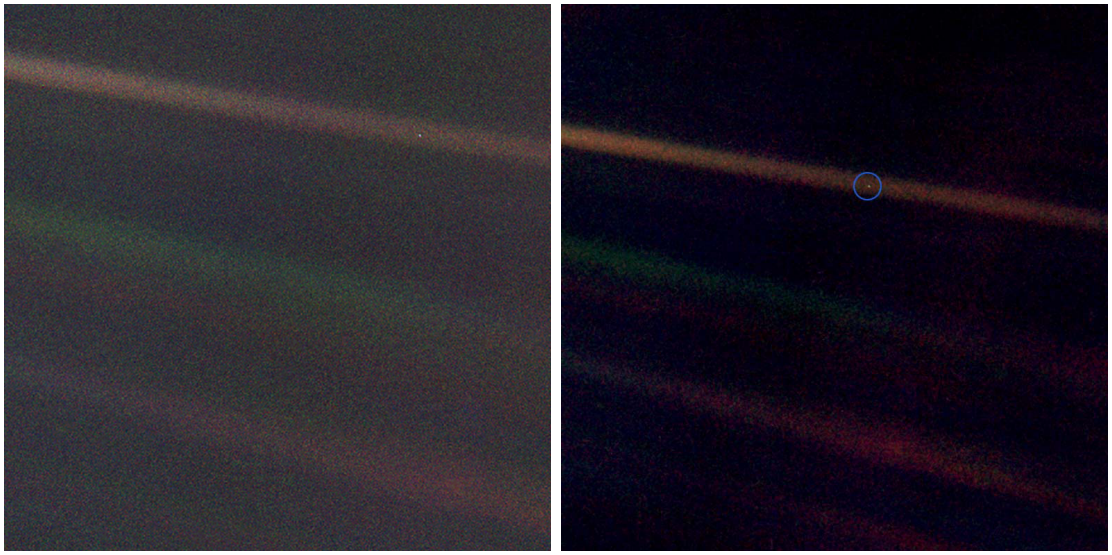

**Figure 1.1:** 'Pale Blue Dot': a photograph taken in 1990 by the *Voyager 1* space probe. The left panel shows the original photograph; the right panel is the same photograph, but with contrast enhanced, and zoomed in to the upper right quadrant. The blue circle encloses, among other pixels, a relatively bright speck of light, 0.12 pixels across: planet Earth, as seen from a distance of 6 billion km. (Public-domain NASA image.)

.

One of the frames comprising the Family Portrait features a tiny, blueish-white speck (just 0.12 px across, in an image comprising 640000 px), suspended precariously among giant bands of sunlight scattered by the optics of *Voyager 1*'s camera; see Fig. 1.1. This blueish-white speck – famously referred to as the 'Pale Blue Dot – is Earth, seen from a distance of approximately 6 billion km.

In a public lecture at Cornell University in 1994, the late Carl Sagan (at whose request the Family Portrait series was captured in the first place) offered eloquent reflections on the significance of this awe-inspiring image:

> 'We succeeded in taking that picture, and, if you look at it, you see a dot. That's here. That's home. That's us. On it, everyone you ever heard of, every human being who ever lived, lived out their lives. The aggregate of all our joys and sufferings, thousands of confident religions, ideologies and economic doctrines, every hunter and forager, every hero and coward, every creator and destroyer of civilisations, every king and peasant, every young couple in love, every hopeful child, every mother and father, every inventor and explorer, every teacher of morals, every corrupt politician, every superstar, every supreme leader, every saint and sinner in the history of our species, lived there – on a mote of dust, suspended in a sunbeam.

> '[...] Our posturings, our imagined self-importance, the delusion that we have some privileged position in the Universe, are challenged by this point of pale light. Our planet is a lonely speck in the great enveloping cosmic dark.
>
> '[...] To my mind, there is perhaps no better demonstration of the folly of human conceits than this distant image of our tiny world. To me, it underscores our responsibility to deal more kindly and compassionately with one another and to preserve and cherish that pale blue dot, the only home we've ever known.'

With these words, Sagan underscored the stark and poignant contrast between the gravity we attach to human life and endeavour, and the objective tininess and fragility of our planet (and by extension, humanity). The 'Pale Blue Dot' photograph has had a deep and lasting cultural impact, and it continues to be celebrated nearly three decades after it was captured. Yet if this humbling vision of our planet as a tiny dot set against the vastness of space did anything to shift humanity's view of Earth's place in the Universe, an utterly profound rethinking was to be precipitated within just a few years.

At the time that *Voyager 1* captured the 'Pale Blue Dot' image, humans didn't know whether there *were* any planets orbiting stars other than the Sun (i.e. exoplanets). Despite our millenia-old quest to understand our place in the cosmos, the only planetary system we'd ever managed to observe or study was our own Solar System. For some perspective, even until 2015, we had little idea what the surface of dwarf planet Pluto looked like (see Fig. 1.2), and Pluto is five thousand times times closer to Earth than Proxima Centauri, the closest star to the Sun.

Then, in 1992, the first confirmed detection of a planet outside the Solar System was made, with the discovery of two low-mass bodies around PSR B1257+12, a pulsar some 2300 ly from the Sun (Wolszczan and Frail, 1992). The first confirmation of an exoplanet orbiting a *main-sequence* star followed in 1995, when a giant planet was found on a 4 d orbit around the Sun-like star 51 Pegasi, just 50.9 ly from Earth (Mayor and Queloz, 1995). And thus the exoplanet revolution began.

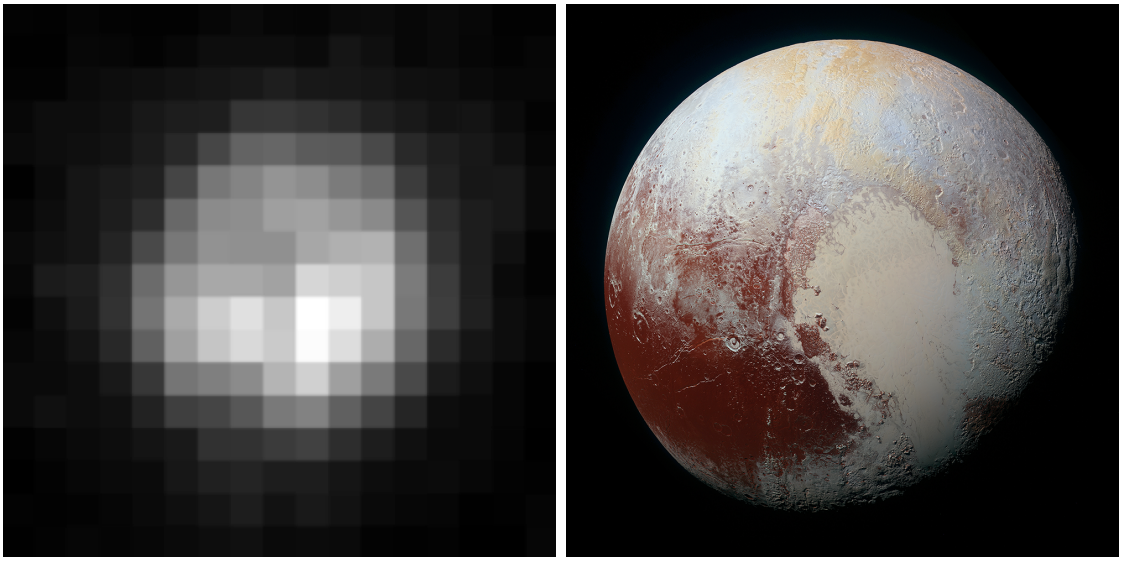

**Figure 1.2:** The highest resolution images of Pluto circa early 2015 were unable to resolve features smaller than several hundred km across, as in the Hubble Space Telescope image on the left; yet even Pluto is thousands of times closer to Earth than the nearest possible exoplanet! Knowledge of Pluto's surface geology and composition came only in July 2015, with the New Horizons spacecraft's flyby. The colour-composite image on the right was captured about 12,500 km from Pluto, and in full 64-mp resolution resolves details on scales of $\sim 1$ km. (Public-domain NASA images.)

.

Fast forward to 2017, and quite remarkably, we now know of more than 3500 exoplanets orbiting more than 2600 different stars.[2] Moreover, exoplanets are no longer simply being discovered: we're even characterising their atmospheres (Seager and Deming, 2010), starting to infer spatial inhomogeneities on their surfaces (Demory et al., 2016; Cowan and Fujii, 2017), hunting for exomoons (Kipping et al., 2015), and more. Suffice it to say that this breathtaking progress has been predicated on equally remarkable advances in scientific instrumentation, astrophysical knowledge, computing power, and tools for statistical inference.

To date, the two most productive techniques for discovering and characterising exoplanets have been transit photometry and Doppler spectroscopy, also known as the radial velocity (RV) method. The former method is based on very regular, high-precision monitoring of the brightness of distant stars; an orbiting exoplanet

[2]Based on counts as of 1 June 2017 from the NASA Exoplanet Archive, available online at `exoplanetarchive.ipac.caltech.edu`.

periodically eclipsing a star will cause minute, periodic dips in the brightness of that star. The latter method is based on observing tiny but periodic Doppler shifts in the light emitted by distant stars; the Doppler shifts are a consequence of the gravitational tug of one or more massive bodies (e.g. exoplanets, other stars) orbiting an observed star.

For many years, the RV method was responsible for far more exoplanet detections than any other technique. More recently, however, transit photometry has taken the lead, thanks to the spectacular success of NASA's *Kepler* satellite. Nevertheless, the two methods are strongly complementary: in particular, the RV method is generally needed to constrain the masses (and by extension, likely compositions, scale heights of potential atmospheres, etc.) of planet candidates detected via transit photometry.

Suffice it to say that humanity's ideas about our place in the Universe would be profoundly and irrevocably changed if we could confirm that our galaxy harbours a population of planets just like the Earth around stars similar to the Sun — to say nothing of the seismic shifts that would be triggered by discovering signatures of extra-terrestrial life in the Universe. Given the limitations of the techniques and instruments used to discover exoplanets, however, the overwhelming majority of known exoplanets are much larger than the Earth, and orbit very close to their host stars. Thus they are fundamentally different to the planets of our Solar System, and true Earth-analogue planets remain elusive.

Modern telescopes, photometers, and RV spectrographs do make it possible, in principle, to detect even the exquisitely tiny signals of Earth-like exoplanets. Unfortunately, however, stars themselves present us with enormous obstacles to detecting and studying these exoplanets.

Far from being static objects waiting patiently to be observed, typical stars are moving relative to the Earth, oscillating (e.g. on account of turbulent convection in their outer layers), covered in complex and continually-shifting patterns of convective cells, and violently ejecting plasma into space. All stars rotate; the faster rotators have stronger magnetic fields, which in turn lead to so-called 'stellar activity' phenomena (i.e. phenomena related to the presence of magnetic fields)

such as dark spots and bright faculae, which seem to appear stochastically, then grow, persist for some time, and decay again. Long-term, cyclical variations in magnetic activity levels complicate the picture yet further.

The upshot is that signals intrinsic to stars themselves can not only drown out but – especially in the case of activity signals – even mimic the signatures of genuine exoplanets. Thus it is imperative that we learn to understand and model all of these stellar signals as well as possible if we are to detect and study the smallest exoplanets, including potential populations of Earth-like exoplanets.

This, then, establishes the broad context for this thesis — the primary objective of which is the development and application of improved techniques for modelling and understanding the signals of 'active' stars (meaning, to a lesser or greater extent, almost all stars), specifically to enhance our ability to detect and characterise exoplanets around these stars.

## 1.2 Contribution

Much of the work in this thesis draws heavily on a type of stochastic process model known as a *Gaussian process*, or GP; an in-depth introduction to GPs is provided in the next chapter of this thesis. Importantly, GPs turn out to provide a very convenient and powerful framework for placing probability distributions on *functions*, rather than just on (say) the parameters of a given function.

Thus, if we're able to formulate physically motivated prior distributions over functions, then we can infer 'optimal' functions or models from real data. This facilitates a far more principled, data-driven approach to modelling and model selection than simply assuming a parametric model that might be grossly inadequate or even somewhat arbitrary, and then optimising the parameters of that *ad hoc* model. Unfortunately, that's exactly the approach to modelling that most exoplanet hunters have, until recently at least, had to adopt when confronted with stellar activity signals in their data. In many cases, these *ad hoc* models have been as crude as treating stellar activity signals as additive white (uncorrelated) noise.

Given that stellar signals have been known to be mistaken for planetary signals, the assumption of e.g. a white noise model is manifestly unfounded.

The most substantial scientific contribution contained in this thesis is a novel framework I developed, based on an advanced application of GPs, for modelling stellar activity signals jointly with exoplanetary signals in RV and other time series data. My framework offers a number of advantages over more traditional modelling approaches, including (1) eschewing *ad hoc* parametric models, and instead allowing marginalisation over an infinitely-large family of competing activity models; (2) allowing prior physical knowledge of a given star's activity properties to be incorporated into the modelling; (3) allowing multiple different types of observational data to be modelled jointly, e.g. RV, photometry, and activity-sensitive spectral line profile diagnostics; and, ultimately, (4) often allowing exoplanet signals to be unambiguously disentangled from stellar ones, where existing modelling approaches would've failed. In short, my framework represents a principled and effective approach to detecting smaller planets around a wider variety of stars than was previously possible.

The thesis also presents a number of significant applications, in each case focused primarily on RV data, of the aforementioned GP framework.

In a first application, I use the framework (along with various other analytical tools) to demonstrate that Alpha Centauri Bb – until early 2016, thought to be the closest exoplanet to Earth, and also the lowest minimum-mass exoplanet around a Sun-like star – is, in fact, almost certainly an astrophysical false positive. This result has been of significant interest to the broader exoplanet community, and has been featured widely in the popular media.

In a second application, I use the framework to re-characterise the Kepler-10 system, heretofore known to host at least two significant exoplanets: Kepler-10b, the first confirmed rocky exoplanet, and Kepler-10c, which in turn has proven something of an enigma. Initial characterisation of the latter planet suggested a remarkably large mass, given its radius, and appeared to point to a new class of denser planets with longer orbital periods. However, subsequent observations

and analysis could not confirm Kepler-10c's high mass; moreover, masses derived using RVs from two different RV spectrographs were strikingly incompatible at a $3\sigma$ level. Through careful application of my GP framework, I present a plausible resolution of the Kepler-10c mass discrepancy, fully reconciling the RVs from different instruments, and incidentally, strengthening existing evidence for a third planet in the Kepler-10 system.

In a third and final application, I use the framework to help detect and characterise various exoplanets, *viz.* HD 175607b, Kepler-37d, and Kepler-538b. HD 175607b is a non-transiting planet whose orbital period appears identical to the rotation period of its host star; using the GP framework allows the stellar signal to be identified, leading to a secure detection of the planet. At the time of this analysis, HD 1175607 held the distinction of being the most metal-poor Sun-like star known to host a planet, and the discovery could have implications for planet formation theories. Kepler-37d and Kepler-538b, on the other hand, are both transiting super-Earths; I employ my GP framework to facilitate the characterisation of these planets in the face of apparent activity signals from their host stars. Both stars thus characterised contributed to the relatively small number of long-period transiting planets with accurately constrained masses.[3]

A theme that features strongly (albeit unexpectedly) in the first and second applications outlined above is the relevance of *sampling*, i.e. the conversion of a continuous signal into a discrete one through a finite number of observations. In both the $\alpha$ Cen B and Kepler-10 cases, I conclude that principled modelling of stellar signals is by itself not enough to permit meaningful (let alone accurate) planet detection and characterisation; a thorough understanding of the effects of a discrete observing window (especially in the case of multiple interfering signals, e.g. stellar plus planetary signals) is often also essential, as is rigorous model comparison. The tests I suggest and the recommendations I make on the basis of these findings have

[3]Over the course of my doctoral studies, I also found various applications for the framework I deemed not interesting or significant enough for inclusion in this thesis, such as straightforward verifications of known planet parameters, as e.g. seemed to be the case when I applied the framework to Proxima Centauri RVs (Anglada-Escudé et al., 2016).

relevance to the analysis of existing time series data – not necessarily exoplanet-related – as well as to the design of future observing programmes.

Finally, stellar activity and sampling considerations aside, the thesis outlines a few other promising applications of GPs in the context of modelling stellar signals and studying exoplanets: (1) using GPs to study stellar rotation; (2) using GPs to generate realistic synthetic observations, e.g. to aid in devising optimal observing strategies; and (3) using GPs to extract RVs directly from observed stellar spectra, without requiring cross-correlation with a template spectrum. Given length restrictions, and that all of these applications remain 'works in progress,' each application is sketched only very briefly. I expect that when developed to maturity, the third application, in particular, will have the potential to make a significant impact on the exoplanet field.

## 1.3 Thesis structure

The next chapter, Chapter 2, presents some background material relevant to all that follows: an overview of exoplanetary science in general, and the radial velocity technique in particular; a discussion of stellar signals, and how they confound attempts to detect and characterise exoplanets; and a fairly detailed, tutorial-style introduction to GPs.

Chapter 3 forms the core of this thesis, in which I develop and present a new formalism, based on GPs, for the joint modelling of stellar activity and exoplanet signals. This chapter is supplemented by two appendices appearing at the very end of the thesis. Appendix A provides some additional mathematical details underlying the GP framework, while Appendix B provides answers to some frequently asked questions regarding GPs in general, and more specifically the framework in Chapter 3.

Chapters 4, 5, and 6 are 'application' chapters, with each chapter featuring the framework from Chapter 3 (supplemented by various statistical tools and analyses) applied in a different manner. Chapter 4 focuses on the $\alpha$ Cen B system, and demonstrates that planet candidate $\alpha$ Cen Bb is almost certainly

a false positive; Chapter 5 re-characterises the Kepler-10 system, and resolves the mass discrepancy involving Kepler-10c; and Chapter 6 sees various planets characterised for the first time.

Chapter 7 outlines a few other promising applications of GPs in the context of modelling stellar signals and studying exoplanets, and finally, Chapter 8 concludes.

Chapters 3 to 6 and Appendix A are based on material drawn from a number of peer-reviewed papers (either published or, in one case, currently *in prep.*); the details of the corresponding paper(s) are given at the start of each chapter. The remaining material in this thesis is unpublished.

## 1.4 A personal note

Some of my earliest memories are of being perched as a toddler on my grandparents' back porch in South Africa, and staring, utterly transfixed, at darkening twilight skies. I stared and stared, and marvelled as glittering pinpricks of light started to appear in the firmament – a 'heaventree of stars hung with humid nightblue fruit,' as Joyce would have it. My parents and grandparents must have thought me a peculiar child, but they humoured me, and helped me identify the Moon, planets (which did not twinkle), and distant stars (which did). I had a vague sense that I was witnessing something sublime, and as night fell, they usually had to drag me inside.

A couple of years later, I remember being mesmerised by an issue of Time magazine I chanced upon while waiting to see a dentist. The magazine contained stunning photographs – serene planets, glorious nebulae, baroque spiral galaxies – beamed down from the then recently-repaired Hubble Space Telescope. I hated visiting the dentist, but on that occasion I was adamant that I had to return as soon as possible, so that I could glimpse those incredible photographs once more. Seeing my enthusiasm, and perhaps eager to get rid of me with minimum fuss, the dentist suggested that I keep the magazine.

The years wore on, and alas, there was little in my school education to convince me that physics or mathematics (boxes sliding down frictionless planes? tangents to circles?) were particularly interesting subjects. For a long time, I could probably

have listed ten things in which I'd sooner have specialised than astronomy. Then one day, quite unexpectedly, I encountered the 'Pale Blue Dot' photograph (Fig. 1.1), and read Carl Sagan's accompanying words. The photograph moved me profoundly, and stirred in me the same visceral awe I'd last felt when poring over the TIME magazine years before. The image screamed at me: *This* puts everything into perspective! How could anything be more interesting than *this*? And just think: there could be an infinity of *worlds* out there!

By and by, I found myself studying astronomy, and I was absolutely thrilled to learn not only of the genesis of exoplanetary science, but also of the breakneck speed at which its frontiers of knowledge were (and still are) advancing. This thesis embodies the small – though I hope useful – contribution I have made to the field of exoplanetary science over the past few years.

Today, I can still scarcely believe how fortunate I am to have the opportunity of carrying out research in one of the most awe-inspiring fields in science: a field in which I find myself filled almost daily with the same child-like wonder I experienced when sitting on my grandparents' porch all those years ago. If there are other 'pale blue dots' out there in the vastness of the cosmos, I very much hope to learn about them in my lifetime.

*Nescire autem quid ante quam natus sis acciderit, id est semper esse puerum.*

*To be ignorant of the past is to be forever a child.*

— Marcus Tullius Cicero, from *Orator Ad M. Brutum*, chapter XXXIV, section 120 (46 BCE).

# 2 Background material

## Contents

---

This chapter presents a selection of background material relevant to all that follows: a broad overview of exoplanetary science in general, and Doppler spectroscopy in particular; a discussion of stellar activity signals, and how they confound attempts to detect and characterise exoplanets through Doppler spectroscopy; and a tutorial-style introduction to Gaussian processes. Given the importance of Gaussian processes throughout the rest of this thesis, the latter introduction is fairly detailed.

Length considerations and the relatively broad scope of this thesis mean that there are inevitably some subjects on which this thesis draws (e.g. linear algebra, Bayesian inference, basic stellar astrophysics) but that cannot be covered as part of this chapter's background material. Similarly, there are a number of inherently interesting subjects (e.g. theories of planetary formation and evolution) that are connected only indirectly to the research presented in this thesis, and are therefore also not covered in this chapter. Fortunately, there are many excellent books and review papers covering all of these topics. Instead of padding this thesis with pallid derivatives of these works, I refer the interested reader to a few of the resources I have found most relevant and useful, listed below.

**Exoplanetary science in general.** The books *Exoplanets* (Seager, 2010, with chapters contributed by 34 authors) and *The Exoplanet Handbook* (Perryman, 2011) provide detailed reviews of all the major research themes in modern exoplanetary science, from instrumentation and detection methods to planet formation, interiors and atmospheres, dynamics, and more. Both books are now a few years old, though I am not aware of any more recent works of similar scope. The NASA Exoplanet Archive[1] is an indispensable and regularly-updated online resource that catalogues all published exoplanet detections, including links to data and literature relevant to any given exoplanet. It also contains various useful visualisation tools.

[1] Available online at `exoplanetarchive.ipac.caltech.edu`

**Stellar astrophysics.** *An Introduction to Modern Astrophysics* by Carroll and Ostlie (2013) covers many topics in stellar structure and evolution, and modern astrophysics more broadly. The book *Solar and Stellar Magnetic Activity* (Schrijver and Zwaan, 2000) provides a comprehensive review and synthesis of the origin, evolution, and effects of magnetic fields in the Sun and other cool stars. The journal *Living Reviews in Solar Physics*[2] is an excellent resource for general reviews and articles on Solar and stellar activity.

**Bayesian inference, data modelling, etc.** The canonical resource on most things Bayesian (and probability theory more broadly) is the tome *Probability Theory: The Logic of Science* by Jaynes (2003). A more accessible, hands-on treatment of many useful topics may be found in the book *Bayesian Logical Data Analysis for the Physical Sciences* by Gregory (2005), which contains numerous worked examples, usually with an astronomical flavour. The book *Data Analysis: A Bayesian Tutorial* by Sivia and Skilling (2006) is more concise, but adds a useful section on nested sampling.

**Numerical methods.** The first book I consult on numerical methods is almost always *Numerical Recipes: The Art of Scientific Computing* by Press (2007), which covers an enormous breadth of topics, including numerical analysis, signal processing, basic data modelling and statistical analysis, linear algebra, and even some basic machine learning. I find this book to be extremely useful and accessible, and given its $> 100,000$ citations as of 2017, I cannot be alone in this judgement!.

## 2.1 Exoplanets

How does one define a planet? According to the International Astronomical Union (IAU), a planet is a celestial body that: (i) is in orbit around the Sun; (ii) has sufficient mass to assume hydrostatic equilibrium (i.e. a nearly round shape); and

[2]Published by the Max Planck Institute for Solar System Research (MPS), and available online at `solarphysics.livingreviews.org/`.

(iii) has 'cleared the neighbourhood' around its orbit.[3] A non-satellite body fulfilling only the first two of these criteria, e.g. Pluto, is classified a 'dwarf planet'.

An extrasolar planet or *exoplanet*, then, may be loosely defined as a body in orbit around a star other than the Sun, and satisfying criteria (ii) and (iii) above. Given the difficulty in verifying observationally whether observed celestial bodies do satisfy these two criteria, however, the IAU has to date adopted only a working definition for an exoplanet, agreed in 2003, and subject to change as knowledge improves (Boss et al., 2007). According to this working definition, objects with true masses below the limiting mass for thermonuclear fusion of deuterium ($\sim 13$ $M_J$ for objects with Solar metallicity) orbiting stars or stellar remnants are to be regarded as exoplanets, regardless of their formation history; the minimum mass and size for an extrasolar body to be considered an exoplanet should be the same as used in the Solar System. Substellar objects with true masses above the limiting mass for thermonuclear fusion of deuterium are 'brown dwarfs,' regardless of how they formed or where they are located. Finally, free-floating objects in young star clusters with masses below the limiting mass for thermonuclear fusion of deuterium are not planets but rather 'sub-brown dwarfs' (or whatever analogous name is most appropriate).

Other authors favour a simpler definition based on formation history, e.g. 'a planet is an end product of disc accretion around a primary star or sub-star' (Soter, 2006).

While devising a precise definition for an exoplanet is clearly a nontrivial matter, within the confines of this thesis we can afford to be a little more cavalier, and can regard an exoplanet simply as a planet orbiting a star other than the Sun, without worrying too much about what we mean by 'planet' or 'star'. At any rate, all of the exoplanets we shall study in this thesis have masses which are not exotic by the standards of the Solar System, and all will orbit main sequence stars.

With definitions out of the way, we now turn to planets themselves, starting with a historical perspective.

[3]This somewhat contested definition was agreed at the IAU's 2006 General Assembly in Prague.

### 2.1.1 Exoplanets: a brief history

The earliest recorded evidence of debates over cosmic pluralism – also known as the 'plurality of worlds,' or simply pluralism, i.e. the idea that there may be other worlds like our own – comes from the time of Anaximander (c. 610—546 BCE), a pre-Socratic Greek philosopher who supposed that worlds appeared and disappeared for a while, and that some were born when others perished. Support for pluralism came from a few other prominent Ancient Greek and Roman philosophers and writers – including such luminaries as Anaxagoras, Democritus, Aristarchus of Samos, Epicurus, and Lucretius – yet pluralism was rejected with greater effect by Plato, Ptolemy and Aristotle, who argued that the Earth is unique and that there could be no other systems of worlds. Their followers echoed these geocentric views throughout antiquity and into the Middle Ages: a stance which dovetailed neatly with later Christian ideas, with the result that pluralism was effectively suppressed for over a millenium.

Some medieval Islamic and Chinese Scholars endorsed the idea of cosmic pluralism. Imam Muhammad al-Baqir (676—733 CE), for example, wrote:

> 'I swear by God that God created thousands and thousands of worlds and thousands and thousands of humankind';

while the *The Adventures of Bulukiya*, an Islamic Golden Age tale from the *One Thousand and One Nights* (or *Arabian Nights*) depicted a cosmos comprising multiple worlds, some larger than Earth, each with its own inhabitants.

Eventually pluralism was reasserted in the West, first tentatively by scholastics, and then more seriously in the 16th century by followers of Copernicus, who had formulated a heliocentric model of the Universe (likely independently of Aristarchus of Samos, who had devised the same model eighteen centuries earlier), albeit one preserving the Ptolemaic idea of the stars being fixed on a stationary sphere surrounding the Sun. Proponents of Copernicus' model included Kepler and Galileo, although most educated Catholics of the time continued to subscribe to geocentric models.

Giordano Bruno (1548–1600), an Italian friar and polymath, was the first person in the West to argue forcefully that objects theretofore conceived of as 'fixed stars' were in fact other stars like the Sun, potentially with other bodies orbiting them, and receiving light and heat. Bettex (1965) quotes Bruno:

> 'There are countless suns and countless earths all rotating round their suns in exactly the same way as the seven planets of our system. We see only the suns because they are the largest bodies and are luminous, but their planets remain invisible to us because they are smaller and non-luminous. The countless worlds in the Universe are no worse and no less inhabited than our Earth. For it is utterly unreasonable to suppose that those teeming worlds which are as magnificent as our own, perhaps more so, and which enjoy the fructifying rays of a sun just as we do, should be uninhabited and should not bear similar or even more perfect inhabitants than our Earth. The unnumbered worlds in the Universe are all similar in form and rank and subject to the same forces and the same laws. Impart to us the knowledge of the universality of terrestrial laws throughout all worlds and of the similarity of all substances in the cosmos! Destroy the theories that the Earth is the center of the Universe! Crush the supernatural powers said to animate the world, along with the so-called crystalline spheres! [...] Take comfort, the time will come when all men will see as I do.'

More than four centuries later, Bruno's impassioned words still seem remarkably prescient. Yet his views were to find little favour during his lifetime, and in fact he was tried on several charges for heresy during the Catholic Inquisition, found guilty, and burned at the stake.

In the centuries that followed Bruno's tragic death, especially during the Enlightenment, cosmic pluralism increasingly grew to be regarded as a mainstream possibility, and was championed by philosophers such as John Locke, astronomers like William Herschel, and even politicians including John Adams and Benjamin Franklin. French astronomer Nicolas Camille Flammarion was a prominent proponent of cosmic pluralism during the latter half of the nineteenth century, and his first book, *La pluralité des mondes habités* (Flammarion, 1862) was a great success, seeing 33 editions published in its first 20 years. Flammarion was one of the first people to advance the idea that hypothetical extraterrestrial beings could be genuinely alien, rather than being simple variations of Earthly creatures.

By the early twentieth century, the term 'cosmic pluralism' had grown archaic, and the debates dominated for millenia by theological or philosophical speculation had grown to be informed primarily by astronomical and biological insights. Moreover, earnest scientific experiments seemed well within reach of providing, for the first time in human history, direct answers to questions about the existence of worlds beyond our Solar System.

Unfortunately, exoplanetary science suffered from a number of false starts. Early claims based on astrometric observations were made for the discovery of invisible, substellar companions to 70 Ophiuchi (Jacob, 1855; See, 1895), 61 Cygni (Strand, 1943), and Barnard's star (van de Kamp, 1963), the closest star to Earth after the stars in the $\alpha$ Cen system. All of these and similar early claims were sooner or later shown to be false positives, however. The claim of a planet around the 70 Ophiuchi binary star system was disproved on the basis that a three-body system with the claimed orbital parameters would have been highly unstable (Moulton, 1899), while the signals of the latter two 'planets' were shown to be artefacts caused by the telescope and photographic plates used to observe the stars in question (Hershey, 1973; Heintz, 1978).

In an influential 1952 paper, Otto Struve proposed a campaign of spectroscopic measurements to detect the RV reflex motion of stars induced by one or more orbiting planets (Struve, 1952). He noted that a Jovian-mass planet orbiting just 0.02 AU from a 1 $M_\odot$ star would induce a reflex motion of 0.2 km $s^{-1}$, with period 1 d: a signal that, at a stretch, might just have been detectable using the most powerful spectrographs then in existence.[4] Struve's proposal led to the initiation of several RV monitoring surveys over the next decades, including those of Campbell and Walker (1979, Campbell et al., 1988), Latham et al. (1989b), and Marcy and Butler (1992).

Despite the efforts of these RV observers, the first confirmed exoplanet discovery came courtesy of pulsar timing in 1992, when Aleksander Wolszczan and Dale

[4]Struve also noted that planets with edge-on orbits, as seen from Earth, would eclipse their parent stars, leading to a periodic photometric signature. The two detection methods proposed in Struve's paper, *viz.* Doppler spectroscopy (the RV method) and transit photometry, are today by far the most productive methods for detecting exoplanets.

Frail observed minute variations in the regular signals from the millisecond pulsar PSR 1257+12. Based on these variations, they inferred the presence of two Earth-mass bodies orbiting the neutron star (Wolszczan and Frail, 1992). Their discovery was quickly confirmed (Wolszczan, 1994), and today, Wolszczan and Frail are widely regarded as having made the first confirmed detection of exoplanets, albeit planets orbiting a degenerate stellar remnant rather than a main sequence star.

In fact, a few substellar objects we now believe to be genuine exoplanets were discovered by the RV method in the late 1980s and early 1990s, but were not recognised as planets at the time (Walker, 2012); these objects include $\gamma$ Cephei Ab and $\epsilon$ Eridani b (Campbell et al., 1988), and $\beta$ Geminorum b (Hatzes and Cochran, 1993; Hatzes et al., 2006). An invisible companion to HD 114762, discovered in 1989 (Latham et al., 1989a; Cochran et al., 1991), was found to have a minimum mass of $m \sin i \sim 11$ $\mathrm{M_J}$, though as of 2017 its orbital inclination and thus true mass remain unknown, so its status as an exoplanet remains undecided.

A major breakthrough for RV planet searches and exoplanetary science more generally came in 1995, when Michel Mayor and Didier Queloz used the ELODIE spectrograph (Baranne et al., 1996) to discover the first exoplanet orbiting a main-sequence star, *viz.* 51 Pegasi b, orbiting the G5V, Sun-like star 51 Pegasi (Mayor and Queloz, 1995). Here, at last, was the first irrefutable answer to the question of whether other worlds existed around stars like our own.

At the time of its discovery, 51 Pegasi b was regarded as something of an exotic planet. With a minimum mass of $\sim 0.5$ $\mathrm{M_J}$, and a 4.23 d orbit just 0.05 AU from its parent star, this 'hot Jupiter' planet challenged the conventional wisdom that giant planets formed beyond the so-called 'snow line' where volatile compounds such as water and ammonia would condense into solid ice grains. Lin et al. (1996) later showed that the planet probably formed by gradual accretion of solids and capture of gas at a much larger distance from the star ($\sim 5$ AU), and that it subsequently migrated inwards through interactions with the remnants of the circumstellar disc. Today, numerous other hot Jupiters have been discovered, and the paradigm

of formation at a distance followed by inward migration (as opposed to *in situ* formation) continues to hold currency (Triaud, 2016).

51 Pegasi b was to be the first of many planets discovered using the RV method, and for many years the RV method was responsible for far more planet discoveries than any other technique, until it was quite rapidly and dramatically surpassed by the transit method, thanks to the unprecedented success of the *Kepler* space observatory (Basri et al., 2005; Koch et al., 2010).

The first exoplanetary transit to be observed was that of HD 209458 b, in 1999 (Henry et al., 2000; Charbonneau et al., 2000), which had previously been discovered through Doppler spectroscopy. Charbonneau et al. (2002) subsequently detected the atmosphere of this planet, using in-transit spectroscopy which showed a deeper transit around the sodium $D_1$ and $D_2$ Fraunhofer lines. Soon thereafter it was also found to have an extended, evaporating upper atmosphere consisting of hydrogen, carbon, and oxygen (Vidal-Madjar et al., 2003). HD 209458 b thus earned the simultaneous distinctions of being the first known transiting exoplanet, the first planet detected through more than one method, the first observed to have an (evaporating) atmosphere, and the first known to have an atmosphere containing oxygen and carbon.[5]

It took until 2002 for another exoplanetary transit to be observed, and indeed for the first exoplanet to be *discovered* through transit photometry, when OGLE-TR-56b was discovered incidentally as part of a gravitational microlensing campaign (Udalski et al., 2002; Konacki et al., 2003). The first planets discovered through microlensing (OGLE-2003-BLG-235Lb, or MOA-2003-BLG-53Lb: Bond et al., 2004) and direct imaging (2M1207b: Chauvin et al., 2004) followed soon thereafter.

Since the discovery of 51 Pegasi b in 1995, the number of known exoplanets has increased *exponentially* over time, doubling every 27.3 months (as per 'Mamajek's Law': Mamajek, 2016). See Fig. 2.1. As of June 2017, there are around 3500 confirmed exoplanets orbiting some 2600 different stars, according to the NASA

[5]It later also became the first planet to have its orbital speed measured, determining its mass directly (Snellen et al., 2010).

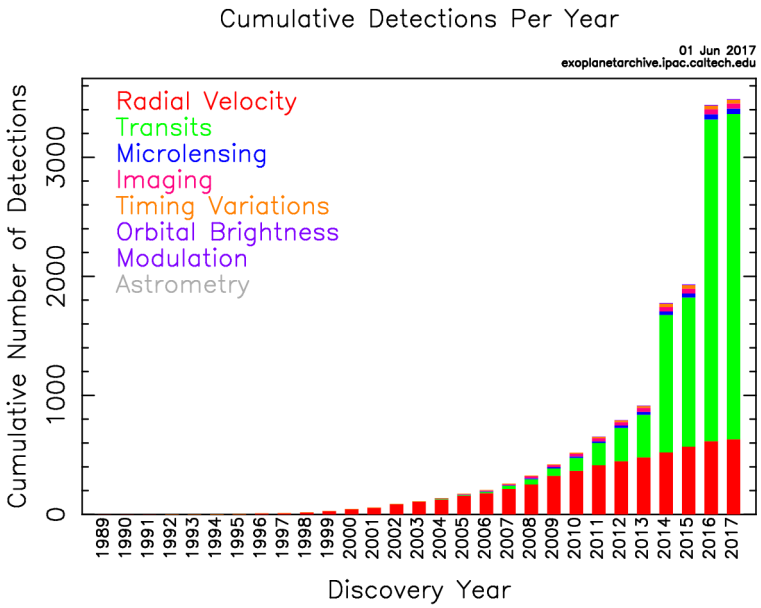


**Figure 2.1:** Exponential growth over the past two decades in the number of known exoplanets. (Image credit: NASA Exoplanet Archive.)

Exoplanet Archive; the Extrasolar Planets Encyclopaedia[6] puts the count at a little over 3600 confirmed exoplanets orbiting some 2700 different stars.[7]

Space does not permit a detailed review of significant exoplanets or planet-hosting systems discovered over the past twenty-odd years – such a review could easily fill an entire book. It is instructive, however, to consider some of the extrema among known exoplanets and their stellar hosts, to gain a rough impression of the remarkably wide range of properties they span. Table 2.1 summarises a few such extrema for known exoplanets, and Table 2.2 for their stellar hosts.

### 2.1.2 Methods for detecting and studying exoplanets

Depending on one's definitions, there are some ten or fifteen distinct methods which have led to the succesful detection of at least one exoplanet. Of these methods, just a few are responsible for the overwhelming majority of exoplanet discoveries; in fact, just two methods are responsible for some 95% of all known exoplanets.

[6] Available online at `exoplanets.eu`.

[7] The Exoplanet Archive includes only published detections of bodies with minimum masses $\leq 30\ M_J$. The Extrasolar Planets Encyclopaedia uses a stricter mass cut-off of $20\ M_J$, but includes detections that have been announced though not yet published in peer-reviewed journals.

**Table 2.1:** A few reliably-known extrema (e.g. most distant, least massive, etc.) among confirmed exoplanets, as of 1 June 2017; data from the NASA Exoplanet Archive.

| Characteristic | Planet(s) | Star(s) | Data | Reference(s) and notes |
|---|---|---|---|---|
| Most distant | SWEEPS-04, or SWEEPS-11 | SWEEPS J175853.92-291120.6, or SWEEPS J175902.67-291153.5 | $\sim 27,700$ ly | Sahu et al. (2006); Galactic bulge transits. Microlensing event PA-99-N2 suggests a possible exoplanet in M31, at a distance of $2.54 \pm 0.11$ Mly (Ingrosso et al., 2011). |
| Least distant | Proxima Cen b | Proxima Cen | 4.224 ly | Anglada-Escudé et al. (2016); RV detection. |
| Most massive | DENIS-P J082303.1-491201 b | DENIS-P J082303.1-491201 | $28.5 \pm 1.9$ $M_J$ | Sahlmann et al. (2013); astrometric detection. Most massive exoplanet in NASA Exoplanet Archive, though may actually be a brown dwarf. |
| Least massive | PSR B1257+12 b (Draugr) | PSR B1257+12 | $0.020 \pm 0.002$ $M_\oplus$ | Wolszczan and Frail (1992); Wolszczan (1994); pulsar timing. Probable mass based on assumption of coplanarity with two outer planets. |
| Largest radius | HAT-P-67 b | HAT-P-67 | $2.09 \pm 0.09$ $R_J$ | Zhou et al. (2017); transiting hot Saturn. |
| Smallest radius | Kepler-37 b | Kepler-37 | $0.354 \pm 0.014$ $R_\oplus$ | Barclay et al. (2013); Marcy et al. (2014); size constrained via asteroseismology. |
| Hottest | Kepler-70b | Kepler-70 | $> 6000$ K | Charpinet et al. (2011); planet likely evaporating, and is on on a 0.006 AU orbit around a post-red giant star (effective stellar temperature $\sim 27730$ K). |
| Coldest | OGLE-2005-BLG-390Lb (Hoth) | OGLE-2005-BLG-390L | $\sim 50$ K | Beaulieu et al. (2006); microlensing. |
| Longest orbital period | 2MASS J2126-8140 | TYC 9486-927-1 | $\sim 9 \times 10^5$ yr | Deacon et al. (2016); also widest known orbit around a single star ($> 4500$ AU). Directly-imaged. |
| Shortest orbital period | PSR J1719-1438 b | PSR J1719-1438 | 0.090706293(2) d | Bailes et al. (2011); also smallest orbit around a single star (0.004 AU). Pulsar timing. |
| Most eccentric orbit | HD 20782 b | HD 20782 | $0.97 \pm 0.01$ | Jones et al. (2006) |

**Table 2.2:** A few reliably-known extrema among confirmed planet hosts (stars or stellar systems), as of 1 June 2017; data from the NASA Exoplanet Archive.

| Characteristic | System(s) | Planet(s) | Data | Reference(s) and notes |
|---|---|---|---|---|
| Highest stellar metallicity | HD 47536 | HD 47536 b | $0.69 \pm 0.02$ dex | Setiawan et al. (2003) |
| Lowest stellar metallicity | Kepler-271 | Kepler-271b, c, d | $-0.951$ dex | Rowe (2014); Morton et al. (2016) |
| Highest stellar mass | o UMa | o UMa b | $3.09 \pm 0.07$ | Sato et al. (2012) |
| Lowest stellar mass | 2MASS J04414489+2301513 | 2MASS J04414489+2301513 b | 0.019 $M_\odot$ | Todorov et al. (2010); the star is a brown dwarf. The main-sequence planet host with lowest mass is probably TRAPPIST-1 ($0.0802 \pm 0.0073$ $M_\odot$; Gillon et al., 2017). |
| Largest stellar radius | $\beta$ Cnc | $\beta$ Cnc b | $47.2 \pm 1.3$ $R_\odot$ | Lee et al. (2014); star is a red giant. |
| Smallest stellar radius | PSR J1719-1438 | PSR J1719-1438 b | 0.04 $R_\odot$ | Bailes et al. (2011); star is a pulsar. The main-sequence planet host with smallest radius is probably TRAPPIST-1 ($0.1170 \pm 0.0036$ $R_\odot$; Gillon et al., 2017). |
| System with most confirmed planets | HD 10180, HD 219134, Kepler-90, TRAPPIST-1 | Planets b–h | 7 confirmed planets each | Lovis et al. (2011); Motalebi et al. (2015); Cabrera et al. (2014); and Gillon et al. (2017), respectively |
| System with most stars | Kepler-64 | Kepler-64b (PH1b) | 4 stars | Schwamb et al. (2013) |

Transit photometry, or the transit method, has led to the detection of approximately 2700 exoplanets (as of June 2017, some three quarters of all known exoplanets), more than 85% of which were discovered by the *Kepler* space observatory alone. The underlying principle is that whenever an exoplanet crosses (transits) in front of its parent star's disc, as seen from Earth, the observed disc-integrated brightness of the parent star will dip by a small amount. The duration and depth of the brightness dip in the star's light curve will be determined by (amongst other parameters) the radius of the planet relative to its parent star, while the period between successive dips will be determined by the planet's orbital period. Advantages of studying transiting planets include that it makes it possible to determine planets' radii and, in some cases, even to study their atmospheres, by studying starlight passing through the upper atmospheres of planets, and/or by studying light observed during secondary eclipses when planets are blocked by their parent stars (Seager and Deming, 2010). Disadvantages include the requirement of near-perfect alignment between a planet's orbital plane and the observer's line-of-sight (i.e. the planet's orbit should be essentially perpendicular to the plane of the sky), necessitating the observation of large numbers of stars; a danger of false positives due to variability intrinsic to stars (e.g. stellar pulsations); and the method's inability to constrain the mass of a planet (Seager, 2010).

The next most productive technique for discovering exoplanets is the RV method, i.e. Doppler spectroscopy, which has led to the detection of roughly 20% of all known exoplanets. The method is based on observing tiny but periodic Doppler shifts in the light emitted by distant stars, caused by the gravitational tug of one or more orbiting bodies, e.g. exoplanets. A primary advantage of the RV method is its ability to constrain a planet's mass: if the inclination angle $i$ of a planet's orbital plane relative to the sky is unknown, the RV method yields the planet's minimum mass, $m \sin i$. In the case of transiting planets, however, where $i$ is tightly constrained, the RV method yields the *true* mass of a planet.[8] A planet's mass, in turn, provides information about its likely composition, formation history, atmosphere scale height,

[8]True masses may also sometimes be inferred via stability analyses of multi-planet systems, or in cases where a planet's spectral lines can be distinguished from those of its host star.

and more. Therefore, apart from being an important technique for discovering planets, the RV method is also indispensable for confirming the planetary nature of candidates yielded by other techniques, including transit photometry (Bouchy et al., 2005). Perhaps the biggest disadvantage of the RV method is that, as is the case with transit photometry, variability intrinsic to stars themselves confounds attempts to discover or study planets using RV data. The RV method is described in more detail in the next section of this thesis.

In terms of productivity, the RV method is followed by gravitational microlensing ($\gtrsim 45$ detections), direct imaging ($\gtrsim 40$ detections), and transit timing variations ($\gtrsim 15$ detections), with all other methods combined contributing fewer than about 25 confirmed exoplanets. Despite the relatively small number of planets contributed by these detection methods, each is able to complement the RV and transit methods in various ways. For example, direct imaging is particularly sensitive to bright planets with face-on orbits (rather than edge-on orbits, as for the RV and transit method) at large distances from their host stars, although in practice imaging any planet directly entails overcoming enormous technical challenges (Traub and Oppenheimer, 2010). Gravitational microlensing – based on the observation of a mass (e.g. exoplanet plus host star) acting as a lens and warping the trajectory of light from a background star – has the advantage of not requiring the observation of any radiation from an exoplanet or its host star, enabling the detection of small planets on large orbits around faint stellar remnants, extremely distant planetary systems, and even free-floating ('rogue') planetary-mass objects (Sumi et al., 2011).

## 2.2 Doppler spectroscopy

The orbital motion of a single planet around a star causes the star to undergo reflex motion about the system's barycentre (centre of mass). This results in the periodic perturbation of three observable properties of the star: its radial velocity, angular (astrometric) position on the sky, and arrival time on Earth of some periodic reference signal (e.g. pulsar beam). This section focuses on stellar radial velocities.

As for all 2-body ('Keplerian') orbits, the orbit of the planet around the system's barycentre will take the form a conic section in a two-dimensional plane: i.e., an ellipse, parabola, or hyperbola. Assuming the orbit is closed, the orbit will form an ellipse, which may be described in polar coordinates by:

$$r(t) = \frac{a(1-e^2)}{1+e\cos\nu(t)}, \tag{2.1}$$

where $r$ is the planet-barycentre separation; $a$ is the *semi-major axis* of the ellipse; $e$ is the *eccentricity* of the orbit; and $\nu(t)$ is the *true anomaly*, the angle between the periapsis (point of minimum distance from the barycentre) and the position of the planet at any given time, as measured from the barycentric focus of the ellipse. $T_0$ defines the time at which the planet is located at a particular reference point, usually taken to be periapsis. Additionally, $a$ determines the planet's orbital period $P$ via Kepler's third law, i.e. $P^2 = 4\pi^2 a^3/G(M_\star + M_\mathrm{p})$. It may be shown (see e.g. Perryman, 2011) that the true anomaly $\nu(t)$ will be given by

$$\tan\frac{\nu(t)}{2} = \sqrt{\frac{1+e}{1-e}}\tan\frac{E(t)}{2}, \tag{2.2}$$

where $T_0$ defines the time at which the planet is located at a particular reference point, usually taken to be periapsis, and the so-called *eccentric anomaly*, $E(t)$, can be found by numerical solution of the following non-linear equation (usually referred to as *Kepler's equation*):

$$E(t) - e\sin E(t) = 2\pi(t-T_0)/P. \tag{2.3}$$

A planet's Keplerian orbit in *three* dimensions is fully specified by six parameters: $a$, $e$, $T_0$, $i$, $\Omega$, and $\omega$. The three angular parameters $(i, \Omega, \omega)$ define the projection of the two-dimensional orbit into the observed orbit, and depend solely on the orientation of the observer relative to the orbit. The *orbit inclination*, $0 \leq i \leq \pi$, defines the inclination of the orbit with respect to a reference plane, with $i = 0$ corresponding to a face-on orbit. The *longitude of the ascending node*, $\Omega$, specifies the longitude of the point where the planet crosses the reference plane while moving away from the observer. Finally, the *argument of periapse*, $\omega$, specifies the coordinate of the

planet's periapsis point relative to its ascending node, measured in the direction of motion in the orbital plane.

The Keplerian parameters for the stellar reflex orbit around the barycentre will be identical, except its $a$ will be reduced by a factor of $M_\mathrm{p}/M_\star$ compared to the planetary value, and its $\omega$ will differ from the planet's value by $\pi$ radians.

Geometric considerations reveal (see e.g. Lovis and Fischer, 2011) that the star's velocity along the observer's line-of-sight, i.e. its radial velocity, will take the form

$$\mathrm{RV}(t) = K\left[\cos\left(\omega + \nu(t)\right) + e\cos\omega\right], \tag{2.4}$$

with the RV *semi-amplitude*, $K$, given by

$$K = \left(\frac{2\pi G}{P}\right)^{1/3} \frac{M_\mathrm{p}\sin i}{\left(M_\star + M_\mathrm{p}\right)^{2/3}} \frac{1}{\left(1-e^2\right)^{1/2}}. \tag{2.5}$$

If we assume a circular orbit with $M_\star \ll M_\mathrm{p}$, we find $K \approx 12.5$ m s$^{-1}$ for a Jovian-mass planet orbiting a Sun-like star, and $K \approx 9$ cm s$^{-1}$ for an Earth-mass planet orbiting a Sun-like star.

Note that there are five observables related to a star's Keplerian orbit which can be fit for a single orbiting planet, on the basis of RV measurements alone: $e$, $P$, $T_0$, $\omega$, and the combination $K = K(a, e, P, i, M_\star, M_\mathrm{p})$. A systemic velocity parameter, $\gamma$, is usually also included to describe the constant component of the star system's barycentre relative to the Solar System barycentre, and sometimes also a linear trend parameter, $d$, to accommodate e.g. instrumental drifts and contributions from massive, long-period companions: i.e., $\mathrm{RV}(t) \rightarrow \mathrm{RV}(t) + \gamma + d(t - T_0)$.

In the case of $n > 1$ planets orbiting a star, the star's total RV signal can usually be very well approximated as a linear sum over $n$ Keplerian terms of the form in equation 2.4, leading to a total of $5n + 2$ parameters to be fit (including $\gamma$ and $d$). This approximation only breaks down in the case of relatively strong gravitational interactions *between* planets, in which case $n$-body integrations will need to be performed,[9] with long-term dynamical stability enforced (e.g. Deck et al., 2014)

[9]The arising *non-Keplerian* dynamics includes a wealth of interesting phenomena such as orbital resonances, transit timing and transit duration variations, and chaos (Fabrycky, 2010). Such non-Keplerian effects become more significant with very high cadence data, e.g. space-based photometry.

The equations above provide a basic though fairly general framework for describing Keplerian orbits, and for relating observed stellar radial velocities to physical quantities of interest, e.g. stellar and planetary masses. For a useful overview of various germane topics such as numerical challenges associated with fitting Keplerians to RV data, RV planet detectability and selection effects, RV observation scheduling, etc., refer to Perryman (2011).

### 2.2.1 Inferring RVs from stellar spectra

An instantaneous measurement of a star's RV is given by the small, systematic wavelength (Doppler) shift of the lines contained in the star's spectrum. If a star is receding with velocity $v$ in the observer's reference frame, at an angle $\theta$ relative to the line-of-sight, the change in wavelength $\Delta\lambda := \lambda_{\rm em} - \lambda_{\rm obs}$ is related to the recession velocity by the following expression for relativistic Doppler shift:

$$\lambda_{\rm obs} = \lambda_{\rm em} \frac{(1 + \beta \cos\theta)}{\sqrt{1 - \beta^2}}, \tag{2.6}$$

where $\lambda_{\rm obs}$ and $\lambda_{\rm em}$ are the observed and emitted wavelengths, respectively, and $\beta := v/c$. For $v \ll c$ and $\theta \ll \pi/2$, the expression reduces to (see Lang 1974, equation 2.226):

$$\mathrm{RV} = v \cos\theta \approx \left(\frac{\Delta\lambda}{\lambda_{\rm em}}\right) c, \tag{2.7}$$

with positive values conventionally indicating recession. It is worth noting that special and even general relativistic corrections to the above expressions become significant at the sub-m $\mathrm{s}^{-1}$ level, at which point wavelength shifts start to become dependent, e.g., on the transverse stellar velocity and gravitational potential at the point of emission (Lindegren and Dravins, 2003).

Information about the instantaneous Doppler shift is contained in the (typically) many thousands of absorption lines present in high-resolution spectra of Solar-type and M-dwarf stars. Most of this information can be concentrated into just a few parameters by *cross-correlation* of the observed spectrum with a template spectrum,

even in low signal-to-noise ratio (SNR) spectra. Formally, the cross correlation between two functions $f$ and $g$, $(f \star g)(\tau)$ may be defined as

$$(f \star g)(\tau) := \int_{-\infty}^{+\infty} f^*(t)g(t+\tau)dt. \tag{2.8}$$

The resulting cross-correlation function or CCF, $f \star g$, measures the linear correlation between $f$ and $g$ as a function of the lag between them, $\tau$, and is also known as a *sliding dot product*. Thus, cross-correlating two stellar spectra in velocity space allows the RV shift between them to be inferred. The technique was proposed by Fellgett (1955), demonstrated in practice by Griffin (1967), and extended to échelle spectroscopy using the CORAVEL instrument by Baranne et al. (1979). Some of the limitations inherent in using an imperfect and fixed physical template may be overcome by implementing cross-correlation numerically as a box-shaped mask of emission lines (Queloz, 1995; Baranne et al., 1996), with weighting according to relative line depths further optimising the SNR (Pepe et al., 2002b).

As the CCF is computed as an integral over all spectral lines, it may be interpreted, roughly, as a mean or 'concentrated' line on which classic spectroscopic analysis may be performed. In the absence of systematic line asymmetry, the shape of the CCF is usually well approximated by a Gaussian, with any asymmetry reflecting structure in the individual spectral lines. The location of the CCF's peak will yield the stellar RV; the CCF's width will yield the stellar rotation velocity $v \sin i$, and its equivalent width will provide a metallicity estimate, provided the star's effective temperature is known approximately (Queloz et al., 2001; Santos et al., 2002).

In practice, a number of instrumental and data reduction challenges need to be overcome to enable accurate extraction of a stellar RV signal from a spectrum. To put things in perspective, measuring RVs with a precision of 1 m s$^{-1}$ corresponds to measuring wavelength shifts of order $10^{-5}$ Å, which for a modern, high-resolution ($R = 100\,000$) spectrograph represents about $1/1000^{\text{th}}$ of a CCD detector pixel!

Accurate wavelength calibration (using, e.g. iodine gas absorption cells, thorium-argon emission lamps, or laser frequency combs) is an obvious though non-trivial prerequisite for accurate RV extraction. RV contributions from the Earth's rotation

and orbit around the Solar System barycentre (including gravitational perturbations from the other planets) are significant at levels of 0.5 km $s^{-1}$ and 30 km $s^{-1}$, respectively, and must be accounted for. Changes in ambient temperature and pressure in the spectrograph translate into wavelength shifts due to the varying refractive index of air; these shifts may amount to several hundreds of m $s^{-1}$ during a single night, so need to be corrected to high accuracy. Pixel-level detector inhomogeneities, varying spectrograph slit illumination (due e.g. to changing telescope focus or guiding), instrumental thermo-mechanical variations, and other such effects can all lead to apparent RV variability at a level of tens to hundreds of m $s^{-1}$. Finally, all spectral features not related to a target star may induce spurious RV signals. Important examples of spectral contamination include telluric lines from Earth's atmosphere, light from faint companions or background objects close on the sky to the target, and diffuse moonlight, which superimposes a reflected Solar spectrum onto the target star spectrum. For more on these and other technical challenges, refer e.g. to Lovis and Fischer (2011) and Halverson et al. (2016).

### 2.2.2 Modern Doppler spectroscopy

Thanks to a number of technical developments, the precision of RV surveys has been steadily improving. Whereas the spectrographs of fifty years ago produced RV measurements with errors in excess of 1 km $s^{-1}$ (nominal precision per measurement), today's so-called 'second generation' optical spectrographs can identify RV shifts down to a few tens of centimetres per second – a remarkable feat, especially in light of some of the enormous technical challenges alluded to above. Significant examples of such second-generation spectrographs include the HARPS (High Accuracy Radial Velocity Planet Searcher) spectrograph hosted by the European Southern Observatory (ESO) 3.6 m telescope in Chile; its Northern hemisphere component, HARPS-N, on the Italian 3.58 m Telescopio Nazionale Galileo telescope on La Palma Island; and HIRES (High Resolution Echelle Spectrograph) at the Keck telescopes in Hawaii.

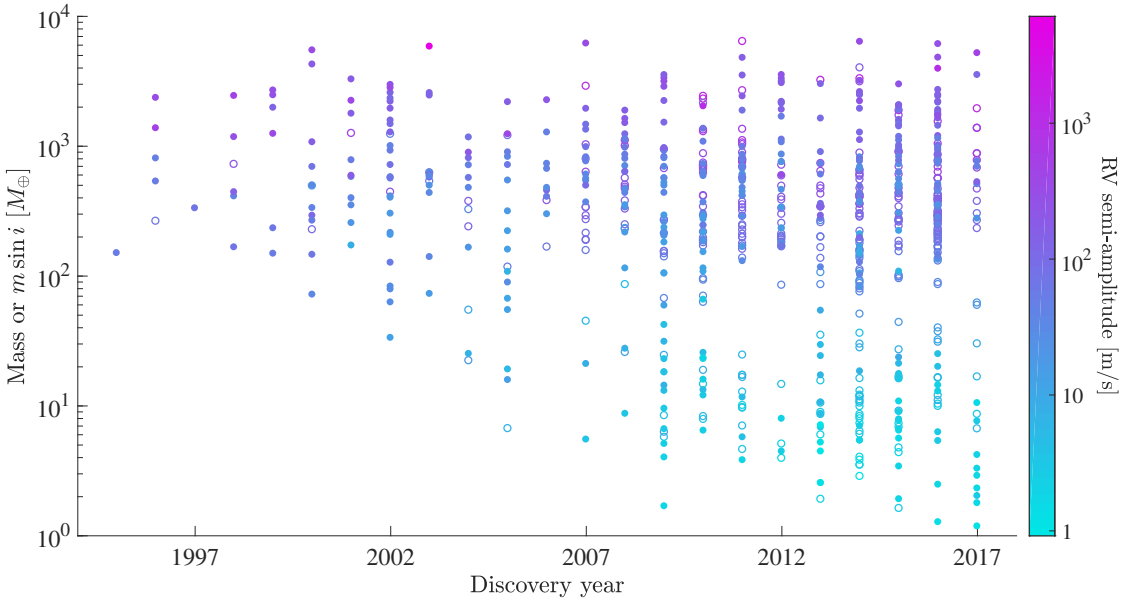


**Figure 2.2:** Between 1995 and 2017, more than 700 exoplanets have been discovered via Doppler spectroscopy. Here, their masses (solid circles) or minimum masses (unfilled circles) are plotted vs. discovery year, with their RV semi-amplitudes represented via the colour scale. Plot based on June 2017 data retrieved from the NASA Exoplanet Archive.

The improving precisions of spectrographs has enabled the discovery of exoplanets with ever-decreasing masses and longer periods; see Fig. 2.2. Indeed, the precision of second-generation spectrographs should, in principle at least, already make it possible to discover many Earth-mass planets orbiting in the habitable zones of their host stars (Pepe et al., 2011a; Lovis and Fischer, 2011). This has motivated intensive, multi-season monitoring campaigns on current state-of-the-art instruments, and results from these campaigns are in turn bolstering the science cases for next-generation instruments to be deployed on the world's largest telescopes.

Third generation spectrographs are expected to have measurement errors below 10 cm $s^{-1}$ (Halverson et al., 2016). Examples of such spectrographs include ESPRESSO (Pepe et al., 2010), scheduled to come online at ESO's Very Large Telescope in 2017; and CODEX (Pasquini et al., 2008), with first light expected circa 2025 on ESO's Extremely Large Telescope, currently under construction.

## 2.3 Stellar nuisance signals and exoplanets

Given that the HARPS and HARPS-N spectrographs (operational since 2003 and 2012, respectively) have both been shown to have short-term precisions of $\sim 30$ cm s$^{-1}$ and sub-m s$^{-1}$ precisions over time-scales of months (Mayor et al., 2003; Cosentino et al., 2014), one would expect that large numbers of Earth-mass planets have already been discovered and/or characterised through Doppler spectroscopy. Yet as of June 2017, as illustrated in Fig. 2.2, only a small handful of exoplanets have had masses or minimum masses measured below $\sim 2$ M$_\oplus$, with the smallest measured RV semi-amplitudes being around $\sim 1$ m s$^{-1}$. So, where *are* all the 'missing' low-mass planets?

Unfortunately, though we live in an age of RV instruments with ever-improving precisions, better instrumentation has meant that exoplanet hunters are also increasingly confronted by an obstacle that they could for many years ignore: the contaminating presence of RV signals from stars themselves. This stellar variability is arguably the most important and vexatious source of noise (or rather nuisance signal) in modern RV data sets. And it is not only RV data sets which are plagued by stellar nuisance signals: the *Kepler* mission's failure to discover the hoped-for numbers of habitable zone, Earth-sized planets is attributable in large part to an underestimation of stellar variability, and its impact on planet detection (Gilliland et al., 2011; Borucki, 2016).

A variety of different stellar phenomena give rise to RV signals on time-scales ranging from minutes to several years. The most significant of these are discussed below, ordered by time-scale, starting with the highest-frequency phenomenon.[10]

### 2.3.1 Stellar p-mode oscillations

Stars with convective outer layers, including the Sun, can excite p-mode (acoustic) oscillations at their surfaces via turbulent convection (Leighton et al., 1962; Bouchy

[10]Some effects which do contribute RV signals have been omitted from this discussion, given their negligible effect on planet searches. For example, coronal mass ejections contribute spike-like RV signatures, while the relativistic gravitational red-shifting of stellar photons will generally vary only weakly over time (Cegla et al., 2012). Neither effect would be liable to be confused with a periodic RV signal due to an orbiting planet.

and Carrier, 2001). The main restoring force at play is a pressure differential, and the dynamics of the standing acoustic waves are determined by variations in the speed of sound within stars; the local sound speed, in turn, depends on the temperature and chemical composition of the stellar atmosphere (Aerts et al., 2010). While studies of these p-mode oscillations can provide asteroseismologists with valuable information about stellar interiors, the same oscillations give rise to RV signals that confound planet-finding efforts.

Solar-like p-mode oscillations of the convective envelopes have typical periods of a few minutes, and typical Doppler amplitudes per mode of a few tens of cm $s^{-1}$ in RV; the superposition of a large number of simultaneously excited modes can lead to RV signals of several m $s^{-1}$ (Schrijver and Zwaan, 2000; O'Toole et al., 2008; Michel et al., 2008). Oscillation frequencies scale with the square root of mean stellar density, while their RV amplitudes scale with stellar luminosity-to-mass ratio (Christensen-Dalsgaard, 2004). Therefore, low-mass, non-evolved stars present easier targets for exoplanet search in terms of lower noise due to p-mode oscillations. However, even in the most favourable cases, obtaining $\sim$ m $s^{-1}$ RV precisions requires averaging out p-mode oscillation signals by integrating exposures over at least a few oscillation periods. Usually, an exposure time of 15 min is sufficient to decrease the p-mode oscillation signal well below 1 m $s^{-1}$ for dwarf stars (Dumusque et al., 2011c; Lovis and Fischer, 2011).

### 2.3.2 Granulation phenomena

Granulation is the photospheric signature of convective flows in the external layers of Sun-like stars. The 'boiling' granulation pattern comprises a large number of cells showing bright upflows surrounded by intergranular lanes of darker[11] downflows, corresponding to hotter plasma welling up from deeper layers and cooler plasma sinking downwards, respectively. See Fig. 2.3.

On the Sun, the typical vertical velocities of these convective flows are 1–2 km $s^{-1}$. However, the large number ($\sim 10^6$) of granules on the visible stellar surface means

[11]Since luminosity increases with the fourth power of temperature, in accordance with the Stefan–Boltzmann law, small temperature differences produce enormous brightness contrasts.

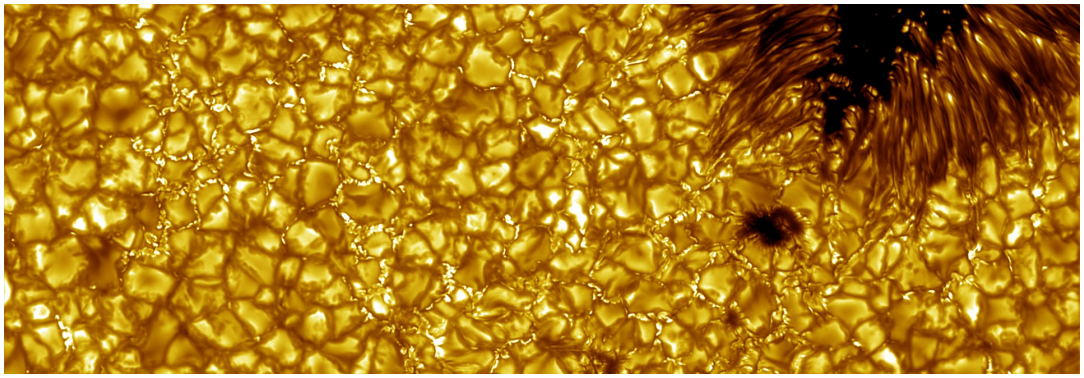

**Figure 2.3:** Granulation on the Solar photosphere, observed at high resolution with the Swedish 1-m telescope at La Palma. A sunspot is visible in the top-right corner, surrounded by convection cells with diameters $\sim$ 1500 km. Image credit: Institute for Solar Physics/Vasco Henriques, `www.isf.astro.su.se/gallery/images/2010/`.

these velocity fields are effectively averaged out to a constant net blueshift of about 200 m s$^{-1}$ (due to the more intense flux associated with large, bright upflows compared to cooler, more compact downflows), with the remaining jitter due to granulation being at the m s$^{-1}$ level for the Sun, and likely less for K dwarfs (Dravins, 1987; Michel et al., 2008; Mathur et al., 2011). The time-scale over which individual granules evolve is about 10 min for the Sun. Apparent larger-scale convection structures referred to as mesogranules and supergranules seem to be characterised by slower plasma flows and evolution on time-scales of a few hours up to about one day, although the nature and origin of these structures is subject to debate and ongoing research (Rieutord and Rincon, 2010).

As with p-mode oscillations, granulation-related phenomena become significant when aiming for sub-m s$^{-1}$ RV precision, and observing strategies to minimise their impact may be employed, e.g. taking 2 or 3 RVs per night, spaced by about 2 hours (Dumusque et al., 2011c).

### 2.3.3 Rotationally-modulated activity phenomena

Maxwell's equations tell us (among other things) that moving charges give rise to magnetic fields. In the Sun and other stars that are not fully radiative, an inner radiative zone and an outer convective zone are separated by a thin layer

called the *tachocline*. Here, strong shearing arises due to the transition between the differentially-rotating convective zone and the uniformly-rotating radiative zone. The complex motion of the electrically-conducting plasma in the tachocline is hypothesised to give rise to, and to sustain, stellar magnetic fields (Tobias, 2002). Moreover, rotation and differential rotation lead to (though so-called omega- and alpha-dynamo effects, respectively) stretching, winding, and twisting of a star's magnetic field lines, all of which results in very complex magnetic field topologies (Schmitt, 1987).

Because magnetic fields, in turn, exert forces on stellar plasma, regions of particularly intense magnetism at the surface of rotating Solar-type stars and M-dwarfs can give rise to inhomogeneities in these stars' atmospheres. These 'active' regions (so-called because they correspond to regions of intense magnetic activity) include dark spots, and bright plages and faculae.

Dark *spots* correspond to areas $\sim$ 500–2000 K cooler than the surrounding stellar photosphere, and tend to come in pairs associated with opposite magnetic field polarities. The lower temperature, and hence luminosity, is a result of localised magnetic suppression of convective energy transport. *Faculae*, on the other hand, are small bright points found in photospheric intergranular lanes, also associated with strong magnetic fields (Keller et al., 2004); they tend to be grouped into networks, and have longer lifetimes than spots. *Plages* are large, bright areas in the chromosphere, and seem to map closely to smaller underlying photospheric faculae, although the exact relationship between faculae and plages is not yet well understood. Spots are always surrounded by faculae and plages, but the latter can exist without spots.[12]

While it is straightforward to imagine how a dark spot, say, might give rise to stellar photometric variability, it is not necessarily as obvious why they would give rise to apparent RV variability.

Consider that as a dark or bright active region moves across the observed stellar disc, it breaks the flux balance between the redshifted and blueshifted

[12]For more on spots, faculae, plages, etc., see Schrijver and Zwaan (2000); Schrijver (2002); Hall (2008).

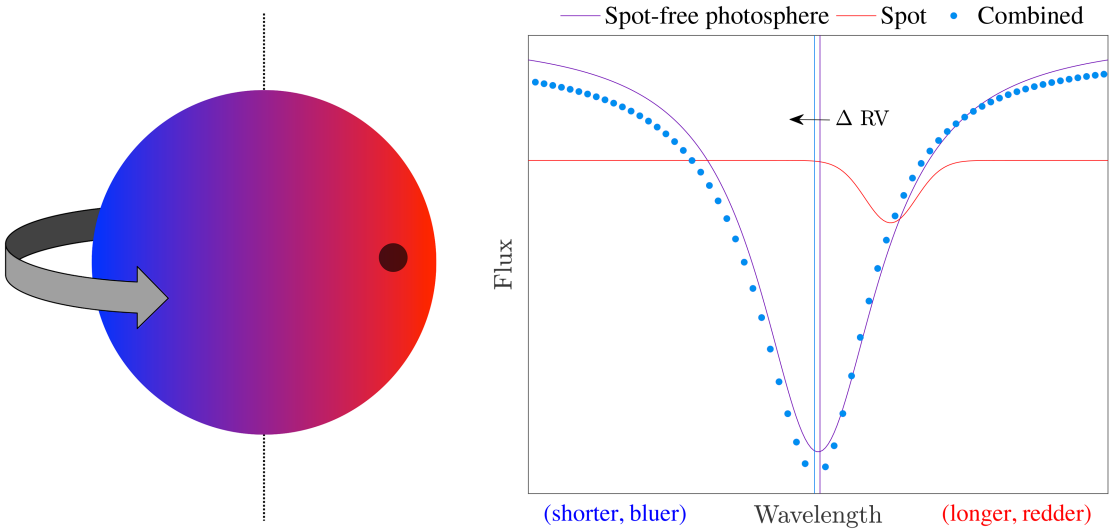


**Figure 2.4:** Illustration of how the flux deficit associated with a single dark spot on a rotating stellar disc (left) induces asymmetries in spectral absorption lines (right), leading to variations in measured RVs. At the instant pictured here, the dark spot is positioned on the receding (redshifted) hemisphere of the stellar disc, leading to a net blueshift.

hemispheres (Dumusque et al., 2011b; Boisse et al., 2012); see Fig. 2.4. This produces an asymmetry in the mean line profile, shifting its centroid by a small amount, and thus generating apparent RV signals on time-scales comparable to the stellar rotation period (Tassoul, 2000; Nielsen et al., 2013).

These surface inhomogeneities appear stochastically, and as they are carried across the observed stellar disc by stellar rotation, individual active regions also migrate across latitudes and evolve over time (grow, persist, decay, and eventually disappear), meaning that their combined signal may be quasi-periodic (QP) rather than strictly periodic, despite the smooth, periodic rotation of the star. Making matters worse, it is known that active regions on other stars can attain sizes more than an order of magnitude larger than those on the Sun, and can even exist near the stellar poles (Strassmeier, 2009).

In fact, the major contribution to RV perturbations arises not directly from the flux excess or deficit in the active regions themselves, but from the strong magnetic fields associated with them, which in turn exert forces on the highly ionised Solar plasma, leading to a localised suppression of the blueshift normally associated with convective cells (Meunier et al., 2010).

The strength of RV perturbations due to activity phenomena varies widely from star to star depending mainly on their mean magnetic activity level, which in turn depends mainly on stellar age for a given spectral type. Solar-type stars undergo a continual decrease in rotational velocity due to magnetic braking as they age, starting with typical rotation period of 1–2 d at 10 Myr, and slowing down to 20–50 d at 5 Gyr. The decline in equatorial angular velocity can be approximated as $\Omega_e \propto t^{-1/2}$, where $t$ is a star's age (Skumanich, 1972; Tassoul, 2000). Accordingly, the magnetic dynamo weakens with age, resulting in reduced surface magnetic activity.

Rotation rate also depends on spectral class, with most main sequence stars of types between O5 and F5 being rapid rotators, with rotation velocity increasing with mass. Typical rotation velocities are of order 190 km s$^{-1}$ for O5 stars, 210 km s$^{-1}$ for B5 stars, 160 km s$^{-1}$ for A5 stars, and 25 km s$^{-1}$ for F5 stars (Allen, 1976). However, as O-, B-, and A-type stars are fully radiative and lack convective envelopes,[13] the onset of surface magnetic activity phenomena is only found in F-type stars, which have shallow convective envelopes, and persists all the way through to the coolest M-type stars, which are fully convective, though generally slowly rotating. Typical M-dwarf rotation periods of a few weeks, and radii between 0.08 $M_\odot$ and 0.7 $M_\odot$, translate into rotation velocities between about 0.1 and 3 km s$^{-1}$ (Jenkins et al., 2009; McQuillan et al., 2013a).

Stellar activity presents major difficulties for exoplanet searches around young ($\lesssim$ 1 Gyr) stars, with rotating active regions causing RV variability of order $\sim$ 10 m s$^{-1}$. Such difficulties have led most RV planet-search surveys effectively to focus on older, slower-rotating, Solar-type stars; still, even chromospherically-quiet Solar-type stars can be expected to have activity-induced RV signals of order 1–2 m s$^{-1}$ (Isaacson and Fischer, 2010), i.e. an order of magnitude larger than the RV signature of an Earth-mass planet orbiting in the habitable zone.

One possible observational approach to mitigating activity-related problems is to conduct exoplanet searches in the infrared and near-infrared domain, where the photospheric contrast of spots is lower than at visible wavelengths (Artigau

[13] A convective envelope is a necessary but not sufficient condition for signficant stellar magnetic activity (Schrijver and Zwaan, 2000).

et al., 2014; Quirrenbach et al., 2014), and where M dwarfs are also more luminous. Another is to obtain dense sampling followed by averaging and binning of data over time-scales similar to the stellar rotation period, since the power spectrum of signals from rotating active regions becomes white at frequencies below the rotation frequency (Lovis and Fischer, 2011). Such an approach has clear disadvantages, however. It requires a large amount of telescope time, essentially throws away large amounts of data in the binning process, and is only suitable for detecting planets with periods much longer than the stellar rotation period. A preferable approach would be to model the QP signals of rotating active regions *jointly* with planetary signals, and to take into account their possible interference; see Sections 2.4 and 2.5.

### 2.3.4 Long-term magnetic activity cycles

The Sun undergoes a 22-year cycle, the Babcock–Leighton Solar dynamo cycle, which corresponds to an oscillatory exchange of energy between toroidal and poloidal Solar magnetic fields (Schrijver and Zwaan, 2000). This cycle gives rise to an 11-year QP cycle characterised by an increasing and decreasing number and size of active regions such as sunspots (as well as changes in the number of coronal mass ejections, the levels of Solar radiation, etc.). During periods in the cycle corresponding to minimum magnetic activity, the Solar disc may appear free of spots for months at a time; during the maximum, up to about 0.5% of the entire Solar disc may be covered in spots (Watson et al., 2011).

Other Sun-like stars exhibit similar long-term magnetic cycles. These long-term modulations of activity levels and numbers of surface magnetic regions can lead to variations in stellar-origin RV perturbations of up to tens of m $s^{-1}$ over the course of many years (Dumusque et al., 2011a; Gomes da Silva et al., 2012; Meunier and Lagrange, 2013).

## 2.4 Spectroscopic activity indicators

While observational strategies can be devised to mitigate nuisance RV signals associated with stellar oscillation and granulation (Dumusque et al., 2011c), activity-

induced signals pose a much bigger challenge. Apart from the nontrivial amplitudes associated with these signals, their periodic or QP nature means they can sometimes *mimic* planetary RV signals, possibly even leading to false-positive planet detections (e.g. Queloz et al. 2001; Desidera et al. 2004; Bonfils et al. 2007; Carolo et al. 2014; Haywood et al. 2014; Santos et al. 2014). Moreover, the associated timescales of these stellar signals can be comparable to those expected for planets: of order days or weeks for rotationally-modulated activity, and of order years for modulations due to long-term magnetic activity cycles. Therefore, in the regime of small-amplitude RV variations associated with lower-mass planets, planetary detection and characterisation rapidly becomes limited not by the quality of available spectrographs, but rather by the presence nuisance signals which can be very difficult to disentangle from putative planetary signals.

Hale (1908) was the first to observe Zeeman splitting of lines produced in spots on the Sun, and thus deduced that these regions corresponded to strong magnetic fields. Indeed, by precise measurement of Zeeman splitting across different regions of the Sun, spatial variations in the strength of the Solar magnetic field may be reconstructed in 'magnetograms' (Babcock, 1953). Unfortunately, given that other stars are usually observed as unresolved point sources, neither photometry-based active region maps nor magnetograms can easily be constructed. Zeeman-Doppler imaging (ZDI) is a tomographic technique which combines time series spectropolarimetry with the principle of maximum entropy to attempt to reconstruct stellar magnetic fields, and yields the simplest magnetic field geometry among the various solutions compatible with data (Marsh and Horne, 1988). However, the technique tends to work best for young, rapidly-rotating, and very active stars featuring significant Doppler broadening and large surface activity features – i.e., unlikely targets for exoplanet searches. Moreover, many fundamental questions remain unaddressed about the reliability of ZDI inversions of complex field topologies on cool stars (Kochukhov, 2016).

Fortunately, a number of activity-sensitive spectral parameters may often be extracted easily and simultaneously with RVs from the stellar spectra. In particular,

the width (full width at half maximum – FWHM) and degree of asymmetry of the spectral lines (commonly measured from the CCF bisector), as well as chromospheric activity indicators (e.g. $\log R'_{\rm HK}$, which measures the amount of emission in the cores of the Calcium II H & K lines), are all sensitive to activity, but not to the presence of planets. These commonly used 'activity indicators' are outlined below.

### The $S$-index and $R'_{\rm HK}$

The Ca II H and K lines, at 3 967 Å and 3 933 Å, are the strongest and broadest lines in the visible Solar spectrum. The detailed formation of the H and K lines is very complicated, given e.g. the nontrivial thermal profiles of stellar atmospheres, though it turns out that emission in the core of these lines is a useful and widely-used proxy for chromospheric magnetic activity levels (Linsky and Avrett, 1970; Schrijver and Zwaan, 2000).[14]

The $S$-index is a dimensionless ratio of the emission in the cores of the H and K doublet to that of two nearby, surrounding continuum bandpasses, and may be expressed thus:

$$S = \alpha \frac{\Psi_{\rm H} + \Psi_{\rm K}}{\Psi_{\rm V} + \Psi_{\rm R}}, \tag{2.9}$$

where $\Psi_{\rm H}$ and $\Psi_{\rm K}$ refer to the fluxes in the H and K line cores, $\Psi_{\rm V}$ and $\Psi_{\rm R}$ refer to the fluxes in the 20-Å bands on the violet and red sides of the H and K lines, and $\alpha$ is a normalisation factor. The flux measured in the reference bands, however, depends on spectral type, so the $S$-index cannot be used to compare activity levels in stars of different type. For the same reason, the $S$-index cannot be used to compare measurements between different instruments with different transmission levels in different bandpasses.

The $R'_{\rm HK}$ index was introduced (Middelkoop, 1982; Noyes et al., 1984b) to quantify activity levels in a way that is independent of spectral type and instrumental

[14] Other fairly common radiative diagnostics of stellar activity include the H-$\alpha$ and Ly-$\alpha$ hydrogen lines, and the Mg II h and k lines. For various reasons these are less widely used than the Ca II H and K lines: e.g., the Mg II h and k lines can only be observed from space, and trace circumstellar and interstellar matter in addition to stellar magnetic activity (Schrijver and Zwaan, 2000).

parameters. It measures the excess fraction of a star's bolometric luminosity radiated as chromospheric H and K emission:

$$R'_{\rm HK} = \frac{\Psi'_{\rm H} + \Psi'_{\rm K}}{\sigma_{\rm B} T_{\rm eff}^4}, \tag{2.10}$$

where $\sigma_{\rm B}$ is the Stefan-Boltzmann constant, $T_{\rm eff}$ is a star's effective temperature, and the primes on the $\Psi$ indicate that the fluxes (measured at the stellar surface, rather than received by the observer) have had the chromospheric contribution of a reference star subtracted. The $R'_{\rm HK}$ is often used in logarithmic units, i.e. as $\log R'_{\rm HK}$, as it will appear in subsequent chapters of this thesis. Stars may be classified, for example, as 'very inactive' ($\log R'_{\rm HK} < -5.1$), 'inactive' ($-5.1 < \log R'_{\rm HK} < -4.75$), 'active' ($-4.75 < \log R'_{\rm HK} < -4.2$), and 'very active' if $\log R'_{\rm HK} > -4.2$ (Maldonado et al., 2010). The Solar value varies between about $-4.85$ and $-5.0$ as the Sun moves between high- and low-activity phases.

**CCF FWHM**

The CCF FWHM indicates the width of the mean spectral line (as summarised by the CCF) at the point where this line achieves half its maximum depth. In the case of relatively rapidly rotating stars, Doppler broadening ensures that the FWHM is determined largely by the projected stellar rotation velocity, i.e. $v \sin i$. Since younger, faster-rotating stars are associated with stronger magnetic activity, the FWHM will give a general indication of the levels of magnetic activity of a star. It may be shown that the FWHM of the CCF will change as e.g. a dark spot crosses the stellar disc, in order to conserve the area enclosed by the line profile (see Fig. 2.4), so that RV perturbations arising from starspot flux deficits on a rotating star will therefore be correlated with variations in the FWHM (Haywood, 2015).

In addition to reflecting the rotation rate of stars, however, the CCF FWHM also reflects thermal, collisional, and turbulent motions in the stellar photosphere, and even Zeeman broadening effects (Desort et al., 2007). Thus, the CCF FWHM in general will also depend on stellar parameters such as temperature, pressure, surface gravity, etc. Rotational broadening ceases to be the dominant contribution

to the CCF FWHM when $v \sin i \ll 10$ km s$^{-1}$(Dempsey et al., 1992), as would be the case e.g. in typical M dwarfs.

**CCF bisector**

Another measure of the shape of the CCF is the bisector, which quantifies the asymmetry of the spectral lines. Perryman (2011) defines the bisector as 'the locus of median points midway between equal intensities on either side of a spectral line, thereby dividing it into two halves of equal equivalent width'. For a Gaussian line profile, this would be a straight vertical line going through the middle of the line profile. However, photospheric granulation in cool stars results in 'C'-shaped bisector, due to the superposition of a blueshifted line profile (due to bright upflowing granules) and a less intense, redshifted line profile (due to darker downflows in intergranular lanes); see Fig. 2.5. Active regions that reduce the net blueshift will thus produce small distortions in the shape of the line bisector.

Many quantities have been defined in relation to the bisector, e.g. the bisector velocity span (Toner and Gray, 1988), the bisector curvature (Hatzes, 1996), and the bisector inverse slope (BIS: Queloz et al., 2001), the latter being used often in subsequent chapters of this thesis.

## 2.5 Treating activity-induced RV variability

Assuming one or more 'activity indicator' time series (e.g. $\log R'_{\rm HK}$, FWHM, BIS, photometry, or perhaps even a direct measurement of disc-integrated magnetic field strength) are available in parallel with stellar RV measurements, it remains to identify any time-correlated variability in these ancillary time series, and thus to disentangle stellar activity-induced variability in the RVs from other e.g. exoplanetary signals.

There has been considerable interest in recent years in developing techniques for doing exactly this. A few of the most commonly used approaches are sketched below, while the focus in the next chapter of this thesis will be the development of a technique that improves on these existing approaches for disentangling activity and planetary signals.

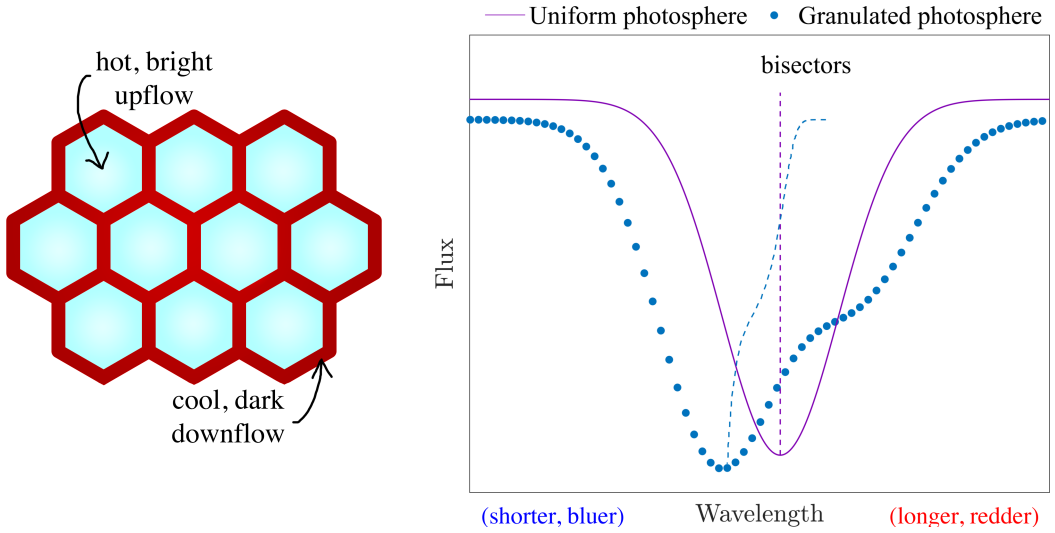


**Figure 2.5:** Illustration of how surface granulation (left) affects the shape of the line bisector (right). The line profile for the granulated star was obtained by summing a blueshifted line profile to a less intense, redshifted one; note how the line bisector curves upwards. Image based on a similar diagram from Haywood (2015).

### 2.5.1 Periodograms, and related methods

The 'classical' or Schuster periodogram (Schuster, 1898), based on the discrete-time Fourier transform (DFT), is perhaps the most widely used of all tools for characterising periodic variations in time series, and an enormous literature exists on the subject. Unfortunately, it requires many samples of a signal with even time sampling, making it unsuitable for e.g. low-cadence data containing sizeable gaps.

The Lomb-Scargle or least-squares periodogram generalises the Schuster periodogram to work with unevenly-sampled signals, by choosing a pair of lagged sinusoidal basis functions that are mutually orthogonal with respect to the uneven sampling (Lomb, 1976; Scargle, 1982). The Generalised Lomb-Scargle (GLS) periodogram generalises this periodogram further by accounting for signals with non-zero means (Zechmeister and Kürster, 2009).

The GLS periodogram is used very widely in exoplanet studies in general, including in Doppler spectroscopy, where it is used to look for both both stellar and planetary signals (e.g. Pepe et al., 2011a,b). For example, GLS analyses are often used to estimate stellar rotation periods from activity-sensitive time

series, e.g. $\log R'_{\rm HK}$, and thus to 'clean' or 'whiten' the RVs. Unfortunately, despite its computational efficiency and widespread currency, the GLS periodogram is a fundamentally unsound tool for studying quasi-periodic stellar activity signals, or even strictly periodic Keplerian signals.

At its core, the GLS periodogram assumes that any signal being studied comprises a *single* sinusoid corrupted only by white, Gaussian noise. That is, it does not just require a strictly periodic signal: it requires that the periodic variations to take the form of a sine wave (a valid restriction in the case of a single planet on a circular orbit). It cannot properly account for time-correlated noise or nuisance signals as simple as, say, a linear trend, let alone more complex signals such as quasi-periodic variations. Moreover, it cannot even account properly for the presence of more than one interfering periodic signal. Of course, in exoplanet studies, multiple, non-sinusoidal periodic signals (due to planets) might be simultaneously present with quasi-periodic stellar activity signals, and all of these signals should be expected to interfere in complex ways, as suggested in Fig. 2.6. Treating a complex interference pattern as a single sine wave is liable to produce very misleading results; see also Section 2.5.2 for more on the dangers of iterative fitting with single sinusoidal models.

Other popular, non-inference based methods for measuring periodic variations include autocorrelation function or ACF analyses (e.g. Edelson and Krolik, 1988; Scargle, 1989; McQuillan et al., 2013b), and wavelet transforms (e.g. Daubechies, 1990; García et al., 2014). While neither method requires sinusoidal signals, ACF methods cannot account well for long-term signal evolution, and sometimes require *ad hoc* devices for period estimation in ambiguous cases; wavelet methods, meanwhile, hinge upon choosing an appropriate mother wavelet to describe a signal over a range of transpositions and scales (Antonini et al., 1992).

A disadvantage of all these methods, including GLS analyses, is that they do not provide any way of modelling correlated nuisance signals (e.g. due to binary star interactions, intra-pixel contamination from other sources, stellar pulsations, instrumental variability, etc.) jointly with the genuine stellar or planetary signals of interest, and they do not provide principled ways of incorporating observational

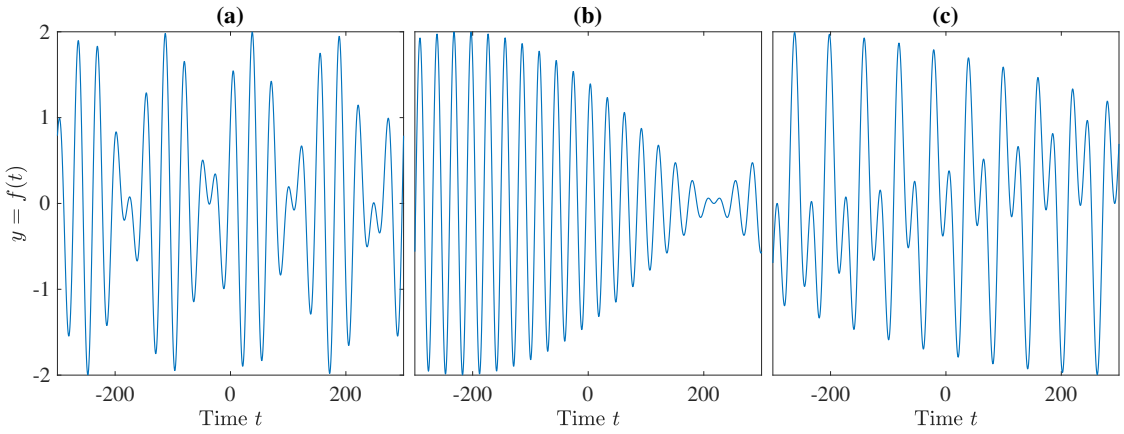


**Figure 2.6:** Interference between periodic signals. The plotted functions have the form $f = \sum_{i=1}^{N} \sin(2\pi t/P_i + \phi_i)$. In all three panels, $N = 2$, and $P_1 = 30$ d; the respective values of $P_2$ are (a) 38 d, (b) 29 d, and (c) 61.5 d. No amount of 'jitter' (independent, identically-distributed white Gaussian noise) could be added to a single sinusoid to explain the observed interference patterns, and trying to model the signals in this way could lead to nonsensical conclusions about the putative sinusoid's amplitude or period. Such an approach could yield even more misleading results when given $N > 2$ interfering signals.

uncertainties into the modelling (Angus et al., 2018). Moreover, they all reduce activity-sensitive time series to point estimates of stellar rotation periods, without providing period posterior probability distributions – nor do they provide any guidelines about how a rotation period point-estimated from one or more activity indicators should best be used to 'clean' an RV time series. Finally, they do not provide any way of disentangling a planetary signal from a stellar signal in cases where their periods are very similar; at best, they might prompt doubt about detecting any planets with periods similar to a rotation period estimated from activity indicator(s).

### 2.5.2 Pre-whitening

Pre-whitening entails carrying out a Fourier (or similar) analysis to find the strongest period in a given signal, then subtracting a sinusoidal fit with this period from the data, and repeating until the noise floor is reached (e.g. Queloz et al., 2009; Hatzes et al., 2010; Hatzes, 2013); it is analogous to the CLEAN method derived by Högbom (1974), which is widely used in radio astronomy. Pre-whitening is a quick and widely-used method to reveal potential periodicities, whether stellar or

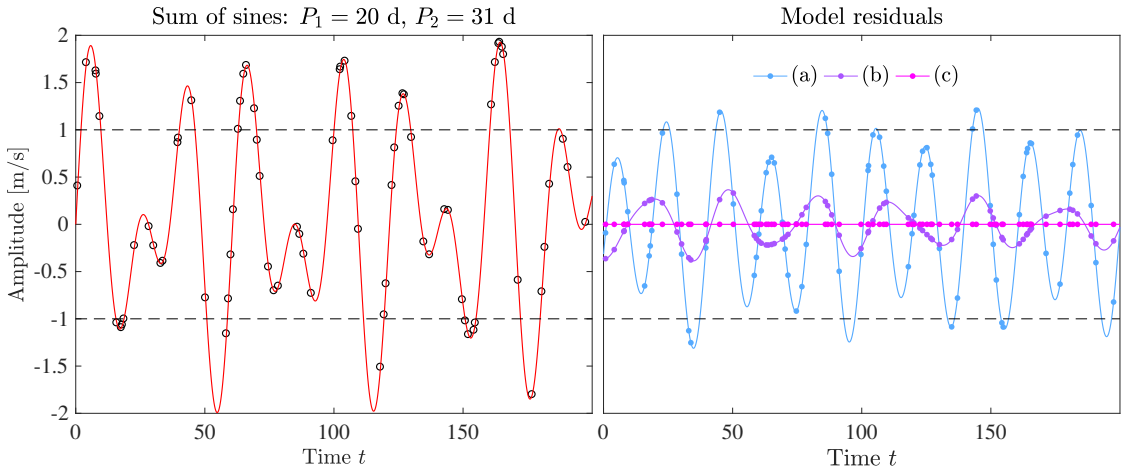


**Figure 2.7:** Toy example of the perils of iterative fitting. Left: a signal comprising two sinusoids with periods 20 d and 31 d, each with 1 m s$^{-1}$ amplitude. The signal is sampled, without noise, at 100 random times (black circles). Right: suppose we know the periods and phases of the signals, but not their amplitudes. After subtracting the best-fitting model for the 31 d signal, the residuals (a) are non-sinusoidal; and even after subtracting the 20 d signal, (b), a ∼30 cm s$^{-1}$ signal remains in the residuals. This is not an 'exoplanet buried in the noise'; it is just an artefact of our modelling. (Similar problems may arise if we did not know the signals' periods, and attempted to infer them iteratively.) If, instead, we fit the 20 d and 31 d signals *simultaneously* rather than iteratively, thus accounting for their mutual interference, the residuals (c) are completely flat.

planetary, present in RV data. For example, pre-whitening may be performed with sinusoids with periods corresponding to an estimated stellar rotation period, and harmonics thereof, thus 'filtering' (in principle) the RVs of stellar activity signals.

Unfortunately, a fundamental flaw in this approach is that it is tantamount to assuming, at each iteration, that the signal to be fitted is in fact just a single sinusoid plus uncorrelated Gaussian noise. Quasi-periodicity and non-sinusoidal variations notwithstanding, if multiple interfering sinusoids are present, the period and amplitude of any fitted sinusoid need not even correspond to a real signal present in the data. Thus, by proceeding according to false assumptions, spurious signals may be introduced into the data. A toy example of how such pre-whitening (or iterative fitting in general) can mutilate even a simple data set is given in Fig. 2.7.

### 2.5.3 Harmonic decomposition

Harmonic decomposition entails modelling stellar activity signals as sinusoids with periods equal to the stellar rotation period (as estimated e.g. via a GLS or ACF method) and varying numbers of harmonics (Boisse et al., 2011; Dumusque et al., 2012). Additionally, a finite coherence time parameter can be included in the modelling to allow, say, short-period signals to be preserved while filtering out lower-frequency signals suspected to be due to stellar activity. The technique is closely related to pre-whitening, although has the advantage of fitting multiple sinusoids jointly rather than iteratively.

While it is true that even the signals of complex (though non-evolving) starspot configurations can usually be modelled well with fewer than about five Fourier terms (Jeffers and Keller, 2009), it is also true that any function at all, even aperiodic ones, can be modelled with a Fourier decomposition, so *ad hoc* strategies need to be devised to choose the number of harmonics used in the fitting. The method also relies on a good prior rotation period estimate, and cannot help distinguish between a stellar and planetary signal with similar periods.

### 2.5.4 Nightly offsets method

Suppose planets with orbital periods of hours or a few days orbit a star with a rotation period of order several weeks, and that multiple RVs are measured on most nights the star is observed. In such a case it is reasonable to assume that on a given night, the stellar activity contribution to the RV signal is roughly constant, and that variations occurring on a time-scale of hours are caused by the orbital reflex motion of planet(s). This being the case, it is possible to fit for nightly offsets due to slowly-varying activity or long-period planets, while also fitting the short-period planetary orbits.

While this method has been used profitably in some systems (e.g. Hatzes et al., 2010, 2011; Pepe et al., 2013), its applicability is clearly limited – to say nothing of the information lost by introducing a large number of free parameters (equal to the

number of observing nights) to force the activity signal to be artificially 'ironed out,' instead of modelling the activity-induced variations jointly with other signals.[15]

### 2.5.5 Linear correlation models

Some authors eschew harmonic analysis and instead fit simple linear or monotonic correlations between the RV measurements and chromospheric activity or line shape indicators (Boisse et al., 2009; Tuomi et al., 2014; Robertson and Mahadevan, 2014) to try to remove stellar activity contamination from the RVs. While this approach has the advantage of not being predicated on assumptions of sinusoidal or periodic variability in the activity diagnostics, and indeed there do exist clear-cuts cases in which strong, nearly-linear correlations between RVs and activity diagnostics have been observed (Queloz et al., 2001; Dumusque, 2012), it is also simple to construct examples of signals where such models will fail completely.

Suppose $y = \sin(t)$ and $z = \sin(t + \pi/2)$; $z$ can straightforwardly be modelled as a phase shifted version of $y$ ($y$ implies $z$, and *vice versa*); yet both the linear and rank correlation between these two signals is zero. Or if $y = f(z)$ where $f$ is an even function, e.g. $y = z^2$, the linear and rank correlation between $y$ and $z$ will again be zero. These are toy examples, but the point is that linear and monotonic correlation can capture only certain very basic types of dependence between observables. Indeed, it is known that the correlations between RV variations and activity indicators need neither be linear nor monotonic; e.g. plots of RVs vs. BIS often display an eight-shaped pattern (Boisse et al., 2011; Figueira et al., 2013).

### 2.5.6 Analytical surface and spot modelling

Detailed analytical modelling of stellar surface features (Lanza et al., 2010; Boisse et al., 2012) has the potential to explain perfectly all observed activity-induced RV variability.

[15] Jaynes (2003) gives numerous insights on the virtues of data 'free of detrending, seasonal adjustment, pre-filtering, smoothing, or any other destructive mutilation,' and offers the withering remark that 'probability theory, correctly applied, is quite capable of extracting all the relevant information from the raw, unmutilated data and does best, with the least total computation, when it is allowed to do so freely.'

Unfortunately, these analytical models may contain enormous numbers (upwards of hundreds) of free parameters,[16] many of which are strongly degenerate (e.g. Russell, 1906; Jeffers and Keller, 2009; Kipping, 2012). Doppler imaging – analogous to the more recently-developed ZDI, but based on mapping simple line distortions rather than Zeeman splitting – may in some cases be able to offer constraints on spot distributions, but the spot maps constructed through Doppler imaging themselves suffer from degeneracy. Thus, robust exploration of spot model parameters is usually not feasible, with model comparisons being essentially impossible.

Simplified spot models (e.g. Lanza et al., 2014, where only two spots are modelled) have produced reasonable results. However, such relatively inflexible approaches sacrifice precision, and remain *ad hoc* models – there is only so much a model constrained to have exactly two spots, say, can possibly tell us about a stellar surface covered in hundreds of evolving active regions on a differentially-rotating star.

### 2.5.7 The $FF'$ method

Aigrain et al. (2012, hereafter A12) showed that there should exist a relatively simple relationship between the photometric brightness and RV variations of a spotted star. This leads to a straightforward method to predict the RV variations from the light curve, with only two free parameters.

In particular, A12 shows that to first order, the RV perturbation $\Delta\mathrm{RV}_{\mathrm{rot}}$ to the star's RVs incurred by the presence of spots on the rotating photosphere can be written

$$\Delta\mathrm{RV}_{\mathrm{rot}} = \frac{d\Psi/dt}{\Psi_0}\left[1 - \frac{\Psi(t)}{\Psi_0}\right]\frac{R_\star}{f}, \tag{2.11}$$

where $\Psi(t)$ is the observed stellar flux, $\Psi_0(t)$ is the stellar flux for an unspotted photosphere, and $R_\star$ is the stellar radius. The parameter $f$ represents the drop in flux produced by a spot at the centre of the stellar disc, and can be approximated as $f \approx 1 - \Psi_{\mathrm{min}}/\Psi_0$, where $\Psi_{\mathrm{min}}$ is the minimum observed flux, i.e. the flux during maximum spot coverage. A12 also showed that the RV perturbation $\Delta\mathrm{RV}_{\mathrm{conv}}$ due

[16] In addition to global stellar parameters such as inclination and rotation period, each spot or active region should have (at minimum) a longitude, latitude, size, temperature and lifetime.

to suppression of convective blueshift associated with spots or faculae – thought to be the dominant source of time-correlated RV signal in the Sun – may be approximated as

$$\Delta \mathrm{RV}_{\mathrm{conv}} = \left[1 - \frac{\Psi(t)}{\Psi_0}\right]^2 \frac{\delta V_{\mathrm{c}} \kappa}{f}, \tag{2.12}$$

where $\delta V_{\mathrm{c}}$ is the difference between the convective blueshift in the unspotted photosphere and within magnetised areas, and $\kappa$ is the ratio of this area to the spot surface (typically $\kappa \gg 1$).

Because the method allows activity-induced RV perturbations (i.e. $\Delta \mathrm{RV}_{\mathrm{rot}} + \Delta \mathrm{RV}_{\mathrm{conv}}$) to be modelled using only the stellar flux and its time derivative as basis functions, this method was named the $FF'$ method. In other words, given a high-cadence photometric time series of a spotted star, the variations in this time series can be used to estimate the RV signal due to the spots (assuming no transiting planets are present, or that their signals have been removed). Thus RVs can be 'cleaned' of stellar contaminants, possibly revealing residual exoplanetary signals.

In the method's favour, it is simple and easy to implement, and has very few free parameters. It also requires little prior knowledge of the star being modelled; in particular, it does not require knowledge of a star's rotation period, or active region configurations. A12 demonstrated the method's success at reproducing Solar RV variations from photometry at the sub-m $\mathrm{s}^{-1}$ level, and at the $\sim$ m $\mathrm{s}^{-1}$-level for Kepler targets. Its performance compares favourably with that of far more complicated spot models.

Nevertheless, the $FF'$ method remains a first-order model, and can only be applied when simultaneous high-precision photometric and RV measurements are available. The method also does not make use of information available in spectroscopic activity indicators, e.g. $\log R'_{\mathrm{HK}}$ FWHM, BIS, etc.

### 2.5.8 Summary

Given the very strong interest in developing ways to mitigate the impact of activity, it is not possible to give here a comprehensive discussion of the very many techniques

that have been developed and applied in recent years. Some of the techniques not discussed above include simple red-noise models (Tuomi et al., 2013; Feroz and Hobson, 2014), apodised Keplerian periodograms (Gregory, 2016), and compressed sensing techniques (Hara et al., 2017); see Dumusque et al. (2017) for a comparison of some of these techniques.

In any case, it should be clear that all of the somewhat commonly-used techniques discussed above suffer from a number of drawbacks, many shared between the methods (and not always explicitly pointed out above for each method), e.g. being based on overly restrictive or unrealistic assumptions, or on the other hand being overly flexible; not allowing principled incorporation and propagation of uncertainties; not allowing rigorous model comparison; relying in various ways on *ad hoc* devices; neglecting to model potentially-relevant information; not being able to disentangle stellar and planetary signals with identical periods; and so on. The method developed in the next chapter of this thesis aims to address these shortcomings.

## 2.6 Gaussian process regression

The preceding discussion about the shortcomings of traditional approaches to modelling or mitigating stellar activity signals leads us very naturally to consider Gaussian processes (GPs).

Whereas traditional regression methods are usually concerned with inferring the parameters of a single function, GPs give us a powerful, non-parametric, and data-driven way to perform Bayesian inferences about *functions* themselves.

Formally, a GP may be defined as a joint distribution over infinitely many random variables, any finite subset of which has a consistent joint Gaussian distribution (Rasmussen and Williams, 2006). As such, a GP is a distribution over functions with a continuous domain, e.g. time or space.

This rather stark definition does not necessarily make it clear why GPs might be useful, say, for modelling stellar activity signals. Therefore, the remainder of

this chapter aims to provide a practical and relatively non-technical introduction[17] to GPs, while elucidating their conceptual basis. We shall see that GPs represent a convenient and powerful approach to modelling a wide variety of physical processes, especially where more traditional or simplistic regression techniques[18] fall short.

### 2.6.1 Parametric vs. non-parametric modelling

Suppose we have a data set of observations obtained as input-output pairs, $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{n} = (\mathbf{x}, \mathbf{y})$; for example, each $x_i$ might represent an observation time, and each $y_i$ might represent the luminance or radial velocity of a star. We concern ourselves now with the regression problem of finding the relationship between the inputs and outputs: this might take the form $y(x) = f(x) + \eta$, where $f()$ is an unknown function and $\eta$ is an additive noise process.

As a concrete example, consider a small set of data samples at $x = 0, 1, 2$, and associated observed values $y = 0, 0.5, 1$. Fitting a simple linear model to the observations, of the form $y = f(x) = ax + b$, gives rise to the curve shown in Fig. 2.8(a). For suitably-chosen values of $a$ and $b$, this model manages to describe our data very well – exactly, in fact. This does not mean the linear model is necessarily 'correct'. Fig. 2.8(b) shows an entire family of curves, generated using a polynomial spline model. As is the case with the linear model, each spline curve passes exactly through each of our observations, yet the curves differ significantly from each other in interpolation and extrapolation, i.e. in regions without observations.

This simple example demonstrates that there will always exist many functions that can explain equally well any data we have observed. How do we go about choosing from the vast array of mathematical functions that give rise to such explanatory curves?

[17] For a comprehensive textbook on GPs, refer to Rasmussen and Williams (2006).

[18] While the focus in the following discussion will be on GPs in the context of *regression* (i.e. inferring relationships between variables), it may be noted that GPs have also found applications in areas such as probabilistic classification, Bayesian quadrature, global optimisation, solution of differential equations, and more (for a diversity of applications, see e.g. Gibbs, 1998; Kuss, 2006; Osborne, 2010).

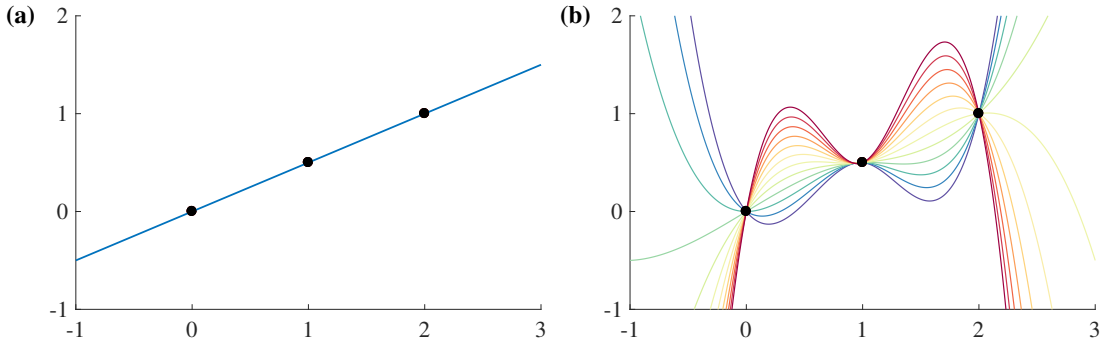


**Figure 2.8:** Two simple examples of curve fitting. (a) Straight line fit to three observations, denoted by black dots. (b) 12 cubic spline curves, each of which fits the observed data as well as the straight line. All of these curves are similar to each other near to the data, but vary significantly in regions far from the observations.

.

Given strong and precise prior knowledge of the system we're observing, it might be possible to reduce this infinite-dimensional function space to a single function or family of functions. For example, suppose applying Newton's second law to a system leads to the equation of motion

$$\frac{d^2y}{dx^2} = -\omega^2 x; \tag{2.13}$$

this equation governs the behaviour of a wide variety of idealised classical and quantum systems, including simple pendula, undamped spring/mass systems, a body undergoing uniform circular motion, resonant electric circuits, the vibrations of atoms in a crystalline lattice, and more (French, 2001). Solution of this differential equation tells us that the system's motion will be described by a sinusoid of the form

$$y(x) = A\cos(\omega x + \phi). \tag{2.14}$$

Thus the problem of choosing a function from an infinite-dimensional family of functions is reduced to the much simpler problem of choosing values (finding the probability distributions) for the parameters $A$, $\omega$, and $\phi$. Prior constraints on some of these parameters may be informed by physical knowledge of the system; additional constraints will come from the extent to which different parameter values lead to a close match between model and data, i.e. from a likelihood function. Such

modelling is referred to as *parametric*, in the sense that a finite number of unknown parameters are to be inferred as part of the data modelling process.

Many simple physical systems are amenable to parametric modelling, and accordingly, there is an enormous body of literature on parametric modelling methods. Nevertheless, there are also many scenarios in which we might have little or no prior knowledge regarding the most appropriate parametric model to use. Or we might have prior knowledge that is motivated directly by physical considerations, but is nevertheless difficult to build into any parametric model.

For example, suppose we are trying to model the photometric (or RV) signature of multiple dark active regions on the rotating, unresolved disc of a given star. We might know that the star takes about one month to rotate about its axis: given persistent star spots, the smooth rotation should lead to an overall smoothness and periodicity in the observed stellar signal. However, we also know that spots can appear stochastically on the stellar disc, grow, persist for a lifetime of perhaps a couple of rotation periods, and then decay until they disappear. Clearly this would lead to a departure from true periodicity, since the changing spot configurations would lead to an evolution of the 'shape' of the photometric signal, despite the smooth underlying rotation. We might also know something about the typical fraction of this star's surface that is covered in spots (which would determine the amplitude of photometric variations due to spots), as well as typical numbers of distinct spots present on the star (which, when combined with effects such as differential rotation and limb darkening, would determine the complexity of the disc-integrated variations we'd observe).

It would certainly be *possible* to build a parametric model that'd fit any observed photometric variations due to evolving spots on this rotating star. However, a number of questions would immediately arise. Should such a model be based on five, or ten, or a hundred spots? If a model with four spots suffices, how would we know that what we're really observing is not eight spots with slightly different sizes or persistence lifetimes – or just three spots, but a more complicated pattern of underlying differential stellar rotation? Which *ad hoc* scheme would we use to decide

if we're over-fitting our data, since a more complex model will generally permit a better fit? How would we go about transforming our imperfect physical knowledge of the star into sensible priors on all of the (possibly hundreds of) parameters in our model? How would we decide which model to favour among arbitrarily many different physical models, if all are equally plausible? Might it be possible that *all* of the models we consider are inadequate? Clearly, in this example at least, any attempt at parametric modelling would throw up a multitude of questions to which there may be no straightforward or defensible answer.

Thankfully, however, it turns out that we *can* work specifically with the infinite space of functions that have characteristics consistent with (in the stellar activity example) our somewhat vague and difficult to parametrise prior physical knowledge. Moreover, we can even contemplate probability distributions over this function space, such that the work of modelling, explaining, and forecasting data is performed by refining these probability distributions, in order to focus on regions of function space that are strong contenders to model our data. As these functions are not characterised by explicit sets of parameters to be inferred, this approach is known as *non-parametric* modelling. A GP is a particularly powerful and convenient example of such a non-parametric model.

In many contexts, and for the purposes of this thesis, it is useful to think of a GP as a convenient way to formulate and work with probability distributions over functions. To gain insight into the idea of probability distributions over *functions* (rather than parameters of a function), as well as into the conceptual basis for GPs specifically, it is useful to turn first to the multivariate Gaussian distribution.

### 2.6.2 Conceptual basis: the multivariate Gaussian

Consider a bivariate Gaussian distribution, defined for random variables $x_1$ and $x_2$. The distribution is fully specified by a mean vector, $\mu$, and a $2 \times 2$ covariance matrix, $\Sigma$, which we may visualise as a covariance ellipse corresponding to equal probability contours of the joint probability distribution $p(x_1, x_2)$. Fig. 2.9 represents an example of a bivariate Gaussian distribution as a series of elliptical contours.

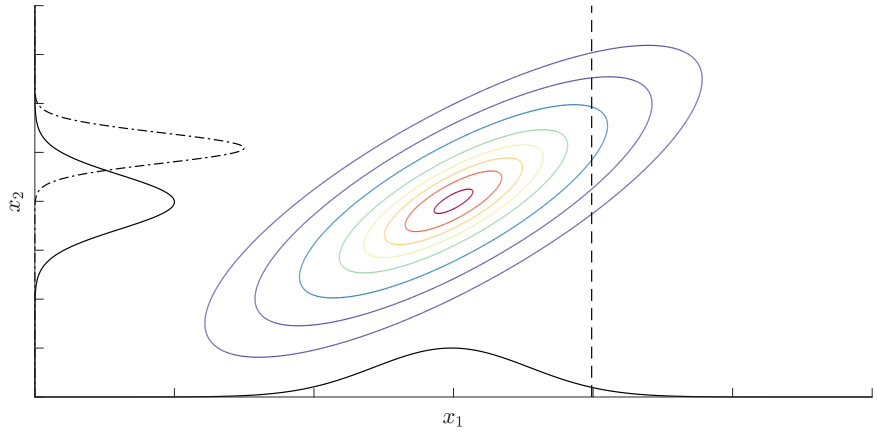


**Figure 2.9:** A joint 2-dimensional Gaussian distribution (covariance ellipses) forms marginal distributions, $p(x_1)$ and $p(x_2)$, that are in some sense 'vague' (black solid lines). Observing $x_1$ at the value indicated by the dashed line changes our beliefs about $x_2$, leading to a conditional distribution (black dash-dotted line). Knowledge of the covariance lets us decrease uncertainty in one variable, based on an observation of the other.

The corresponding marginal distributions $p(x_1)$ and $p(x_2)$ are shown as projections (solid lines) along the $x_1$ and $x_2$ axes. Now suppose that we observe one of the variables, $x_1$, at the location of the dashed vertical line in Fig. 2.9. The resultant conditional distribution $p(x_2|x_1)$, indicated by the dash-dotted curve, now looks very different to the marginal distribution $p(x_2)$. Given the relationship between the variables implied by and encoded in the covariance, knowledge of one variable shrinks our uncertainty in the other.

To build a bridge between this simple example and a time-series analysis problem, let us represent the same effect in a different format. Fig. 2.10 shows the mean (black line) and $\pm\sigma$ (blue shaded region) for $p(x_1)$ and $p(x_2)$. Fig. 2.10(a) depicts our initial state of ignorance, and Fig. 2.10(b) shows the situation after observing $x_1$, which changes the location and width of the distribution over $x_2$.

Now for the key point: there is no reason for us to stop at just two variables. We can extend this toy example to arbitrarily many variables, the relationships between which will be defined by correspondingly large covariance matrices. Fig. 2.11 shows the distribution for a 10-dimensional example in which noisy observations are made at times (or locations) 2, 6, and 8. Fig. 2.11(a) shows the posterior mean and $\pm\sigma$ as in the previous example. Fig. 2.11(b) extends the posterior distribution evaluation

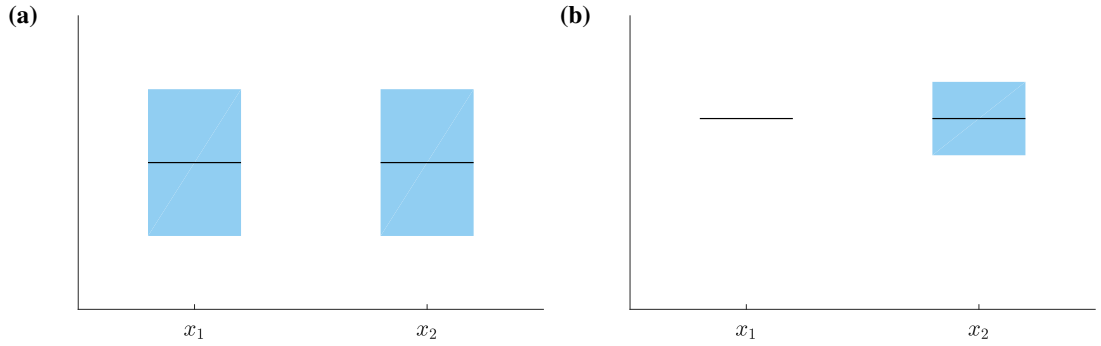


**Figure 2.10:** The change in distributions on $x_1$ and $x_2$, as in Fig. 2.9, but represented now in a form more familiar to time-series analysis. (a) The initial, vague, distributions; the black line denotes the mean and the blue shading $\pm\sigma$. (b) Distributions after observing $x_1$; the distribution over $x_2$ has become less uncertain, and the predicted value of $x_2$ has also shifted.

.

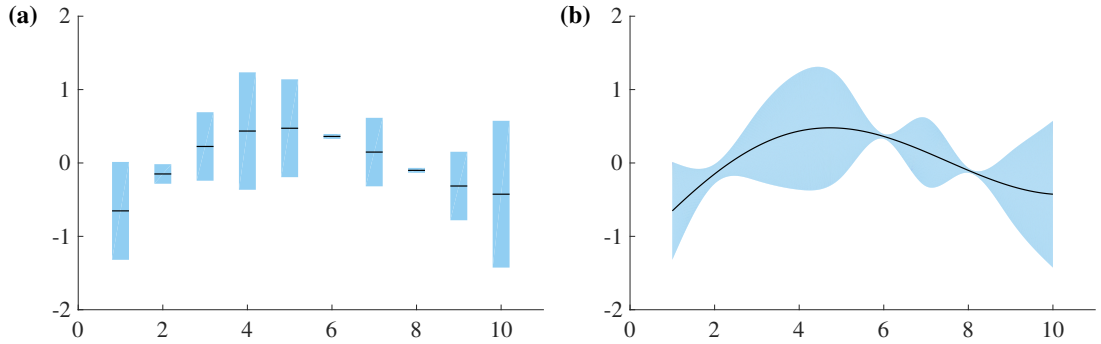


**Figure 2.11:** (a) The posterior distribution (the black line showing the mean and the blue shading $\pm\sigma$) for the case of 10 observations, with observations made at locations 2, 6 and 8. (b) Dense evaluation of the posterior in the interval $[1, 10]$, showing how arbitrarily dense evaluation gives rise to a 'continuous' posterior distribution with respect to time.

.

to an arbitrary number (thousands, say) of locations in the same interval: we note now that the discrete distribution has become rather continuous. We can keep extending this picture to the limit in which the locations of the $x_i$ are infinitely dense (here on $\mathbb{R}$), such that the infinite joint distribution over all of the $x_i$ is equivalent to a distribution over a function space.[19]

As the locations of observations and points at which we wish to investigate the function are not in any way constrained to lie on any predefined sample points, we

[19] After all, a continuous function may be thought of (loosely) as a mapping from an infinitely-large input vector to an infinitely-large ouput vector.

are now working in continuous time (or space) with a GP. Whereas a Gaussian distribution $\mathcal{N}$ is specified by a mean vector and covariance matrix, a Gaussian *process* $\mathcal{GP}$ will be specified by a mean *function* and a covariance *function*: in principle, both infinite-dimensional objects – and a draw from this GP will be a random *function* rather than a finite-dimensional random vector.

In practice, we shall not have to work with infinite-dimensional spaces; rather, it will turn out to be sufficient (thanks to the remarkably expedient properties of Gaussian distributions) to be able to evaluate the probability distribution at any finite number of locations on the real line we may choose, and to incorporate any of the finite number of observations we may have at other points.

To formulate a covariance over arbitrarily large sets, we define a *covariance kernel function*, $k(x_i, x_j)$, which provides the covariance element between any two arbitrary sample locations, $x_i$ and $x_j$. Because the covariance of $x_i$ with $x_j$ must be the same as the covariance of $x_j$ with $x_i$, we require $k(x_i, x_j) = k(x_j, x_i)$; the additional requirement that all linear combinations of random variables have non-negative variances leads to the requirement that $k(x_i, x_j)$ be a positive semi-definite function.

Thus, instead of having to specify an infinitely large covariance matrix, we shall simply work with a 'recipe' for computing covariances of interest. For a set of locations $\mathbf{x} = x_1, x_2, \ldots, x_n$, we may define the *covariance matrix* thus:

$$\mathbf{K}(\mathbf{x}, \mathbf{x}) = \begin{pmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{pmatrix}. \tag{2.15}$$

Provided $k(x_i, x_j)$ is a valid i.e. positive semi-definite covariance function, $\mathbf{K}$ will be a valid covariance matrix, i.e. $\mathbf{K}$ will be symmetric and will have non-negative eigenvalues.

### 2.6.3 Learning an unknown function from data

Let us now return to the problem posed in Section 2.6.1, where we have a set of observations $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n = (\mathbf{x}, \mathbf{y})$, and we wish to find the relationship

between the inputs and outputs. Let us assume this relationship takes the form

$$y_i = g(x_i) + \epsilon_i, \tag{2.16}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, and we do not know the functional form of $g(x)$. If we have some prior knowledge about certain properties of $g$, however, we can place a GP prior on the function $g$, with our prior knowledge about $g$ built into the covariance structure of the GP:

$$g \sim \mathcal{GP}(0, k). \tag{2.17}$$

If we happen to know how to parametrise part of the relationship between $y$ and $x$, such that $g(x) = \mu(x) + f(x)$, we can then place the GP prior on $f$ instead:

$$f \sim \mathcal{GP}(\mu, k); \tag{2.18}$$

we refer to $\mu(x)$ as the mean function of our GP. Note that our GP prior does not depend on the observed ('training') data $\mathcal{D}$; it serves to specify general properties we believe function $f$ to have, and we shall discuss how to choose a covariance function and a mean function in Sections 2.6.4 and 2.6.5 respectively. Lastly, for convenience, we can fold the uncorrelated observational noise into our covariance function by transforming $k(x_i, x_j) \to k(x_i, x_j) + \sigma^2\delta_{ij}$; generalisation to the case of different variances for different observations is straightforwardly achieved by replacing $\sigma^2\delta_{ij}$ with $\sigma_i^2\delta_{ij}$.

By construction, the entire function evaluation associated with the points in $\mathbf{x}$ is now a draw from a multivariate Gaussian distribution,

$$p(\mathbf{y}) = p(\mathbf{f}(\mathbf{x})) = \mathcal{N}(\boldsymbol{\mu}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x})), \tag{2.19}$$

where $\mathbf{f}$ is our unknown function evaluated at locations $x_1, x_2, \ldots, x_n$; $\boldsymbol{\mu}(\mathbf{x})$ is the GP mean function, $\mu$ evaluated at the same locations; and $\mathbf{K} \in \mathbb{R}^{n\times n}$, with $K_{i,j} = k(x_i, x_j)$.

Suppose now we would like to evaluate the posterior distribution at some test location, $x_*$, not included in $\mathbf{x}$. We start by considering the joint distribution of the observed data $\mathcal{D}$, consisting of $\mathbf{x}$ and associated values $\mathbf{y}$, augmented by $x_*$ and $y_*$:

$$p\left(\begin{bmatrix}\mathbf{y}\\ y_*\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}(\mathbf{x})\\ \boldsymbol{\mu}(x_*)\end{bmatrix}, \begin{bmatrix}\mathbf{K}(\mathbf{x},\mathbf{x}) & \mathbf{K}(\mathbf{x},x_*)\\ \mathbf{K}(x_*,\mathbf{x}) & \mathbf{K}(x_*,x_*)\end{bmatrix}\right), \tag{2.20}$$

where $\mathbf{K}(\mathbf{x},x_*)$ is the column vector formed from $k(x_1,x_*),\ldots,k(x_n,x_*)$, and $\mathbf{K}(x_*,\mathbf{x}) = \mathbf{K}(\mathbf{x},x_*)^{\mathrm{T}}$. Following some manipulation,[20] we find that the posterior distribution over $y_*$ is Gaussian, with mean and covariance given by

$$m_* = \mu(x_*) + \mathbf{K}(x_*,\mathbf{x})\mathbf{K}(\mathbf{x},\mathbf{x})^{-1}(\mathbf{y}-\boldsymbol{\mu}(\mathbf{x})), \tag{2.21}$$

and

$$\sigma_* = K(x_*,x_*) - \mathbf{K}(x_*,\mathbf{x})\mathbf{K}(\mathbf{x},\mathbf{x})^{-1}\mathbf{K}(\mathbf{x},x_*). \tag{2.22}$$

By repeating this process not just for a single test datum but an arbitrary set of locations outside our observations, $\mathbf{x}_*$, we may evaluate the posterior distribution $\mathbf{y}(\mathbf{x}_*)$; using standard results for multivariate Gaussians, it may be shown that the posterior mean and variance (conditioned on the known observations $\mathbf{y}$) will be given by

$$p(\mathbf{y}_*|\mathbf{y}) = \mathcal{N}(\mathbf{m}_*,\mathbf{C}_*), \tag{2.23}$$

with

$$\mathbf{m}_* = \boldsymbol{\mu}(\mathbf{x}_*) + \mathbf{K}(\mathbf{x}_*,\mathbf{x})\mathbf{K}(\mathbf{x},\mathbf{x})^{-1}(\mathbf{y}-\boldsymbol{\mu}(\mathbf{x})), \tag{2.24}$$

and

$$\mathbf{C}_* = \mathbf{K}(\mathbf{x}_*,\mathbf{x}_*) - \mathbf{K}(\mathbf{x}_*,\mathbf{x})\mathbf{K}(\mathbf{x},\mathbf{x})^{-1}\mathbf{K}(\mathbf{x},\mathbf{x}_*), \tag{2.25}$$

where the shorthand notation for the covariance, $\mathbf{K}(\mathbf{a},\mathbf{b})$, is defined as:

$$\mathbf{K}(\mathbf{a},\mathbf{b}) = \begin{pmatrix} k(a_1,b_1) & k(a_1,b_2) & \cdots & k(a_1,b_n)\\ k(a_2,b_1) & k(a_2,b_2) & \cdots & k(a_2,b_n)\\ \vdots & \vdots & \ddots & \vdots\\ k(a_n,b_1) & k(a_n,b_2) & \cdots & k(a_n,b_n)\end{pmatrix}, \tag{2.26}$$

[20] The standard formula for conditioning a joint Gaussian is $\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{bmatrix}, \begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right) \rightarrow (\mathbf{x}|\mathbf{y}=\mathbf{a}) \sim \mathcal{N}\left(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\mathbf{a}-\boldsymbol{\mu}_2), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\right)$.

so that $\mathbf{K}(\mathbf{a}, \mathbf{b}) = \mathbf{K}(\mathbf{b}, \mathbf{a})^{\mathrm{T}}$. It is easy to verify, by inspection of the above equations giving the posterior distribution for a specific set of observations, the more general result that placing a GP prior on $f$ leads us to a GP posterior distribution: $f|\mathcal{D} \sim \mathcal{GP}(\mu_{\mathcal{D}}, k_{\mathcal{D}})$, with

$$\mu_{\mathcal{D}}(x) = \mu(x) + \mathbf{K}(x, \mathbf{x})\mathbf{K}^{-1}(\mathbf{y} - \boldsymbol{\mu}(\mathbf{x})), \tag{2.27}$$

and

$$k_{\mathcal{D}}(x_i, x_j) = k(x_i, x_j) - \mathbf{K}(x_i, \mathbf{x})\mathbf{K}^{-1}\mathbf{K}(\mathbf{x}, x_j). \tag{2.28}$$

Equation 2.27 provides us with a best posterior estimate for the function $f$, given the prior $\mathcal{GP}(\mu, k)$ and the observed data $\mathcal{D}$, while equation 2.28 defines our uncertainty in the posterior estimate.[21] We are thus equipped with all the apparatus needed to learn an unknown function from some observed data, to make predictions outside of our observations, and even to place error bars on the function we have learned.[22]

Given the central role of covariance structure in defining the functions a GP may generate, we now turn to the important question of how to choose a covariance kernel.

### 2.6.4 Choosing a covariance function

To know how to choose the functional form of the kernel $k(x_i, x_j)$, let us consider what the covariance elements tell us. In our example in Section 2.6.2 featuring a simple bivariate Gaussian distribution, the diagonal elements of $\Sigma$ tell us the variance of the random variables $x_1$ and $x_2$ (informally, the variances determine how far $x_1$ and $x_2$ are spread out from their mean). The *off-diagonal* elements, on the other hand, determine the *joint* variability of $x_1$ and $x_2$. Independence would imply zero covariance, while if greater values of $x_1$ are associated on average with smaller values of $x_2$, say, their covariance would be nonzero and negative.

[21] Note that the posterior variance $k_{\mathcal{D}}(x_i, x_j)$ is equal to the prior variance, $k(x_i, x_j)$, minus a positive term. The observed data provides us with additional constraints on the model $f$, so the posterior variance is always smaller than the prior variance.

[22] It is worth reflecting that the remarkably expedient properties of multivariate Gaussian distributions have enabled the analytic conversion of a prior distribution over an infinite dimensional function space into marginal and conditional distributions over a finite-dimensional space.

In the context of time series analysis – and much of this thesis will involve time series analysis of astronomical data, especially RV data – a reasonable starting point is to assume that the informativeness of past observations, in explaining current data, should be a function of how long the past observations were made. For example, if we observed a star to be unusually bright half an hour ago, we might reasonably expect that it is likely to be at least somewhat bright at the present moment. Yet its recent behaviour might not tell us very much about its behaviour a hundred years hence, or its behaviour in the very distant past. From this starting point, we arrive at a family of so-called 'stationary' covariance functions that depend only on $|x_i - x_j|$, i.e. the Euclidean distance (in time or space) between observations, the most widely-used example of which is the squared exponential (SE) covariance kernel.

**Squared exponential (Gaussian) kernel**

The SE covariance kernel is given by

$$k_{\mathrm{SE}}(x_i, x_j) = h^2 \exp\left(\frac{-(x_i - x_j)^2}{2\lambda^2}\right); \tag{2.29}$$

the kernel is governed by two so-called 'hyperparameters,' $h$ and $\lambda$, which respectively govern the output and input (length or time) scale of functions. These parameters, and the parameters of GP covariance kernels more generally, are referred to as *hyperparameters* as they govern the properties of a prior distribution, rather than acting explicitly on the functional form of curves drawn from the distribution.

Functions drawn from an SE kernel will be infinitely differentiable, and will exhibit smooth variations with a typical time (or length) scale of $\lambda$. Fig. 2.12 depicts an SE kernel, and a few random functions drawn from it.

Stein (1999) notes that the very strong smoothness assumptions underpinning the SE kernel may be unrealistic for modelling some physical processes, and advocates using the Matérn class (below) instead.

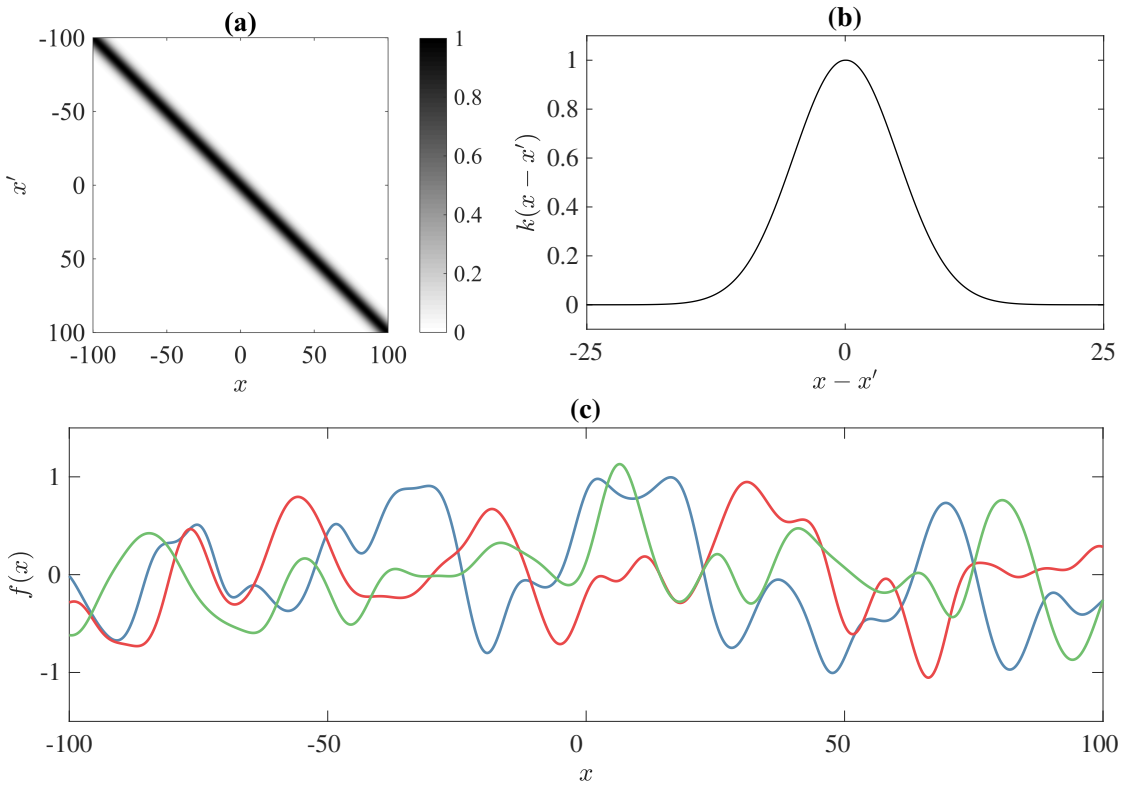


**Figure 2.12:** SE covariance kernel, as defined in equation 2.29, with $h = 1$ and $\lambda = 5$. (a) 2D representation of the corresponding covariance matrix $\mathbf{K}(\mathbf{x}, \mathbf{x})$ with $\mathbf{x} = [-100, 100]$. (b) 1D slice through this covariance matrix. (c) Three random functions drawn from a zero-mean GP with this covariance kernel. Each random function is infinitely differentiable.

.

### Matérn class of kernels

The Matérn class of covariance functions is defined by

$$k_{\mathrm{M},\nu} = h^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{\sqrt{2\nu}|x_i - x_j|}{\lambda} \right)^{\nu} \mathbb{B}_\nu \left( \frac{\sqrt{2\nu}|x_i - x_j|}{\lambda} \right), \tag{2.30}$$

where $h$ is the output scale, $\lambda$ is the input scale, $\Gamma()$ is the standard Gamma function, and $\mathbb{B}()$ is the modified Bessel function of second order. The hyperparameter $\nu$ controls the degree of differentiability of the resultant functions modelled by a GP with a Matérn covariance kerel, such that they are only $\nu + \frac{1}{2}$ times differentiable. As $\nu \to \infty$, the functions become infinitely differentiable, and the Matérn kernel becomes the SE kernel. At the other end of the spectrum, setting $\nu = \frac{1}{2}$ gives us

the Ornstein–Uhlenbeck (OU) or exponential kernel:

$$k_{\mathrm{OU}}(x_i, x_j) = h^2 \exp\left(\frac{-|x_i - x_j|}{\lambda}\right); \tag{2.31}$$

this kernel looks almost identical to the SE kernel, but the functions it models are *not* smooth. Instead, they are only once differentiable, and correspond to the random motion of a massive Brownian particle under the influence of friction (Bhattacharya and Waymire, 2009). Fig. 2.13 depicts an OU kernel, and a few random functions drawn from it.

Rasmussen and Williams (2006) note that in the absence of explicit prior knowledge about the existence of higher order derivatives, in practice it is probably not feasible to use a finite set of noisy data to distinguish between finite values of $\nu \geq \frac{7}{2}$ and $\nu \to \infty$.

**White noise**

White noise with variance $\sigma$ is represented by

$$k_{\mathrm{WN}}(x_i, x_j) = \sigma^2 \delta_{ij}, \tag{2.32}$$

where $\delta_{ij}$ is the Kronecker delta function ($\delta_{ij} = 1$ if $i = j$, otherwise $\delta_{ij} = 0$). This kernel is usually used to account for uncertainty in observed data, and is typically found added to other kernels. The use of a $k_{\mathrm{WN}}$ kernel in isolation, along with a zero-mean function, reduces a GP model to the special case of a simple additive white Gaussian noise (AWGN) model; all functions generated by such a GP will simply be white Gaussian noise.

**Building new kernels from old ones**

It turns out that the sum or product of valid covariance kernels will always yield another valid covariance kernel. This property of covariance functions allow us to move straightforwardly from one dimensional inputs (e.g. time) to a multidimensional set of inputs (e.g. time and space). If we assume our input space has finite dimension and use $x_i^{(d)}$ to denote the value of the input in the $d^{\mathrm{th}}$ dimension at index point $i$,

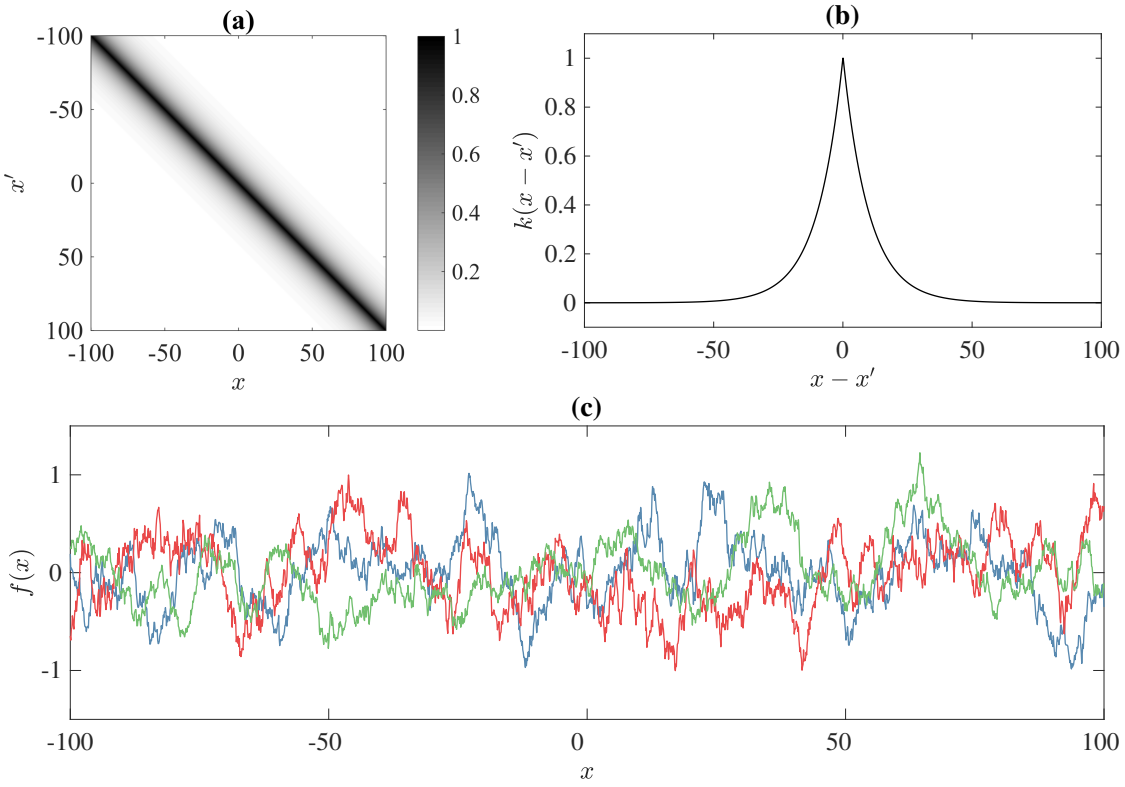


**Figure 2.13:** OU covariance kernel, as defined in equation 2.31, with $h = 1$ and $\lambda = 10$. (a) 2D representation of the corresponding covariance matrix $\mathbf{K}(\mathbf{x}, \mathbf{x})$ with $\mathbf{x} = [-100, 100]$. (b) 1D slice through this covariance matrix. (c) Three random functions drawn from a zero-mean GP with this covariance kernel. Each random function is only once differentiable.

then we can construct a covariance function that is the product of one-dimensional covariances over each input:

$$k(x_i, x_j) = \prod_d k^d(x_i^{(d)}, x_j^{(d)}), \tag{2.33}$$

where $k^{(d)}$ is a valid covariance function over the $d$th input. Because the product of covariances is a valid covariance, the expression in equation 2.33 is a valid covariance over the multi-dimensional input space.

For multidimensional outputs (e.g. a time series comprising measurements of both a star's RV and luminance), we may consider an artificial multidimensional input space, $[l, x]$, with $x$ denoting time, say, and the label $l$ being an index for the outputs. As before, we can then exploit the property that a product of covariance

functions is a covariance function in its own right, and write

$$k([l_m, x_i], [l_n, x_j]) = k_x(x_i, x_j)k_l(l_m, l_n), \tag{2.34}$$

taking covariance function terms over both time and time-series label. This approach allows us to model dependencies between multiple time series (Osborne, 2010), and is used extensively in the next several chapters of this thesis.

Another useful result is that a valid covariance function under any arbitrary (smooth) map remains a valid covariance function. For any function $u : x \to u(x)$, a covariance function $k()$ defined over a range of $x$ gives rise to a valid covariance function $k'()$ over the domain of $u$. We use this property in the next section to transform the SE kernel into a periodic kernel.

**Periodic kernel**

Consider the mapping $u : x \to u(x)$, where

$$u(x) = \left(u^{(a)}(x), u^{(b)}(x)\right) = \left(\cos\left(\tfrac{2\pi x}{P}\right), \sin\left(\tfrac{2\pi x}{P}\right)\right); \tag{2.35}$$

this mapping of a one-dimensional input variable $x$ into the unit circle gives us a simple way to modify simple covariance functions to formulate prior distributions over periodic functions. If we take the covariance over $x$ to be the SE kernel in equation 2.29, and use the identity $(\cos(x) - \cos(x'))^2 + (\sin(x) - \sin(x'))^2 = 4\sin^2\left(\frac{x-x'}{2}\right)$, we obtain

$$k_{\text{per-SE}}(x_i, x_j) = h^2 \exp\left(\frac{-2\sin^2\left(\frac{\pi(x_i - x_j)}{P}\right)}{\lambda^2}\right). \tag{2.36}$$

The covariance kernel $k_{\text{per-SE}}$, first derived by MacKay (1998), models periodic functions with overall amplitude defined by $h$ and period $P$. The hyperparameter $\lambda$ is now a 'roughness' parameter, defined relative to $P$, and sets the harmonic complexity of functions: large values of $\lambda$ will lead to simple, sinusoidal variations, while smaller values will lead to more complicated periodic functions, with more structure within a single period. Thus $k_{\text{per-SE}}$ gives us a means of performing inference about periodic functions with arbitrary complexity, without having to assume a simple and perhaps

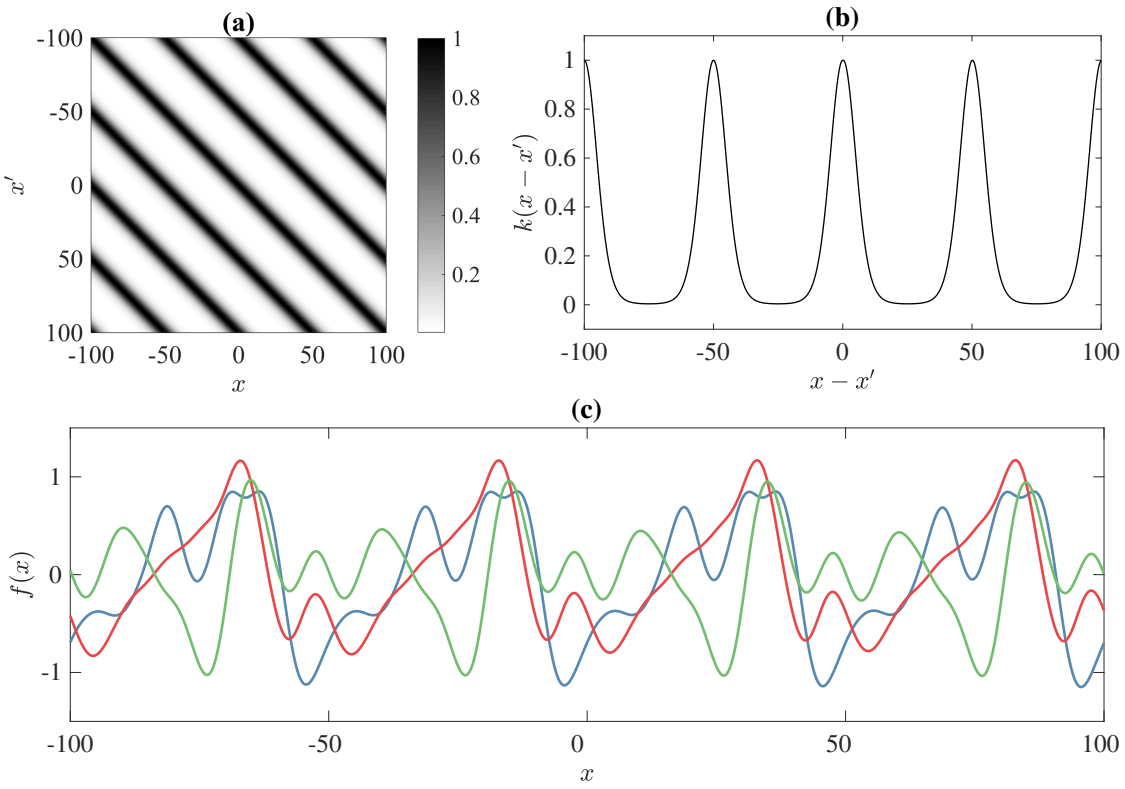


**Figure 2.14:** Periodic covariance kernel, as defined in equation 2.36, with $h = 1$, $P = 50$, and $\lambda = 0.125$. (a) 2D representation of the corresponding covariance matrix $\mathbf{K}(\mathbf{x}, \mathbf{x})$ with $\mathbf{x} = [-100, 100]$. (b) 1D slice through this covariance matrix. (c) Three random functions drawn from a zero-mean GP with this covariance kernel. Each random function is smooth, and satisfies $f(x) = f(x + nP)\ \forall\ x \in \mathbb{R}, n \in \mathbb{Z}$.

.

overly restrictive parametric model such as a sinusoidal model. Fig. 2.14 depicts this periodic kernel, and a few random functions drawn from it.

While $k_{\text{per-SE}}$ was obtained by transforming an SE kernel, it should be noted that a periodic covariance kernel can be constructed from any kernel (e.g. a Matérn kernel) through the mapping in equation 2.35. Nor is this mapping a unique way to construct a periodic covariance, though it is probably the simplest. Valid positive definite covariance kernels depending on $\cos|x - x'|$ rather than $e^{-\sin^2|x-x'|}$ are possible, for example, though the use of a cosine term might not be desirable unless we wanted to model only those periodic functions that alternate between positive and negative autocorrelation.

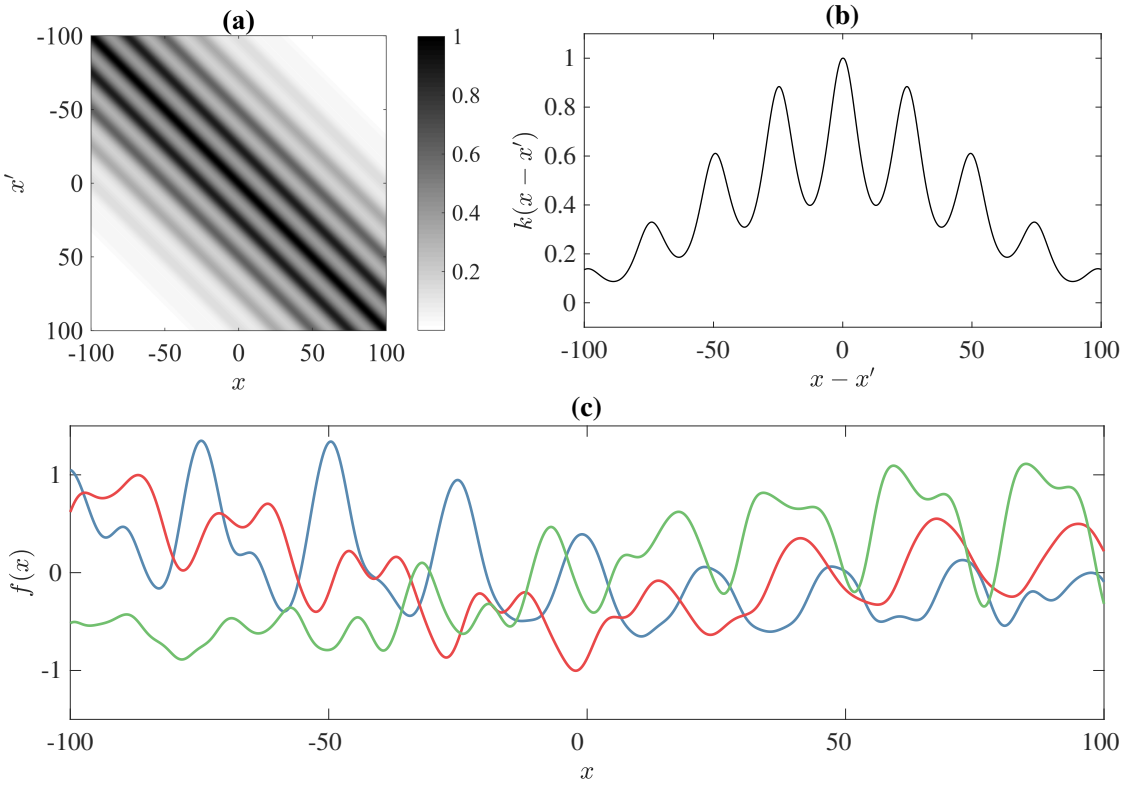


**Figure 2.15:** QP covariance kernel, as defined in equation 2.37, with $h = 1$, $P = 25$, $\lambda_{\mathrm{p}} = 0.75$, $\lambda_{\mathrm{e}} = 50$. (a) 2D representation of the corresponding covariance matrix $\mathbf{K}(\mathbf{x}, \mathbf{x})$ with $\mathbf{x} = [-100, 100]$. (b) 1D slice through this covariance matrix. (c) Three random functions drawn from a zero-mean GP with this covariance kernel. Each random function is smooth, and satisfies $f(x) \approx f(x + nP)\ \forall\ x \in \mathbb{R}, n \in \mathbb{Z} : |nP| \lesssim \lambda_e$.

.

### Quasi-periodic (locally periodic) kernel

We can obtain a kernel for modelling *quasi*-periodic (locally periodic) functions simply by multiplying a periodic kernel with one of the basic stationary kernels discussed earlier, with the latter kernel serving to specify the rate of evolution of the periodic signal. For example, if we multiply $k_{\text{per-SE}}$ by $k_{\text{SE}}$, we obtain

$$k_{\text{QP-SE}}(x_i, x_j) = h^2 \exp\left(-\frac{\sin^2\left(\frac{\pi(x_i - x_j)}{P}\right)}{2\lambda_{\mathrm{p}}^2} - \frac{(x_i - x_j)^2}{2\lambda_{\mathrm{e}}^2}\right); \tag{2.37}$$

the roles of hyperparameters $h$ and $P$ are unchanged from $k_{\text{per-SE}}$; $\lambda_{\mathrm{e}}$ defines how rapidly functions evolve, i.e. decay away from strict periodicity, while $\lambda_{\mathrm{p}}$ controls the harmonic complexity of functions (to simplify notation, $\lambda_{\mathrm{p}}$ has been redefined

to absorb a multiplicative constant).[23] Thus $k_{\text{QP-SE}}$ gives us a means of performing inference about functions which are governed by some sort of overall periodicity, but whose shapes may evolve over time. This kernel is depicted in Fig. 2.15, and will be used extensively in the next few chapters of this thesis, in order to model the signals of evolving active regions on rotating stars.

**Other considerations**

While the preceding discussion touched on only a few examples of covariance kernels, it is worth noting that in general, a broad range of function properties may be conveniently encoded into a covariance kernel, and thus into a GP prior over functions. These properties include typical length or time-scales for variations, stationarity (translation invariance), isotropy or anisotropy (direction independence or dependence), signal variance, noise variance, periodicity or quasi-periodicity, degree of smoothness, abrupt changes in input or output scales, relationships between observations and derivative or integral observations, and more.

### 2.6.5 Choosing a mean function

The mean function of a GP is usually used to account for whatever we expect a function to do in the absence of any stochastic or hard-to-parametrise variation accounted for by the GP's covariance structure. Put another way, we usually choose to build any deterministic, easily-parametrised physics into the mean function of a GP. Thus the mean function of a GP typically accommodates the 'traditional' components of a physical model.

For example, consider the problem of trying to model the RV variations of an active star that we suspect might host one or more orbiting planets (the next chapter of this thesis will have much more to say about this problem). If using a GP model, we might choose as a mean function a linear combination of $N$ Keplerian functions, to account for the stellar reflex motion induced by up to $N$ orbiting planets. Long term RV drifts, perhaps due to a binary stellar companion, might

[23] Section 7.1.3 of this thesis provides a more detailed discussion on how these hyperparameters should be interpreted and constrained.

be accounted for by a polynomial term in the mean function. Or if there are no orbiting planets, a constant mean function might suffice. On the other hand, any knowledge we have of the stochastic stellar activity signals would be more conveniently encoded into the GP's covariance function.

### 2.6.6 Hyperparameter and model inference

The role of inference in GP models is to refine vague distributions over many, very different functions, to more precise distributions focused on functions that are most compatible with our observed data. As the form of these functions is uniquely controlled by the covariance hyperparameters, in practice, inference proceeds by refining distributions over the hyperparameters. A major advantage of GPs over other regression methods is that the hyperparameters (ergo the model function) can be inferred directly and in a principled way from the observed data.

Suppose our covariance kernel function is controlled by some set of hyperparameters $\boldsymbol{\theta}$, and our mean function by parameters $\boldsymbol{\phi}$. From our underlying assumption of Gaussianity, the log marginal likelihood for our observations $\mathbf{y}$, $\log \mathcal{L} = \log p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}, \boldsymbol{\phi})$, obtained by marginalising analytically over unobserved function values, turns out to be given by the straightforward expression (Rasmussen and Williams, 2006):

$$\log \mathcal{L} = -\tfrac{1}{2} \log \det \mathbf{K}(\boldsymbol{\theta}) - \tfrac{1}{2} \mathbf{r}^{\mathrm{T}} \mathbf{K}^{-1}(\boldsymbol{\theta}) \mathbf{r} - \tfrac{N}{2} \log 2\pi, \tag{2.38}$$

where $\mathbf{r} := \mathbf{y} - \mathbf{m}(\boldsymbol{\phi})$ is a vector of residuals. The first term, $-\frac{1}{2} \log \det \mathbf{K}(\boldsymbol{\theta})$, is a complexity penalty term, which measures and penalises the complexity of the model; the second term plays the role of a goodness-of-fit measure, and is the only term which depends on the observations $\mathbf{y}$; and the third term is an (uninteresting) log normalisation term. The trade-off between complexity penalty and goodness-of-fit in GP regression is automatic, so there is no need to resort to any external method to address questions of over-fitting.

As observations in the real world are always limited in accuracy and completeness, deductive reasoning cannot be employed to prove or disprove a hypothesis about a

given system, e.g. a distant star. Fortunately, the tools of Bayesian statistical inference give us a principled framework for assessing the plausibility of one or more competing models, and for estimating the model parameters and their uncertainties (Jaynes, 2003).

To perform inference about $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, we simply combine the likelihood in equation 2.38 with a prior, $p(\boldsymbol{\theta}, \boldsymbol{\phi})$, and explore the posterior probability distribution $p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}, \boldsymbol{\phi})$. Indeed, the likelihood function in equation 2.38 can serve as a straightforward, 'drop-in' replacement for a more traditional likelihood function (which may not account for correlated nuisance signals, say).

Given some 'training' data, when the posterior distribution for a hyperparameter is tightly concentrated around a particular value, authors often informally describe the hyperparameter as having been 'learned' as having that value.[24]

In some applications, fixing $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ to their maximum likelihood values might be justifiable (for example, in the case of tightly peaked likelihoods and relatively flat priors, and where one is not interested in marginalising the parameters to facilitate model comparison) to avoid the computational burden of exploring and working with full hyperparameter posterior distributions. There are potentially serious pitfalls inherent in such approximations, however, and so they should be used with caution (Snelson, 2008).

To compute the Bayesian evidence for a given model $\mathcal{M}_0$ – which we take to mean a combination of a specific covariance kernel function and mean function, not assuming any particular model parameters – we must integrate (marginalise) over all model parameters. If we use the shorthand $\boldsymbol{\Phi}_0$ to denote the combined (hyper)parameters $\boldsymbol{\theta_0}$ and $\boldsymbol{\phi_0}$, the marginal likelihood or evidence for model $\mathcal{M}_0$ will be given by:

$$p(\mathbf{y}|\mathcal{M}_0) = \int p(\mathbf{y}|\boldsymbol{\Phi}_0, \mathcal{M}_0)p(\boldsymbol{\Phi}_0|\mathcal{M}_0)d\boldsymbol{\Phi}_0. \tag{2.39}$$

Approaches to computing such integrals include Markov chain Monte Carlo (MCMC) techniques, Laplace approximations, and Bayesian quadrature (Roberts et al., 2013).

[24]Such terminology has its origins in the field of machine learning; in fact 'learning' GP hyperparameters is no different to *inferring* the parameters of any other model.

The posterior odds ratio for model $\mathcal{M}_0$ against a competing model $\mathcal{M}_1$ requires computing the ratio of marginal likelihoods for the two models, i.e. the so-called Bayes factor (Gregory, 2005):

$$\frac{p(\mathcal{M}_0|\mathbf{y})}{p(\mathcal{M}_1|\mathbf{y})} = \frac{\int p(\mathbf{y}|\Phi_0, \mathcal{M}_0) p(\Phi_0|\mathcal{M}_0) d\Phi_0}{\int p(\mathbf{y}|\Phi_1, \mathcal{M}_1) p(\Phi_1|\mathcal{M}_1) d\Phi_1} \cdot \frac{p(\mathcal{M}_0)}{p(\mathcal{M}_1)}. \tag{2.40}$$

Returning to the earlier example of trying to model the RV variations of an active star we suspect to host one or more orbiting planets, model $\mathcal{M}_0$ might be 'GP with SE covariance kernel for stellar activity signals, and a constant mean function'; $\mathcal{M}_1$ might be 'SE covariance kernel for stellar activity signals, and a Keplerian mean function for a single planet'; $\mathcal{M}_2$ might add a second Keplerian to $\mathcal{M}_1$'s mean function; and so on.

### 2.6.7 An illustration of GP regression

As a simple and concrete demonstration of some of the ideas presented in the preceding discussion, consider the function $f(t)$, defined as follows:

$$f(t) := 2\sin\left(\tfrac{2\pi t}{27}\right)\cos\left(\tfrac{2\pi t}{27}\right) + e^{-|t|/100} \cdot \frac{2}{5}\sum_{j=-5}^{5} \sin(\tfrac{2\pi t}{27+j/2}) + \frac{t}{50}. \tag{2.41}$$

If we take $t$ to be time in days, then $f(t)$ describes the sum of a sinusoidal signal (with period 13.5 d), a linear term, and 11 different sinusoids (with periods uniformly spaced between 24.5 and 29.5 d) each of which has an exponentially-decaying amplitude. For $|t| \lesssim 100$, $f(t)$ will exhibit smooth, quasi-periodic variations; a GLS periodogram of $f(t)$ sampled densely on this domain will show a sharp peak at 13.5 d, and a broad complex of peaks around $27 \pm 5$ d.

Let us now use this function to synthesise a set of noisy observations made at random times, $\mathcal{D} = \{(t_i, y_i)\}_{i=1}^{50}$, with $t_i \sim \mathcal{N}(0, 50)$, $y_i = f(t_i) + \sigma_i$, and $\sigma_i \sim \mathcal{N}(0, 1)$. These synthetic observations are shown in Fig. 2.16(a).

If we had no idea about the functional form of $f(t)$, but nevertheless needed to model the observations $\mathcal{D}$, we could place a GP prior on $f$: $f \sim \mathcal{GP}(\mu, k)$. Hints of periodic variation in our observations suggest using the QP kernel defined by equation 2.37; we could add to our QP kernel a white-noise kernel (equation 2.32)

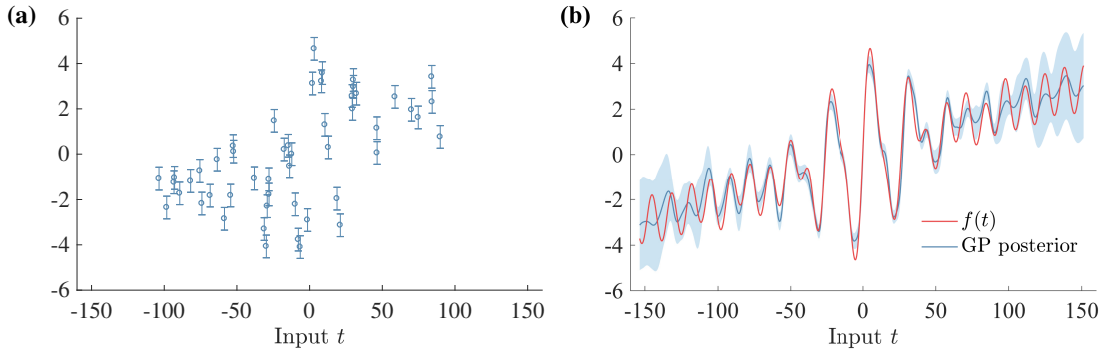


**Figure 2.16:** (a) 50 noisy observations generated using the function defined in equation 2.41. (b) Posterior predictive distribution for a GP conditioned on the noisy observations in panel (a), with the blue line indicating the GP mean function, and the shaded region $\pm\sigma$. The red line shows the actual function $f(t)$ from which the noisy observations were generated.

.

to account for variance due to uncorrelated noise. The observations also suggest some sort of simple, long-term trend, so we might adopt (somewhat arbitrarily) a second-order polynomial mean function.[25]

We place flat priors[26] on all covariance hyperparameters and mean function parameters, and use an MCMC code to sample from the (hyper)parameter posterior distributions. Maximum *a posteriori* hyperparameter values thus obtained suggest an overall period, $P$, of approximately 28 d; an overall (mean-subtracted) amplitude, $h$, of 2.1; a noise variance, $\sigma$, of 0.97; an inverse harmonic complexity, $\lambda_{\rm p}$, of 0.37; and a signal evolution time-scale, $\lambda_{\rm e}$, of about 40 d.

We then use equations and to compute the GP posterior predictive distribution for the function $f$; this is plotted in Fig. 2.16, along with the generative (true) function $f(t)$, as defined in equation 2.41.

We see that the GP posterior mean looks very similar to the generative function, both in regions between observations and when extrapolating beyond $|t| \gtrsim 100$.

[25] A model comparison test would have told us that a simpler, linear mean function would have sufficed.

[26] Of course we could expect better results by using more informative priors; to ensure our model captures some sort of periodicity, we could enforce $P < t_{50} - t_1$ (period shorter than data baseline), and $\lambda_{\rm e} > P$ (function evolution time-scale longer than a single period). The GP covariance amplitude and mean function parameters could also be straightforwardly constrained e.g. by studying the overall scale of variation and zero-offset in $\{y_i\}$.

Of course, the GP model does not correctly reproduce every single bump and wiggle in $f(t)$, although such shortcomings are sensibly reflected in the GP posterior uncertainty. Indeed, it seems remarkable how well our GP model permits inference about $f(t)$, given only a small number of noisy observations, limited prior knowledge, and no assumptions about the parametric form of the complicated function $f(t)$.

### 2.6.8 A few (dis)advantages of GP regression

Before turning in the next chapter to the development and application of a novel, GP-based approach to modelling and mitigating stellar activity signals, it is worth reviewing a few of the potential advantages and disadvantages of GP regression. Some of the advantages have already been touched on in the preceding discussion, and include the following.

1. GPs provide an analytically-tractable framework for formulating physically-motivated prior distributions over functions, and for inferring unknown functions (with error bars) in a data-driven way.
2. GP-based regression reduces the problem of function inference to one of inference about a finite set of covariance hyperparameters. This is especially useful when modelling stochastic and/or otherwise difficult-to-parametrise signals, in which case GP-based modelling represents a principled alternative to treating the signals as white noise, or to adopting *ad hoc* parametric models that may be overly simplistic or unrealistically flexible.
3. GP regression may readily be accommodated within the broader framework fo Bayesian inference; consequently, matters such as uncertainty propagation, avoidance of over-fitting, and model comparison are all handled straightforwardly.
4. GPs allow multiple different types of inputs (time, space, etc.) and outputs to be modelled jointly.
5. GPs facilitate continuous interpolation between, and extrapolation far from, observations.

6. GPs are easily combined with existing techniques: implementing GP regression can be as straightforward as replacing an existing likelihood function with the appropriate GP likelihood (equation 2.38) in a non-linear optimisation or MCMC routine.

Despite their numerous advantages, there are, of course, a few disadvantages to using GPs for regression. Some of these are outlined below.

1. Given $N$ observations, each computation of a GP likelihood involves inverting an $N \times N$ matrix, which via Cholesky factorisation requires $\mathcal{O}(N^3)$ operations (Press, 2007). With this unfavourable scaling behaviour, naïve GP regression quickly becomes unfeasible on a desktop computer when $N$ is much larger than a few thousand data points. Fortunately, a number of sophisticated schemes for making GP regression feasible even on large data sets have already been developed (Lázaro-Gredilla et al., 2010; Chalupka et al., 2012; Ambikasaran et al., 2014); there are even GP-based methods whose computational requirements scale *linearly* with the size of the data set (Reece and Roberts, 2010; Sarkka et al., 2013; Reece et al., 2014; Foreman-Mackey et al., 2017). As of 2017, developing fast methods for GP regression remains an area of intense research focus.
2. GP marginal likelihoods are, in general, *not* convex functions of the kernel hyperparameters, and will instead typically be multimodal hypersurfaces (Tipping et al., 2003). A related point is that there often exist degeneracies between hyperparameters (Gibson et al., 2012). Therefore, great care needs to be taken during the formulation of hyperparameter priors, and during hyperparameter marginalising or posterior sampling. Simple local optimisation schemes in particular are liable to produce misleading results (Snelson, 2008).
3. It usually but not always a straightforward matter to relate covariance hyperparameters to physical quantities of interest (in light e.g. of the aforementioned degeneracies).
4. Despite the enormous flexibility of GPs, the assumption of Gaussian noise is not always justifiable in real-world applications. For more details on workarounds in the case of non-Gaussian noise, refer to Appendix B.

*...man erst nach vielen vergeblichen Versuchen dazu gelangt, und dann meistens auf beschwerlichen künstlichen Wegen, während die einfacheren Methoden lange verborgen bleiben.*

*...one only succeeds after many futile attempts; and even then, it is often by some tedious and artificial process, while the simpler methods may long remain concealed.*

— Carl Friedrich Gauss, in preface to *Mathematische Abhandlungen* (Eisenstein, 1847).

# 3

# A GP framework for modelling stellar activity signals in RV data

## Contents



This chapter forms the core of this thesis, with the next several chapters hinging upon the formalism developed and presented herein. Most of the text in this chapter is based on the paper entitled *A Gaussian process framework for modelling stellar activity signals in radial velocity data* (Rajpaul et al., 2015).

# Abstract

To date, the radial velocity (RV) method has been one of the most productive techniques for detecting and confirming extrasolar planetary candidates. Unfortunately, stellar activity can induce RV variations which can drown out or even mimic planetary signals – and it is notoriously difficult to model and thus mitigate the effects of these activity-induced nuisance signals. This is expected to be a major obstacle to using next-generation spectrographs to detect lower mass planets, planets with longer periods, and planets around more active stars. Enter Gaussian processes (GPs) which, we note, have a number of attractive features that make them very well suited to disentangling stellar activity signals from planetary signals. We present here a GP framework we developed to model RV time series jointly with ancillary activity indicators (e.g. bisector velocity spans, line widths, chromospheric activity indices), allowing the activity component of RV time series to be constrained and disentangled from e.g. planetary components. We discuss the mathematical details of our GP framework, and present results illustrating its encouraging performance on both synthetic and real RV data sets, including the publicly-available Alpha Centauri B data set.


## 3.1 Introduction

Much of the material that originally formed the introduction to this paper appears in expanded form in Section 2.2 of this thesis; to avoid repetition, it has been excised from this chapter. For the sake of continuity, though, we simply mention that the original introduction sketched the role of Doppler spectroscopy in detecting and characterising exoplanets, and noted that signals from stars themselves contaminate RVs, confounding attempts to study exoplanets.

The original introduction emphasised an announcement that caused a particularly significant stir in the exoplanet community: that of a planet around the star $\alpha$ Cen B with an orbital period of $\sim 3.24$ d, and a minimum mass similar to that of the Earth, discovered using the HARPS instrument (Dumusque et al., 2012). If confirmed, this

discovery would have been particularly remarkable not only because the minimum mass of the planet is so low (at the time of writing, the lowest of any planet around a Sun-type star), but also because the host star is the nearest Sun-like star to our own system, and the RV semi-amplitude reportedly induced by the planet (50 cm $s^{-1}$) is the smallest to have been measured to date.

The introduction thus established the broad context for the present work, the focus of which is a GP framework we developed for modelling the activity-induced variations in RV time series, in order to disentangle activity from e.g. planetary contributions

## 3.2 Stellar nuisance signals

This section has also been excised to avoid repeating material that now appears in Section 2.3 of this thesis. For the sake of continuity, we note the corresponding section published in Rajpaul et al. (2015) discussed the main physical phenomena in stars giving rise to RV perturbations, especially those related to magnetic activity. It also outlined existing strategies, and their significant disadvantages, for mitigating stellar activity signals when attempting to study planets.

In the next section we introduce an extension of A12's $FF$' method, already discussed in Section 2.5.7. Our extension uses the same principles, but models the RVs jointly with one or more activity indicators, e.g. $\log R'_{\rm HK}$, BIS, photometry, etc. The various time series are modelled as linear combinations of a single, underlying GP and its time derivative, plus a separate noise term which is specific to each time series.

## 3.3 Proposed formalism

### 3.3.1 Formulation of the physical model

The physical model we propose to use is an extension of the $FF'$ framework introduced in A12, which can be used to estimate RV variations due to stellar activity using photometry. In this framework, the fraction of the visible stellar hemisphere that is covered in spots is represented by $F(t)$. For the simplest case of

a single, small, equatorial spot, $F(t)$ represents the projected area of the spot, and varies as $\cos(2\pi t/P)$, where $P$ is the rotation period. A12 then showed that the RV perturbation due to activity (even for a non-equatorial spot) can be written as:

$$\Delta\mathrm{RV} = V_\mathrm{r}\, F(t)\, \dot{F}(t) + V_\mathrm{c}\, F^2(t), \tag{3.1}$$

where the first term represents the rotational modulation signal, and the second term the suppression of convective blueshift in magnetised regions. This remarkably simple expression arises because the RV of the stellar surface varies as $\sin(2\pi t/P)$, which is proportional to $\dot{F}(t)$, whereas the projection of the convective upwelling along the line of sight varies as $\cos(2\pi t/P)$, which is proportional to $F(t)$. Both terms are multiplied by the projected spot area, which is proportional to $F(t)$. The two constants $V_\mathrm{r}$ and $V_\mathrm{c}$ can be related to physical quantities such as the fractional coverage and contrast of spots and faculae, the stellar equatorial velocity and the convective blueshift velocity, but are treated as free parameters for the present purpose. Equation (3.1) was derived by considering a single active region, but is used to describe the combined effect of all the active regions on the visible hemisphere. Though this is a first-order approximation, A12 carried out a number of tests of this formalism and showed that its performance was similar to that of more sophisticated methods. A caveat, however, is that photometry is insensitive to certain spot distributions, i.e. photometric signals will not always contain sufficient information adequately to predict RV variations. This problem is addressed in this thesis, where we propose to use more than one activity indicator to constrain the activity component in RV variations.

We thus seek similar expressions to estimate RV variations due to stellar activity using activity diagnostics *other than photometry*. We consider the case where we have a chromospheric activity indicator and a line asymmetry measurements available; for the sake of concreteness, we take these to be the $\log R'_\mathrm{HK}$ index, and the inverse slope of the bisector of the cross-correlation function or simply BIS, respectively, as is the case for HARPS data sets.

As a first approximation, it seems reasonable to expect $\log R'_{\rm HK}$ to behave as $F^2(t)$, the convective blueshift suppression term, which is close proxy for the fractional coverage of active regions. On the other hand, like the RVs, the BIS also depends on the velocity of the stellar surface at the location of active regions, so the expression for the BIS should include an additional $F(t)\,\dot{F}(t)$ term:

$$\log R'_{\rm HK} = L_{\rm c}\,F^2(t) \tag{3.2}$$

$$\mathrm{BIS} = B_{\rm r}\,F(t)\,\dot{F}(t) + B_{\rm c}\,F^2(t), \tag{3.3}$$

where $L_{\rm c}$, $B_{\rm c}$ and $B_{\rm c}$ are free parameters. Additionally, each time series will have its own noise term, which is treated as white (this is discussed in more detail in later sections).

Equations 3.2 and 3.3 relate two particular activity-sensitive observables ($\log R'_{\rm HK}$ and BIS) to spot coverage, although analogous relations could be constructed for other activity-sensitive observables. For example, the $S_{\rm HK}$ chromospheric activity index can be related to active region coverage in the same way as the $\log R'_{\rm HK}$ index (Isaacson and Fischer, 2010). Similarly, FWHM data are usually very tightly correlated with (if noisier than) $\log R'_{\rm HK}$ data – especially if there is a tight relation between the size and location of the photospheric spots and that of the chromospheric structures such as faculae – and so the same form of relation could be used.

### 3.3.2 Gaussian process framework

GPs provide a mathematically-tractable and very flexible framework for performing Bayesian inference about functions. They are particularly suitable for the joint modelling of deterministic processes with stochastic processes of unknown functional forms (though with some known properties); they lend themselves very naturally to rigorous error propagation; and they allow us to relax the usually quite restrictive (e.g. linear) relationships which have been assumed in previous attempts at the joint modelling of RV with ancillary times-series (Tuomi et al., 2014).

GP regression enables us to model complex stochastic processes by parametrising the covariance between pairs of data points, rather than writing down an expression

for the data themselves; deterministic components of the model can also be incorporated as the mean of the GP. Mathematically, the prior joint distribution of the outputs $\mathbf{y}$ (observables, in the case of our framework) is taken to be a multivariate Gaussian,

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{y}; \mathbf{m}, \mathbf{K}). \tag{3.4}$$

The mean vector $\mathbf{m}$ is given by

$$\mathbf{m} = m(\mathbf{x}; \boldsymbol{\theta}), \tag{3.5}$$

where $m$ is a mean function of the inputs $\mathbf{x}$ (observation times, for our framework) with parameters $\boldsymbol{\theta}$. The elements of the covariance matrix $\mathbf{K}$ are given by

$$\mathbf{K}_{ij} = k(x_i, x_j; \boldsymbol{\phi}), \tag{3.6}$$

where $k$ is a covariance kernel function with parameters $\boldsymbol{\phi}$. The function $k$ provides the covariance element between any two sample locations or times, $x_i$ and $x_j$.

The elements of $\mathbf{m}$ and $\mathbf{K}$ are, in the strict sense of the term, the parameters of the model; however, the values of these parameters are controlled by a small number of *hyperparameters* $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, and in practice will never be inferred directly.

Gaussian distributions obey many convenient mathematical identities, which allow us trivially to marginalise over unobserved function values – even in the common case where there are an infinite number of such values. This remarkable property enables us to write down a marginal likelihood $\mathcal{L}_{m,k}(\boldsymbol{\theta}, \boldsymbol{\phi})$ for the data, given a GP model:

$$\log\left[\mathcal{L}_{m,k}(\theta, \phi)\right] = -\frac{1}{2}\mathbf{r}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{r} - \frac{1}{2}\log\left(\det\mathbf{K}\right) - \frac{n}{2}\log\left(2\pi\right), \tag{3.7}$$

where $\mathbf{r} = \mathbf{y} - \mathbf{m}$ is the vector of residuals of the data after the mean function has been subtracted, and $n$ is the number of data points. This process is known as GP regression, and is described in more detail in textbooks such as Bishop (2006) and Rasmussen and Williams (2006), as well as in the previous chapter of this thesis. The free *hyperparameters* $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ can then be varied to maximise $\mathcal{L}_{m,k}$; this process

is known as Type-II maximum likelihood, or marginal likelihood maximisation (Gibson et al., 2012). In so doing we refine vague distributions over many, very different functions – the forms of which are controlled by $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ – to more precise distributions which are focused on functions that best explain our observed data.

To compare models with different mean and covariance functions ($m$ and $k$, respectively), we must compute the model evidence for each model, which involves numerical integration of the likelihood $\mathcal{L}_{m,k}$ and priors over hyperparameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. In practice, the computational cost of doing this can be extremely expensive. Furthermore, formulating physically motivated, statistically proper priors for each hyperparameter can be challenging. On the other hand, it is often both theoretically-defensible and computationally-convenient (see Gibson et al., 2012) first to maximise $\mathcal{L}_{m,k}$ with respect to all hyperparameters, next to fix the 'uninteresting' ones (typically those in $\boldsymbol{\phi}$, which are often also degenerate) to their MAP values, and lastly to marginalise fully over the remaining parameters of interest (typically those in $\boldsymbol{\theta}$). This approximation avoids having to perform matrix inversion at every step of marginalisation. In practice, this entails marginalising over planet parameters, for example, but simply optimising the parameters of covariance kernels. We adopt this approximation throughout this chapter, though discard the approximation in later chapters of this thesis.

### 3.3.3 GP model for multiple time series

Again for the sake of concreteness, we consider the case of a HARPS-like data set, where we might wish to model jointly three observables: RV perturbations $\Delta$RV, an activity index $\log R'_{\rm HK}$, and the BIS. Further, to illustrate the way deterministic components can be incorporated into the framework, we assume – as is the case for the $\alpha$ Cen B data set, analysed in Section 3.4.4 – that we wish to model dynamical effects in the RVs (binary motion, a possible planet, etc.).

In this case, we have $N$ observations of one independent variable, time $t$, and three dependent variables: $\Delta$RV, $\log R'_{\rm HK}$, and BIS. In order to treat all three as different manifestations of a single underlying process, we re-arrange the data into

$3N$ observations of a two-dimensional input $\mathbf{x} = [t, l]$, and of the corresponding output $y$, where $y^{(l)}(t) = \Delta\mathrm{RV}$ if $l = 1$, $\log R'_{\mathrm{HK}}$ if $l = 2$, and BIS if $l = 3$.

The baseline mean function is taken to be a second order polynomial, which in this example would suffice to represent the motion of $\alpha$ Cen B around the centre of mass of the $\alpha$ Cen binary system (Dumusque et al., 2012), for the RV data, and a constant (DC offset) for the other time series,

$$m(\mathbf{x}) = m^{(l)}(t), \tag{3.8}$$

$$m^{(1)}(t) = a + bt + ct^2, \tag{3.9}$$

$$m^{(2)}(t) = d, \tag{3.10}$$

$$m^{(3)}(t) = e, \tag{3.11}$$

where $\boldsymbol{\theta} = [a, b, c, d, e]$ are free hyperparameters. One or more Keplerian terms can be added to the mean function to represent planetary signals, but this baseline model has no planet.

A white-noise component (to encapsulate observational noise, and also activity-induced and other stellar signals not captured by our model, arising e.g. from pulsation) is incorporated in the covariance function as follows:

$$k(\mathbf{x}_i, \mathbf{x}_j) = k_{\mathrm{act}}^{(l_i,l_j)}(t_i, t_j) + \sigma_{l_i}^2\, \delta(t_i - t_j) \tag{3.12}$$

where $\sigma_{l_i}$ is the standard deviation of the noise associated with each type of observation, and $\delta(x)$ is the Kronecker delta function.

We next introduce a new latent (unobserved) variable, $G(t) := F^2(t)$, which represents the activity term, and is described by a zero-mean GP with covariance function $\gamma$. Conveniently, we can then re-write equations (3.1), (3.2), and (3.3) as

$$\Delta\mathrm{RV} = V_{\mathrm{c}}\, G(t) + V_{\mathrm{r}}\, \dot{G}(t), \tag{3.13}$$

$$\log R'_{\mathrm{HK}} = L_{\mathrm{c}}\, G(t), \tag{3.14}$$

$$\mathrm{BIS} = B_{\mathrm{c}}\, G(t) + B_{\mathrm{r}}\, \dot{G}(t), \tag{3.15}$$

where a factor 2 has been incorporated into the constants $V_\mathrm{r}$ and $B_\mathrm{r}$. We now make use of the fact that any linear combination of a GP and its derivative[1] is also a GP; following Osborne (2010), the covariance between an observation of $G$ at time $t_i$ and an observation of $\dot{G}$ at time $t_j$ is

$$\gamma^{(G,dG)}(t_i, t_j) = \frac{\partial}{\partial t}\gamma^{(G,G)}(t, t_j)\Big|_{t=t_i}, \tag{3.16}$$

where $\gamma^{(G,G)}(t_i, t_j)$ is used to denote the covariance between (non-derivative) observations of $G$ at times $t_i$ and $t_j$. Similarly, the covariance between two observations of $\dot{G}$ at times $t_i$ and $t_j$ is

$$\gamma^{(dG,dG)}(t_i, t_j) = \frac{\partial}{\partial t'}\frac{\partial}{\partial t}\gamma^{(G,G)}(t, t')\Big|_{t=t_i}\Big|_{t'=t_j}. \tag{3.17}$$

Using the above notation, the individual covariance functions over the various inputs can then be related to $\gamma$ as follows (in the notation below, $k^{(12)}_\mathrm{act} = \mathrm{cov}\,(\Delta\mathrm{RV}, \log R'_\mathrm{HK})$, $k^{(31)}_\mathrm{act} = \mathrm{cov}\,(\mathrm{BIS}, \Delta\mathrm{RV})$, etc.):

$$\begin{aligned}
k^{(11)}_\mathrm{act}(t_i, t_j) &= V_\mathrm{c}^2\,\gamma^{(G,G)}(t_i, t_j) + V_\mathrm{r}^2\,\gamma^{(dG,dG)}(t_i, t_j) \\
&\quad + V_\mathrm{c}\,V_\mathrm{r}\left(\gamma^{(G,dG)}(t_i, t_j) + \gamma^{(dG,G)}(t_i, t_j)\right), && (3.18)\\
k^{(22)}_\mathrm{act}(t_i, t_j) &= L_\mathrm{c}^2\,\gamma^{(G,G)}(t_i, t_j), && (3.19)\\
k^{(33)}_\mathrm{act}(t_i, t_j) &= B_\mathrm{c}^2\,\gamma^{(G,G)}(t_i, t_j) + B_\mathrm{r}^2\,\gamma^{(dG,dG)}(t_i, t_j) \\
&\quad + B_\mathrm{c}\,B_\mathrm{r}\left(\gamma^{(G,dG)}(t_i, t_j) + \gamma^{(dG,G)}(t_i, t_j)\right), && (3.20)\\
k^{(12)}_\mathrm{act}(t_i, t_j) &= V_\mathrm{c}\,L_\mathrm{c}\,\gamma^{(G,G)}(t_i, t_j) + V_\mathrm{r}\,L_\mathrm{c}\,\gamma^{(dG,G)}(t_i, t_j), && (3.21)\\
k^{(13)}_\mathrm{act}(t_i, t_j) &= V_\mathrm{c}\,B_\mathrm{c}\,\gamma^{(G,G)}(t_i, t_j) + V_\mathrm{r}\,B_\mathrm{r}\,\gamma^{(dG,dG)}(t_i, t_j) \\
&\quad + V_\mathrm{c}\,B_\mathrm{r}\,\gamma^{(G,dG)}(t_i, t_j) + V_\mathrm{r}\,B_\mathrm{c}\,\gamma^{(dG,G)}(t_i, t_j), && (3.22)\\
k^{(23)}_\mathrm{act}(t_i, t_j) &= L_\mathrm{c}\,B_\mathrm{c}\,\gamma^{(G,G)}(t_i, t_j) + L_\mathrm{c}B_\mathrm{r}\,\gamma^{(G,dG)}(t_i, t_j). && (3.23)
\end{aligned}$$

The final covariance matrix over the new input space is formed as the following block matrix:

$$\mathbf{K}_\mathrm{act} = \begin{pmatrix} \mathbf{k}^{(11)}_\mathrm{act} & \mathbf{k}^{(12)}_\mathrm{act} & \mathbf{k}^{(13)}_\mathrm{act} \\ \mathbf{k}^{(21)}_\mathrm{act} & \mathbf{k}^{(22)}_\mathrm{act} & \mathbf{k}^{(23)}_\mathrm{act} \\ \mathbf{k}^{(31)}_\mathrm{act} & \mathbf{k}^{(32)}_\mathrm{act} & \mathbf{k}^{(33)}_\mathrm{act} \end{pmatrix}. \tag{3.24}$$

[1] In fact, *any* affine operator (including linear, derivative, and integral operators) applied to a GP yields another GP. This enables our framework to exploit the well-known analytic identities for conditioning and marginalising Gaussian distributions, and makes the problem computationally tractable.

Provided $\gamma$ is a valid covariance function, $\mathbf{K}_{\rm act}$ is guaranteed to be a valid covariance matrix: in particular, it will be symmetric and positive semi-definite. In practice, then, it is not necessary to compute the blocks $\mathbf{k}_{\rm act}^{(21)}$, $\mathbf{k}_{\rm act}^{(31)}$, and $\mathbf{k}_{\rm act}^{(32)}$ explicitly.

Thus, equipped with input vector $\mathbf{x}$, output vector $\mathbf{y}$, and covariance matrix $\mathbf{K}_{\rm act}$, we are in a position to use the standard machinery of GP inference to model all three time series jointly as manifestations of a single underlying GP. However, it remains for us to choose the form of the covariance function, $\gamma$, defining this underlying stochastic process.

### 3.3.4 Choice of latent covariance function

We now seek the simplest form for the covariance kernel function, $\gamma$, that satisfies our beliefs and ignorance about the activity process we are modelling.

#### Squared-exponential covariance

For many time series, we expect the informativeness of past observations in explaining current data to be a function of how long ago the past observations were made (for example, we might expect two observations made minutes apart to have similar values, but would not necessarily expect any connection between observations made months apart); if this is the case, we can restrict our consideration to *stationary* covariance functions which are dependent solely on $|t - t'|$. Such covariance functions can be represented as the Fourier transform of a normalised probability density function (via Bochner's Theorem: Rasmussen and Williams, 2006), which can be interpreted as the spectral density of the process.

The most widely used covariance function of this class – indeed, probably the most widely-used covariance function in GP inference – is the squared-exponential or Gaussian kernel function, given by

$$\gamma^{(G,G)}(t, t') = h^2 \exp\left[-\frac{(t - t')^2}{2\lambda^2}\right]. \tag{3.25}$$

The hyperparameter $h$ governs the output scale (gain/amplitude) of our function, and $\lambda$ controls the time-scale. In our problem the amplitude parameter $h$ is not

strictly necessary since $G(t)$ is never directly observed, and $h$ is absorbed by the model parameters[2] $V_\mathrm{c}$, $V_\mathrm{r}$, $L_\mathrm{c}$, $B_\mathrm{c}$, and $B_\mathrm{r}$, as in equations 3.1–3.3.

The popularity of this covariance function stems not only from its simplicity, but also the fact that it is infinitely differentiable, meaning that a GP with this covariance function will generate functions with no sharp discontinuities.

The squared-exponential covariance function is usually flexible enough to model a wide range of smoothly-varying stochastic processes. However, as discussed in Section 3.1, we expect RV data and ancillary time series to show evidence for activity-induced variations on up to three distinct time-scales: years (magnetic activity cycle), weeks/months (rotation), and days (oscillation, granulation). Thus it makes sense to generalise the covariance function in equation (3.25) to explicitly allow for multiple evolutionary time-scales:

$$\gamma_\mathrm{SE}^{(G,G)}(t,t') = \sum_{i=1}^{N} h_i^2 \exp\left[-\frac{(t-t')^2}{2\lambda_i^2}\right]; \tag{3.26}$$

here, the $h_i$'s control the relative amplitude of the $N \geq 1$ components, each with evolutionary time-scale $\lambda_i$.

We initially experimented (using both real and synthetic data; see Section 3.4) with $N = 3$ separate additive terms, to describe possible variations on the short, medium, and long time-scales discussed in Section 3.2. However, we rapidly found that the longer time-scale term was invariably unnecessary: any reasonable choice for the medium time-scale term had to be fairly flexible, since rotational signals generally evolve from year to year, and so the longer term variations can be captured by the medium time-scale term. On the other hand, we found a separate short time-scale term was generally required: without it, the residuals in all time series considered often exhibited obvious correlations.

Appendix A gives expressions for the covariance between observations of $G$ and its derivative $\dot{G}$, and between two observations of $\dot{G}$, for the case when $G$ is defined by a squared-exponential covariance function, as in equation (3.26).

[2]Note that because each of these five parameters only appears in the final covariance matrix either squared or multiplied by one of the remaining four parameters, it is convenient to fix the sign of one, say $V_\mathrm{c} > 0$, to remove a degeneracy with respect to the signs of these parameters.

### Quasi-periodic covariance

While the squared-exponential covariance function given in the previous section is both flexible and an obvious starting point when not given much information about the stochastic process we are modelling, we can do better. In particular, we have a physical reason to expect a degree of periodicity in the activity signals, since they are modulated by the periodic rotation of the star. We therefore consider the following QP covariance function for the latent process $G(t)$ – as previously considered by A12 to model observed variations in the Sun's total irradiance, and by Haywood et al. (2014) to model correlated noise in RV residuals following an application of the $FF'$ method – formed by multiplying[3] a stationary kernel with a periodic one:

$$\gamma_{\mathrm{SE}}^{(G,G)}(t,t') = \exp\left\{-\frac{\sin^2\left[\pi(t-t')/P\right]}{2\lambda_{\mathrm{p}}^2} - \frac{(t-t')^2}{2\lambda_{\mathrm{e}}^2}\right\}, \tag{3.27}$$

where $P$ and $\lambda_{\mathrm{p}}$ correspond to the period and length scale of the periodic component of the variations, and $\lambda_{\mathrm{e}}$ is an evolutionary time-scale. While $\lambda_{\mathrm{e}}$ has units of time, $\lambda_{\mathrm{p}}$ is dimensionless, as it is relative to $P$. Once again there is no need to introduce an amplitude parameter, as this will be controlled for various time series by the model parameters $V_{\mathrm{c}}$, $V_{\mathrm{r}}$, $L_{\mathrm{c}}$, $B_{\mathrm{c}}$, and $B_{\mathrm{r}}$, as in equations 3.1–3.3.

Examples of functions drawn from a GP with this quasi-periodic covariance function are given in Fig. 3.1.

A word is in order about the interpretation of the hyperparameters. While they could be related to spot persistence lifetimes, complexities of spot configurations, autocorrelation function decay time-scales, etc., such direct interpretations are rarely straightforward or well-grounded theoretically, especially because of degeneracies between the hyperparameters. In the case of modelling differential rotation and/or spot evolution, for example, the hyperparameter $P$ might become almost irrelevant, with a large value of $\lambda_P$ and a very small value of $\lambda_e$ allowing for complex, rapidly evolving waveforms without any obvious characteristic periodicity. In future, an in-depth study of how the values of these hyperparameters translate into physically

[3]As described in Rasmussen and Williams (2006), a valid covariance function under any arbitrary (smooth) map remains a valid covariance function. Valid covariance functions can also be constructed by adding or multiplying simpler covariance functions.

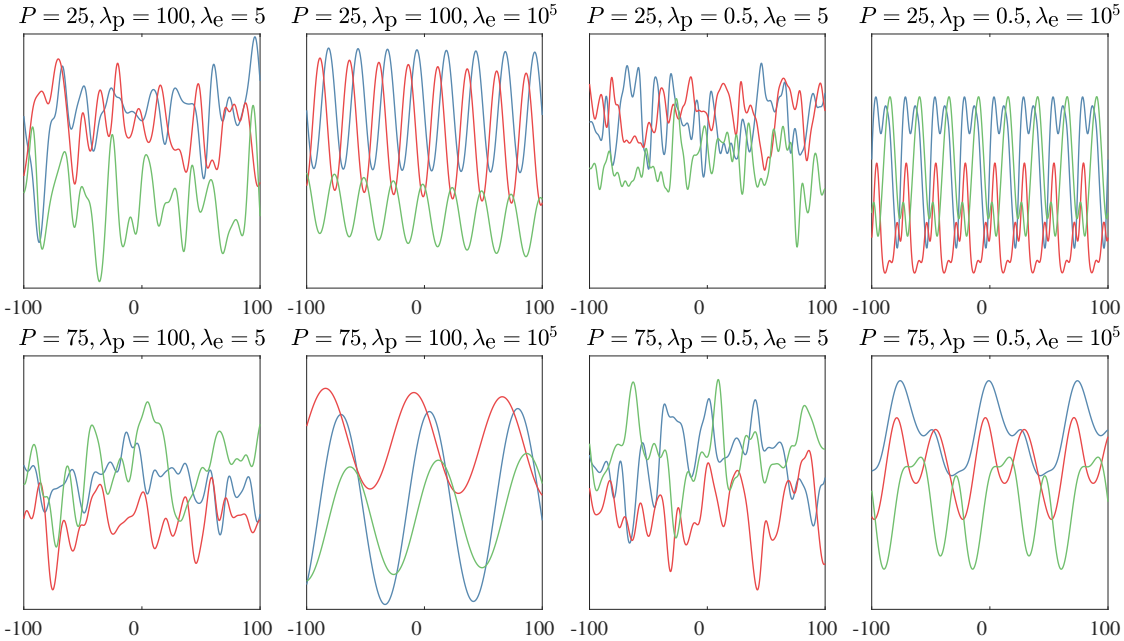


**Figure 3.1:** Examples of functions drawn from a GP with a quasi-periodic covariance function, as defined in equation 3.27, in each panel with different hyperparameter values (three function draws per panel). The hyperparameter $P$ sets a characteristic period for the function's variations, $\lambda_P$ controls the complexity of variations or harmonic content within one period, and $\lambda_e$ controls the time-scale over which the (quasi-)periodic component evolves; variations become strictly periodic as $\lambda_e \to \infty$. In all cases, the function amplitudes and offsets are arbitrary, with constant mean function.

interpretable properties, for example using the large sample of Kepler light curves considered by Aigrain et al. (2015) – which in turn highlights the difficulty of distinguishing between differential rotation and spot evolution – might be useful.[4]

In any event, the QP covariance kernel function defined in equation 3.27 is physically-motivated (albeit not always easy to interpret straightforwardly) and has only three hyperparameters; moreover, in tests with both real and simulated data, we found it to produce as good or better results than a squared-exponential covariance function with a similar number of hyperparameters. Therefore in all subsequent discussion the QP covariance function will be assumed by default. (We do however discuss the squared-exponential covariance function because it is conceivable that there will be cases, though not considered here, where it might

[4]Section 7.1.2 of this thesis provides a more detailed discussion on how the QP covariance hyperparameters should be interpreted and constrained.

be a more appropriate choice than the QP covariance; alternatively, it might be appropriate to generalise the function to include more than one QP component.)

Appendix A.2 gives expressions for the covariance between observations of $G$ and its derivative $\dot{G}$, and between two observations of $\dot{G}$, for the case when $G$ is defined by a squared-exponential covariance function as in equation (3.27).

**Other possible covariance functions**

There exists a wide variety of functions that can be used to specify the correlation between pairs of inputs, and thence to generate a covariance matrix over a set of observations and predicants. These functions can further be combined and modified in a multitude of ways, giving one great flexibility in how functions are modelled. In the preceding subsections, we considered two simple such functions that were deemed adequate for the cases considered in this thesis chapter; however, it would certainly be possible to replace these and use, within our same framework, other covariance functions (e.g. Rasmussen and Williams, 2006).

For example, as an alternative to using a sum of squared-exponential kernels, one might use a rational-quadratic kernel, given by:

$$\gamma_{rq}^{(G,G)}(t,t') = h^2\left(1+\frac{(t-t')^2}{\alpha\lambda^2}\right)^{-\alpha}; \tag{3.28}$$

it may be shown that this is equivalent to a scale mixture of squared-exponential kernels with different length scales, with the latter being distributed according to a Beta distribution with parameters $\alpha$ and $\lambda^{-2}$.

On the other hand, if one has reason to expect that the latent process being modelled is not as smooth as functions drawn from a GP with squared-exponential kernel, one could consider the Matérn class of covariance functions; see equation 2.30. Though a Matérn covariance function with $\nu < \infty$ might be deemed more physically realistic than a squared-exponential one, its use does of course come with the penalty of some additional mathematical and computational complexity. For the purposes of this work, then, we did not carry out any tests with these more sophisticated covariance kernels; their use might, however, feature in future work.[5]

[5]Initial tests with the rational-quadratic kernel function led to very similar results to those

### 3.3.5 Summary and appraisal of the framework

The key aspects of the above framework are sketched schematically in Fig. 3.2. Though it may appear quite mathematically complex, especially to those unfamiliar with GP regression, its underlying ideas are quite simple. To summarise:

1. we assume that the underlying stochastic process giving rise to activity signals in observables (RVs and ancillary time series) can be described by a GP, with suitably-chosen covariance function;[6]
2. we can then use physically-motivated though semi-empirical models to build analytical links between this unobserved GP and available observables;
3. finally, with the addition of noise and deterministic components (e.g. dynamical effects for the RVs), all observables can be modelled jointly as GPs, with the ancillary time series serving to constrain any activity component of the RVs.

Our framework offers a number of advantages over existing approaches. First, modelling all available activity-sensitive time series (line widths, asymmetries, chromospheric activity indices etc.) jointly with RVs should allow tighter constraints to be placed on activity signals in RVs, compared to exploiting only simple correlations between RVs and (typically) one of these other time series. Secondly, our general framework is flexible in that it uses GP draws (and derivatives thereof) as basis functions to model available time series, rather than e.g. sinusoids or other simple parametric models, the inappropriate use of which could lead to the inadvertent introduction of correlated signals into model residuals. In the case our GP framework, the properties of basis functions are informed in a more principled way by the observables. Thirdly, our framework facilitates smooth interpolation between observables, as well as extrapolation to future times (prediction). Lastly, being based on GPs, the entire framework can be accommodated very naturally

when using the squared-exponential kernel. A rigorous approach to choosing between models with different kernel functions would require computation of Bayes factors for all models considered, an extension to this work we defer to future consideration.

[6]Given that a GP can in principle approximate arbitrary unknown functions, and that we can encode whatever limited knowledge we *do* have of a stochastic activity process into a GP's covariance structure, this is not a restrictive starting point.

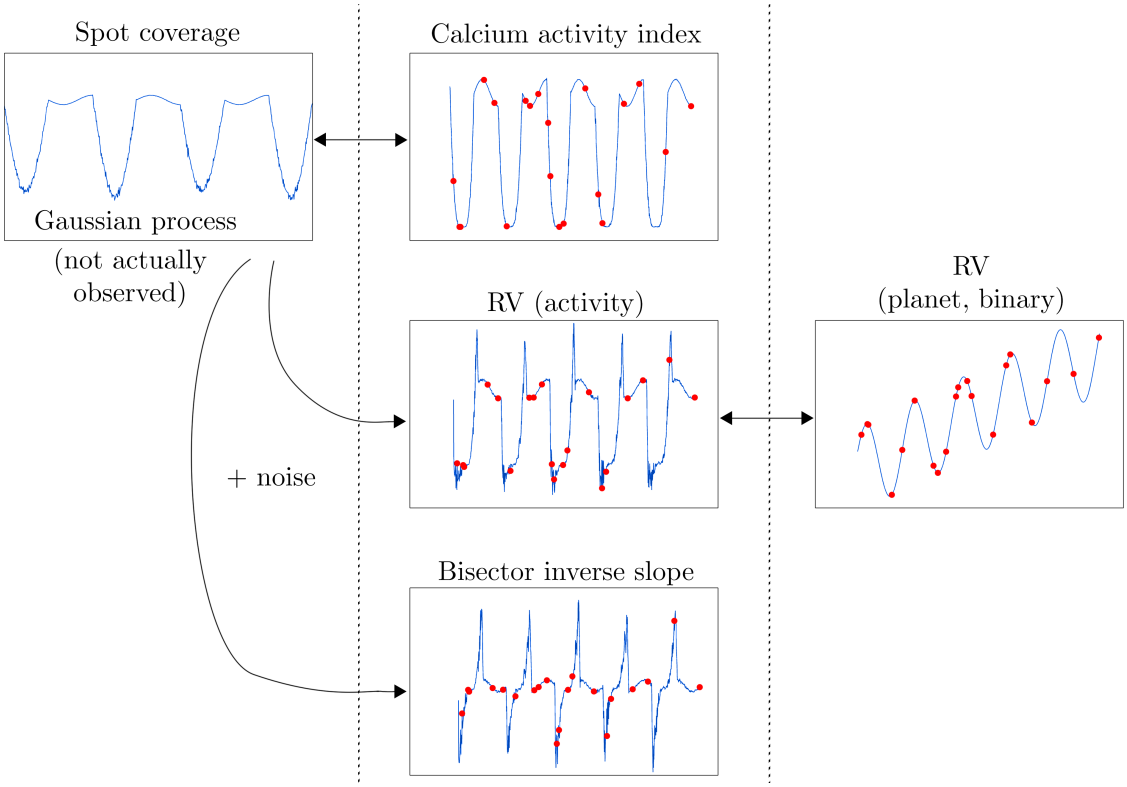


**Figure 3.2:** Simplified, schematic sketch of our GP-based scheme for the joint modelling of an RV time series with ancillary activity diagnostics. The unobserved, stochastic process giving rise to activity-related RV variations (left panel) is assumed to be describable by a GP. The way in which this propagates into observable time series (middle panels) is computed analytically, using a simple physical model, and so all observed time series can be modelled jointly. A noise model and deterministic physical component (right panel, e.g. Keplerian model to describe exoplanet-induced RV variations) completes the basic framework.

within the broader framework of Bayesian inference, allowing uncertainties to be handled in a principled way: overly-complex models are automatically penalised, nuisance parameters can be marginalised over, and rigorous model comparisons can be performed (Rasmussen and Williams, 2006).

A possible disadvantage of our framework is that, in its current form, it can only use the latent 'activity' GP and linear transformations thereof (including derivatives) to generate the activity signals in observables; this is sufficient for linking e.g. RVs and photometry via the $FF'$ method, but the latter is only a first-order model. Allowing for more general relationships between the latent GP and the observables might be possible, although it is not immediately clear that

doing so would result in a computationally or mathematically tractable framework. Another potential disadvantage is that 'vanilla' GP regression is not well-suited to modelling large data sets ($N \gg 1000$ observations): GP regression hinges on matrix inversion, with naïve inversion algorithms scaling as $N^3$. Fortunately there do exist a number of approximation techniques which hold great promise for allowing efficient GP regression on larger data sets (Lázaro-Gredilla et al., 2010; Chalupka et al., 2012; Ambikasaran et al., 2014; Foreman-Mackey et al., 2017). In the case of time series analysis, there exist very efficient $\mathcal{O}(N)$ approaches for GP regression based on approximate re-interpretation as filters (Reece and Roberts, 2010; Sarkka et al., 2013; Reece et al., 2014).

It should be noted that this is not the first instance of GPs being used to model temporally-correlated stellar nuisance signals: Hou (2014) presented a framework for using GPs to model stochastic stellar oscillations in RV data, while Barclay et al. (2014) used GPs to model photometric granulation signals in a Kepler-91 data set. The only published instances (to our knowledge) of GPs being used specifically to model stellar *activity* signals are the work of Haywood et al. (2014), who used a GP to mitigate the effects of active regions without photometric counterparts in the CoRoT-7 system, and the work of Grunblatt et al. (2015), who used a GP to model activity in Kepler-78 data sets. In both cases, GPs with QP kernels were first trained on photometric data, after which a GP model with the same covariance properties was used to model available RV data. The framework we have presented may be seen as a more general and flexible version of the aforesaid approach, in the sense that our framework allows *all* available activity-sensitive time series to be modelled simultaneously (rather than iteratively), and does not necessarily require simultaneous RV and photometry data. When both RV and photometric data are available, the $FF'$ formalism is built directly into the modelling; unlike the $FF'$ method, however, our framework would not require all active regions to have photometric signatures in order to constrain their RV contributions, provided at least one other activity-sensitive time series (e.g. BIS) were available.

## 3.4 Tests and applications

### 3.4.1 Simulated data using naïve physical model

As a first test of our GP framework, we used a number of different functions (e.g. polynomials; sinusoids; wavelets; function draws from GPs; etc.) to generate a range of synthetic time series to represent possible latent, activity-driving processes; we included periodicities and correlations on many different time-scales, to test our model's ability to deal simultaneously with the three different time-scales for variations discussed in Section 3.2. These time series were densely-sampled and noise-free. We then used the relations in equations (3.13), (3.14), and (3.15) to simulate $\Delta$RV, $\log R'_{\mathrm{HK}}$ and BIS 'observables,' computing derivatives numerically using finite-difference approximation where applicable. In other words, we generated the observables we would obtain for arbitrary activity processes, assuming the physical model we used to link the activity process to observables corresponded exactly to reality.

Using a number of different covariance functions (including those discussed in Section 3.3.4), our general findings were the following.

1. By performing unconstrained optimisation of the free parameters in model likelihood functions $\mathcal{L}_{m,k}(\boldsymbol{\theta}, \boldsymbol{\phi})$, all observed time series could be modelled very accurately, with fits that appeared 'exact' upon visual inspection (though in fact with small residuals consistent with numerical error due to finite precision arithmetic, matrix inversion, derivative approximations, etc.).[7]
2. In all such cases, the fitted model's latent process (GP) was also correspondingly close to the latent process used to generate the observables. In cases where the latent process used to generate the observables was itself a GP, the hyperparameters of this latent process could be recovered accurately.

[7]Multi-modality did not seem to pose a significant problem with the likelihood functions in question: a single run of a simple Nelder-Mead ('downhill simplex') algorithm often sufficed to locate globally-optimal solutions, though in a few cases the algorithm did benefit from multiple starting locations.

3. Imposing constraints on the hyperparameters of chosen mean and covariance functions led to the expected changes in fits. For example, constraining the evolutionary time-scale $\lambda_e$ (for the QP covariance kernel) to be longer than a certain time prevented physically unrealistic, short time-scale 'contortions' in the fits. Similarly, forcing a finite period parameter $P$ (again in the QP covariance kernel), with long evolutionary time-scale $\lambda_{\rm e}$, led to poor fits when the simulated observables contained no actual periodicities.
4. When adding white Gaussian noise to the simulated observables, the quality of fits remained good – as measured by a reduced chi-square statistic – provided that the estimated amplitudes of the added noise were reasonable.

In sum, all aspects of the modelling worked as expected; we established that the analytical aspects of (if not the physical assumptions underlying) our framework were watertight, and that when this framework was implemented numerically, we could indeed jointly model multiple time series generated by an unobserved stochastic process and its derivative.

### 3.4.2 Simulated data from SOAP 2.0

SOAP (Spot Oscillation and Planet) 2.0 is a software package developed by Dumusque et al. (2014a) that simulates the effect of stellar spots and plages on RV and photometry; it extends an older code, SOAP, developed by Boisse et al. (2012). The tool is available online at `http://www.astro.up.pt/soap2`.

In brief, the original SOAP code works by dividing up a simulated stellar disc into tens of thousands of resolution elements, with a CCF (representing the typical spectral line) for each element then being modelled by a Gaussian. Each CCF is Doppler-shifted according to the projected rotational velocity, and weighted by a linear limb-darkening law. In this way, the non-spotted emission of the visible disc may be computed in photometry and in RV. SOAP then calculates the positions of rotating surface inhomogeneities, which are defined by their latitudes, longitudes, and sizes; the weighted CCF and flux for cells inside the inhomogeneities are added to (in the case of bright spots or plages) or removed from (in the case of dark spots)

those of the non-spotted star. Finally, SOAP delivers the integrated spectral line, the flux, the RV, and the BIS as a function of stellar rotational phase.

SOAP 2.0 extends SOAP by considering, additionally, the inhibition of convective blueshift inside active regions, the limb brightening effect of plages, a quadratic limb darkening law, a realistic active region contrast, and the resolution of the spectrograph used for the (simulated) observations.

A number of stellar, spot, and spectral parameters can be provided as inputs to the SOAP 2.0 code, and the code allows up to four active regions (each specified by a size, brightness, latitude and longitude) to be included in one simulation.

The SOAP 2.0 code does incorporate simplifications in its modelling: for example, the observed spectrum of a plage is considered to be the same as that of a spot, and some of the stellar physics included in the modelling is informed only by Solar photosphere and spot spectra, which are assumed to be representative of all quiet photosphere regions and active regions. It also does not include long-term magnetic activity cycles in its simulations. Its output nevertheless provides a far more realistic set of test cases than those considered in Section 3.4.1 – especially because it does not impose any *ad hoc* constraints on the functional relationship between photometric, RV and BIS time series. Moreover, Dumusque et al. (2014a) showed that SOAP 2.0 manages to reproduce the activity-induced variation on the early-K dwarfs HD 189733 and $\alpha$ Cen B; modelling the activity-induced variation on the latter star with our new GP framework is considered in Section 3.4.4.

Given the wide range of possible outputs we could generate from SOAP 2.0 (and then modify in various ways), we adopted the following scheme in order to make a systematic study of how our GP modelling framework could be applied to realistic data.

1. Generate noise-free, densely-sampled outputs for a number of different spot configurations, and perform fits to these outputs (modelling photometric flux in place of $\log R'_{\rm HK}$, i.e. using the $FF'$ model; see Section 3.3.1). Based on these results, select a few qualitatively different configurations – including the configuration for which the poorest fits are obtained – for further study.

2. Study the residuals from the fits in the selected noise-free cases, and examine how much additive white noise – i.e. the $\sigma_{l_i}$ terms in equation (3.12) – should be allowed in the various time series to ensure that activity is accurately constrained in RVs.
3. Repeat the fitting as above (now allowing for additive white noise in the GP model), but this time modifying the SOAP code's output to include observational noise, realistic time sampling, etc.
4. Repeat the fitting as above, but with the injection of a Keplerian signal into the RVs, and study whether the Keplerian signal can be disentangled from the activity signal.

It should be emphasised that although this resembles a principled approach to studying our modelling framework using the SOAP 2.0 tool, we did not attempt to make a comprehensive study of (amongst other things) the inter-related effects of different levels of observational noise, different time samplings, different Keplerian amplitudes, periods or phases, etc. Apart from being beyond the scope of this work, such an all-encompassing study would likely have been misguided, given that both our model and the SOAP 2.0 tool incorporates simplifying assumptions which are not easily quantified. As such, discrepancies between SOAP output and our GP models would not necessarily represent shortcomings of our model; conversely, good agreement between SOAP output and our model would not necessarily imply equally good agreement could be expected with real observational data. Rather, the purpose of these tests was to ascertain whether our framework would work at all with fairly realistic simulated data, to understand how the $\sigma_{l_i}$ terms in our model should be constrained, and to get a conservative though quantifiable estimate of the possible shortcomings of our model.

We found that by allowing proportionately more additive noise (the $\sigma_{l_i}$ terms in equation 3.12) in the ancillary time series than the RV time series – in effect, attaching less weight in the fitting to these time series, though still using them to constrain the latent process' hyperparameters – we were generally able to obtain RV residuals that resembled white Gaussian noise. In other words, we were only

interested in modelling activity-induced variations in $\Delta$RV accurately, and we could achieve this goal provided we 'loosened' the assumed relationships between $\Delta$RV and ancillary, activity-sensitive time series (see Section 3.3.3). When we did not allow for any white-noise in any of the time series, or set equal upper bounds on the white noise in all time series (effectively giving equal weight to all time series), we generally ended up with better fits to the ancillary time series, poorer fits for $\Delta$RV, and correlated residuals in all time series.

In particular, we deduced the following approximate upper bounds to allow on the $\sigma_{l_i}$ terms, if we wanted to ensure white Gaussian RV residuals:

- $\sigma_{\Delta\mathrm{RV}}$: up to 5% of rms($\Delta$RV);
- $\sigma_{\mathrm{flux}}$: up to 10% of rms(flux);
- $\sigma_{\mathrm{BIS}}$: up to 20% of rms(BIS).

Constraining the additive white noise in terms of the root mean square (rms) variability of a given time series reflects the fact that there are certain features in these time series that our model cannot capture accurately (e.g. an inflection point in the flux time series, when none exists in $\Delta$RV), regardless of the details of the scaling, offsets, etc. of the time series in question. These constraints are simply upper bounds on $\sigma_{l_i}$, and serve to quantify the extent to which our framework fails to model all three time series simultaneously; in practice, the values of $\sigma_{l_i}$ will usually be smaller than these upper bounds, and will be determined by optimisation or marginalisation, as for any of the other free parameters in our framework.

To illustrate our findings, we consider the two representative configurations of a single spot (spot radius 10% of the stellar radius) at latitude $\phi = 45°$, and of four equally-sized spots (each with radius 5% of the stellar radius) at $\phi = 0°, 30°, 45°, 60°$, equally spaced in longitude. These configurations are representative of cases that were particularly easy and difficult, respectively, to model with our GP framework. In both cases, the observed star in the simulation had Solar parameters (5778 K effective temperature, 5115 K spot temperature, 25.05 d rotation period, Solar limb darkening coefficients, etc., as per Dumusque et al., 2014a), and the simulated

instrument resolution (115,000) corresponded to a HARPS-like instrument. The observations we simulated were noise-free, and spanned three rotation periods (approximately 75 d), with uniformly randomised temporal sampling.

We used a QP covariance kernel, and placed non-informative priors on all free model parameters. Uniform priors were used for parameters with known scales (e.g. $P$, expected to correspond to a stellar rotation period; or $V_\mathrm{r}$ and $V_\mathrm{c}$, which set the amplitude for RV variations), while scale-invariant Jeffreys priors were used for all other parameters (e.g. $\lambda_\mathrm{P}$, $\lambda_\mathrm{e}$). We then studied maximum *a posteriori* (MAP) fits of our GP model to the simulated data. Typical computation time (on a modern laptop) was of the order of a few minutes to obtain a MAP solution, using a single run of a Nelder-Mead (downhill-simplex) algorithm. We established (probable) global optimality by making sure downhill-simplex runs from different starting points converged to the same optimum.

Fig. 3.3 shows results for the 'easy' (one-spot) configuration, while Fig. 3.4 shows results for the 'difficult' (four-spot) configuration. In both cases, the $\Delta$RV time series could be modelled extremely accurately (rms of residuals of the order of 0.01 m s$^{-1}$ in both cases). The ancillary time series were, necessarily, modelled less accurately – especially in the case of the four-spot configuration, where the equatorial spot contributed the most problematic features to the ancillary time series. The magnitude of the $\sigma_{l_i}$ terms (additive white noise) used in the modelling is indicated by the size of the $\pm\sigma$ posterior uncertainties plotted in these figures.

We next considered using more realistic time sampling and adding noise to the simulated observations. To do so we used temporal sampling taken directly from the publicly-available $\alpha$ Cen B data set from Dumusque et al. (2012), which featured nearly four years' worth of data, with gaps of several months between observing seasons. We also added white Gaussian noise to the SOAP 2.0 output: for each time series, the noise was scaled to be the same fraction of the rms variation of the time series itself as was the corresponding $\alpha$ Cen B noise estimate.

Again we found that we could model $\Delta$RV very accurately: for both spot configurations, the rms of $\Delta$RV residuals was marginally smaller than the rms of the

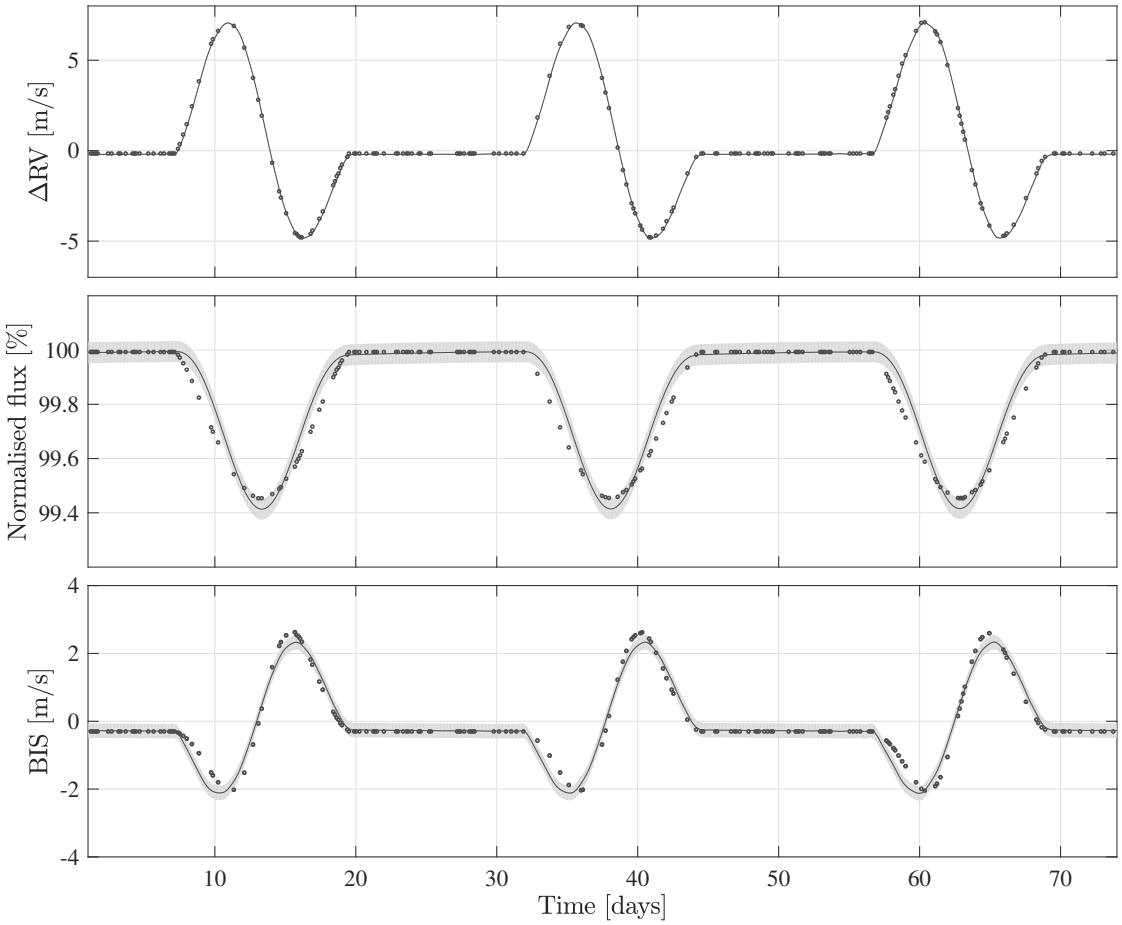


**Figure 3.3:** GP model MAP fit to noise-free SOAP 2.0 data, based on a simulation of a single rotating spot at latitude $\phi = 45°$, and radius 10% of the stellar radius. The dots indicate the simulated observations; the solid lines are model posterior means, and the shaded regions denote $\pm\sigma$ posterior uncertainty. The posterior uncertainty in each time series is directly related to the amount of additive noise ($\sigma_{l_i}$ terms) favoured in the MAP model.

observational noise, indicating that our fits were limited only by our (simulated) noise floor. Figs. 3.5 and 3.6 show results for the 'easy' and 'difficult' spot configurations, as before. Despite the gaps in the observational coverage, the fitted models accurately recovered the structure of the periodic variations in all time series (with the exception of a few outlier points, reflecting our model's inability to reproduce very sharp changes in one time series when another changes less sharply, since all time series are modelled with combinations of a single function and its first derivative). It is worth emphasising that when the data to be modelled include observational noise, the $\sigma_{l_i}$ terms in our model serve to account for both this noise *and* any features

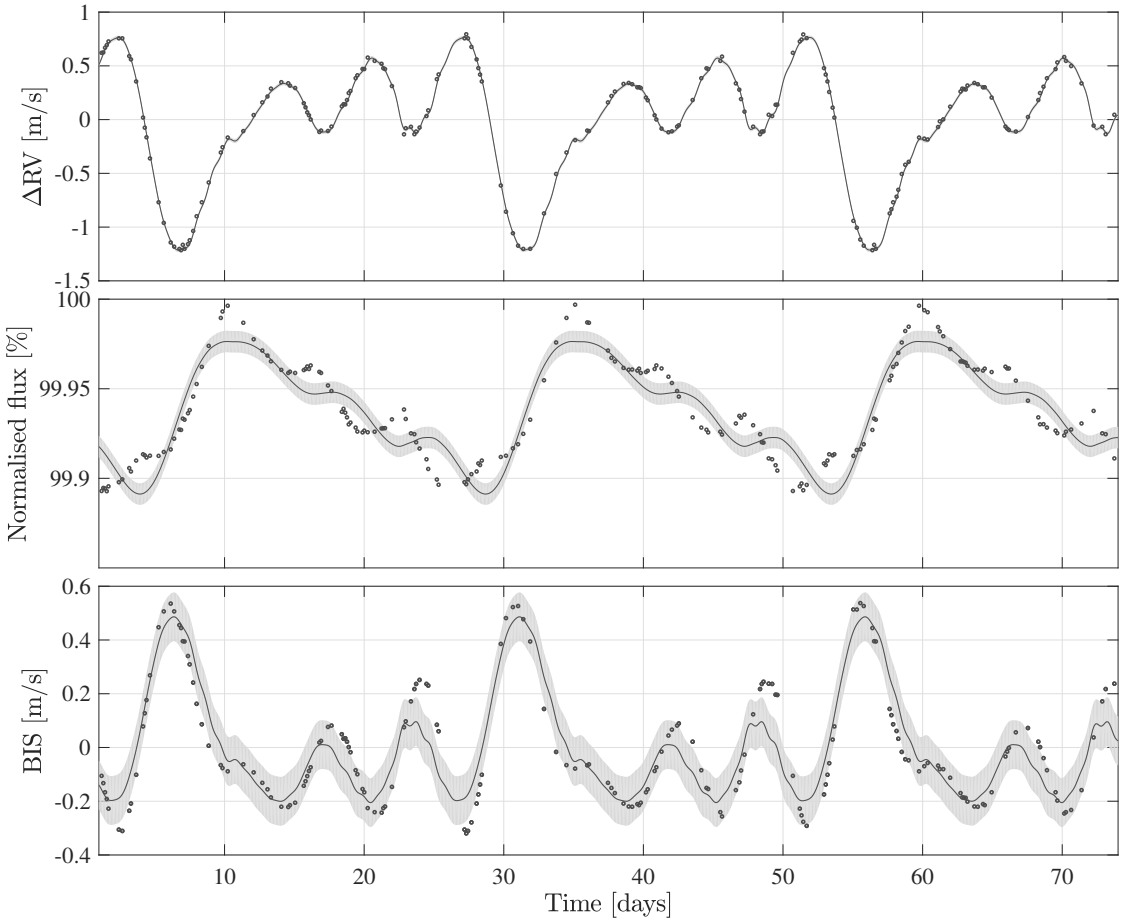


**Figure 3.4:** GP model MAP fit to noise-free SOAP 2.0 data, based on a simulation of four rotating spots (each with radius 5% of the stellar radius) at latitudes $\phi = 0°, 30°, 45°, 60°$, equally spaced in longitude. The dots indicate the simulated observations; the solid lines are model posterior means, and the shaded regions denote $\pm\sigma$ posterior uncertainty. The posterior uncertainty in each time series is directly related to the amount of additive noise ($\sigma_{l_i}$ terms) favoured in the MAP model.

in the ancillary time series that our model cannot accurately reproduce. We could artificially split the white-noise terms in the covariance matrix into 'observational' and 'model imperfection' components, although this would not affect in any way the practical implementation of our model.

Finally, having ascertained that our framework allowed activity-induced RV variations to be modelled accurately in isolation, we turned our attention to planet injection tests. We wanted to answer the question: Could we use our framework accurately to *disentangle* activity-induced RV variations from other injected (non-activity) signals, e.g. planetary signals? Our general finding was that, as expected,

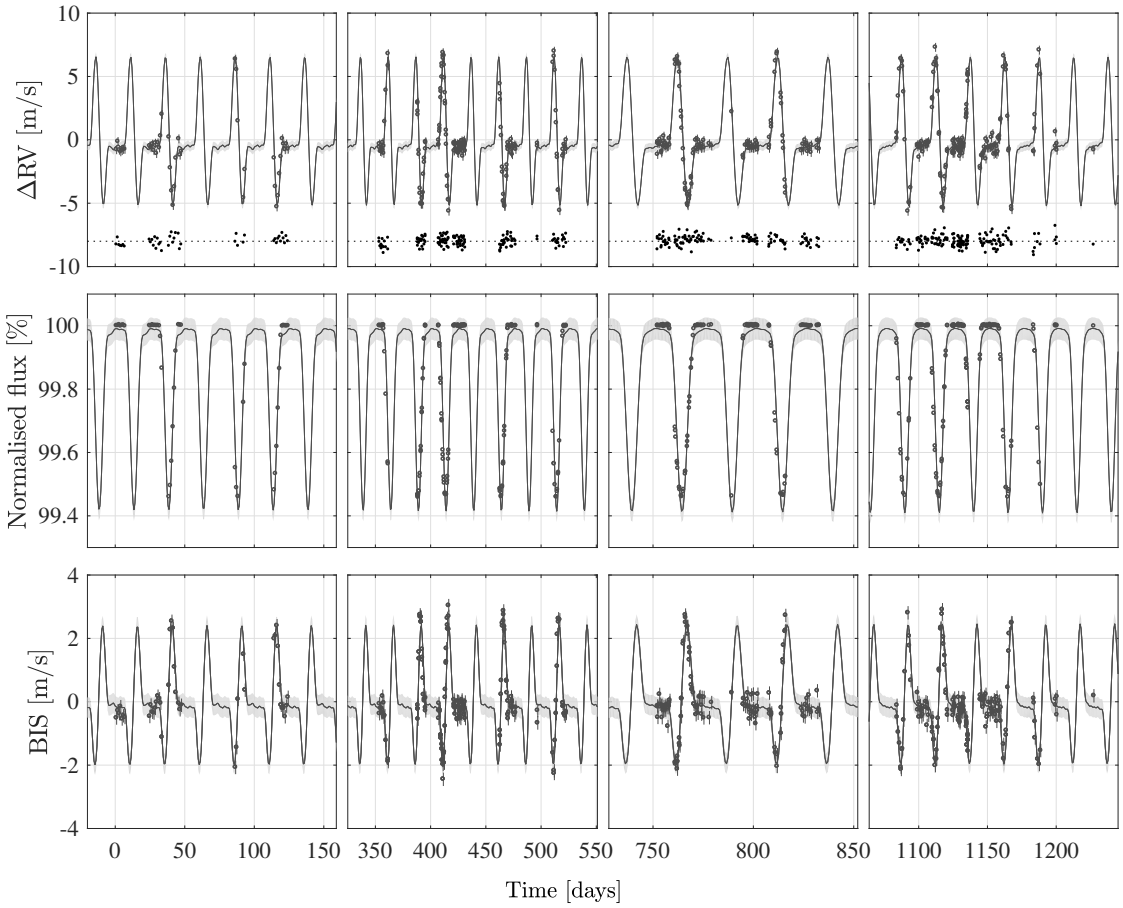


**Figure 3.5:** GP model MAP fit to SOAP 2.0 data, based on the same one-spot configuration as in Fig. 3.3, but this time with time sampling and noise levels taken from a real HARPS data set. Each time series is split into four panels, to avoid plotting large gaps where no data are present. Residuals in the ΔRV time series are plotted below the simulated data and fitted model, but for the sake of clarity, with an arbitrary vertical offset from the main time series.

modelling the ancillary, activity-sensitive time series in conjunction with ΔRV did indeed allow the activity component of ΔRV to be very well constrained, but – crucially – without the non-activity RV signals being subsumed by the same GP model. Broadly speaking, we found that it became more and more difficult to disentangle injected Keplerian signals from activity signals as the amplitude of the injected signals grew smaller relative to the activity signals (and, of course, relative to observational noise). Having a Keplerian signal with a period very similar to that of the activity signal generally did not make disentangling activity and injected signal any more difficult, provided the amplitude of the Keplerian signal

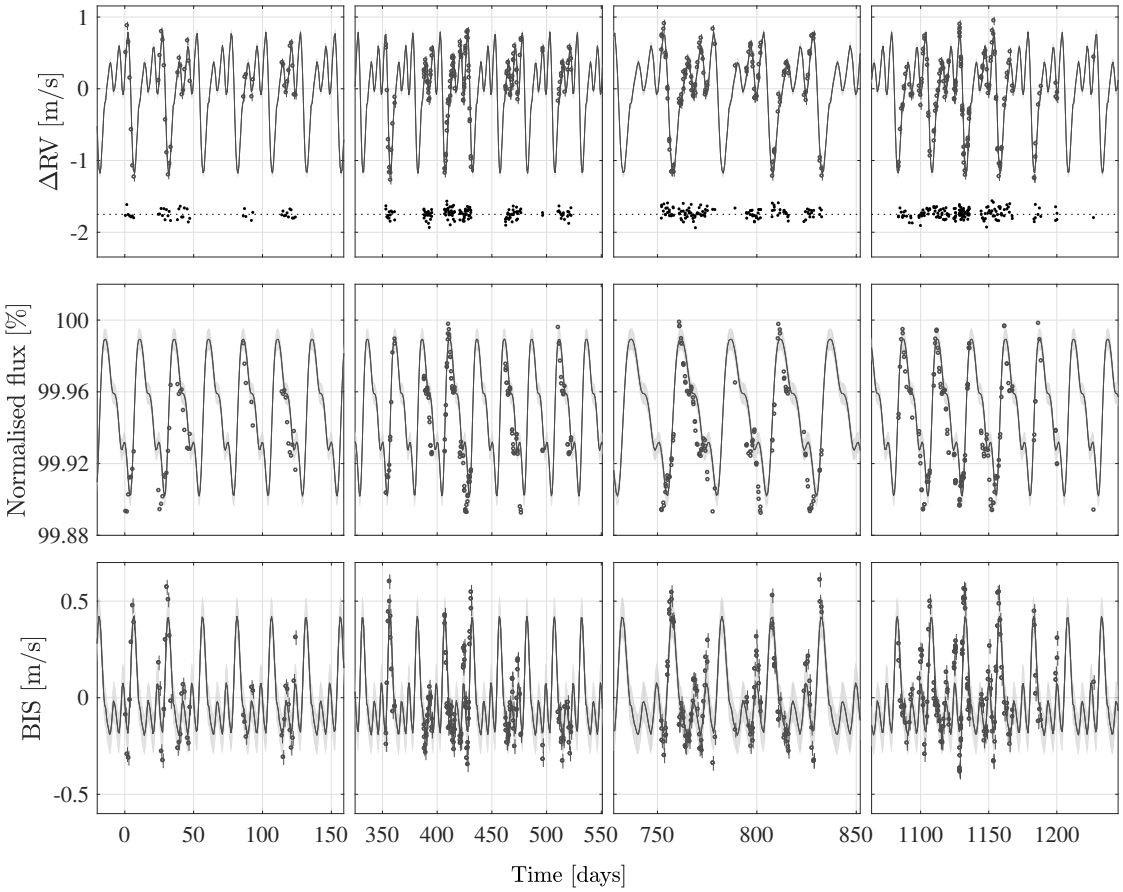


**Figure 3.6:** GP model MAP fit to SOAP 2.0 data, based on the same four-spot configuration as in Fig. 3.4, but this time with time sampling and noise levels taken from a real HARPS data set. Each time series is split into four panels, to avoid plotting large gaps where no data are present. Residuals in the ΔRV time series are plotted below the simulated data and fitted model, but for the sake of clarity, with an arbitrary vertical offset from the main time series.

was not also much smaller than the activity signal.

To illustrate these findings, we present as examples three qualitatively-different Keplerian signals injected into the ΔRV time series arising from the 'difficult' four-spot configuration considered earlier. The activity-induced ΔRV variations arising from this configuration had a semi-amplitude of 1.65 m $s^{-1}$, and a period of 25.05 d. We simulated observational noise using the $\alpha$ Cen B template as before, and we used time sampling corresponding to only one season of the $\alpha$ Cen B data, spanning approximately half a year (such that the simulated data sampled about seven stellar rotations).

The injected Keplerian signals we consider have the following properties (where $K$ is the RV semi-amplitude, and $P_K$ is the period of the variations; we use $P_K$ to distinguish this period from the period hyperparameter $P$ in our QP covariance kernel): (i) $K = 1.4$ m s$^{-1}$, $P_K = 10.0$ d; (ii) $K = 1.4$ m s$^{-1}$, $P_K = 25.05$ d; and (iii) $K = 0.28$ m s$^{-1}$, $P_K = 25.05$ d. Other orbital parameters were identical in all three cases; in particular, eccentricity was fixed at $e = 0.1$. Case (i) corresponds to an orbiting planet which induces an RV signal with amplitude comparable to the activity signal, and with a period that is not close to the stellar rotation period or any of its harmonics. Case (ii) adds the complication of the planet having a period very similar – identical, in fact – to the period of the rotational activity signal, while case (iii) goes one step further by representing a planetary signal with period identical to, but amplitude five times smaller than, the activity signal.

Fits to data including these injected signals are presented in Figs. 3.7, 3.8, 3.9 (the Keplerian component was included in our GP model through its mean function; non-informative priors were placed on all Keplerian orbital parameters). MAP parameter estimates for selected Keplerian orbital elements in the planet injection tests, along with $\pm\sigma$ posterior uncertainties, are presented in Table 3.1. All parameter inference was performed using the MultiNest nested-sampling algorithm (Feroz and Hobson, 2008; Feroz et al., 2009, 2013), with the GP hyperparameters first fixed at their MAP values, as per the computational approximation motivated in Gibson et al. (2012). Typical computation time to obtain posterior distributions for physical parameters, using MultiNest on a modern laptop, and using MultiNest's default convergence criteria, was of the order of tens of minutes. Posteriors were generally unimodal, although in general, of course, this would depend on the model in question, the data sampling, etc.

In cases (i) and (ii), the fitted Keplerian amplitudes and periods agree with the true values to within a small fraction of a percent. In case (iii), the absolute errors in the fitted values are larger (e.g. the recovered amplitude is a little smaller than the true amplitude), but of course one could never expect perfect fits given an imperfect

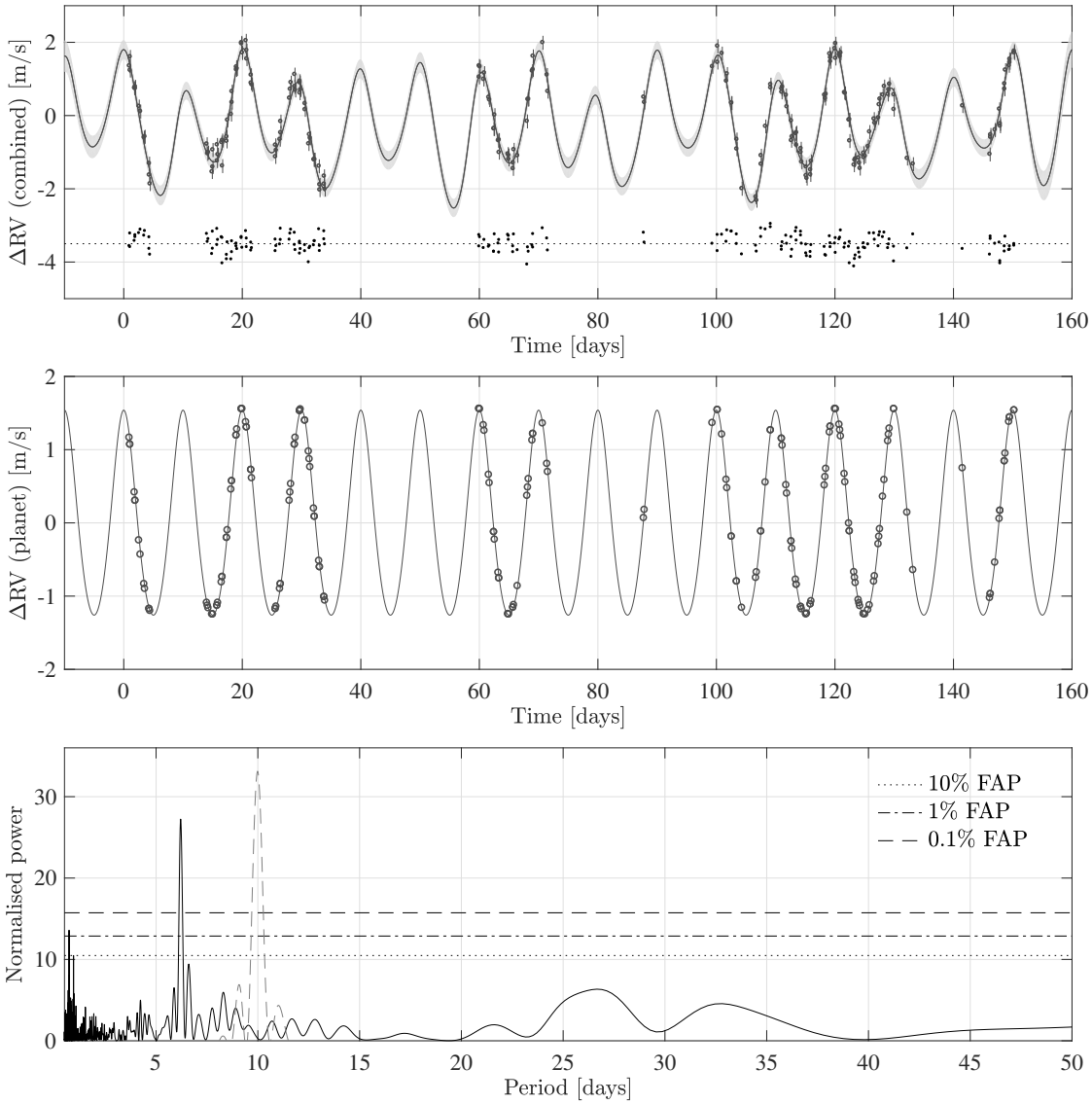


**Figure 3.7:** GP model MAP fit to SOAP 2.0 data, based on the same four-spot configuration as in Fig. 3.4, with time sampling and noise levels taken from one season of a real HARPS data set, plus an injected Keplerian signal. Here the injected signal has an amplitude comparable to, but period different from, the rotationally-modulated activity signal. The top panel shows the model fit to the ΔRV time series, including residuals (fits to ancillary time series not shown here); the middle panel shows the Keplerian component of the fit (open circles represent injected signal *before* noise was added); and the bottom panel shows the normalised LS periodogram of the ΔRV residuals, along with false alarm probability thresholds. The dashed grey peak in the periodogram around 10.0 d indicates the significant excess power that remains in the ΔRV residuals when performing a fit *without* a planet (the rest of the power spectrum remains qualitatively similar, and for clarity is not included in full).

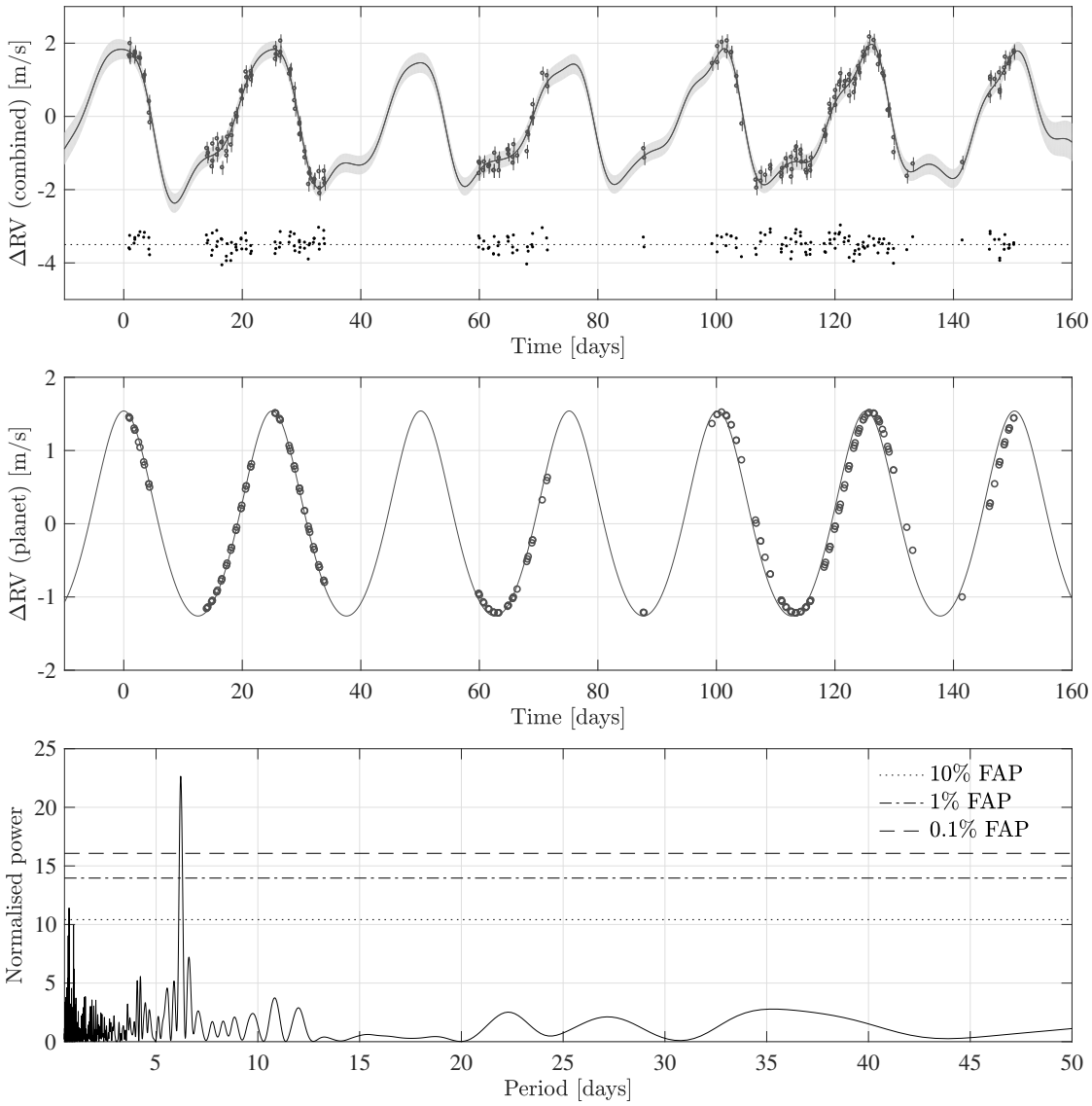


**Figure 3.8:** As for Fig. 3.7, but with the injected planetary signal now with period identical to that of the rotationally-modulated activity signal.

model, noise, and discrete time sampling. Nevertheless, it is reassuring that the true values do still lie within the $\pm 2\sigma$ credible intervals around the MAP values.

In all cases, the rms of the $\Delta$RV residuals was of the order of 0.2 m s$^{-1}$. The LS periodogram[8] of the $\Delta$RV residuals Figs. 3.7, 3.8, and 3.9 all show significant peaks at 6.5 d: we interpret this as the fourth harmonic of the activity signal, as this feature

[8]We computed normalised LS periodograms based on the implementation described by Press (2007); false alarm probabilities were estimated by randomly permuting the original data, keeping observation times fixed.

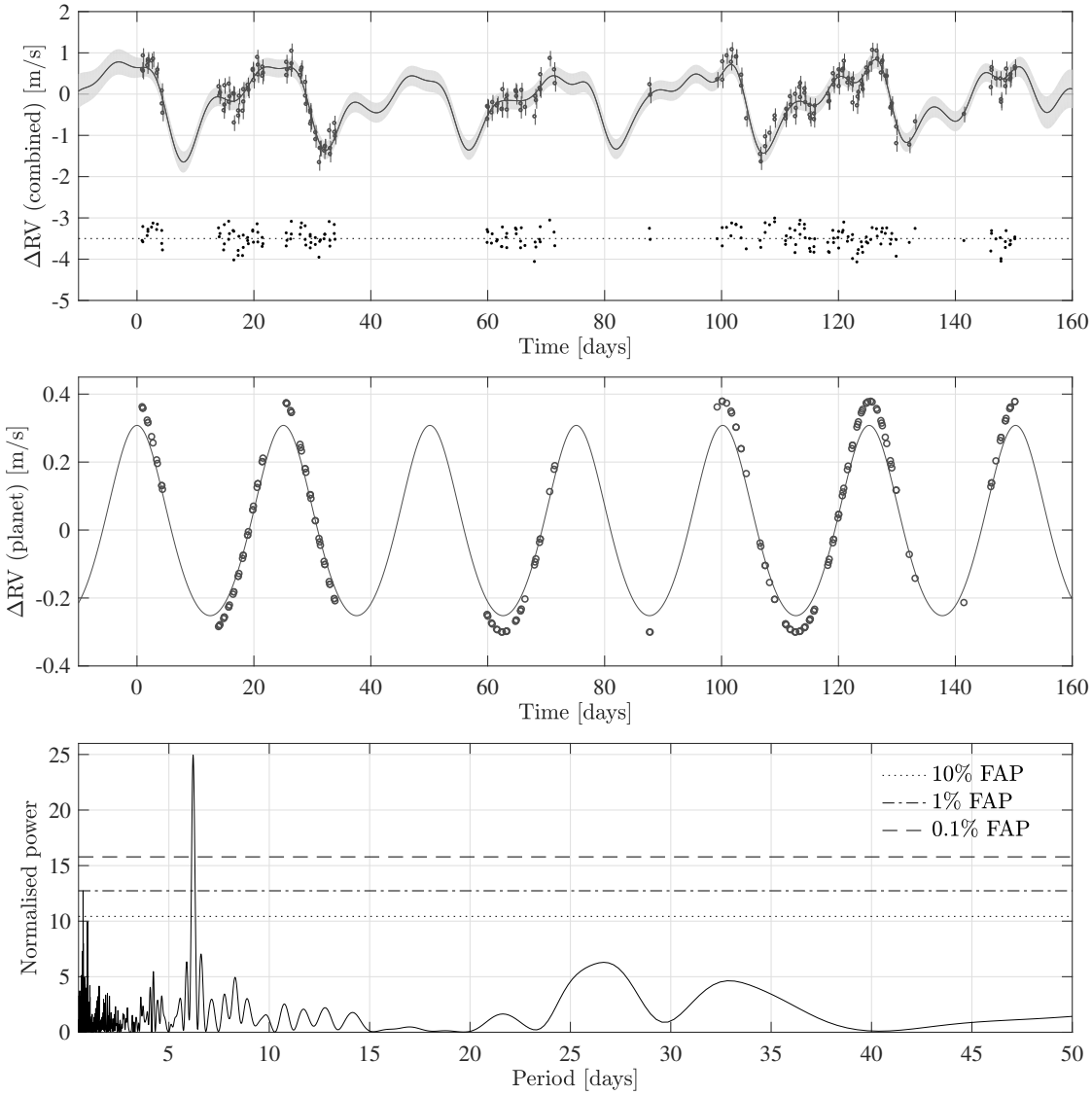


**Figure 3.9:** As for Fig. 3.8, but with the amplitude of the injected planetary signal now several times smaller than that of the activity signal (periods once again identical).

remained present in the ΔRV residuals even when the injected Keplerian signal had a period that is not similar to the stellar rotation or any of its harmonics, e.g. as in case (i). This is a reflection of our GP activity model's inability to reproduce the activity signal in its full harmonic complexity; in informal terms, the GP model does not capture perfectly every last 'bump and wiggle' of the activity signal. Nevertheless, by using the ancillary activity indicators to aggregate information on the activity signal's period, shape, and phase, it succeeds at reproducing the RV signal of the activity very well, and indeed to much better than first order.

The only other significant peaks in the periodogram of the residuals, clustered around 1 d, arise from the window function (time sampling) in the $\alpha$ Cen B data set.

In cases (i), when we did not include a Keplerian component in the modelling, the power spectrum of the $\Delta$RV residuals contained significant excess power at 10.0 d, i.e. the planet's period - a telltale sign that the periodic $\Delta$RV variations could not be explained by rotational activity alone. When not including a Keplerian component in cases (ii) and (iii), visual inspection of the periodogram of residuals did not by itself point clearly to the presence of a planet with period 25.05 d (the same as the stellar rotation period); however, the reality of the Keplerian components could be inferred by the significant improvement in the likelihoods of the models (approximated e.g. using Bayesian Information Criteria – BIC; see Section 3.4.3) that included a planet.

Thus we have demonstrated that our GP framework is able to disentangle activity signals from Keplerian signals, even when the Keplerian signal has a period identical to that of the activity signal, and an amplitude much smaller than the activity signal (close to the noise floor, in fact). Although the example we presented featured imperfect removal of the activity signal from the $\Delta$RV data (since the residuals contained a signature of one of the activity signal's harmonics), the ancillary time series served the role of constraining very tightly the activity component in the $\Delta$RV time series, allowing the Keplerian signal to be modelled accurately. Presumably, a more physically-realistic incarnation of our GP framework (see Section 3.5) will allow activity signals to be even more tightly constrained by the ancillary time series. We also note that the harmonic of the activity signal was not present in the $\Delta$RV residuals when modelling the signal arising from the same configuration but with more data, as in Fig. 3.6, where nearly four years' worth of data were included in the modelling. Nevertheless, the presence of a correlated, periodic artefact of the activity signal in the RV residuals in this test underscores the importance of accurate activity modelling, and the dangers of hastening to a planetary interpretation of signals in residuals. It would be worth checking, in

future work, whether a Bayesian model comparison test would clearly penalise and disfavour a model that included a second Keplerian component to explain this signal.

We conclude the discussion of our SOAP 2.0 tests by noting the following shortcomings of these tests, along with possibilities for future extensions.

1. Though the covariance kernel we used can naturally accommodate QP signals (see Fig. 3.1), our simulations did not include any spot evolution, e.g. migration across latitudes, growth, and decay; the SOAP 2.0 code does not directly facilitate such evolution. In principle, though, activity signals arising from different spot configurations could be combined, perhaps with large gaps in between the different signals, to simulate evolving active regions. Long-term magnetic cycles could be simulated in a similar fashion.
2. Non-informative (uniform and Jeffreys) priors were placed on all model parameters/hyperparameters. Improved fitting could be expected if we were to use priors informed by a better understanding of e.g. the hyperparameters of our QP covariance kernel, or more physically-motivated priors for Keplerian signals (see also Section 7.1.2 of this thesis).
3. We did not include any plages in the active regions simulated by SOAP 2.0. At the time of writing, the SOAP 2.0 code's output appeared to include a non-trivial amount of high-frequency noise when including plages in the simulations. Preliminary tests with a smoothing filter applied to the SOAP 2.0 output suggest that even when plages are included, the activity components in $\Delta$RV time series can still be modelled accurately, although the large residuals in the ancillary time series in these tests underscore the simplicity of the physical model used in our framework. Complications arising from the use of smoothing filters notwithstanding, these tests suggest the need for a better understanding of the shortcomings of the physical model (ultimately to be replaced, hopefully, by a more sophisticated one) currently incorporated into our GP framework.

**Table 3.1:** MAP parameter estimates for selected Keplerian orbital elements in the planet injection tests, along with $\pm\sigma$ posterior uncertainties. Parameter inference was performed using the MultiNest nested-sampling algorithm, with non-informative priors placed on all GP mean function (Keplerian) parameters.

| | Semi-ampl. $K$ [m s$^{-1}$] | | Period $P_K$ [d] | | Eccentricity $e$ | |
|---|---|---|---|---|---|---|
| | Model | Truth | Model | Truth | Model | Truth |
| (i) | $1.400 \pm 0.045$ | 1.40 | $9.9941 \pm 0.0086$ | 10.0 | $0.118 \pm 0.016$ | 0.10 |
| (ii) | $1.372 \pm 0.070$ | 1.40 | $25.09 \pm 0.26$ | 25.05 | $0.1050 \pm 0.0081$ | 0.10 |
| (iii) | $0.334 \pm 0.047$ | 0.28 | $25.10 \pm 0.89$ | 25.05 | $0.1051 \pm 0.0080$ | 0.10 |

### 3.4.3 The planet around Gliese 15 A

Howard et al. (2014) – hereafter H14 – reported the discovery of low-mass planet orbiting Gl 15 A, part of the Gl 15 binary system, based on RVs from the Eta-Earth survey using HIRES. They characterise Gl 15 Ab as a planet with minimum mass $M \sin i = 5.35 \pm 0.75$ M$_\oplus$, based on an RV semi-amplitude of $K = 2.94 \pm 0.28$ m s$^{-1}$. The planet has an orbital period $P_K = 11.4433 \pm 0.0016$ d, and an orbit that is consistent with circular. The detection and characterisation of Gl 15 Ab was based on data with a 15-yr baseline (January 1997 through December 2011).

Gl 15 A itself is a modestly-active, M2 V dwarf. Over their 15-yr baseline, H14 detected a $9 \pm 2.5$ yr activity cycle with a semi-amplitude of $\sim 0.05$, in the dimensionless units of $S_{\rm HK}$ (derived similarly to $\log R'_{\rm HK}$), which represents a $\sim 10\%$ fractional change; this variation may be a magnetic cycle analogous to the solar cycle. On shorter time-scales, they detected a strong periodicity in $S_{\rm HK}$ measurements with period near 44 d, which they interpreted as rotationally-modulated activity. The fact that they also detected these $\sim 44$ d modulations in optical photometry and RV time series supported this interpretation. On the other hand, the 11.44 d RV signal was not detected in photometry or $S_{\rm HK}$, strengthening the planetary interpretation of the 11.44 d signal.

Since the 11.44 d period of Gl 15 Ab could easily be mistaken for the fourth harmonic of the $\sim 44$ d stellar rotation period, H14's publicly-available data set provides an interesting test case for our GP model: an obvious (though perhaps

unfounded) concern in this case would be that the GP activity model would simply absorb the planetary signal.

Accordingly, we modelled the 15 years of data published by H14, comprising 117 RV and $S_{\rm HK}$ measurements, 59 of which were taken in the 2011 season (BJD >2,455,500). As for our earlier tests, we used a QP covariance kernel for the latent process driving activity; we also included a linear RV trend in both our single-planet and planet-free models, to account for the $\sim 2$ m s$^{-1}$ yr$^{-1}$ acceleration of Gl 15 A by its companion Gl 15 B.

When we did not include a Keplerian component in the mean function for the $\Delta$RV time series, the LS periodogram of the residuals contained multiple significant peaks, including one at around 11.4 d, indicating that the putative planetary signal was not absorbed by the activity model: since this periodic modulation was not present in the $S_{\rm HK}$ time series, our model did not allow it in the $\Delta$RV time series. On the other hand, when including a Keplerian component with circular orbit in the $\Delta$RV mean function, the quality of the fit improved significantly, and an LS periodogram of the residuals indicated that no significant periodicities remained. Using BIC as a first-order approximation to Bayes factors (Berger and Pericchi, 1996; Raftery, 1999), our single planet model was strongly preferred with $\Delta$BIC = 27; H14 arrived at the same conclusion, with $\Delta$BIC = 24.

Our GP model, one-planet fit to H14's data is presented in Fig. 3.10, while detailed diagnostics of the $\Delta$RV residuals (rms: 0.69 m s$^{-1}$) are presented in Fig. 3.11. Our MAP estimate for the $\Delta$RV semi-amplitude of the planetary signal was $K = 2.22 \pm 0.36$ m s$^{-1}$, and our estimate for its period $P_K = 11.4431 \pm 0.0010$ d. H14 obtained $K = 2.94 \pm 0.28$ m s$^{-1}$, and $P_K = 11.4433 \pm 0.0016$ d; thus our approach to modelling H14's data set led us to detect a planet with period and amplitude consistent with those derived by H14, to within our respective 95% ($2\sigma$) credible intervals for these parameters. The MAP estimate of the period hyperparameter in our QP covariance kernel was $P = 43.8 \pm 0.9$ d, which is consistent with the $\sim 44$ d stellar rotation period derived by H14.

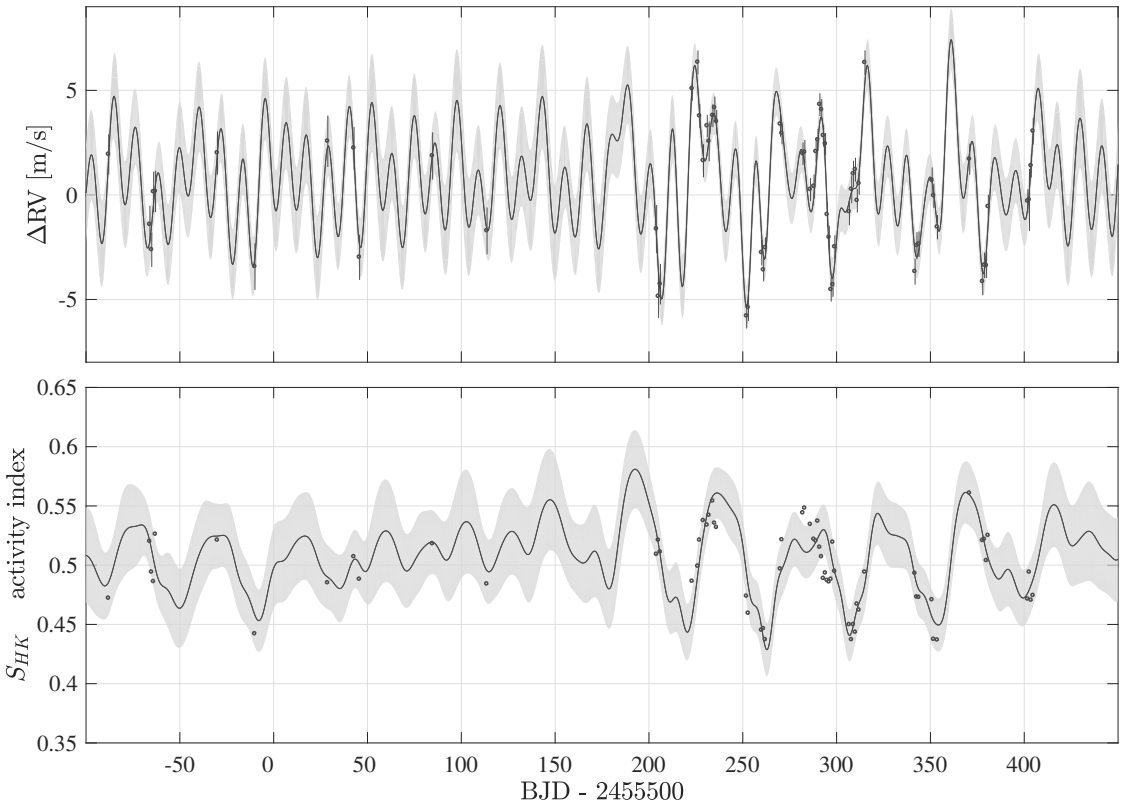


**Figure 3.10:** GP model MAP fit to the publicly-available Gl 15 A data from H14. The fit was performed to the full 15 years of data, comprising 117 observations; since most of the observations were made during the 2011 season (BJD >2,455,500), however, only data from this season are plotted here. The dots indicate H14's data (with estimated errors, where applicable); the solid lines are model posterior means, and the shaded regions denote $\pm\sigma$ posterior uncertainty.

We did not consider any multiple-planet models. We defer such analyses of this and other data sets to future work, in which we aim to leverage a better understanding of appropriate priors over our GP model's hyperparameters to perform more rigorous Bayesian model comparison.

### 3.4.4 The Alpha Cen B data set

The announcement by Dumusque et al. (2012) of the detection of a planet around the modestly-active, K1 V star $\alpha$ Cen B caused a major stir in the exoplanet community: at the time of announcement, the claimed planet $\alpha$ Cen Bb would have been the closest (distance: 1.34 pc) exoplanet to Earth ever discovered, and the lowest-minimum-mass planet detected around a Solar-type star. In particular,

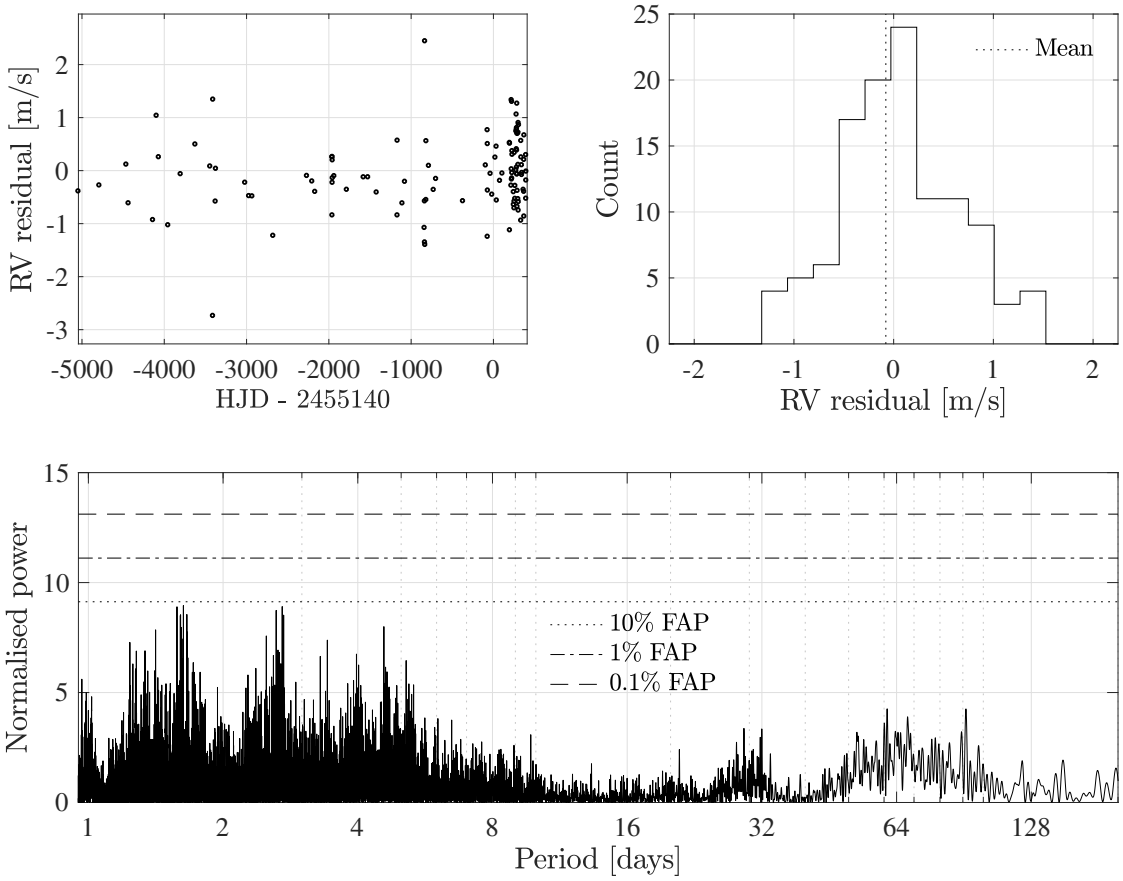


**Figure 3.11:** Diagnostics of the ΔRV residuals for Gl 15 A, after subtracting our MAP GP model (stellar activity, plus binary orbit and one planet). The top left panel contains the residuals plotted as a function of time, spanning some 15 yr; the top right panel contains a histogram of the residuals (rms: 0.69 m $s^{-1}$); and the bottom panel shows the normalised LS periodogram of the residuals, along with false alarm probability thresholds.

the planet was claimed to have an orbital period of 3.24 d, and a minimum mass of $1.13 \pm 0.09$ $M_\oplus$.

Dumusque et al. (hereafter D12) obtained 459 HARPS RV data points, along with ancillary FWHM, BIS and $\log R'_{\rm HK}$ time series, over a period of 4 years. The D12 RVs are dominated by a long-term linear trend, which is due to the orbit of $\alpha$ Cen B around the centre of mass of the $\alpha$ Cen binary system. Once this trend is subtracted, a gradual rise and fall over the 4-yr span of the observations is evident, as well as variability on shorter time-scales. D12 interpreted the gradual rise and fall as a signature of the star's activity cycle, and the QP variations on time-scales of a few weeks as caused by the rotational modulation of star spots. Lastly, there

are short-term variations evident in all time series. These variations are correlated in time, with a time-scale of a few days, as well as with each other (across all three time series), albeit in a somewhat complex way. It is unclear whether these variations might be due to activity, observational noise, or a combination of both. Of course, in the case of the RV data in isolation, a part of this variability could also be explained by a planetary companion with a relatively short orbital period.

In any event, D12 used a variety of mathematical transformations to try to 'filter' out many sources of RV variance – including starspots, photospheric granulation, and binary motion due to the presence of companion star $\alpha$ Cen A – to isolate the putative planet's RV signal. The amplitude of the binary-motion signal they removed was of the order of hundreds of m s$^{-1}$; the combined amplitude of the nuisance signals ascribed to stellar activity they removed was of the order of 10 m s$^{-1}$; and the final, claimed planetary signal they isolated had a semi-amplitude of about $\sim 0.5$ m s$^{-1}$. For comparison, the long-term precision of HARPS is 0.8 m s$^{-1}$ (Mayor et al., 2003).

However, the detection was a contentious one, with a number of authors calling into question the planet's existence (see e.g. Hatzes, 2012, 2013). Some of the possible reservations about the approach used by Dumusque et al. include the following.

1. Short-term stellar rotational activity was dealt with by fitting sinusoidal waves at the rotational period of the star, and a variable number of harmonics. The constraint that all activity-induced variations should be strictly sinusoidal is very hard to motivate physically (Lanza et al., 2001; Brinkworth et al., 2005), and subtracting multiple sinusoids from the $\Delta$RV time series opens the possibility of inadvertently introducing periodic signals (or harmonics thereof) that were not present in the original data. See also Fig. 2.7 on pg. 46.
2. Available activity-sensitive time series (BIS and $\log R'_{\rm HK}$) were not jointly modelled with the $\Delta$RV time series – even though it is in the $\log R'_{\rm HK}$ time series that the rotational signal was most prominent. Instead, a low-pass smoothing filter was applied to the $\log R'_{\rm HK}$ time series; it was then simply noted that the smoothed time series looked similar to the $\Delta$RV time series, and the shape of

this smoothed time series was then used to fit the RV variations in order to mitigate the effects of long-term activity.

3. Each of the four observing seasons was fitted independently of the others, thus ignoring any information about possible long-term trends/correlations in the time series, and resulting in a model with almost four times as many degrees of freedom as might have been the case if all seasons were modelled simultaneously.
4. The model used to fit the $\Delta$RV time series alone, *without* a planet, contained 23 free parameters, and the combined fit for activity and a planet contained at least 26 free parameters. The question of possible over-fitting was not addressed in either case. For comparison, our baseline/planet-free GP model – discussed below – models multiple time series simultaneously, yet contains significantly fewer free parameters than D12's model.[9]
5. No rigorous model comparison was performed to quantify the evidence for the existence of the planet, i.e. the extent to which a planetary model is (or is not) preferred.

Given that our GP framework for modelling activity would allows us, in principle, to avoid all of these limitations, we decided to apply it to the publicly-available D12 data set.

As in all previous tests, we used a QP covariance kernel for our latent process. We modelled all 459 data points in each of the $\Delta$RV (corrected for binary motion, and with residual DC offset removed for each season), $\log R'_{\rm HK}$ and BIS time series jointly using this single latent process and its derivative.[10] Error estimates were published by D12 for the $\Delta$RV and $\log R'_{\rm HK}$ time series; these estimates were included in our modelling. Though not directly provided, BIS errors were estimated (following D12's prescription) using provided photon-noise estimates. Finally, as before, non-informative priors were placed on all model (hyper)parameters.

[9] A more meaningful application of Occam's razor would require marginalisation over the full volumes of respective model parameter spaces.

[10] The FWHM data are noisier than, but very tightly correlated with, $\log R'_{\rm HK}$, and thus are unlikely to contain useful extra information. Both can be seen as proxies for the integrated spot coverage of the visible hemisphere. On the other hand, the BIS also depends on the velocity of the stellar surface at the location of the spots (or active regions).

Our baseline GP model did not incorporate any Keplerian component in the mean function for the RVs. In total, the combined model for all three time series contained 14 free (hyper)parameters – in particular, it should be noted that all three time series across all four observing seasons were fitted using this same model. This stands in contrast to the approach used by D12, where the activity signal in each of the four seasons was fitted independently, since the model they used required that stellar features did not evolve significantly during the time considered, whereas spots and plages and spots are known to evolve on a time-scale of a few rotations. The MAP fit using this model is presented in Fig. 3.12, while detailed diagnostics of the $\Delta$RV residuals are presented in Fig. 3.13.

The $\Delta$RV residuals appear to be normally-distributed, with rms 0.70 m s$^{-1}$. The power spectrum of the $\Delta$RV residuals in Fig. 3.13, devoid of any significant peaks save for a complex of peaks tightly clustered around 1.00 d, suggests that the $\Delta$RV variations can be explained as arising from activity alone. As noted in Section 3.4.2, the peaks around 1.00 d arise from the window function in the $\alpha$ Cen B data set. In particular, there are peaks at $f_1 = 0.97$ d$^{-1}$ and $f_2 = 1.03$ d$^{-1}$; these peaks appear in the power spectra of both the model and the observations, and may be interpreted as the first-order aliases of the 1/37.8 d$^{-1}$ rotational frequency. A third peak appears only in the $\Delta$RV residuals, but not in the model or the observations: this may be interpreted as a beat frequency arising during the computation of residuals, on account of interference between the other two closely-spaced frequencies.[11]

When we did include a Keplerian component in the mean function for the RVs, we were unable to extract any significant (non-activity) signal. Even when forcing the Keplerian signal to have a period of close to 3.24 d, the MAP solution favoured a Keplerian signal with an arbitrarily small (within the bounds of a Jeffreys prior) amplitude. Conversely, when forcing the Keplerian signal to have an amplitude of about $\sim 0.5$ m s$^{-1}$, the MAP solution favoured either a very short ($P_K \ll 1$ d) or very long ($P_K \gg 1$ yr period), so that this signal essentially became a DC

[11] To understand the beat phenomenon, it is useful to recall the trigonometric identity: $\sin(2\pi f_1 t) \pm \sin(2\pi f_2 t) = 2\sin\left(2\pi \frac{f1 \pm f2}{2} t\right)\left(2\pi \frac{f1 \mp f2}{2} t\right)$.

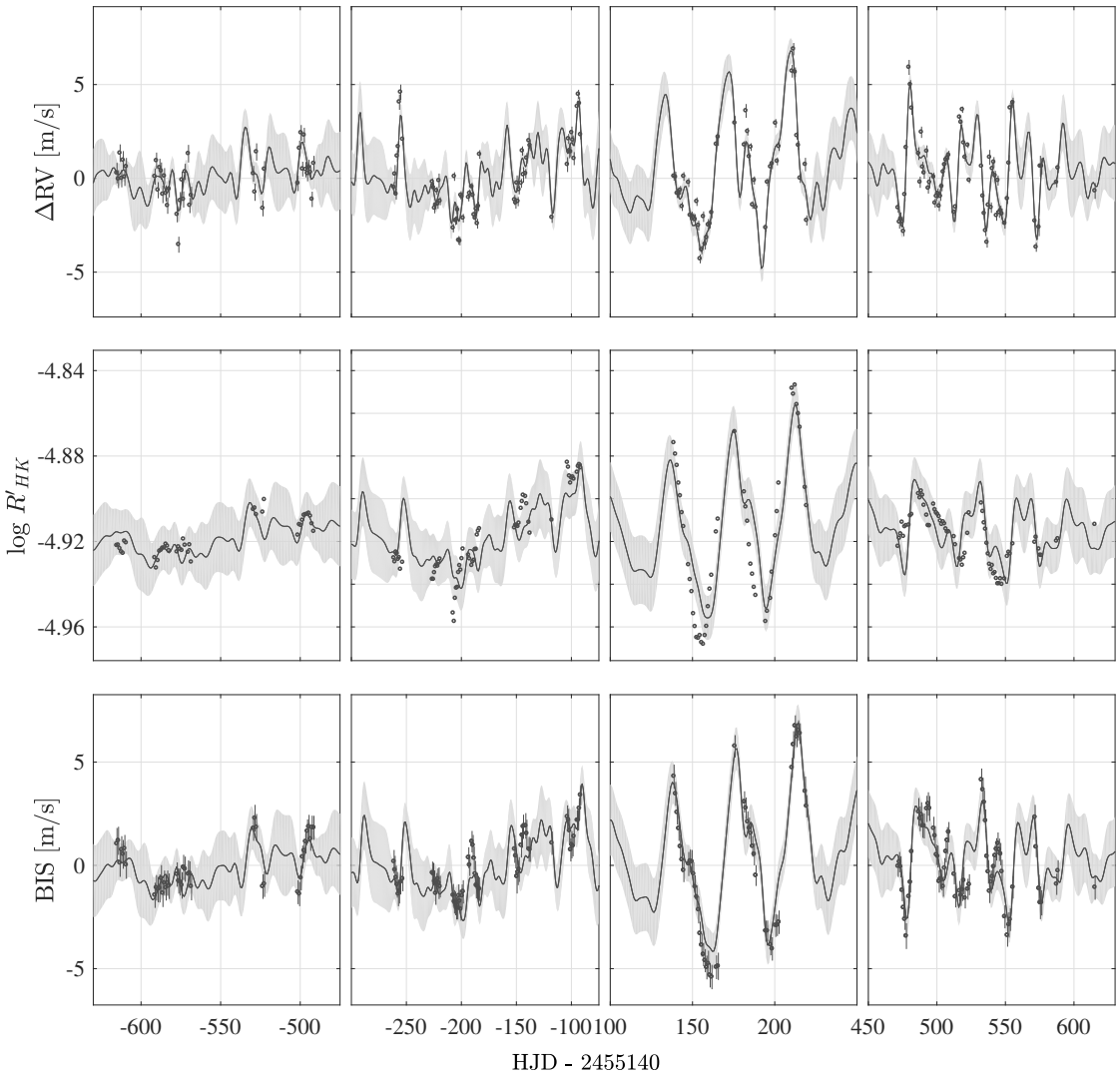


**Figure 3.12:** GP model MAP fit to the publicly-available $\alpha$ Cen B data from D12. All four seasons, comprising some 459 observations, were fit simultaneously, i.e. using a single latent process and set of (hyper)parameters. The dots indicate D12's data (with estimated errors, where applicable); the solid lines are model posterior means, and the shaded regions denote $\pm\sigma$ posterior uncertainty.

offset to the $\Delta$RV time series. Furthermore, our non-detection of the planetary signal was also robust against different choices of covariance kernel functions (e.g. squared exponential, rational quadratic). Nor did allowing for somewhat tighter or looser links between the RVs and the activity-sensitive $\log R'_{\rm HK}$ and BIS time series, by modifying the priors on the three $\sigma_{l_i}$ terms in our model, allow us to detect a 3.24 d Keplerian signal in the data.

In summary, using the activity-sensitive $\log R'_{\rm HK}$ and BIS time series to constrain

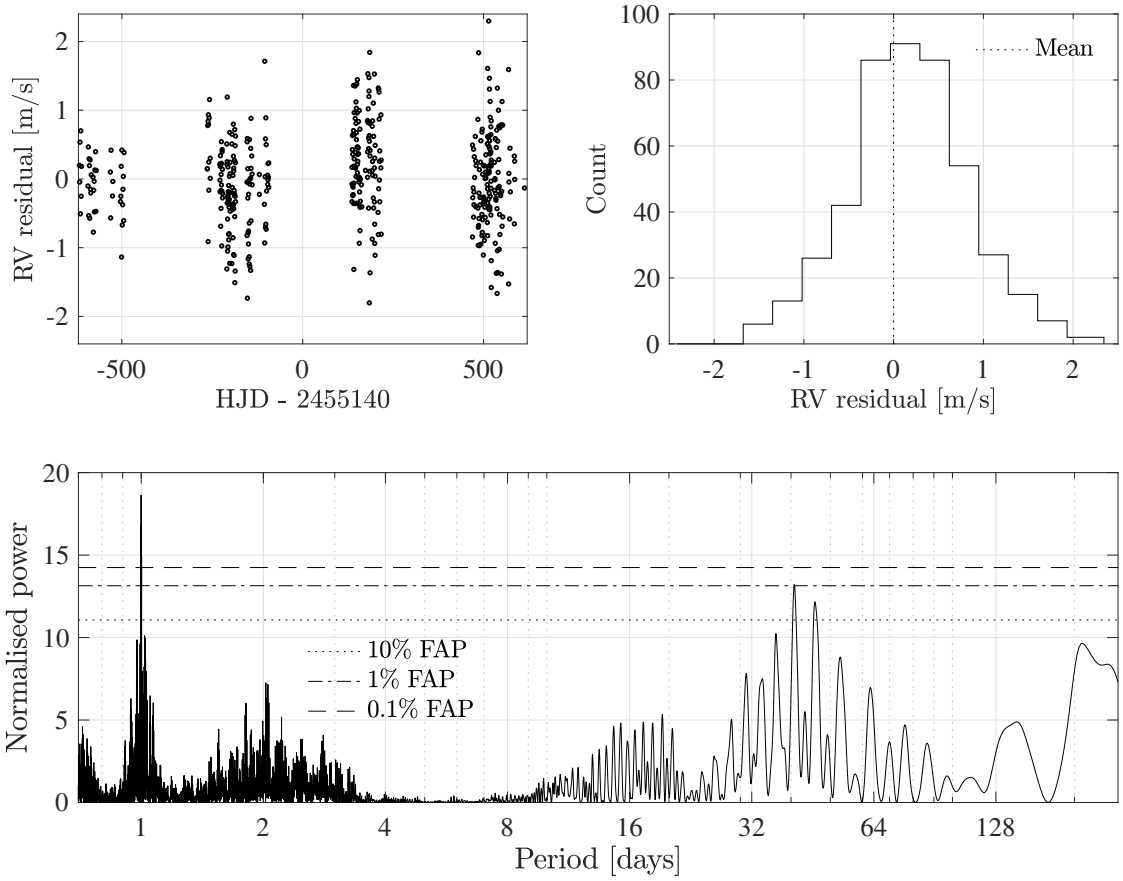


**Figure 3.13:** Diagnostics of the ΔRV residuals for $\alpha$ Cen B, after subtracting our MAP GP model (stellar activity plus binary orbit). The top left panel contains the residuals plotted as a function of time; the top right panel contains a histogram of the residuals (rms: 0.70 m $s^{-1}$); and the bottom panel shows the normalised LS periodogram of the residuals, along with false alarm probability thresholds.

the activity component of the RV time series suggests that all of the significant variation in the RV time series can be explained as being induced by stellar activity, without requiring any additional Keplerian component.

We do not claim that this finding proves that the claimed planet $\alpha$ Cen Bb definitely does not exist. However, the difficulty we have in recovering the signal as something unrelated to activity underscores just how important it is to model activity carefully and robustly, if we are to detect and characterise planets at the sub-m $s^{-1}$ level. In subsequent work, we aim to make a more rigorous and comprehensive study of D12's $\alpha$ Cen B data set. We would like to investigate in which situations our planet-free models lead to a signal with a 3.24 d period in the RV residuals.

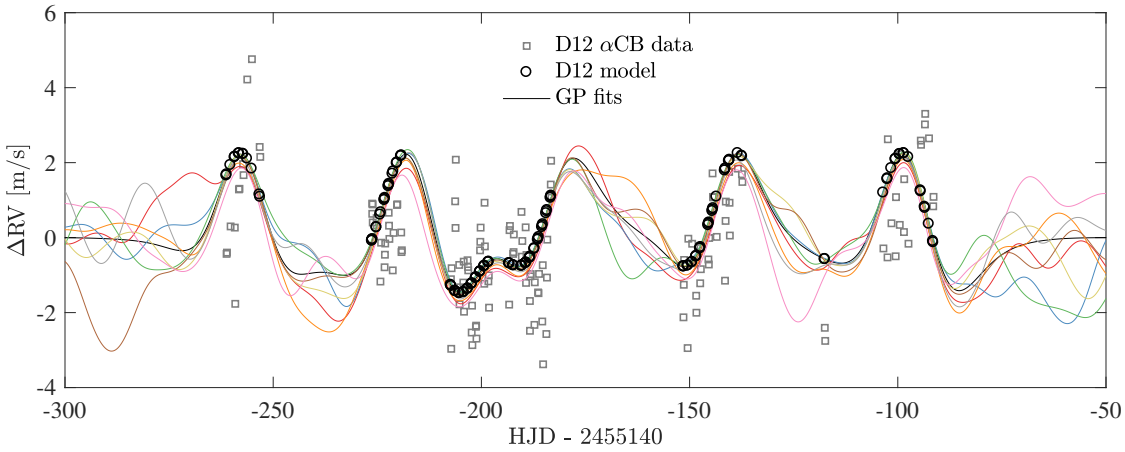


**Figure 3.14:** Demonstration that the predictions of a semi-parametric GP model can subsume those of D12's model as a special case. The grey squares correspond to the publicly-available $\alpha$ Cen B data from D12, after removing binary and long-term magnetic cycle contributions, while the black circles are predictions from D12's published activity model, for the same times as the real observations. Draws from a GP model (a few examples of which are plotted) fitted to the noisy data can be found such that they essentially interpolate D12's model predictions; for clarity, only one observing season is shown here, but the interpolation holds true for the other seasons as well. Indeed, it is straightforward (though not necessarily useful or desirable, and ignoring considerations of over-fitting) to find hyperparameters that will force a GP model to agree exactly with D12's activity model, for arbitrarily-many observing times.

This should happen at least in some cases, since our GP framework should be able to reproduce the model of Dumusque et al. as a special case – even if only with a non-optimal model (see Fig. 3.14), and/or when loosening significantly the links between the ancillary, activity-sensitive time series and the RVs. Concurrently, we would like to undertake detailed studies of planetary detection limits using both our GP framework and a model akin to that used by Dumusque et al., for synthetic data similar to the $\alpha$ Cen B data set. Lastly, we would like to perform rigorous Bayesian model comparison (planet versus no-planet models) in order to draw definitive conclusions about whether the data really suggest the presence of a planetary signal. These analyses will form the focus of the next chapter of this thesis.

## 3.5 Discussion and conclusions

Stellar activity can induce RV variations that can drown out or even mimic planetary signals, and it is notoriously difficult to model and thus mitigate the effects of these activity-induced nuisance signals. This is expected to be a major obstacle to using next-generation spectrographs to detect lower and lower mass planets, planets with longer periods, and planets around more active stars. We have presented here a new framework for modelling RV time series jointly with one or more ancillary, activity-sensitive time series (including photometry, line widths, chromospheric activity indices, line asymmetries, etc.), with a view to better constraining activity signals in RVs, and thus better detecting and characterising possible planets.

Our framework treats the underlying stochastic process giving rise to activity signals in all available observables (RVs and ancillary time series) as being described by a GP, with suitably-chosen covariance function. We then use physically-motivated and empirical models to link this GP to the observables. With the addition of noise and deterministic components (e.g. dynamical effects for the RVs), all observables can be modelled jointly as GPs, with the ancillary time series thus serving to constrain the activity component of the RVs.

We demonstrated the performance of our framework using both synthetic (Sections 3.4.1 and 3.4.2) and real (Sections 3.4.3 and 3.4.4) data. We started by noting that our framework can model all available time series jointly, and exactly, in the simplest case where the physical/empirical relationships between all time series holds exactly. Next, using more realistic data simulated using the SOAP 2.0 tool (including noise and realistic time sampling), we showed that our framework does a good job of constraining activity signals in RV data, provided we allow for additive white-noise terms in each time series we model; these additive noise terms may be interpreted as accounting for the extent to which our assumed relationships between the time series does not hold exactly, observational noise notwithstanding. We also showed that the framework can be used to disentangle activity and planetary signals – even when the planetary signal is weaker than and has a period identical to the activity signal – thus serving its intended purpose.

Moving to real data, we applied our framework to the Gl 15 A system: we were able to disentangle activity and planetary components in a HIRES RV data set, and obtained a fitted planetary model which was consistent with one published by Howard et al. (2014). Finally, we turned our attention to the much-discussed $\alpha$ Cen B data set: we showed that, contrary to the analysis of Dumusque et al. (2012), we were able to attribute the observed RV variations to stellar activity, without requiring a planet.

Although our framework hinges, at least partly, upon a number of approximations and empirical relationships, its performance appears promising, and it offers many advantages over existing approaches to mitigating activity in RV data sets. It is flexible; though algebraically non-trivial, it is conceptually simple, and makes few assumptions about the properties of the underlying processes inducing activity signals in observables; it facilitates smooth interpolation between observables, as well as extrapolation to future times; and lastly, the entire framework is accommodated very naturally within the broader framework of Bayesian inference. As such, we hope that our new framework will form a useful addition to the toolbox of those confronted with mitigating stellar activity signals in RV data sets.

An investigation of ways in which our framework could be improved would represent a natural extension to this work. Some possibilities worth studying include the following.

1. Developing a more realistic relationship between RV and ancillary time series, especially the BIS time series – perhaps based on theoretical considerations, or perhaps through empirical means, e.g. by studying simulated data. Preliminary tests suggest the inclusion of a curvature term (second derivative of the latent process) could be helpful for more accurately coupling the RV and BIS time series. Though it would add significant algebraic complexity, it would only add one free parameter to the overall model.
2. Performing more comprehensive studies of planet detection rates and false-positive rates under our framework.
3. Extending the SOAP 2.0 tests to include plages and spot evolution.

4. Investigation of a more physically-motivated GP covariance function, e.g. a function from the Matérn class of covariance functions, which are differentiable only a finite number of times, and have the squared exponential function as a smooth limiting case (Rasmussen and Williams, 2006).
5. Investigation of more physically-motivated priors for all model (hyper)parameters, with a view to computing Bayes factors and performing rigorous model comparison. The latter will require marginalisation over GP hyperparameters, rather than fixing them at their MAP values.
6. Using raw, high-resolution spectra to construct a quantity that is a more sensitive proxy to stellar activity than ones currently in widespread use (BIS, $\log R'_{\mathrm{HK}}$, etc.). An artificial neural network might be useful for machine-learning such a quantity from a large data set.
7. Investing including a GP component to model correlated instrumental noise, rather than forcing such noise to be artificially absorbed by an additive white-noise component, as is the case in our current framework.

In tandem with the above, as discussed in Section 3.4.4, we would like to present a far more detailed and rigorous analysis of the much-discussed $\alpha$ Cen B data set. We would like to investigate in which situations our fitted planet-free models lead to a signal with a 3.24 d period in the RV residuals (this should happen at least in some cases, since our framework should be able to reproduce the model of Dumusque et al. as a special case); we would like to undertake detailed studies of planetary detection limits using both our GP framework and a model akin to that used by Dumusque et al., for synthetic data similar to the $\alpha$ Cen B data set; and we would like to perform Bayesian model comparison (planet versus no-planet models) in order to draw definitive conclusions about whether the data really suggest the presence of a planetary signal. These analyses form the focus of the next chapter of this thesis.

*It is far harder to kill a phantom than a reality. She was always creeping back when I thought I had despatched her. Though I flatter myself that I killed her in the end, the struggle was severe...*

— Virginia Woolf, from a 1931 paper read to The Women's Service League, published in the collection *The Death of the Moth, and Other Essays* (1942).

# 4

# 'Killing' a planet: the case of Alpha Cen B

## Contents



This chapter is based on the Letter entitled *Ghost in the time series: no planet for Alpha Cen B* (Rajpaul et al., 2016) – as of June 2017, the most-cited output from my doctoral research. My analysis depends directly on the GP framework presented in the previous chapter, and builds on the analysis in Section 3.4.4, though space constraints meant this fact was mentioned only somewhat parenthetically the original letter. My GP framework is first used to learn the properties of activity signals associated with Alpha Cen B, and is thence used to generate realistic synthetic data for exposing the shortcomings of the model used to claim the existence of the planet Alpha Cen Bb. Given the significance of former planet candidate Alpha Cen Bb, the results presented in this chapter have also enjoyed wide coverage in the press and popular media.

# Abstract

We re-analyse the publicly available radial velocity (RV) measurements for $\alpha$ Cen B, a star hosting an Earth-mass planet candidate, $\alpha$ Cen Bb, with 3.24 d orbital period. We demonstrate that the 3.24 d signal observed in the $\alpha$ Cen B data almost certainly arises from the window function (time sampling) of the original data. We show that when stellar activity signals are removed from the RV variations, other significant peaks in the power spectrum of the window function are coincidentally suppressed, leaving behind a spurious yet apparently-significant 'ghost' of a signal that was present in the window function's power spectrum *ab initio*. Even when fitting synthetic data with time sampling identical to the original data, but devoid of any genuine periodicities close to that of the planet candidate, the original model used to infer the presence of $\alpha$ Cen Bb leads to identical conclusions: viz., the $3\sigma$ detection of a half-a-metre-per-second signal with 3.236 d period. Our analysis underscores the difficulty of detecting weak planetary signals in RV data, and the importance of understanding in detail how every component of an RV data set, including its time sampling, influences final statistical inference.


## 4.1 Introduction

Astronomical observations of the K1 V star $\alpha$ Cen B have been conducted for over 150 years. Astrometric observations date back to the 1800s, and the RV velocities of both $\alpha$ Cen A and $\alpha$ Cen B have been tabulated since 1904 (Pourbaix et al., 2002). The announcement by Dumusque et al. (2012) of the detection of a planet around the modestly-active $\alpha$ Cen B generated exceptional excitement in the astronomy community: at the time of announcement, the claimed planet $\alpha$ Cen Bb would have been the closest exoplanet to Earth ever discovered (distance: 1.34 pc), and the lowest-minimum-mass planet detected around a Solar-type star.[1] The claimed

[1] As noted in Section 2.1.1, as of June 2017, the closest known exoplanet to Earth is Proxima Cen b, whose discovery was announced almost a year after the results in this chapter were first published online.

planet has an orbital period of $3.2357 \pm 0.0008$ d, and an estimated minimum mass of $1.13 \pm 0.09$ $M_\oplus$, inferred from an RV semi-amplitude of $0.51 \pm 0.04$ m s$^{-1}$.

Dumusque et al. (hereafter D12) obtained 459 HARPS RV data points, along with ancillary line width (full width at half maximum), bisector inverse slope (BIS) and chromospheric activity ($\log R'_{\rm HK}$) time series, over a period of 4 years. The RVs are dominated by a long-term linear trend, due to the orbit of $\alpha$ Cen B around the centre of mass of the $\alpha$ Cen binary system. Once this trend is subtracted, a gradual rise and fall over the 4-yr span of the observations is evident, as well as variability on shorter time-scales.

D12 used a variety of mathematical transformations to try to filter out many sources of RV variance, including starspots, photospheric granulation, and binary motion due to the presence of companion star $\alpha$ Cen A. The amplitude of the binary-motion signal they removed was of the order of hundreds of m s$^{-1}$; the combined amplitude of nuisance signals ascribed to stellar activity was of order a few m s$^{-1}$; and the final, putative planetary signal they isolated had a semi-amplitude of $\sim 0.5$ m s$^{-1}$. For comparison, HARPS' long-term precision is 0.8 m s$^{-1}$ (Mayor et al., 2003).

The detection, however, was a contentious one, with some voicing reservations about the modelling approach used to infer the planet's presence (e.g., Hatzes, 2012, 2013). For example, short-term rotational activity was modelled using sinusoidal waves at the rotational period of the star, and a varying number of harmonics – essentially, a truncated Fourier series. This is not necessarily a good approximation in the case of an *evolving* rotation signal (e.g., Lanza et al., 2001), and subtracting such a model from the $\Delta$RV time series opens the possibility of inadvertently introducing periodic signals, or harmonics thereof, into the power spectrum. Secondly, available activity-sensitive time series (BIS and $\log R'_{\rm HK}$) were not jointly modelled with the $\Delta$RV time series, even though the rotational signal was strongest in the $\log R'_{\rm HK}$ time series. Instead, a low-pass smoothing filter was applied to the $\log R'_{\rm HK}$ time series; this smoothed signal was noted to look similar to the $\Delta$RV time series, and then used to mitigate the effects of long-term activity in the $\Delta$RV time series. Thirdly, each of the four observing seasons was fitted independently of the others,

possibly discarding information about long-term correlations in the time series, and resulting in a more complex model than would likely have been the case if fitting were performed for all seasons simultaneously. The model used to fit the $\Delta$RV time series alone, without a planet, contained 23 free parameters, which raises the question of possible over-fitting. These potential reservations aside, however, the signal D12 extracted certainly did seem to bear many of the hallmarks of a real planetary signal, including a coherent phase over four observing seasons, and low false alarm probability (FAP).

In the previous chapter, we attempted (Rajpaul et al., 2015, hereafter R15) to address some of the aforementioned reservations by applying to the D12 data set a GP framework that we developed to model RV time series jointly with ancillary, activity-sensitive time series, allowing the activity component of an RV signal to be disentangled from dynamical components. This modelling framework uses GP draws and derivatives thereof as basis functions to model available time series in a flexible, data-driven way, rather than e.g. sinusoids or other simplistic parametric models. Moreover, the entire framework is accommodated very naturally within the broader framework of Bayesian inference, allowing uncertainties to be handled in a principled way: overly-complex models are automatically penalized, nuisance parameters can be marginalized over, and rigorous model comparisons can be performed. Using this framework, we showed that the observed RV variations could be satisfactorily modelled as stellar activity alone, without requiring a planet. Yet we did not claim that $\alpha$ Cen Bb does *not* exist, in part because that analysis did not provide any definitive explanation for the origin of the 3.24 d signal uncovered by D12.

We present here a re-analysis of D12's data, and argue that the 3.24 d signal almost certainly arises from the window function (time sampling) of the original data.

## 4.2 An apparent planetary signal

To probe the origins of the planetary signal claimed by D12, we implemented exactly the planet-free model published by D12, and used it to fit the original observations. We optimised model parameters using multiple runs of a Nelder-Mead

algorithm, and obtained marginally better goodness-of-fit statistics than those published by D12 (likely because D12 approximated the optimisation problem as a linear one; X. Dumusque, *pers. comm.,* 2015). Nevertheless, our RV residuals corresponded almost exactly to those published by D12, and the 3.24 d signal appeared, with the expected $> 3\sigma$ significance.

Studying the power spectra of the *periodic* components of D12's model (see Section 4.4), in combination with varying levels of noise, proved inconclusive; none of these components appeared to give rise to 3.24 d periodicities. This was consistent with D12's own bootstrap tests to ascertain whether the 3.24 d signal was introduced during the process of eliminating stellar signals. We next ran a Bayesian model comparison test of D12's baseline model versus D12's planet model. We placed uniform priors on all parameters, with prior domains set arbitrarily to 10 times the $\pm 1\sigma$ confidence intervals published by D12; model evidences ($\mathcal{Z}$) were computed using the MultiNest nested-sampling algorithm (Feroz and Hobson, 2008; Feroz et al., 2009, 2013). The planet model was favoured very strongly over the simpler baseline model, with a Bayes factor of the order of $10^5$. Moreover, $\log \mathcal{Z}$ for the planet model increased approximately linearly when artificially increasing the number of data points used in the modelling (up to a maximum of 459).

However, some further tests suggested – without proving outright that $\alpha$ Cen Bb does not exist – that a planetary origin for the 3.24 d signal is unlikely. These tests are sketched below.

First, we injected into D12's RV observations different 0.51 m s$^{-1}$ sinusoidal signals, with random phases, and periods uniformly distributed between $3.24 \pm 2.0$ d, excluding $3.24 \pm 0.2$ d. We considered $100,000$ such signals. In 87% of cases, the normalised least-squares/Lomb-Scargle (LS) power at the period of the injected signal was, after first subtracting the dominant binary signal from the data, higher than the power of the strongest signals between $3.24 \pm 0.02$ d. (For a given period $P$, the LS periodogram of a signal $f(t)$ represents a test statistic for the corresponding problem of testing between two hypotheses: *viz.* that $f(t)$ represents

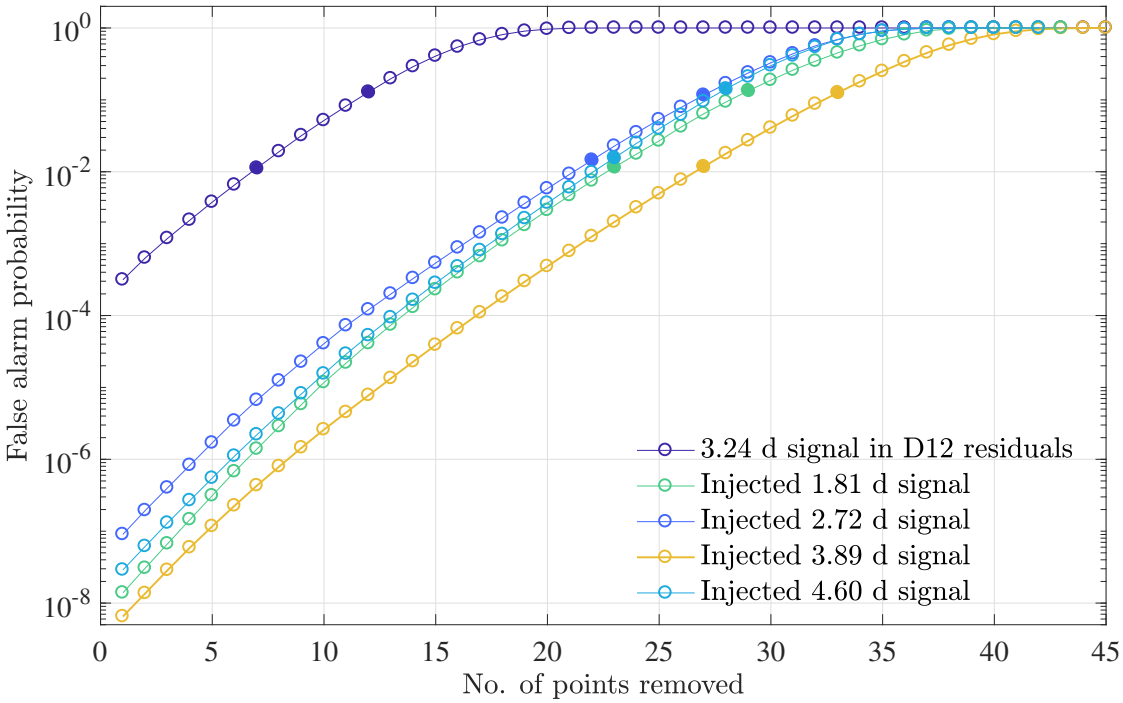


**Figure 4.1:** Illustration of a few 0.51 m $s^{-1}$ signals, injected into D12's published residuals, which are more robust to removal of data points than D12's signal. Solid markers correspond to points where signals first exceed 1% and 10% false alarm probability (FAP) thresholds.

white Gaussian noise, versus a coherent sinusoidal signal with period $P$, based on a fit using a least-squares metric.)

Secondly, we considered 100,000 sets of random Gaussian noise, drawn at times identical to D12's observations, and with standard deviations on each point corresponding to D12's HARPS error estimates. We found that 18% of these noise signals contained more normalised LS power specifically at $3.24 \pm 0.02$ d than was present in D12's binary-subtracted observations. Moreover, *all* of these time series of noise contained at least one signal between 2 d and 10 d with lower FAPs than any signals between $3.24 \pm 0.02$ d in D12's binary-subtracted data.

Thirdly, we found that the statistical significance of D12's 3.24 d signal could be destroyed by removing as few as 12 (non-random) points from the residuals. These points were distributed through all four observing seasons, and happened to correspond to unusually large amplitude residuals ($> 2.5$ m $s^{-1}$). On the other hand, it was a straightforward exercise to find injected signals with amplitude and

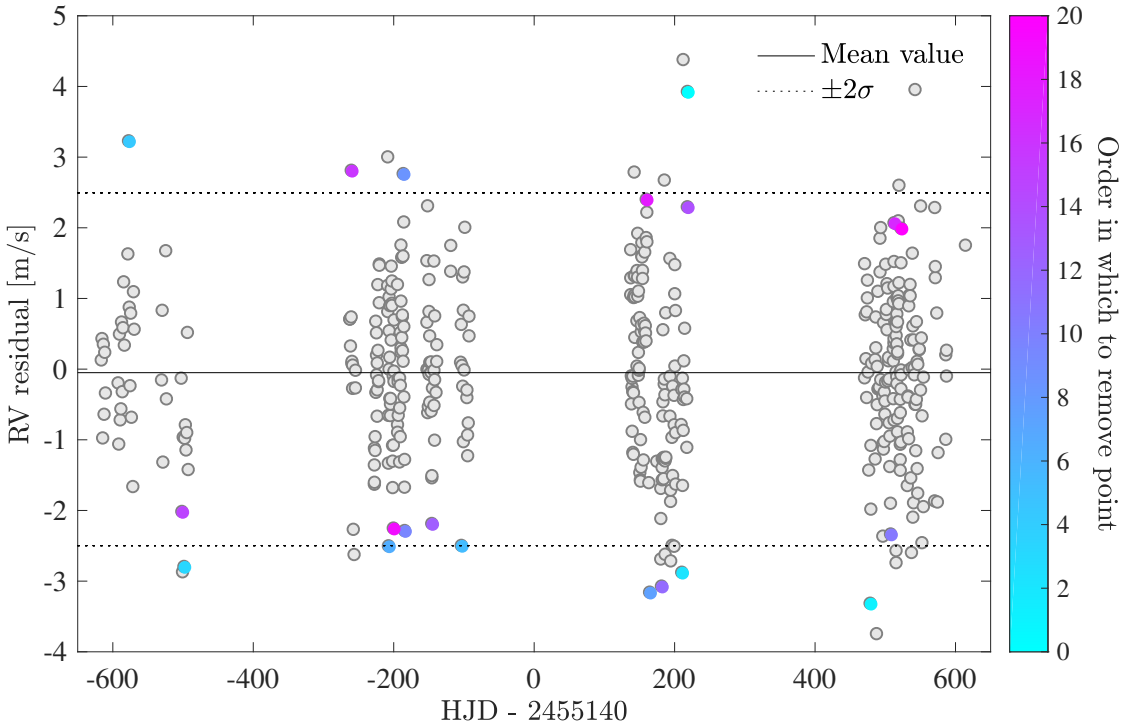


**Figure 4.2:** Illustration of the points which, when removed from D12's observations, 'destroy' the 3.24 d signal in the published residuals. The colour-coded order in which points are to be removed corresponds to Fig. 4.1; e.g. removing points 1–12 will suffice to boost the signal's FAP above 10%.

period similar to D12's signal, but which were twice as robust to data points being removed. See Fig. 4.1 and Fig. 4.2; cf. Hatzes (2013).

Fitting the original data using the framework described in R15 proved puzzling. Our best-fitting GP models had no 3.24 d signal in the residuals; curiously, however, we found that the presence of the 3.24 d signal seemed to be sensitive to how we fitted the long-term binary signal in the data (e.g. subtracting it before fitting the rest of the model, versus fitting everything simultaneously), as well as to whether we binned observations to one point per night (by default, we did *not* bin any data). This motivated us to look beyond our activity models – and in particular, to make a closer study of the window function for D12's observations.

## 4.3 A closer look at the window function

If a signal is sampled at discrete times, the so-called window function (i.e. vector of discrete sampling times) imprints periodicities on the signal, and the LS periodogram of the window function gives us information about the significance of these periodicities associated with sampling. At first glance, the LS periodogram of the window function for D12's data set is devoid of interesting features. It is dominated by a tight cluster of peaks around 1 d, arising due to periodicities in the observing times imposed by the Solar and sidereal day. The strongest feature (at 0.9978 d) has an LS FAP tens of thousands of times lower than any other peak at periods shorter than a few months.

However, when masking off this 'over-saturated' region of the power spectrum, the following become apparent. First, the strongest feature between (just over) 1 d and approximately two weeks is a complex of peaks around 3.249 d, including a peak at 3.236 d. Secondly, on time-scales between (just over) 1 d and approximately three months, the three strongest peaks in the power spectrum are consistent with $P_{rot} \pm 2\sigma$ or $P_{rot}/2 \pm \sigma$, where $P_{rot}$ is a stellar rotation period for $\alpha$ Cen B derived by D12, and $\sigma$ is the uncertainty on the period (D12 derived slightly different rotation periods for different observing seasons, which they attributed to differential rotation of $\alpha$ Cen B). See Fig. 4.3.

It seems a remarkable coincidence that the peaks around 3.249 d correspond (within 20 min) to the period of D12's planet candidate, *and* that they happen to be the most significant peaks in a sizeable region of the power spectrum. Ordinarily, the formal FAPs ($> 99\%$) of these peaks might lead to them being ignored. However, it is a further coincidence that the power spectrum contains significant structure at periods consistent with the stellar rotation period, $P_{rot}$, and its first harmonic. By 'whitening' the power spectrum and removing these periodicities (e.g. by fitting and removing sinusoids at these periods, exactly as was done with D12's model), the statistical significance of the peaks at 3.24 d can be enhanced. Indeed, as Fig. 4.4 confirms, the $\sim 3.24$ d signals can readily be boosted above a $3\sigma$ significance threshold simply by suppressing 'rotational activity' signals in the window function.

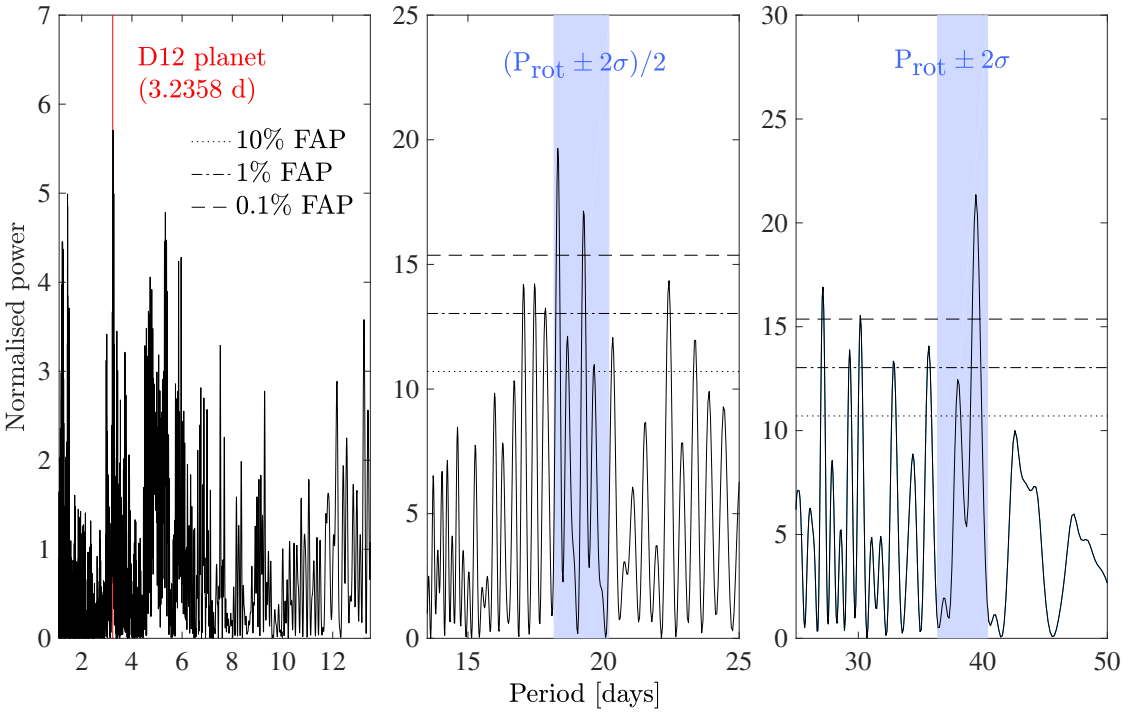


**Figure 4.3:** LS power spectrum of the window function of D12's observations. The arg max of the power spectrum on three non-trivial domains corresponds to the period of D12's planetary candidate (left panel), the estimated rotation period of $\alpha$ Cen B (right), and the first harmonic of the rotation period (middle).

Of course, the window function contains no real rotational activity signals, except as an unfortunate numerical coincidence.

To probe further the extent to which the window function gives rise to $\sim 3.24$ d periodicities, we computed the LS power spectra for $10,000$ polynomials with degree $< 10$, and uniformly-random coefficients, all with time sampling identical to D12's observations. In *every* case, the strongest periodicity between about 1.1 d and 13.5 d was found at 3.249 d. When fitting and subtracting a polynomial from one of these randomly-generated polynomial 'observations,' the power spectrum was flattened out (as expected, since exact fits were possible). However, when first adding non-white noise to the polynomials (e.g. random walk noise, periodic and quasi-periodic signals, GP draws, or combinations thereof) and then fitting and subtracting a polynomial, we observed different behaviour.

In cases where the injected noise signals 'drowned out' the polynomial signals (i.e. the window function imprinted on a polynomial), the power spectra of the residuals

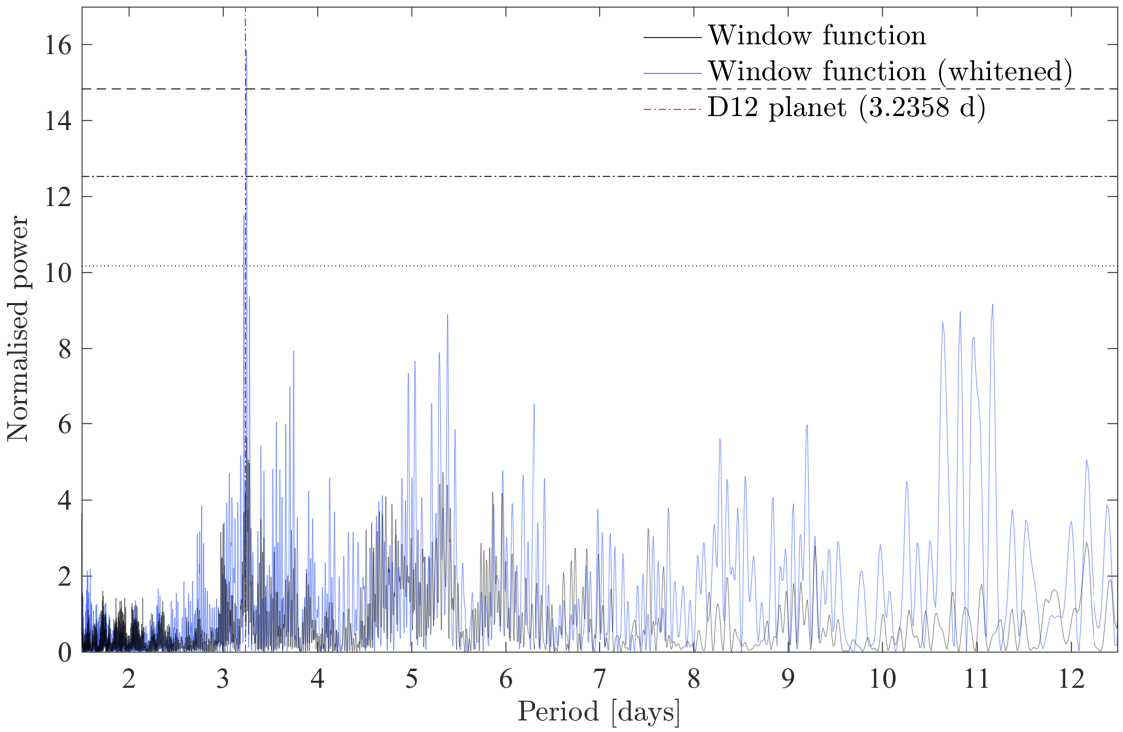


**Figure 4.4:** LS power spectrum of the 'whitened' window function, obtained by removing from the window function four fitted sinusoids with periods constrained to lie between $P_{rot} \pm 2\sigma$ or $P_{rot}/2 \pm \sigma$ (ranges taken from D12's best-fitting model), and unconstrained amplitudes and phases. For comparison, D12's activity model included eight different sinusoidal components. The dotted/dashed black lines correspond, from top to bottom, to 0.1%, 1%, and 10% FAP thresholds, as in Fig. 4.3.

tended to match closely those of the injected signals. However, in other cases where the injected signals' contribution to the total signal variance was much smaller than the polynomial contribution, the original polynomials could be fit well, but *not* perfectly. This meant that the residuals still contained a small signal bearing the imprint of the window function, so very often in these cases the most significant period (usually between 1.1 d and 13.5 d) in the residuals turned out to be 3.249 d, or a period very close to (within 1% of) 3.249 d. This was not surprising, as our earlier test confirmed that virtually *any* polynomial with the same sampling as D12's observations would have in its power spectrum a 3.249 d signal as a local maximum. Slight ($< 1\%$) deviations from 3.249 d in the latter tests could be attributed to biases in the power spectrum introduced by the injected signals. Furthermore, when suppressing other, more significant components of the window function (as before;

see Fig. 4.4), the significance of the $\sim 3.24$ d signals could be boosted.

## 4.4 The true origin of the planetary signal

D12's baseline model used to fit the RVs contained three components, which modelled, respectively: (1) the binary signal; (2) long-term magnetic activity; and (3) stellar rotational activity. Component (1) comprised a polynomial, second order (quadratic) in time, with the same three coefficients used for all observing seasons. Component (2) was proportional to the low-pass filtered $\log R'_{\rm HK}$ series; the only free parameter here was a scaling factor to convert the $\log R'_{\rm HK}$ to $\Delta$RV variations. Component (3) included sinusoidal terms, different for each season; in total, it contained 19 free parameters. D12 also considered a model with an added sinusoid term, (4), to describe a possible zero-eccentricity planetary contribution to the $\Delta$RV time series. This took the form $\Delta{\rm RV}_{\rm planet} = K_1 \sin(2\pi t/P) + K_2 \cos(2\pi t/P)$; equivalently, $\Delta{\rm RV}_{\rm planet} = K \sin(2\pi t/P + \phi)$. All components of the model were fitted simultaneously.

When computing the LS power spectrum of component (1) of the best-fitting model, i.e. the polynomial component, the normalized LS power at $3.25 \pm 0.1$ d was found to be 99.3% of the normalized power from the raw $\Delta$RV observations (which included a binary signal, long-term magnetic activity variations, rotational activity, and unknown contributions). In fact, we had to inject into the raw $\Delta$RV observations sinusoids with semi-amplitudes of the order of tens of m s$^{-1}$ (assuming periods of the order of a few days) to obtain comparable LS power to that observed at $3.25 \pm 0.1$ d.

This suggests the following scenario: a signal at $\sim 3.24$ d was present in the original observations; however, it arose not from a planet, but from the window function. As with the tests described at the end of Section 4.3, D12's imperfect model meant that the imprint or 'ghost' of the window function remained in the model residuals. Moreover, because other significant components of the window function happened to be suppressed when removing stellar activity, i.e. through model component (3), the $\sim 3.24$ d signals end up being *boosted* in the window function's imprint on the residuals.

To test this scenario, we generated synthetic observations as follows. Components (1) and (2) of D12's original model, with free parameters set at their MAP values, were taken to represent (a) binary, and (b) long-term magnetic variations. We then used the framework from R15 to train a GP with quasi-periodic covariance kernel on the 'observed' rotational activity ($\Delta$RV observations minus binary and magnetic activity models), in tandem with $\log R'_{\rm HK}$ and BIS observations. We took the posterior mean of the GP for the $\Delta$RV series to represent rotational activity, (c). Finally, components (a), (b), and (c) were co-added with white Gaussian noise to form a set of synthetic observations.

We avoided using component (3) of D12's model to synthesise the rotational activity: otherwise the synthetic observations would have been generated entirely by D12's original model, and it would have been naïve to expect anything but an excellent fit, with white, Gaussian residuals (see Section 4.3). Instead, using a GP allowed us to synthesise observations with the same covariance properties – (quasi)periodicities, evolution time-scales, smoothness, noise levels, etc. – as the observed rotational activity, regardless of the extent to which D12's simple harmonic model might have failed to describe such variations. Furthermore, the GP framework meant that any periodicities not also present in the ancillary time series (i.e. $\log R'_{\rm HK}$, BIS) would *not* be captured by the $\Delta$RV model. We confirmed that the power spectrum of component (c) of our model was free of any significant periodicities on time-scales of a few days; and it certainly contained no signals near 3.24 d.

We next used D12's model to fit our baseline set of synthetic observations – to be quite clear, a set of observations synthesized only from non-planetary components of D12's model, a conservative activity model for the original observations, and white noise. As before, uninformative priors were placed on all parameters in D12's model, and parameter inference was performed using MultiNest.

The strongest feature in the power spectrum of the residuals from the MAP fit was a significant peak at $\sim 3.236$ d, with estimated FAP of 0.04%; see Fig. 4.5. When adding a zero-eccentricity Keplerian (sinusoidal) component to D12's overall model, we obtained an RV semi-amplitude for the Keplerian of $K = 0.46 \pm 0.03$ m s$^{-1}$, and

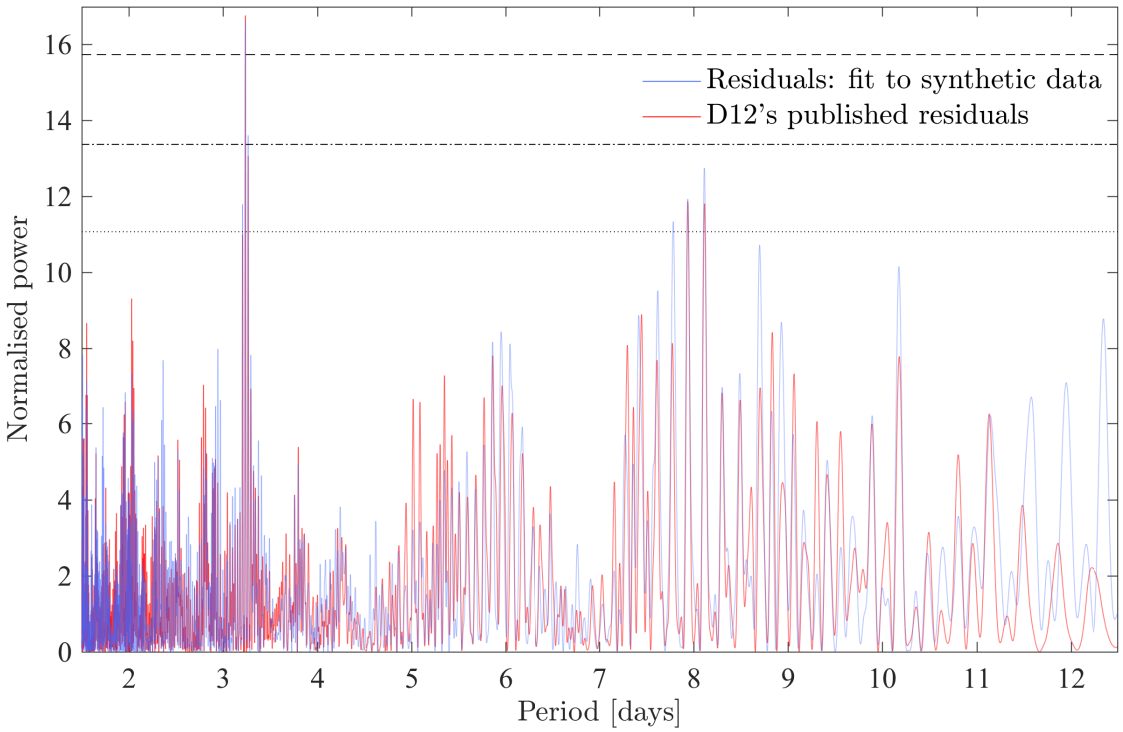


**Figure 4.5:** LS power spectrum of the residuals from synthetic, planet-free data (with realistic covariance properties, and time sampling identical to D12's data), obtained from a MAP fit with D12's original model. In spite of all, a strong signal appears at 3.2358 d, corresponding to D12's planet candidate. The dotted/dashed black lines correspond, from top to bottom, to 0.1%, 1%, and 10% FAP thresholds, as in Fig. 4.4.

a period of $3.2358 \pm 0.0002$ d. As before, the more complex sinusoidal model was strongly favoured over the baseline model, with a Bayes factor of $\sim 10^5$.

We repeated these tests 10 times using different 'observations,' synthesized by perturbing the MAP parameters from components (1) and (2) of D12's model, within their $1\sigma$ confidence intervals; by using different GP draws for the rotational activity model, within a $1\sigma$ envelope of the GP posterior mean function; and by using different seeds when generating additive white Gaussian noise. In each case, a significant signal appeared at $\sim 3.236$ d in the residuals, and when including a Keplerian in the model, a MAP period consistent with 3.2358 (within $1\sigma$) was obtained. However, a range of semi-amplitudes was obtained – the smallest was $K = 0.33 \pm 0.03$ m s$^{-1}$, and the largest, $K = 0.53 \pm 0.04$ m s$^{-1}$. For comparison, the planet claimed by D12 had best-fitting parameters $K = 0.51 \pm 0.04$ m s$^{-1}$, and $P = 3.2357 \pm 0.0008$ d.

It is exceptionally implausible that synthetic data simply with the same time sampling and activity-related covariance properties as D12's data set, but no planet, would consistently lead to the detection of the same planet claimed by D12, if indeed D12's planet were real. We thus conclude that the 3.24 d signal that D12 originally recovered was most likely nothing more than a 'ghost' of the window function, amplified during the modelling process to non-trivial significance levels.

## 4.5 Discussion and conclusions

We have demonstrated that the 3.24 d signal observed in the $\alpha$ Cen B data almost certainly arises from the window function of the original data. When stellar activity signals are filtered out from the RV variations, other significant peaks in the power spectrum of window function are coincidentally suppressed, leaving behind a spurious yet apparently-significant 'ghost' of a 3.24 d signal that was present in the window function's power spectrum *ab initio*.

We showed that even when fitting synthetic data that have the same time sampling as the original data, but that are by design devoid of any genuine periodicities close to the period of the planet candidate, the original model used to infer the presence of $\alpha$ Cen Bb leads to identical conclusions: *viz.*, the $3\sigma$ detection of a half-a-metre-per-second signal with 3.236 d period. Central to our analysis was the use of R15's GP framework, for generating realistic synthetic data for fitting with D12's model.

Our analysis underscores the difficulty of detecting weak planetary signals in RV data, and the importance of understanding in detail how every component of an RV data set, including its time sampling, influences final statistical inference. In future attempts at detecting low-mass exoplanets in RV data sets, it would be prudent to study the window function of the observations carefully. A planet candidate should be treated with extreme circumspection if its period happens to coincide closely with an arg max of the window function's power spectrum on some non-trivial domain; in the case of the D12 data, 3.24 d happened to be the arg max of the window function's power spectrum on an interval between about 1 d and

2 weeks. Additionally, D12's data set was particularly pathological because the window function happened to contain periodicities that coincided with the stellar rotation period of $\alpha$ Cen B, and its first harmonic; when these signals were filtered out, the significance of the 3.24 d signal was preferentially boosted.

We alluded to a number of other tests we believe worth carrying out when considering the reliability of planet detections from noisy, discretely-sampled signals. These include using the same model used to detect the planet instead to fit synthetic, planet-free data (with realistic covariance properties, and time sampling identical to the real data), and checking whether the 'planet' is still detected; comparing the strength of the planetary signal with similar Keplerian signals injected into the original observations; performing Bayesian model comparisons between planet and no-planet models; and checking how robust the planetary signal is to data points being removed from the observations. As this study has demonstrated, simply considering the FAP of a signal in a model's residuals, or (somewhat equivalently) checking whether a signal has a coherent phase over a full set of observations, is clearly insufficient. Even rigorous, Bayesian model comparisons do not necessarily shed light on the *origins* of a planet-like signal.

We propose to carry out, in the near future, a systematic study of the properties of spurious yet coherent signals that can arise in noisy data sets with real (e.g. HARPS-like) observing calendars. This will hopefully inform best practices for identifying or mitigating such signals; it might also prove relevant to already-published cases of planets whose existence was supported by rigorous model comparison tests, yet not by follow-up observations. Certainly, when carrying out future observations of $\alpha$ Cen B, it would be worth ensuring that the window function of the proposed observations did not contain significant power at 3.24 d, or $P_{rot} \pm 2\sigma$ and harmonics thereof.

*As knowledge advances, we are able to invent better and better models, which reproduce more and more features of the real world, more and more accurately. Nobody knows whether there is some natural end to this process, or whether it will go on indefinitely.*

— Edwin Thompson Jaynes, from *Probability Theory: The Logic of Science* (Jaynes, 2003).

# 5

# Reweighing a planet: Kepler-10c

## Contents



This chapter is based on the Letter entitled *Pinning down the mass of Kepler-10c: the importance of sampling and model comparison* (Rajpaul et al., 2017), due to be published in MNRAS Letters. Here, the GP framework from Chapter 3 is used to re-characterise the well-studied Kepler-10 system, and to resolve a mystery surrounding the mass of planet Kepler-10c.

## Abstract


Kepler-10c has proven something of an exoplanetary enigma. Initial RV characterisation suggested a mass of $\sim 17$ $M_\oplus$, which was remarkably high for a planet with radius 2.32 $R_\oplus$; further observations and subsequent analysis hinted at a (possibly much) lower mass, but masses derived using RVs from two different spectrographs

(HARPS-N and HIRES) were incompatible at a $3\sigma$ level. We demonstrate here how such mass discrepancies may readily arise from sub-optimal sampling and/or neglecting to model even a single coherent signal (stellar, planetary, or otherwise) that may be present in RVs. We then present a plausible resolution of the mass discrepancy, and ultimately characterise Kepler-10c as having mass $7.37^{+1.32}_{-1.19}$ $M_\oplus$, and mean density $3.14^{+0.63}_{-0.55}$ g cm$^{-3}$.

## 5.1 Introduction

Kepler-10 (KOI-72; hereafter K-10 for short) is a slowly-rotating, Sun-like star that exhibits little stellar activity (Dumusque et al., 2014b, hereafter D14). It is known to host at least two planets, *viz.* Kepler-10b, and Kepler-10c.

Stony-iron world Kepler-10b (hereafter K-10b) – with orbital period 0.84 d, radius 1.48 $R_\oplus$, and mass $\sim 4$ $M_\oplus$ – was the first unambiguously rocky exoplanet to be discovered, and also the first super-Earth discovered around a Sun-like star (Batalha et al., 2011).

Kepler-10c (hereafter K-10c) – with orbital period 45.29 d, and radius 2.32 $R_\oplus$ – has proven more enigmatic. Following its discovery and statistical validation as a planet (Batalha et al., 2011; Fressin et al., 2011), D14 used 148 HARPS-N radial velocities (RVs) spanning two observing seasons to infer a mass of $17.2 \pm 1.9$ $M_\oplus$. Given K-10c's radius, this was a striking result. Most planets with radii 2.0–2.5 $R_\oplus$ have masses significantly lower than 17 $M_\oplus$, with a weighted mean mass of 5.4 $M_\oplus$ (Weiss and Marcy, 2014); Weiss and Marcy's empirical mass-radius relation for planets between 1.5 and 4 $R_\oplus$, *viz.* $M_p/M_\oplus = 2.69(R_p/R_\oplus)^{0.93}$, predicts a mass of 5.8 $M_\oplus$ for K-10c. D14 interpreted the composition of K-10c as being mostly rock by mass, and regarded the planet as being the first evidence of a class of more massive solid planets with longer orbital periods.

Weiss et al. (2016, hereafter W16) built on the work of D14, adhering closely to the techniques employed by the latter authors, but adding 72 RVs from Keck-HIRES to the analysis, resulting in a combined RV baseline of 6 years. Since it has been well established that both the HIRES and HARPS-N spectrometers are independently

capable of accurate and precise measurement of low-amplitude planetary signals (c.f. e.g. Howard et al., 2013; Pepe et al., 2013), it was a great surprise when W16 inferred a mass for K-10c of $5.69^{+3.19}_{-2.90}$ $M_\oplus$ (fitted RV semi-amplitude $K_c = 1.09 \pm 0.58$ m s$^{-1}$) using the HIRES RVs alone, which was incompatible with D14's estimate of $17.2 \pm 1.9$ $M_\oplus$ ($K_c = 3.26 \pm 0.36$ m s$^{-1}$) using the HARPS-N RVs alone.

W16 concluded that some additional, time-correlated signal, possibly from stellar activity or undetected planets, was present and led to the discrepant mass estimates for K-10c. This claim was supported by (i) the fact that masses inferred using RVs from *either* instrument were found to be time-dependent, and (ii) $> 5\sigma$ evidence for transit timing variations (TTVs) of K-10c (Kipping et al., 2015). W16 found that dynamical solutions including a third planet candidate, KOI-72.X, were very strongly favoured over a two-planet solution (based on Bayesian Information Criterion differentials); the TTVs and RVs were consistent with KOI-72.X having an orbital period of 24, 71, or 101 d, with 101 d being strongly favoured over the other periods. W16 inferred a likely mass of $\lesssim 7$ $M_\oplus$ for KOI-72.X, based on the best solutions from a partial exploration of the dynamical parameter space, with the parameters of K-10b fixed.

Even when including a third planet in their models, however, W16 were not able to reconcile the HIRES and HARPS-N masses for K-10c, so settled on a 'compromise' mass for K-10c of $13.98 \pm 1.79$ $M_\oplus$. We suggest, at the outset, that the observed $3\sigma$-incompatibility between the HARPS-N and HIRES estimates for K-10c's mass points to an inadequate model under which at least one (if not both) of the inferred masses is incorrect, and that the true mass need not lie in the middle of the two incompatible mass posteriors.[1]

[1]To take a more intuitive example, consider the forecasts (i) that $13 \pm 2$ mm of rain will fall tomorrow, and (ii) that only $1 \pm 1$ mm of rain will fall tomorrow. No matter how much or little it actually rains, at least one (though possibly both) of these mutually-incompatible claims must turn out to be both overconfident and incorrect, suggesting a need for a better weather model(s).

## 5.2 Double trouble: imperfect model meets inadequately sampled signal

To shed light directly on the effects of (i) sub-optimal sampling and (ii) inference based on an imperfect physical model, consider synthetic RV data sets $\{(t_i, y_i)|i = 1, 2, \ldots, N\}$ generated as follows:

$$y_i = \sum_{j=1}^{M} K_j \sin(\tfrac{2\pi t_i}{P_j} + \phi_j); \tag{5.1}$$

$y_i$ may be interpreted as the combined RV signal at time $t_i$ due to $M$ planets on zero-eccentricity orbits around a star. For planet $j$, the associated RV amplitude $K_j$ would be determined by the planet's mass and inclination (assuming known stellar mass); $P_j$ would correspond to the planet's orbital period; and $\phi_j$ would be determined by the planet's orbital phase in some coordinate system.

Suppose we set $M = 3$, and let the periods of the three mock planets be $P_1 = 0.84$ d, $P_2 = 45.29$ d, and $P_3 = 101.36$ d; $P_1$ and $P_2$ correspond to the known orbital periods of K-10b and K-10c, respectively, while $P_3$ corresponds to the most likely orbital period (per W16) for planet candidate KOI-72.X. We further set $K_1 = K_2 = 2.5$ m s$^{-1}$, and $K_3 = 1.0$ m s$^{-1}$; the values for $K_1$ and $K_2$ are roughly the average of the RV semi-amplitudes of both K-10b and K-10c variously reported in the literature, while the value for $K_3$ is based on the most likely mass for KOI-72.X reported by W16. Lastly, instead of fixing the mutual phases of the planets, let us generate synthetic data by evaluating equation (5.1) over a grid of phases $(\phi_1, \phi_2, \phi_3) \in \Phi^3$ where $\Phi = [0, 0.1, 0.2, \ldots, 2\pi]$.

Now consider the problem of using such synthetic data to infer the value of $K_2$, i.e. the RV semi-amplitude of the second mock planet, using both a 2-planet and a 3-planet model, and with the synthetic signals sampled discretely using the following calendars:

1. the real HIRES observing calendar ($N = 72$) for K-10;
2. the real HARPS-N observing calendar ($N = 148$) for K-10;
3. the combined HARPS-N/HIRES observing calendars; and

4. 220 uniformly-spaced observations with a $\sim 6$ yr baseline (see Section 5.3 for more details).

Thus we will produce an estimate for $K_2$, $\hat{K}_2$, using 2 different models (of orders $M = 2$ and $M = 3$), 4 different sampling patterns, and many different input values for $\phi_j$. In each case we assume that the periods and orbital phases of the planets are tightly constrained, as if performing RV follow-up of transiting planets, but that the planets' RV semi-amplitudes are not known *a priori*. For the 2-planet model we assume only the presence of the $P_1 = 0.84$ d and $P_2 = 45.29$ d planets, while for the 3 planet model we also assume the presence of a $P_3 = 101.36$ d planet.

We place flat priors on all free parameters, and to simplify computation, reduce the more general problem of finding posterior distributions to one of maximum likelihood (ML) estimation; our priors render ML estimates equivalent to MAP estimates. As a further convenience, we assume independent, identically-distributed Gaussian measurement errors on our synthetic data (with small, constant standard deviation), then use a standard downhill simplex algorithm with multiple starting points to locate ML parameters through least-squares fitting.

In Fig. 5.1 we present the results of this fitting exercise, showing ML estimators of $K_2$, $\hat{K}_2$, for a range of values of $\phi_2$ and $\phi_3$. We fix $\phi_1$ to permit visualisation in two dimensions.[2] The following striking conclusions emerge.

1. When using an inadequate model (i.e. $M = 2$) to fit the data, the inferred mass for the second mock planet is very sensitive to the mutual phases of that planet and the unobserved third planet, with $\hat{K}_2$ varying between $\sim 1.5$ and 3.5 m s$^{-1}$.
2. For almost all possible mutual phases of the second and the unobserved third planet, $\hat{K}_{2,\mathrm{HIRES}}$ and $\hat{K}_{2,\mathrm{HARPS}}$ differ[3] by up to $\sim \pm 1$ m s$^{-1}$ (in the worst cases) and typically by $\sim \pm 60$ cm s$^{-1}$. For about 60% of possible phase

[2]Given its short period $P_1 \ll P_2 < P_3$, the phase $\phi_1$ of the innermost mock planet does not have a significant effect on our inference about the properties of the other signals in our synthetic data. $K_1$ can be estimated more accurately and precisely than $K_2$, regardless of the phases of the other signals, and of the choice of observing calendar.

[3]In a few cases where we chose to explore full posterior distributions for $K_2$ (rather than just find ML estimators), estimates of $K_2$ obtained with the HARPS-N vs. HIRES sampling often disagreed at a $2\sigma$ level.

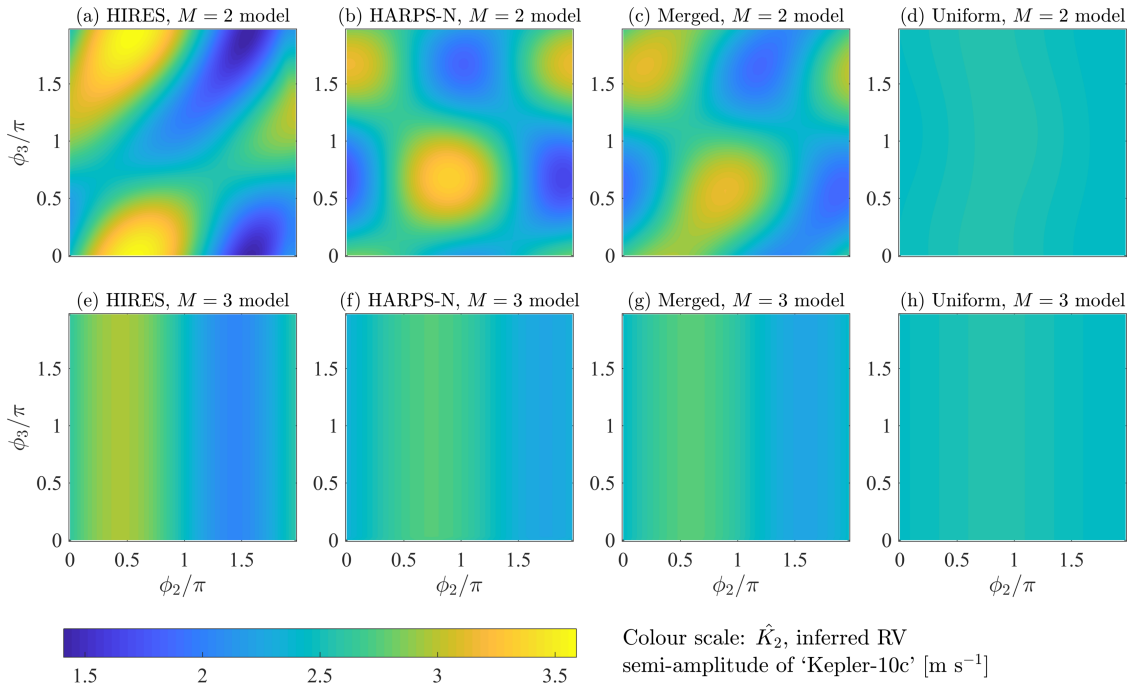


**Figure 5.1:** ML estimates for $K_2$ based on synthetic data comprising three sine waves, with 4 different sampling patterns (left to right, corresponding to sampling patterns 1–4 listed on pg. 140) and fitted with a 2-sine model (upper panels) and a 3-sine model (lower panels).

configurations, the HARPS-N and HIRES observing calendars either both result in overestimation or in underestimation of $K_2$ (i.e., the true value is *not* bracketed); for about 40% of configurations, one calendar will lead to $\hat{K}_2 > K_2$ while the other leads to $\hat{K}_2 < K_2$.

3. When combining the HIRES and HARPS-N observations, $\hat{K}_2$ interpolates the values predicted by the separate data sets, yet may still differ from $K_2 = 2.5$ m s$^{-1}$ by up to $\sim \pm 65$ cm s$^{-1}$.
4. Notably, even when using the inadequate $M = 2$ model but the uniform observing cadence, $|\hat{K}_2 - K_2| < 5$ cm s$^{-1}$, $\forall\ \phi_j$.
5. Equally notably, when using the correct $M = 3$ model, the inferred mass for the second planet is relatively *insensitive* to the observing calendar. Now we find $|\hat{K}_{2,\mathrm{HIRES}} - K_2| < 37$ cm s$^{-1}$ and $|\hat{K}_{2,\mathrm{HARPS}} - K_2| < 22$ cm s$^{-1}$; with the combined HARPS-N and HIRES observations, $|\hat{K}_2 - K_2| < 25$ cm s$^{-1}$; and with the calendar where $t_{i+1} - t_i =$ constant, $|\hat{K}_2 - K_2| < 4$ cm s$^{-1}$, $\forall\ \phi_j$.

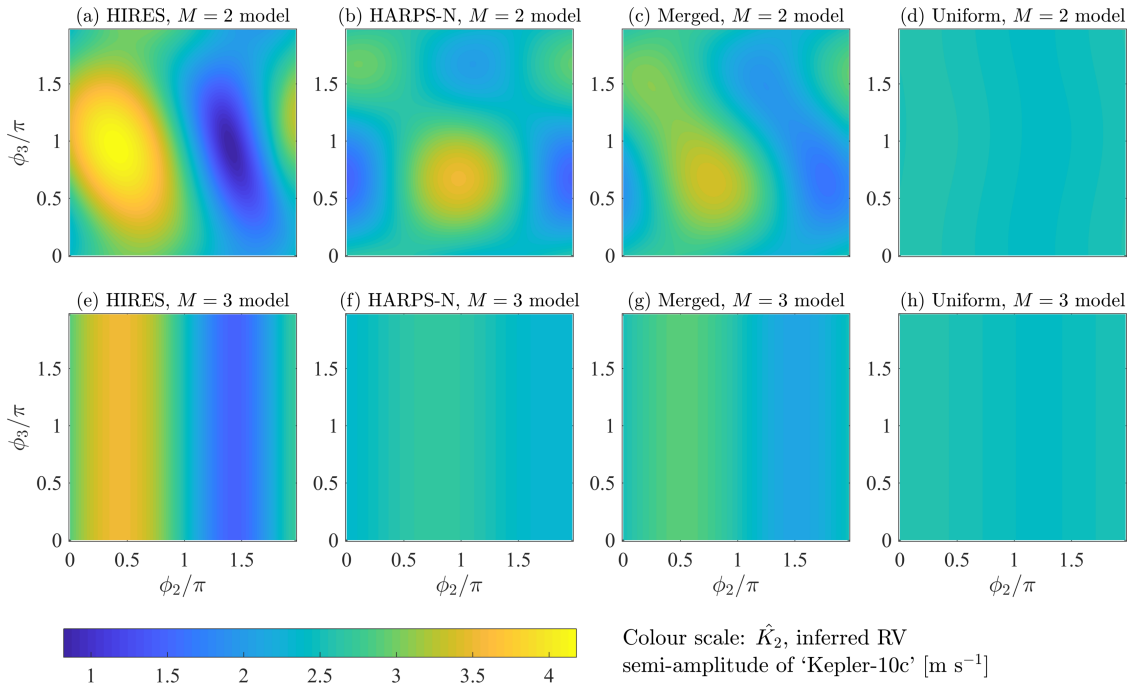


**Figure 5.2:** As for Fig. 5.1, but now with synthetic data comprising four rather than three sine waves (see text on pg. 143).

Suppose we add another signal into the synthetic RV data, with amplitude $K_4 = 1.0$ m s$^{-1}$ and period $P_4 = 55$ d, as a simplistic representation of a non-evolving stellar activity signal (we could use a GP to synthesise a quasi-periodic signal instead, but such sophistication is not required for the present illustration), then repeat the exercise of trying to estimate $K_2$. Our results are shown in Fig. 5.2.

Now, neither the $M = 2$ nor the $M = 3$ model is adequate in that neither accounts for a fourth periodic signal present in the data. Accordingly, we find that inference about $K_2$ becomes even more sensitive to sampling. $\hat{K}_{2,\mathrm{HIRES}}$ and $\hat{K}_{2,\mathrm{HARPS}}$ now differ by up to 2 m s$^{-1}$ under the $M = 2$ model, and by up to 1 m s$^{-1}$ under the $M = 3$ model. As before, however, when using the uniform observing cadence, the resultant uniform phase coverage allows remarkably robust inference about $K_2$ to be made: e.g., $|\hat{K}_2 - K_2| < 7$ cm s$^{-1}$ $\forall$ $\phi_j$, even under the $M = 2$ model.

Thus far we have established the prevalence of sizeable differences in $\hat{K}_2$ when fitting *simplistic* synthetic signals with HARPS-N vs. HIRES sampling, yet even larger differences will result when including in our synthetic data such details as

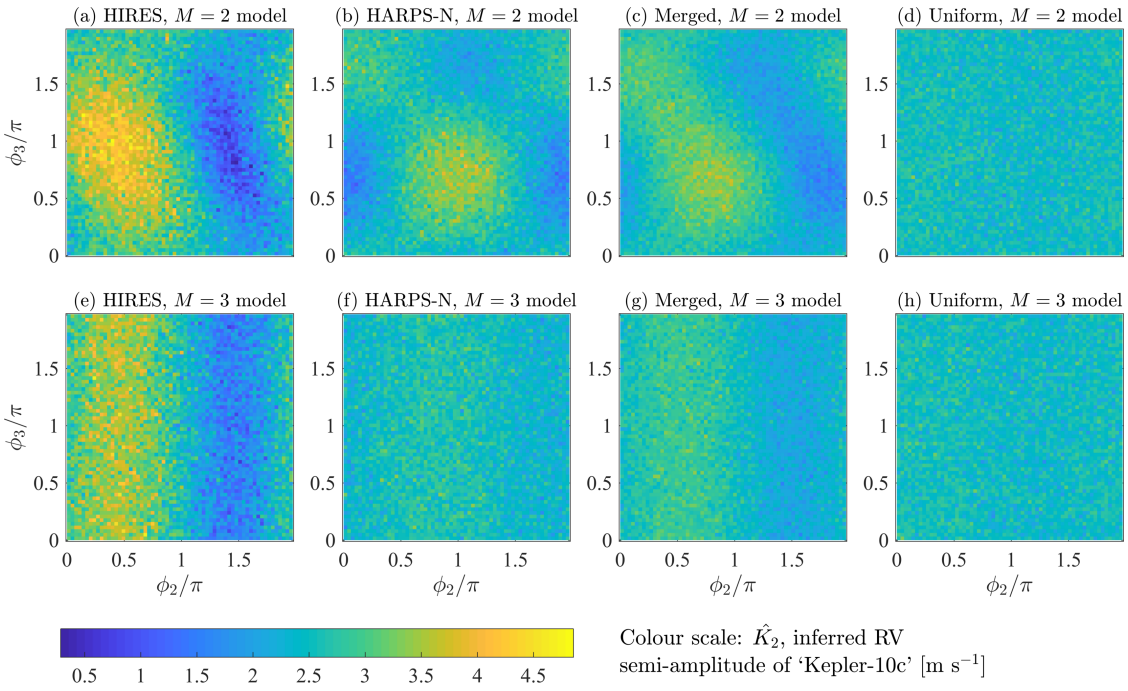


**Figure 5.3:** As for Fig. 5.2, but now with 2 m s$^{-1}$ photon noise added to the synthetic data (see text on pg. 144 for more details).

photon noise, quasi-periodic stellar activity signals, instrumental noise, multiple undetected planets, planets with non-circular orbits, possible dynamical interactions between planets, and more. For example, adding to our synthetic data white Gaussian noise at a level consistent with that estimated for the HARPS-N data set ($\sigma \sim 2$ m s$^{-1}$), then repeating the previous test, results in HARPS-N/HIRES discrepancies for $\hat{K}_2$ of up to 3 m s$^{-1}$ under the $M = 2$ model, and up to 2.3 m s$^{-1}$ under the $M = 3$ model; see Fig. 5.3.

The upshot is that using an inadequate physical model, and/or suboptimal sampling, can lead to incorrect conclusions about the masses even of planets whose other properties are well constrained – and even when we have hundreds of RVs at our disposal. Moreover, through our choice of real sampling patterns, and realistic values for $K_j$ and $P_j$, we have provided a plausible explanation for why W16 obtained discrepant masses for K-10c using HIRES vs. HARPS-N RVs. Specifically, the real K-10 RVs likely contained not only K-10b and K-10c's signals, but one or more other coherent signals (KOI-72.X, a stellar signal, etc., as indeed adduced by

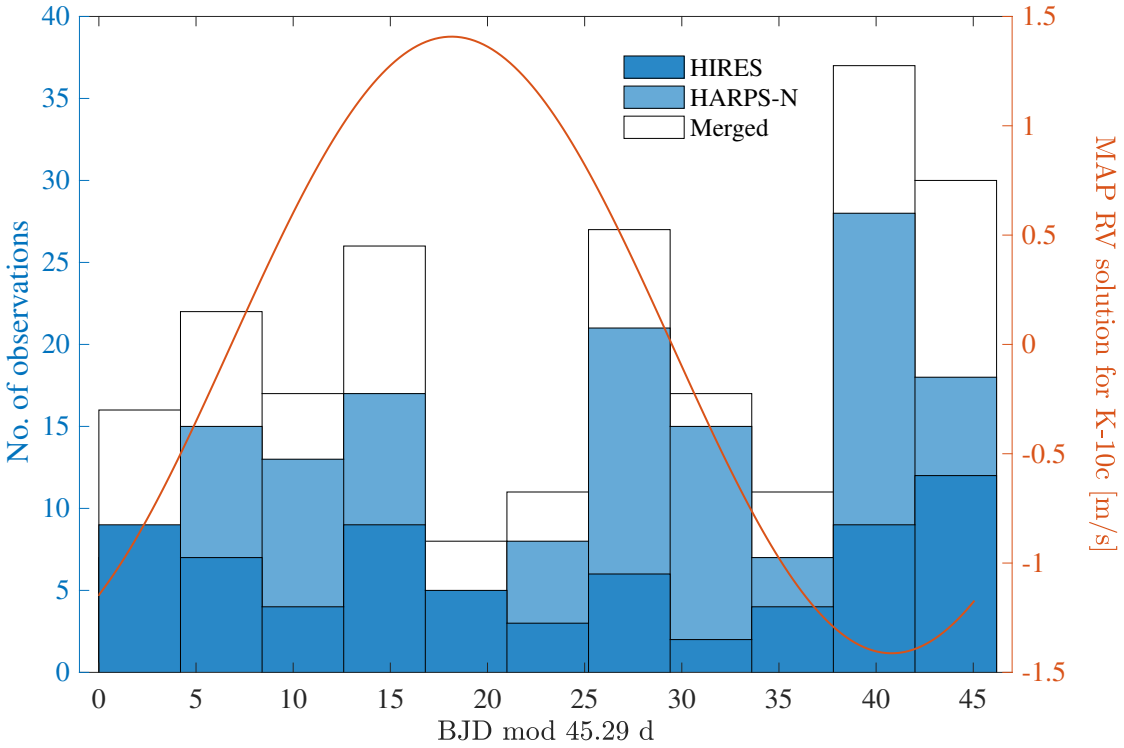


**Figure 5.4:** The uneven coverage of K-10c's orbital phase provided by existing HARPS-N and HIRES observations.

W16) which interfered constructively or destructively with the signals of the known planets. The sub-optimality of the phase coverage is easily checked by phasing the HIRES or HARPS-N observation times to the orbital period of K-10c; see Fig. 5.4.

In principle, accounting for the other signals jointly with those of the known planets (i.e., using a more appropriate physical model), and/or obtaining more observations to provide more complete phase coverage of K-10c's signal, could have mitigated the discrepancy.

## 5.3 On uniform cadence and phase coverage

Before turning to a possible resolution of the K-10c mass discrepancy, we now make a brief digression to illuminate the connection between uniform observing cadence and uniform observational coverage of a planet's orbital phase.

Consider first the simple case where a planet with known orbital period $P$ is observed $N$ times over the course of one full orbit. By the assumption of

uniform cadence, and through suitable choice of origin for the time coordinate, the observation times may be written $\{\frac{P}{N}, \frac{2P}{N}, \ldots, P\}$. If we now phase the observation times to the period of the planet, by computing the observation times modulo $P$, i.e. the remainder after division by $P$, we find $\{\frac{P}{N}, \frac{2P}{N}, \ldots, P\} \bmod P = P\left[\{\frac{1}{N}, \frac{2}{N}, \ldots, 1\} \bmod 1\right] = P\left[\{0, \frac{1}{N}, \ldots, 1 - \frac{1}{N}\}\right]$. Thus we see, trivially, that the phased observations will cover the planet's orbit uniformly; a $\sqrt{N}$-bin histogram (say) of the phased observations should contain $\sqrt{N}$ observations in each bin.[4]

Suppose now we extend this to the case of observations covering $M \in \mathbb{N}$ full orbits, again with uniform cadence, and $N$ observations per orbit. The $MN$ observation times are now $P\{\frac{1}{N}, \frac{2}{N}, \ldots, 1, \ldots, M\}$, which we can write as $P \cdot \bigcup_{i=1}^{M} T_i$, where $\bigcup$ indicates set union, and the $T_i$ are multisets[5] such that $T_1 = \{\frac{1}{N}, \frac{2}{N}, \ldots, 1\}$, and $T_{i+1} = T_i + 1$. If we phase these observation times to the period $P$, we find

$$\left[P \cdot \bigcup_{i=1}^{M} T_i\right] \bmod P = P\left[\bigcup_{i=1}^{M} T_i \bmod 1\right] = P \cdot \bigcup_{i=1}^{M} \{0, \tfrac{1}{N}, \ldots, 1 - \tfrac{1}{N}\}, \qquad (5.2)$$

since $T_j \bmod 1 = T_1 \bmod 1 = \{0, \frac{1}{N}, \ldots, 1 - \frac{1}{N}\}$. Thus the phased observation times will be uniformly-distributed over a full orbit, but with multiplicity $M$, so that a $\sqrt{N}$-bin histogram should contain $M\sqrt{N}$ observations in each phase bin.

Now we can consider two generalisations of the above cases. First, suppose the planet's orbital period is *not* known, but that the observing cadence is uniform, and that the observing baseline is significantly longer than a single orbital period. We can then truncate the observation times to just the $M \in \mathbb{N}$ full orbits covered by the observations, and apply the above arguments to show that this large subset of observations will be evenly-distributed when phased to the orbital period of the planet. The discarded observations covering an incomplete orbit *will* contribute to slight non-uniformity, but the effect should be small since $M \gg 1$. Importantly, the

[4] In this analysis, we argue that uniform phase coverage is *preferable* to very uneven phase coverage for constraining Keplerian signal amplitudes; this does not mean, however, that uniform phase coverage is necessarily optimal, as it might be possible to obtain even tighter constraints on amplitudes e.g. by sampling preferentially where the signals are largest. Further work is required to investigate these possibilities for Keplerian signals.

[5] A multiset generalises the concept of a set to allow multiple instances of the same element. For example, $\{1, 1, 2\}$ and $\{1, 2\}$ are different multisets though they are the same set.

approximately uniform phase coverage will apply to *all* observed planets, provided all have periods significantly shorter than the observing baseline.

Second, suppose the observational cadence is *not* strictly uniform. If the uniformity of orbital phase coverage is studied by making a histogram of the phased observation times, provided the deviations from uniform cadence are significantly smaller than the histogram bin widths, the number of items in each bin should not change compared to the case of strictly-uniform cadence.[6]

We have shown that (approximately) uniform observing cadence will lead to (approximately) uniform orbital phase coverage. Note that the latter, however, does not require or imply the former. As a counterexample, consider again the first case of observations made at times $\{\frac{P}{N}, \frac{2P}{N}, \ldots, P\}$ for a planet with orbital period $P$. It is easy to verify that translating any number of observation times by $nP$ where $n \in \mathbb{Z}$ will lead to identical phase coverage, even though the new observation times will no longer themselves be uniformly-spaced in time.

In case (iv) considered in Section 5.2, i.e. '220 uniformly-spaced observations with a 6 yr baseline,' the observation times were for simplicity spaced 6 yr/220 $\sim$ 10 d apart, with no retrospective consideration given to weather, K-10's visibility, etc. Given the long baseline and the relatively short (known) orbital period of K-10c, i.e. 45.29 d, it would be straightforward to use the above considerations to devise a more realistic observing programme that ensured equivalent phase coverage.

More generally, it should be noted that the above considerations do not necessarily apply to the case of planets with long orbital periods, say of the order several months or more. If a planet's orbital period is long but known, it should still be straightforward to plan observations, possibly spanning more than one observing season, to provide uniform phase coverage. If the orbital period of a planet is unknown *and* comparable to the baseline of observations, however, it will not be possible to guarantee uniform phase coverage *a priori*. Finally,

[6]One could derive criteria under which this approximate-uniformity would be ensured, although it is much easier to verify numerically that synthetic observation times with approximately-uniform cadence phased to an unknown period will lead to approximately uniform phase coverage, provided the observing baseline is much longer than the period.

uniform phase coverage is not possible with an observing baseline shorter than a planet's orbital period.

## 5.4 Reconciling the mass estimates

Returning to the K-10 analysis, we noted $> 3\sigma$ evidence for linear correlations ($\rho \sim 30\%$) between the published HARPS-N RVs and (i) $\log R'_{\rm HK}$ index and (ii) BIS (bisector inverse slope) measurements; we did not find any similarly-significant correlations in the HIRES RVs. Whereas the models of D14 and W16 did not accommodate possible stellar activity signals in the RVs, we opted to use the GP framework from Chapter 3 of this thesis (published as Rajpaul et al., 2015, hereafter R15) to model jointly all available RV, $\log R'_{\rm HK}$ (in the case of HARPS-N) or $S_{\rm HK}$ (in the case of HIRES), and BIS time series, for a total of 660 data points.

As in R15, we adopted a QP covariance kernel, and placed non-informative priors on all GP hyper-parameters. GP amplitude parameters were also constrained to be smaller than the total variation seen in a given time series, and of the overall GP period we required $P > 20$ d (based on D14's lower limits on K-10's stellar rotation period). We additionally allowed at least 2 but up to 5 possible planetary signals in the RVs, modelled with Keplerian functions. We constrained the periods and periapsis passage times of two of the Keplerians to be consistent with the most precise values inferred from K-10b and K-10c's transits (Morton et al., 2016; Holczer et al., 2016), but left the other parameters free, with priors identical to those in W16's eccentric 2-planet model. We adopted analogous uninformative priors for all parameters of the additional possible planets, ensuring only that planet periods did not overlap. Finally, we used the MultiNest nested-sampling algorithm (Feroz and Hobson, 2008; Feroz et al., 2009, 2013) to obtain a full joint posterior distribution for each model's parameters (and marginal posteriors for parameters of interest), and to compute a Bayesian evidence ($\mathcal{Z}$) for each model.

Of the numerous models we considered, we found only one in which MAP estimates for *all* planet parameters were consistent within $1\sigma$ between the HARPS-N, Keck-HIRES, and merged data sets: *viz.* a model including three planets, all with

**Table 5.1:** Summaries of marginal posteriors for selected planet and GP parameters from our favoured model (three planets plus correlated noise); the planet parameters are the same as those from W16, while the GP parameters are as defined in Chapter 3, with $K_{\rm GP} = \sqrt{V_{\rm r}^2 + V_{\rm c}^2}$. Periapsis passage times are omitted: for K-10b and K-10c, these were effectively fixed in our models *a priori* via known transit times, while a periapsis passage time for non-transiting planet candidate KOI-72.X could not be well constrained, given its apparently-circular orbit.

| | | | HIRES RVs only | | HARPS-N RVs only | | Merged RVs | |
|---|---|---|---|---|---|---|---|---|
| Parameter | Description | Units | Median | $\pm\sigma$ | Median | $\pm\sigma$ | Median | $\pm\sigma$ |
| $K_{\rm b}$ | RV semi-amplitude | m s$^{-1}$ | 2.39 | $^{+0.30}_{-0.28}$ | 2.33 | $\pm0.16$ | 2.32 | $^{+0.21}_{-0.18}$ |
| $P_{\rm b}$ | Period | days | 0.83748 | $\pm0.00003$ | 0.837501 | $\pm0.000005$ | 0.837501 | $^{+0.000005}_{-0.000004}$ |
| $\sqrt{e_{\rm b}}\cos\omega_{\rm b}$ | Eccentricity, periapse arg. | - | 0.000 | $\pm0.003$ | 0.000 | $\pm0.003$ | 0.000 | $\pm0.004$ |
| $\sqrt{e_{\rm b}}\sin\omega_{\rm b}$ | Eccentricity, periapse arg. | - | 0.000 | $\pm0.003$ | 0.000 | $\pm0.003$ | 0.000 | $\pm0.004$ |
| $m_{\rm b}$ | Mass | $M_\oplus$ | 3.33 | $^{+0.40}_{-0.42}$ | 3.25 | $^{+0.22}_{-0.23}$ | 3.24 | $\pm0.28$ |
| $\rho_{\rm b}$ | Mean density | g cm$^{-3}$ | 5.65 | $^{+0.94}_{-0.85}$ | 5.51 | $^{+0.73}_{-0.64}$ | 5.48 | $^{+0.78}_{-0.68}$ |
| $K_{\rm c}$ | RV semi-amplitude | m s$^{-1}$ | 1.27 | $^{+0.42}_{-0.35}$ | 1.64 | $^{+0.42}_{-0.34}$ | 1.41 | $^{+0.25}_{-0.23}$ |
| $P_{\rm c}$ | Period | days | 45.2948 | $\pm0.0008$ | 45.2940 | $^{+0.0008}_{-0.0007}$ | 45.2946 | $\pm0.0008$ |
| $\sqrt{e_{\rm c}}\cos\omega_{\rm c}$ | Eccentricity, periapse arg. | - | 0.1 | $\pm0.2$ | 0.0 | $\pm0.1$ | 0.0 | $\pm0.1$ |
| $\sqrt{e_{\rm c}}\sin\omega_{\rm c}$ | Eccentricity, periapse arg. | - | 0.0 | $\pm0.2$ | 0.1 | $\pm0.1$ | 0.0 | $\pm0.1$ |
| $m_{\rm c}$ | Mass | $M_\oplus$ | 5.87 | $^{+2.20}_{-1.82}$ | 8.59 | $^{+2.19}_{-1.79}$ | 7.37 | $^{+1.32}_{-1.19}$ |
| $\rho_{\rm c}$ | Mean density | g cm$^{-3}$ | 2.50 | $^{+0.98}_{-0.78}$ | 3.66 | $^{+0.98}_{-0.80}$ | 3.14 | $^{+0.63}_{-0.55}$ |
| $K_{\rm X}$ | RV semi-amplitude | m s$^{-1}$ | 1.30 | $^{+0.51}_{-0.45}$ | 0.84 | $^{+0.16}_{-0.14}$ | 0.85 | $^{+0.24}_{-0.14}$ |
| $P_{\rm X}$ | Period | days | 102 | $^{+8}_{-7}$ | 101 | $^{+6}_{-5}$ | 102 | $\pm1$ |
| $\sqrt{e_{\rm X}}\cos\omega_{\rm X}$ | Eccentricity, periapse arg. | - | $-0.1$ | $\pm0.2$ | $-0.1$ | $\pm0.1$ | $-0.1$ | $\pm0.1$ |
| $\sqrt{e_{\rm X}}\sin\omega_{\rm X}$ | Eccentricity, periapse arg. | - | $-0.1$ | $\pm0.2$ | 0.0 | $\pm0.1$ | 0.0 | $\pm0.1$ |
| $m_{\rm X}$ | Mass | $M_\oplus$ | 8.93 | $^{+3.50}_{-3.15}$ | 5.80 | $^{+1.20}_{-1.03}$ | 5.90 | $^{+1.70}_{-1.01}$ |
| $K_{\rm GP}$ | GP semi-amplitude (RV) | m s$^{-1}$ | 0.09 | $^{+0.22}_{-0.06}$ | 1.46 | $\pm0.17$ | 1.68 | $\pm0.25$ |
| $P$ | GP period | days | 63 | $\pm10$ | 55 | $\pm1$ | 55.5 | $\pm0.8$ |
| $\lambda_{\rm p}$ | GP inverse harmonic complexity | - | 1.3 | $^{+0.6}_{-0.3}$ | 0.33 | $^{+0.04}_{-0.02}$ | 0.32 | $^{+0.02}_{-0.01}$ |
| $\lambda_{\rm e}$ | GP evolution time-scale | days | 330 | $\pm100$ | 86 | $\pm4$ | 90 | $\pm6$ |

orbits consistent with circular, plus correlated noise. Significantly, this model was also favoured above others by Bayesian model comparison tests, and the fitted period for the third planet in our model was $102 \pm 1$ d: in accord with the W16's favoured period for KOI-72.X (based on both analytical considerations and dynamical modelling), despite us not including this as prior information in our modelling. We summarise the marginal posteriors for this favoured model's planet parameters in Table 5.1; masses (for all three planets) and mean densities (for the transiting planets) were derived using the same stellar mass and planet radii as in W16.

Additionally, we note the following.

1. For the HIRES, HARPS-N and merged data sets, 3-planet models were strongly favoured over 2-planet models ($\Delta \ln \mathcal{Z} > 10$), which were in turn favoured over 4- and 5-planet models.
2. We obtained consistent parameters for all planets when splitting either the HIRES or HARPS data sets in two; presumably W16 found discrepant results because neither a third planet nor a nuisance signal model was included when performing the same test.
3. A zero-amplitude GP component was favoured for the HIRES RVs ($\Delta \ln \mathcal{Z} \sim 3$), whereas a non-zero GP amplitude was favoured ($\Delta \ln \mathcal{Z} \gtrsim 10$) for the HARPS-N and merged RVs. The latter two cases suggested a GP period of $P = 55 \pm 1$ d.

We interpreted (iii) as evidence of the HARPS-N RVs being confounded by at least one semi-coherent though not strictly periodic, $\gtrsim 1$ m s$^{-1}$ nuisance signal that could not be ascribed to a planet (for which a simpler Keplerian model would have sufficed, rather than a GP; a planetary signal should also have been simultaneously present in the HIRES RVs). Given the correlation observed between HARPS-N RVs and activity indicators, at least part of this signal could be due to stellar activity; it is unclear, though, whether $P = 55 \pm 1$ d corresponds to a stellar rotation period. An instrumental component to the signal also cannot be ruled out.[7]

[7]The small posterior uncertainty of $\pm 1$ d may simply indicate that 55 d is the only GP period that does a reasonable job of modelling some (possibly complex) combination of nuisance signals. Regardless, given the similarity of the 55 d period to K-10c's orbital period, the nuisance and planet signals interfere strongly over time-scales of several months (an envelope with period 248 d would be predicted if the nuisance signal were sinusoidal).

Either way, it appears that when *not* accounting for a correlated nuisance signal, and an apparent third planet (KOI-72.X), the amplitude of the signal ascribed to K-10c is forced to inflate artificially to absorb some of this appreciable variability in the discretely-sampled RV signal.

Our mass and mean density estimates for K-10b are consistent with those of D14 and W16. Our mass estimate for K-10c ($7.37^{+1.32}_{-1.19}$ $M_\oplus$), however, is significantly lower than those from D14 and W16 ($17.2 \pm 1.9$ $M_\oplus$ and $13.98 \pm 1.79$ $M_\oplus$, respectively); accordingly, we also infer a significantly lower mean density of $\rho_c = 3.14^{+0.63}_{-0.55}$ g cm$^{-3}$. This implies a composition which is either consistent with a low-density solid planet with a significant fraction of volatiles in the form of e.g. water or methane, or a planet with a dense core and an extended gaseous envelope. K-10c would thus join a region of parameter space in the mass-radius diagram occupied by a number of other exoplanets with radii between 2.0 and 2.5 $R_\oplus$ that have similar mean densities to K-10c; see Fig. 5.5.

Finally, our inferred mass of $5.90^{+1.70}_{-1.01}$ $M_\oplus$ for KOI-72.X is compatible with W16's point estimate of $\sim 7$ $M_\oplus$, though it remains to be established whether this is a genuine planet. We used the parameters from our Keplerian solutions as inputs to numerical N-body integrations using TTVFast (Deck et al., 2014); we noted that the maximum difference between the RVs from a full dynamical simulation and from our Keplerian solution was of the order 1 cm s$^{-1}$ over 101 d. As this is two orders of magnitude below the RV noise floor, we concluded that full dynamical modelling would have yielded no constraints beyond those already derived by W16.

## 5.5 Instrumental considerations

W16 detrended the HIRES RVs by removing correlations between RVs and instrumental parameters, RV uncertainties, and spectrum signal-to-noise ratio. The RVs published in W16 are these *detrended* RVs; the published RV uncertainties also already have jitter applied. We obtained both the pre-detrending RVs and the uncertainties without jitter from Weiss (*pers. comm.*), and re-ran the analyses described in Section 5.4. As before, we ended up favouring a 3-planet plus correlated

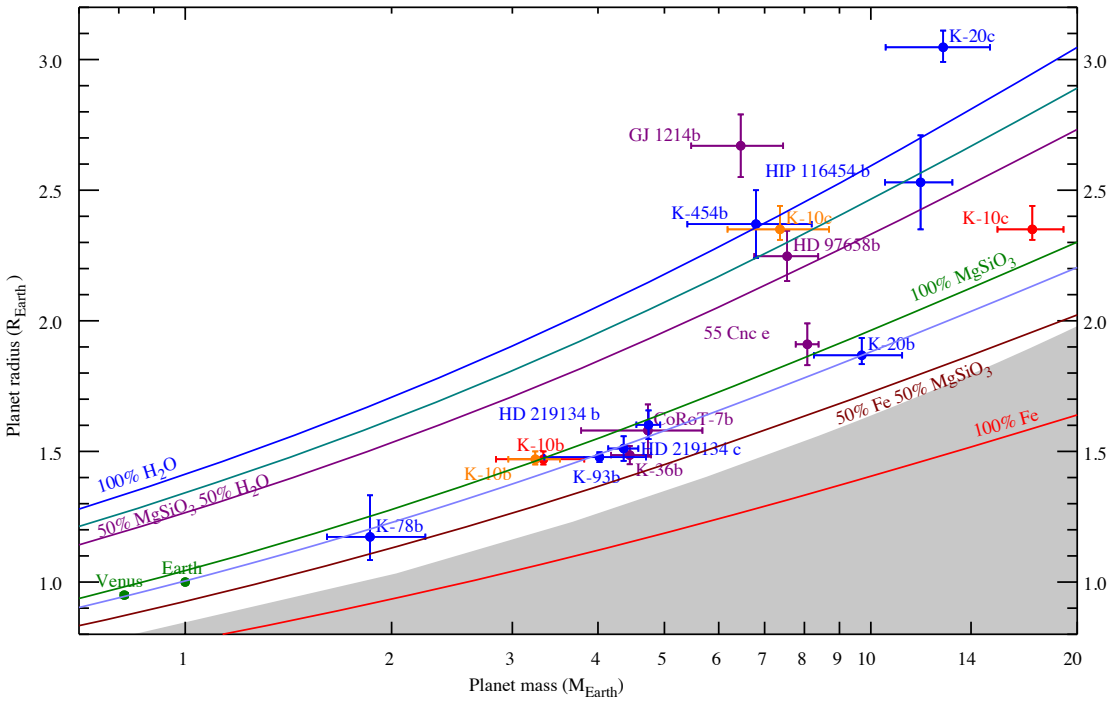


**Figure 5.5:** Mass-radius relation for planets smaller than 3.2 $R_\oplus$ and mass determinations better than 20% precision, as of June 2017. The shaded region denotes where planets would have an iron content exceeding the maximum value predicted from collisional stripping models. The solid curves are theoretical models for planet with a composition consisting of 100% $H_2O$ (blue), 25% silicate and 75% $H_2O$ (teal), 50% silicate and 50% $H_2O$ (purple), 100% silicate (green), 70% silicate and 30% iron, consistent with an Earth-like composition (light blue), 50% silicate and 50% iron (brown), and 100% iron(red) (Zeng and Sasselov, 2013). The blue points indicate planets with masses measured using the HARPS-N spectrograph; purple points are from other sources. The red and orange points correspond to the confirmed K-10 planets, as characterised by D14 and this study, respectively.

noise model strongly over all competing models, and the posterior distributions for the parameters of all three Keplerians were consistent ($< 1\sigma$) with those obtained when using the detrended HIRES RVs.

To explore the possibility of the HARPS-N data reduction pipeline contributing to the discrepancy, we applied a novel, template- and mask-free approach we are developing (paper *in prep.*; see also Section 7.3 of this thesis) for extracting RVs from observed spectra. We model each observed spectrum non-parametrically, with shifts between all possible pairs of spectra included as parameters in the modelling (in addition to possible telluric, stellar activity, and instrumental effects).

Interestingly, we found that when modelling HARPS-N RVs extracted with our own pipeline vs. the HARPS-N pipeline, our inferred RV semi-amplitude for K-10b was unchanged, but we reliably inferred $K_c < 2$ m s$^{-1}$ even *without* a correlated noise (GP) component in our model. This suggests the possibility that at least part of the signal confounding K-10c's signal might be instrumental rather than stellar (and would explain why the same nuisance signal is absent from the HIRES RVs). Given the preliminary nature of our pipeline, however, further investigation is required.

## 5.6 Discussion and conclusions

Previous studies (e.g. Dawson and Fabrycky, 2010) have explored the impact of irregular time sampling on planet *period* estimation; here we have demonstrated that a failure to account for one or more coherent signals (whether of stellar, planetary, or instrumental origin) in RV data, and/or uneven phase coverage, can confound attempts to infer the *masses* of planets with known periods. We used synthetic data with sampling based on real observations to demonstrate how such difficulties could arise when characterising planets in a system analogous to K-10; tests such as the ones we presented may readily be applied to other systems, to test the sensitivity of planet characterisation to sampling and model selection.

By accounting for a time-correlated (stellar or instrumental) signal present in the HARPS-N K-10 RVs, as well as a likely third planet in the system, we were able to achieve full consistency between the Keplerian solutions for the HIRES, HARPS-N, and combined RVs. The third planet included in our model has properties consistent with K-10c's TTVs; and although our model is more complex than the one used by W16 to model the RVs, it was nevertheless favoured over simpler models in Bayesian model comparison testing. While our proposed resolution of the K-10c mass discrepancy is a plausible one, it appears that (many) more RVs will be required for a definitive characterisation of the K-10 system.

Whereas W16 suggested a strategy of employing a long observing baseline compared to time-correlated noise influences, we suggest it is also important to focus on obtaining more complete *phase coverage* of the relevant signals. As we

demonstrated in Section 5.2, good phase coverage can permit robust inference about known planets, even when using a demonstrably-inadequate physical model. While uniform cadence might not be feasible or desirable, e.g. to avoid aliasing, a long observing baseline and approximately-uniform cadence would lead to good phase coverage even of planets with unknown orbital periods. And while W16 suggested that a long baseline would help to average out spurious signals that may arise from stellar activity, we suggest it is strongly preferable to *model* these nuisance signals, as it is difficult to know *a priori* how these nuisance signals might interfere with signals of interest. Baselines and cadence aside, it seems all but essential to implement a variety of physical models (to account for varying numbers of possible planets, nuisance signals, etc.), and to compare systematically the evidence for the competing models.

Our findings may also have relevance to archival RV data sets, and indeed, this is not the first example of a system where inference has turned out to be extremely sensitive to both sampling and model choice, despite the availability of a large number of RVs (Rajpaul et al., 2016). Then again, K-10 might have been a relatively pathological case. As W16 noted, there *were* various hints (TTVs, K-10c mass discrepancy, etc.) that existing characterisations of the system were inadequate.

Looking to the future, with a new generation of RV spectrographs with expected instrumental precisions of 10 cm $s^{-1}$ soon to come online, it is clear that optimised sampling strategies and careful model selection will both be essential if we are to use these spectrographs for accurate characterisation of small planets, especially those in potentially multi-planet systems. Moreover, it would be prudent to coordinate observations made by different teams using different telescopes, in order to minimise 'redundant' observations that do not contribute to improved coverage of a given planet's orbital phases.

*On a holiday, so many miles*
*Looking for a place to stay*
*Near some friendly star, he found this mote*
*And now we wonder where we are...*

— From the Pixies' song *Motorway to Roswell* (1991).

# 6

# Discovering and characterising new planets

## Contents



This chapter presents a few relatively straightforward applications of the GP framework from Chapter 3 (straightforward, at least, when compared to the preceding analyses focused on the $\alpha$ Cen B and Kepler-10 systems).

In the first application, in Section 6.1, the GP framework is used to discover a sub-Neptune planet in a tricky case where the planet's orbital period appears identical to its host star's rotation period. The discussion in this section is adapted from parts of a paper I co-authored (Mortier et al., 2016).

In a second and third application, the GP framework is used to model stellar signals in RVs in order to infer precise mass (i) for Kepler-37 d, a transiting super-Earth in the well-studied Kepler-37 system, known to host at least 4 planets (one of which, Kepler-37 b, is currently the smallest known exoplanet), and (ii) for Kepler-538 b, another transiting super-Earth. The masses of these planets were hitherto unknown; these analyses add Kepler-37d and Kepler-538b to a relatively small number of planets with long orbital periods, and precise mass measurements. The analyses are unpublished (paper *in prep.*) so are presented here only in outline form. Further, given the similarity between the analyses of the two systems, they are presented in parallel.

## 6.1 HD 175607 b

### 6.1.1 Introduction and context

Soon after the first exoplanets were discovered, it was suggested that higher-metallicity stars have a higher probability of hosting Jovian planets than stars with lower metallicities (Gonzalez, 1997). This result was confirmed in a number of subsequent studies (e.g. Santos et al., 2001; Fischer and Valenti, 2005; Johnson et al., 2010; Mortier et al., 2013). Taken at face value, this relation favours planet formation theories based on a core-accretion model (e.g. Pollack et al., 1996; Mordasini et al., 2009, 2012), according to which dust and grains coagulate to form planetesimals and combine to make larger cores and thus planets. Metal-rich stars and discs can form these cores more quickly, so they have time to accrete gas before the disc dissipates, resulting in more gas giants.

For lower-mass planets (mass $< 0.1$ $M_J$), e.g. Neptunes and (super-)Earths, the same correlation is not observed, and the planet occurrence rate appears to be independent of the host-star metallicity (e.g. Udry and Santos, 2007; Sousa et al., 2011b; Buchhave and Latham, 2015). Planet synthesis simulations based on the theories of core-accretion and planet migration showed that the correlation may even be reversed in the case of Earth-sized planets, where stars with sub-Solar metallicities are favoured for harbouring an Earth-sized planet (Mordasini et al., 2012).

For these reasons, a sample of 109 metal-poor stars was chosen for an extensive RV survey with the HARPS spectrograph to search for Neptunes and (super-)Earths (Santos et al., 2014). The targets in this survey are bright, chromospherically quiet FGK dwarfs with metallicities [Fe/H] $\in [-2.0, -0.4]$ dex. More details about this programme can be found in Santos et al. (2014).

At the time of this study, no low-mass planets had been detected in this metal-poor HARPS sample. Jenkins et al. (2013) reported on the existence of two super-Earths orbiting one star in this metal-poor sample, HD 41248, close to a 7:5 mean motion resonance. However, further data and subsequent analysis could not confirm the existence of the planets; one of the signals of the claimed planets is clearly due to stellar activity, and the other may be due to the time sampling of the data and/or differential rotation (Santos et al., 2014; Jenkins and Tuomi, 2014).

This study reports on the discovery of at least one Neptune around HD 175607, one of the stars in the metal-poor HARPS survey.

### 6.1.2 Observations

This analysis draws on a total of 119 HARPS spectra of HD 175607, taken on 110 nights between July 2004 and October 2013, drawn from three different observing programmes. Most spectra were obtained with an integration time of 15 minutes, to average out nuisance signals from short-term stellar oscillations (e.g. Santos et al., 2004). In the few cases where more than one spectrum was taken on a given night, the observations were separated by several hours to reduce granulation effects (following the observing strategies advocated by Dumusque et al., 2011c), and then averaged. The spectra have a mean signal-to-noise ratio of 104 around 6200 Å.

RVs were derived homogeneously using the HARPS Data Reduction Software (DRS). The DRS pipeline cross-correlated the observed spectra with a mask representing a G8 dwarf (the spectral type of HD 175607 is G6V). By fitting a Gaussian to the cross-correlation function (CCF), the value and uncertainty of the RV is determined (e.g. Baranne et al., 1996; Pepe et al., 2002a). We end up with 110 RV measurements with a mean error of 0.95 m $s^{-1}$, including photon,

calibration, and instrumental noise contributions. The DRS pipeline also produced measurements of various stellar activity indicators: CCF FWHM, BIS, and contrast; chromospheric activity indicator $\log R'_{\rm HK}$; and an H$\alpha$ index. Error bars for the CCF parameters (FWHM, BIS, and contrast) were scaled from the RV error, following Santerne et al. (2015).

### 6.1.3 Stellar properties

HD 175607 is a bright dwarf star of spectral type G6V. It is located 45.27 pc from the Sun (van Leeuwen, 2007), and has mass $0.71 \pm 0.01$ M$_\odot$ and radius $0.70 \pm 0.01$ R$_\odot$ (Torres et al., 2010; Mortier et al., 2016). With a spectroscopic metallicity of [Fe/H] $= -0.62 \pm 0.01$ dex, the star is indeed metal poor (Sousa et al., 2011a).

It is an old star ($10.32 \pm 1.58$ Gyr: da Silva et al., 2006; Bressan et al., 2012), and rotates slowly. Glebocki and Gnacinski (2005) reported a projected rotational velocity of $v \sin i = 0.9$ km s$^{-1}$, while estimates based on the $B - V$ colour index and CCF FWHM (Santos et al., 2002) suggest a projected rotational velocity of $v \sin i = 1.3$ km s$^{-1}$.

The weighted mean value of $\log R'_{\rm HK}$ over all 110 measurements is $-4.92$, indicating a relatively inactive star. This is consistent with the star's old age and slow (projected) rotation. Using empirical relationships between $\log R'_{\rm HK}$ and the $B - V$ colour index (Noyes et al., 1984a; Mamajek and Hillenbrand, 2008), the star's rotational period may be estimated to be $\sim 29$ d. The star's true rotation period has not, however, been measured directly.

See Mortier et al. (2016) for a detailed table of stellar parameters for HD 175607, along with details of how the properties were measured or derived.

### 6.1.4 First-pass analyses

My co-authors of the Mortier et al. (2016) paper carried out a number of preliminary, 'first-pass' analyses of the HD 175607 data. I summarise these very briefly below.

First, they computed and studied Generalised Lomb-Scargle (GLS) periodograms for the RVs and the available activity indicators. The most significant peak in the

RV time series was found around 29 d, which corresponded to the rotation period estimated from the star's activity level. None of the CCF activity indicators showed significant periodicities, although the H$\alpha$ index did show significant peaks at 24.5 and 48 days, and the $\log R'_{\rm HK}$ time series showed a significant peak around 36 d; this strengthened the hypothesis that the stellar rotation period was of the order one month. Additionally, some long-term variability in the $\log R'_{\rm HK}$ and CCF contrast appeared to be simultaneously present in the RVs. No significant linear or rank correlations were found between the RV time series and the activity indicators.

Secondly, they computed a number of other spectroscopic activity indicators as alternatives to the BIS, designed to probe line profile variations more accurately in cases of low signal-to-noise ratios or weak RV-activity correlations (Nardetto et al., 2006; Boisse et al., 2011; Figueira et al., 2013). None of these alternative indicators showed significant periodic variation, or correlations with the RVs.

Thirdly, they performed an iterative pre-whitening analysis using the GLS periodogram, and found that fitting and then subtracting sinusoids with periods of 29.03 d and $\sim$ 700 d (or 1400 d) sufficed to render the residual RV variation devoid of any statistically significant periodicities. A multi-frequency periodogram (see e.g. Baluev, 2013) also favoured the combination of 29 d and $\sim$ 700 d periods.

These findings suggested that the periodic variations observed in the RVs may have been due to one or more planets rather than stellar activity. Thus they decided to fit the RVs with a sum of Keplerians, plus a jitter term (added in quadrature to the observational errors).

Non-informative (uniform or log uniform) priors were placed on all parameters, and EMCEE (Foreman-Mackey et al., 2013), which implements an affine invariant MCMC ensemble sampler, was used to sample from parameter posteriors. A genetic algorithm (Ségransan et al., 2011) provided some though not all of the starting points for the MCMC walkers used by EMCEE, and thermodynamic integration[1]

[1]My own experience is that evidences estimated via thermodynamic integration using EMCEE should be treated with caution. I have found that on Keplerian-fitting problems, EMCEE's evidence estimates often diverge significantly from those computed by techniques better suited to computing evidences in the case of multimodal posteriors and/or high model dimensionality, e.g. MULTINEST or POLYCHORD (Handley et al., 2015). The latter techniques also have the advantage of providing

was used to compute approximate model evidences.

Parameter posteriors for their one- and two-planet models are summarised in Table 6.1 on pg. 162. They found that a two-planet (two-Keplerian) model was favoured over a one-Keplerian model, with an estimated Bayes factor of over $10^6$.

### 6.1.5 Gaussian process analysis

The first-pass analyses in the previous section did not model any of the activity indicators jointly with the RVs – nor did they address the possibility that the periodic signal(s) in the RVs may have a stellar rather than planetary origin. Therefore I chose to model the RV, $\log R'_{\mathrm{HK}}$, and BIS time series jointly using the framework from Chapter 3 (published as Rajpaul et al., 2015, hereafter R15) with a quasi-periodic covariance kernel.

The FWHM data were not included in my modelling since they were noisier than but very tightly correlated with $\log R'_{\mathrm{H}K}$, and thus did not provide useful extra information. Non-informative priors were placed on all GP covariance and Keplerian orbital parameters, as for the MCMC procedure discussed in the previous section; see Mortier et al. (2016) for more details. Parameter posterior sampling was performed using the MultiNest nested-sampling algorithm (Feroz and Hobson, 2008; Feroz et al., 2009, 2013)

My findings were as follows. When not including a planetary component in the GP's mean function for the RV time series, the MAP value of the GP period hyper-parameter, $P_{\mathrm{GP}}$, ended up being $29.0 \pm 0.1$ d: because the 29.0-d signal was so significant in the $\Delta\mathrm{R}V$ time series, the GP was forced to absorb this, whilst all but ignoring and thus failing to fit the ancillary time series. The semi-amplitude of variations absorbed by the GP in the RV time series was 3.8 m $\mathrm{s}^{-1}$.

On the other hand, when including a Keplerian component, the hyper-parameter $P_{\mathrm{GP}}$ ended up being $29.9 \pm 0.4$ d, with a significant part of the 29-d signal being absorbed by the Keplerian component; under this model, the semi-amplitude of

quantitative uncertainty estimates for model likelihoods.

the RV variations absorbed by the GP was reduced to 1.9 m $s^{-1}$, with the semi-amplitude of the Keplerian signal being 1.8 m $s^{-1}$. This is significant because whereas a GP can in principle model an arbitrarily-complex signal arbitrarily well, a Keplerian function is far simpler, and is always strictly periodic. Therefore, the fact that the simpler, less flexible Keplerian model sufficed to fit a significant part of the RV variability with 29 d period indicated that this variability had a coherent phase over the *entire* data set (with $\sim 9$ yr baseline), strengthening the planetary interpretation of the 29-d signal.[2]

The planet parameters inferred using the GP framework are presented in Table 6.1, alongside those from the Keplerian-only fits; see also Fig. 6.1. The MAP fit from the GP plus one-planet model is presented in Fig. 6.2, with a close-up in Fig. 6.3.

Because the framework from R15 forces the same GP basis functions to explain variability simultaneously in RVs and activity indicator time series, the 1.8 m $s^{-1}$ signal modelled by the GP may be identified as a stellar activity signal modulated by stellar rotation with a $\sim 30$ d period. The evolution time-scale for the apparent activity signal, 67 d, would correspond to just over two rotation periods.

After subtracting the MAP GP plus one planet model, the residual time series appeared white and normally-distributed, with no significant power on time-scales shorter than one year, and with rms 0.95 m $s^{-1}$. This suggests that all of the RV variation (at least on time-scales smaller than one year) can be explained fully with the planet plus activity model. See Fig. 6.4. The $\log R'_{HK}$ and BIS residuals contained no significant periodicities on any timescales.

I used the sample size-adjusted Akaike information criterion (AIC; Burnham and Anderson, 2002) to select between models with different numbers of Keplerian components.[3] The AIC value for the zero-planet model was $-25.44$, and the

[2]The 29 d signal did not correspond to any significant peaks in the power spectrum of the observing window, in which the only significant peak was at 1.00 d.

[3]The AIC seeks to quantify the trade-off between the goodness of fit of a model and the complexity of the model. It is similar to the Bayesian Information Criterion (BIC), though the AIC has some practical and theoretical advantages over the BIC (Burnham and Anderson, 2002). I used the AIC in lieu of Bayes factors (which are used elsewhere in this thesis) as, at the time of this particular analysis, I'd not yet ironed out bugs in the code I was using to compute GP model posterior probabilities.

**Table 6.1:** Summaries of marginal posteriors from three different models used to fit the HD 175607 observations. The first two models are simple Keplerian plus jitter fits to the RVs; the third model fits the RV, $\log R'_{\rm HK}$ and BIS time series simultaneously with a quasi-periodic GP, using the framework of R15, and includes a Keplerian term in the mean function for the RVs only. GP parameters are as defined in R15, with $K_{\rm GP} = \sqrt{V_{\rm r}^2 + V_{\rm c}^2}$.

| Parameter | Description | Units | 1 planet plus jitter | | | 2 planets plus jitter | | | GP plus 1 planet | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Median | $+\sigma$ | $-\sigma$ | Median | $+\sigma$ | $-\sigma$ | Median | $\pm\sigma$ |
| $\gamma$ | Systemic velocity | m s$^{-1}$ | $-91889.69$ | 0.22 | 0.22 | $-91890.41$ | 0.29 | 0.28 | $-91889$ | 1 |
| $K_1$ | RV semi-amplitude | m s$^{-1}$ | 2.21 | 0.33 | 0.33 | 2.37 | 0.29 | 0.30 | 1.8 | 0.4 |
| $P_1$ | Period | days | 29.03 | 0.03 | 0.03 | 29.01 | 0.02 | 0.02 | 29.0 | 0.2 |
| $m_1 \sin i$ | Minimum mass | $M_\oplus$ | 8.26 | 1.25 | 1.25 | 8.98 | 1.10 | 1.10 | 6.7 | 1.5 |
| $e_1$ | Eccentricity | - | 0.16 | 0.14 | 0.11 | 0.11 | 0.09 | 0.07 | 0.16 | 0.15 |
| $\omega_1$ | Periapse argument | - | $0.55\pi$ | $0.36\pi$ | $0.34\pi$ | $0.79\pi$ | $0.29\pi$ | $0.29\pi$ | $1.0\pi$ | $0.40\pi$ |
| $T_{0,1}$ | Periapse time | BJD | 2455528 | 5 | 5 | 2455532 | 4 | 4 | 2453219 | 6 |
| $K_2$ | RV semi-amplitude | m s$^{-1}$ | – | – | – | 2.86 | 0.51 | 0.51 | – | – |
| $P_2$ | Period | days | – | – | – | 1337 | 103 | 46 | – | – |
| $m_2 \sin i$ | Minimum mass | $M_\oplus$ | – | – | – | 34.97 | 6.93 | | – | – |
| $e_2$ | Eccentricity | - | – | – | – | 0.42 | 0.15 | 0.14 | – | – |
| $\omega_2$ | Periapse argument | - | – | – | – | $0.08\pi$ | $0.10\pi$ | $0.09\pi$ | – | – |
| $T_{0,2}$ | Periapse time | BJD | – | – | – | 2455244 | 63 | 74 | – | – |
| $\sigma_{RV}$ | RV jitter | m s$^{-1}$ | 2.01 | 0.17 | 0.19 | 1.40 | 0.16 | 0.17 | – | – |
| $K_{GP}$ | GP semi-amplitude (RV) | m s$^{-1}$ | – | – | – | – | – | – | 1.87 | 0.16 |
| $P_{GP}$ | GP period | days | – | – | – | – | – | – | 29.9 | 0.2 |
| $\lambda_{\rm p}$ | GP inv. harmonic complexity | - | – | – | – | – | – | – | 0.16 | 0.02 |
| $\lambda_{\rm e}$ | GP evolution time-scale | days | – | – | – | – | – | – | 67 | 11 |

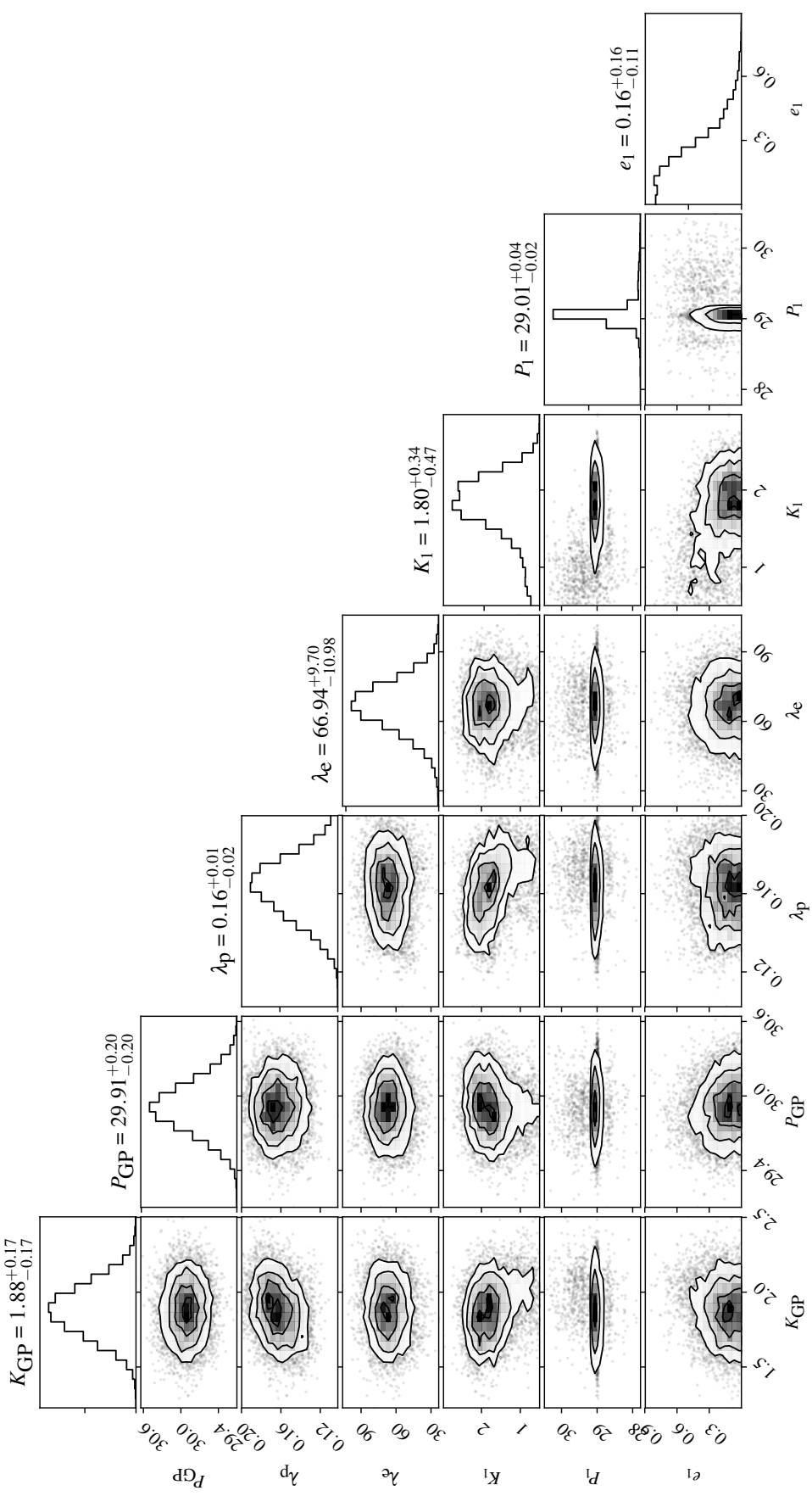


**Figure 6.1:** 1D and 2D marginal posterior distributions for a few of the main parameters in the GP plus 1-planet model fit to the HD 175607 observations; refer to Table 6.1 for parameter descriptions. Plot made using `corner.py` (Foreman-Mackey, 2016).

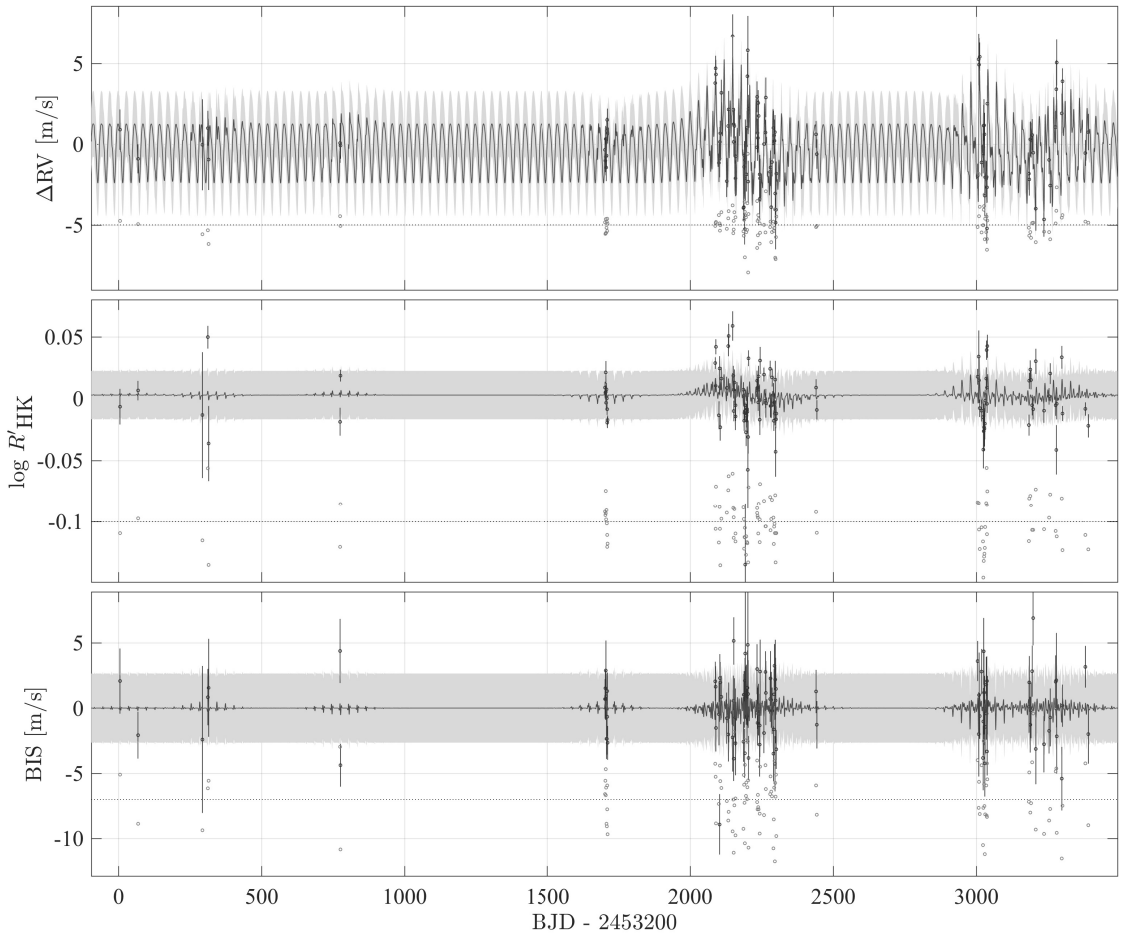


**Figure 6.2:** GP plus planet model MAP fit to the mean-subtracted HD 175607 data. Black dots indicate observed data, with estimated errors; solid lines are model posterior means; and shaded regions denote $\pm\sigma$ posterior uncertainty. Residuals are plotted as grey dots below the observed data and fitted model, but for the sake of clarity, with arbitrary vertical offset from the main time series (denoted by a dotted line).

corresponding value for the one-planet model $-33.06$, indicating that the planetary explanation was favoured by about a factor of ten. A GP plus two-planet model (fitted period for outer planet $P_2 > 1000$ d) led to an AIC value of $-32.72$, indicating a lack of evidence for a second planet. Given the approximations inherent in the AIC, however, it would be unwise to conclude that the long-period signal is definitely not a planet. In any case, one may conclude that the available data constrain the properties of this putative planet only very weakly.

Two further analyses shored up the planetary interpretation of the 29 d signal. First, when starting with only a subset of the 110 observations, then artificially

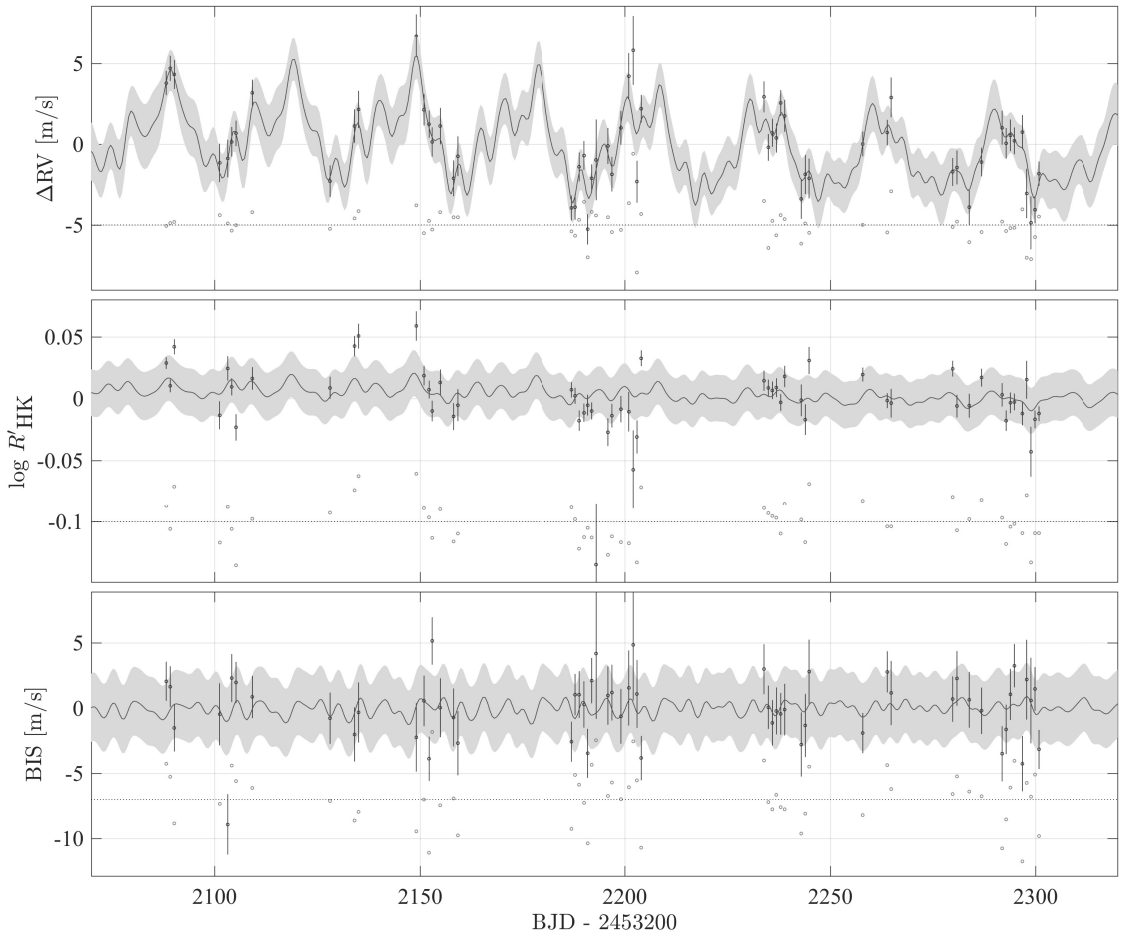


**Figure 6.3:** As for Fig. 6.2, but with a close-up view of the region with densest time coverage (57 observations over about seven months).

adding further observations, the statistical significance of the 29 d signal (as modelled by a sinusoid) increased monotonically. If the signal did not have a coherent phase over long baselines (as would be the case for an activity signal), the statistical significance would instead be expected to fluctuate as more observations were added. Secondly, we injected into the data a noise-free Keplerian signal with the same semi-amplitude and eccentricity as the MAP model for the 29 d signal, but slightly different period (21 d). The GLS power of this 21 d signal was found to be almost identical to that of the 29 d signal, and joint GP plus Keplerian modelling led to this signal being recovered with about the same precision as the putative 29 d planetary signal.

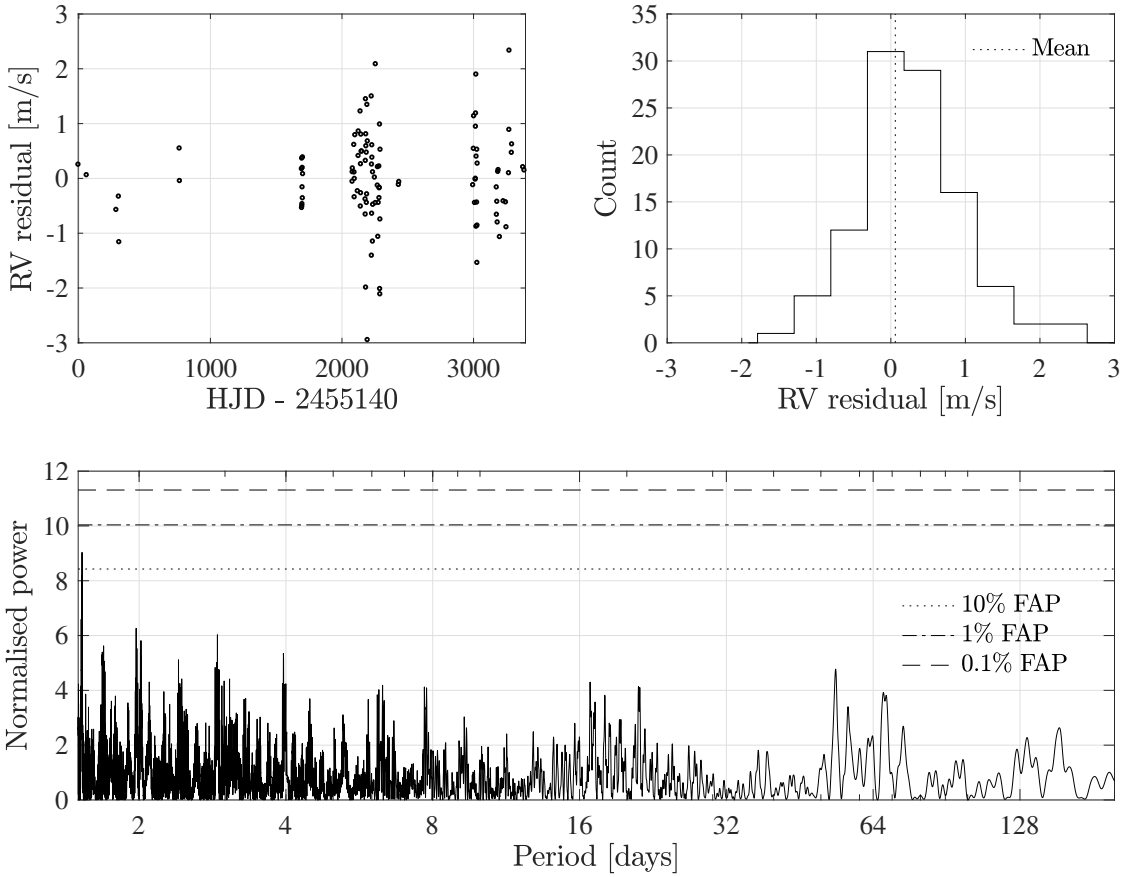


**Figure 6.4:** Diagnostics of the RV residuals for HD 175607, after subtracting the MAP GP model (stellar activity, plus polynomial trend and one planet). The top left panel contains the residuals plotted as a function of time, spanning some 9 yr; the top right panel contains a histogram of the residuals (RMS: 0.97 m $s^{-1}$; *cf.* mean RV error of 0.95 m $s^{-1}$); and the bottom panel shows the normalised Lomb-Scargle periodogram of the residuals, along with false alarm probability thresholds.

### 6.1.6 Discussion and conclusions

We analysed the HARPS RVs of HD 175607, a metal-poor ([Fe/H] = −0.62) G6V dwarf. The RVs show significant periodicities around 29 d, and around 700 or 1400 d.

If the longer-period signal is due to a planet, this planet would have a minimum mass of about 35 $M_\oplus$ or 0.1 $M_J$, making it a large Neptune. Unfortunately, the available RVs cover this putative planet's orbital phase very unevenly, so it is not possible to draw firm conclusions about the properties of this signal. Further observations are needed for reliable characterisation of this signal.

The 29 d signal has a statistical significance many orders of magnitude higher

than the longer-period signal. However, the star's estimated rotational period is also consistent with 29 d, so careful analysis was required before the observed periodicity could be attributed to a planet. Our joint analysis of the RVs with activity indicators (using a GP framework to model variability of stellar origin) led us to conclude that HD 175607 is indeed orbited by a planet, HD 175607 b, with orbital period 29 d, and minimum mass $8.98 \pm 1.10$ $M_\oplus$, which would make it a small Neptune. This conclusion was supported by a number of considerations, including (i) the lack of simple linear or rank correlations between the RVs and stellar activity indicators; (ii) the coherence of the 29 d signal over nine years, and the related fact that a simpler Keplerian model was favoured over a far more flexible GP model for this signal; and (iii) the consistency of the 29 d signal's behaviour with that of a synthetic Keplerian signal injected into the observations. The GP analysis ultimately allowed us to disentangle a coherent, 1.8 m $s^{-1}$ planetary signal from a quasi-periodic, 1.9 m $s^{-1}$ stellar signal, even though both appear to be characterised by an overall period of $\sim$ 29 d.

There are other known cases of a planet with orbital period the same as its host star's rotation period, such as CoRoT-11 b (Gandolfi et al., 2010) or XO-3 b (Hébrard et al., 2008). However, these are hot Jupiters orbiting rapidly-rotating stars, where the synchronous planetary orbit may arise from tidal locking with the host star (e.g. Lanza et al., 2011; Bolmont et al., 2012). The 29 d period of mini-Neptune HD 175607 b makes it implausible that the planet would have synchronised its orbit with its host star's rotation, as timescales for such synchronisation scale as $(a/R_*)^5 \cdot 1/M_p$ (e.g. Dobbs-Dixon et al., 2004; Brown et al., 2011). The planet *could* be tidally locked to the star, but there is no way of verifying this without knowing the planetary spin period.

The discovery of HD 175607 b was made as part of a large HARPS survey for Neptunes around metal-poor FGK dwarfs; HD 175607 b is the first Neptune-mass planet discovered in this survey. In fact, at the time of this analysis, HD 175607 was the most metal-poor FGK dwarf known to have an orbiting Neptune.[4] Giant

[4] As of June 2017, and according to the NASA Exoplanet Archive, there are at least three FGK dwarfs that are known to host planets *and* have lower metallicities than HD 175607.

planets are also rare around metal-poor stars, and it has been proposed that a lower metallicity limit ([Fe/H] $\gtrsim -0.7$ dex) could exist for the formation of giant planets (Mortier et al., 2012). It is still unclear if the same might hold true for Neptunes, or if we are simply limited in our abilities to detect such lower-mass planets. Either way, the discovery of HD 175607 b may have important consequences for planet formation and evolution theories.

## 6.2 Kepler-37d and Kepler-538b

### 6.2.1 Introduction

Kepler-37 (also known as KIC 8478994 and KOI-245; hereafter K-37) is a G-type main sequence star located 66 pc from Earth, with radius $0.77 \pm 0.03$ $M_\odot$, and mass $0.80 \pm 0.07$ $M_\odot$. The star is known to host at least four planets, all of which have been studied extensively: K-37b (Barclay et al., 2013; Marcy et al., 2014; Van Eylen and Albrecht, 2015), K-37c (*ibid.;* Holczer et al., 2016; Morton et al., 2016), K-37d (*ibid.;* Hadden and Lithwick, 2014), and K-37 e (Hadden and Lithwick, 2014). The inner three planets were detected through transit photometry, while the outer planet, K-37e, was detected through transit timing variations.

K-37b has orbital period 13.368 d, and radius $0.354 \pm 0.014$ $R_\oplus$ (Van Eylen and Albrecht, 2015); K-37c has orbital period 21.302 d, and radius $0.68^{+0.03}_{-0.02}$ $R_\oplus$ (Morton et al., 2016); K-37 d has orbital period 39.792 d, and radius $1.922 \pm 0.024$ $R_\oplus$; and K-37 e has orbital period 51.196 d, while its radius is not known. The periods of the inner three periods are close to (within 1% of) a 5 : 8 : 15 mean-motion resonance relationship. As of June 2017, K-37b holds the distinction of being the smallest known exoplanet. Its radius is just slightly larger than that of the Moon, and its mass cannot be more than a few times that of the Moon (which would lead to unphysically high densities: Marcy et al., 2014).

Given the orbital periods and radii of the K-37 planets, the RV signatures of K-37b and K-37c would be expected to be $< 10$ cm s$^{-1}$, and therefore not detectable with current instrumentation. (It might be the height of optimism to hope to detect the RV signature of the smallest and possibly also lightest known exoplanet!) The RV

signature of K-37d would be expected to be of order 1–2 m $s^{-1}$, and therefore likely detectable. K-37e would also be detectable if it were a mini-Neptune or gas giant.

The aim of the first strand of our analysis is to use RV measurements of K-37 to infer the mass of K-37d (and K-37e, if possible), and thus to obtain a more complete characterisation of the K-37 system.

Kepler-538 (also known as KIC 11623629 and KOI-365; hereafter K-538) is another is a G-type main sequence star; it is located $165^{+8}_{-7}$ pc from Earth, with radius $0.900^{+0.040}_{-0.1070}$ $M_\odot$, and mass $0.960^{+0.033}_{-0.040}$ $M_\odot$. K-538 is known to host at least one transiting planet, K-538b Holczer et al. (2016), which has orbital period 81.738 d, and radius $2.26^{+0.10}_{-0.09}$ $R_\oplus$, suggesting a likely super-Earth on a relatively long-period orbit.

The aim of the second strand of our analysis is to use RV measurements of K-538 to infer the mass of K-538b.

### 6.2.2 Observations

In the case of K-37, we incorporated the following observations into our analysis.

- 112 HARPS-N RVs (plus the usual activity indicators provided by the DRS pipeline, e.g. BIS, FWHM, $\log R'_{\rm HK}$, etc.), taken over two observing seasons, and spanning the date range 2014 Apr 15 to 2015 Nov 7. The median and mean time between observations was 1.01 d and 5.15 d, respectively. The rms variation of these RVs is 3.61 m $s^{-1}$, with median estimated error 0.96 m $s^{-1}$.
- 33 Keck-HIRES RVs (not accompanied by activity indices), taken over three observing seasons, and spanning the date range 2010 Apr 26 to 2012 Sep 4. The median and mean time between observations was 2.12 d and 26.93 d, respectively. The rms variation of these RVs is 2.88 m $s^{-1}$, with median estimated error 2.39 m $s^{-1}$.

In the case of K-538, we incorporated the following observations into our analysis.

- 83 HARPS-N RVs (plus activity indicators), taken over two observing seasons, and spanning the date range 2014 Apr 15 to 2015 Nov 7. The median and mean time between observations was 1.01 d and 6.17 d, respectively. The rms variation of these RVs is 4.19 m $s^{-1}$, with median estimated error 1.84 m $s^{-1}$.

- 26 Keck-HIRES RVs (not accompanied by activity indices), taken over five observing seasons, and spanning the date range 2010 Jul 25 to 2014 Jul 11. The median and mean time between observations was 14 d and 58 d, respectively. The rms variation of these RVs is 4.11 m $s^{-1}$, with median estimated error 1.31 m $s^{-1}$.

### 6.2.3 Analysis

A Lomb-Scargle analysis of the K-37 data revealed that the most signifiant sinusoidal periodicity in the RVs was at 29 d; this coincided with a significant complex of peaks in the BIS and FWHM power spectra at around 29 d. In the case of the K-538 data, the most significant peak in the RV data was found at 31 d, which coincided with the most significant peak in the BIS (but not FWHM) power spectrum at around 30 d. These findings pointed tentatively to rotation periods of approximately one month for both stars.[5]

In both the K-37 and K-538 data sets, we observed a strong correlation between the RV and FWHM observations: for K-37, we obtained a linear correlation coefficient of $\rho = 0.42$, with $p$-value for zero correlation of $p = 10^{-7}$; for K-538, the correlation coefficient was $\rho = 0.23$, with $p = 0.02$. In both cases this pointed to significant stellar activity contamination fo the RV time series. We observed weaker but still significant correlations between the RVs and BIS observations, while we found no statistically significant correlation between the RVs and $\log R'_{\rm HK}$ observations in either data set.

Thus, for both data sets, we elected to use the framework from R15 (see Chapter 3) to model jointly the RV, FWHM and BIS time series.

We modelled possible planetary signals in the RV data using Keplerian functions, with orbital elements as defined in Seager (2010). We placed non-informative priors on all parameters related to the activity components of the GP framework (see R15); Keplerian orbital parameters, incorporated into the GP's mean function, were

[5]The most significant peak in the window functions of both the K-37 and K-538 was at 1.00 d. Both window functions did contain peaks 34 d and $36 \pm 1$ d; however, the peaks at these periods were distinct from the most significant ones in the RVs and activity indicators, and indeed aliases of the latter periods appeared at the locations expected given the sampling window.

**Table 6.2:** Posterior probability summaries for the main GP covariance and mean function parameters (activity plus one planet model) for our favoured K-37 d fit. Keplerian orbital elements are as defined in Seager (2010), except for $\Delta T_{\rm tr}$, which is computed from the periapsis passage time $T_0$, but with origin set by the K-538 d mid-transit time published by Morton et al. (2016), which we define to have a value of $\Delta T_{\rm tr} = 0$.

| Parameter | Description | Median | $\pm\sigma$ |
|---|---|---|---|
| $K_{\rm GP}$ | GP RV semi-amplitude | 2.68 m s$^{-1}$ | $\pm$0.52 m s$^{-1}$ |
| $P$ | GP period | 30.29 d | $\pm$0.24 d |
| $\lambda_{\rm p}$ | GP inv. harmonic complexity | 0.43 | $\pm$0.04 |
| $\lambda_{\rm e}$ | GP evolution time-scale | 61 d | $^{+1.19}_{-0.73}$ d |
| $\gamma$ | System velocity | $-$30.651 km s$^{-1}$ | $\pm$3 m s$^{-1}$ |
| $K_d$ | RV semi-amplitude | 1.02 m s$^{-1}$ | $\pm$0.15 m s$^{-1}$ |
| $P_d$ | Period | 39.792217 d | $\pm$4 s |
| $e_d$ | Eccentricity | 0.028 | $^{+0.028}_{-0.012}$ |
| $\omega_d$ | Periapsis longitude | 1.21$\pi$ | $^{+0.58}_{-0.95}\pi$ |
| $\Delta T_{\rm tr,d}$ | $\Delta$ Transit mid-point | +7 s | $^{+35}_{-31}$ s |
| $M_d$ | Mass | 4.68 M$_\oplus$ | $\pm$0.72 M$_\oplus$ |
| $\rho_d$ | Mean density | 3.52 g cm$^{-3}$ | $^{+1.30}_{-0.92}$ g cm$^{-3}$ |

constrained using available information from transits (Morton et al., 2016; Holczer et al., 2016). We performed all parameter inference using the MultiNest nested-sampling algorithm (Feroz and Hobson, 2008; Feroz et al., 2009, 2013). Typical computation time to obtain posterior distributions for models including between zero and a few planets, and using MultiNest's default convergence criteria, was of the order of several hours on a modern laptop.

### Kepler-37

For the 'combined' (HARPS-N plus HIRES) K-37 data set, we obtained an integrated likelihood (model evidence) of $\log \mathcal{Z} = -163.31 \pm 0.01$ for a 0-planet model, and $\log \mathcal{Z} = -158.37 \pm 0.01$ for a model including K-37d. Thus, the 1-planet model was favoured by a factor of more than 100 over the 0-planet model, indicating very strong evidence in support of the more complex planetary model.

Posterior distributions for parameters of interest are summarised in Table 6.2. Note that the periapsis longitude is poorly constrained, as this parameter is only well-defined for solutions with $e > 0$, and our MAP solution is consistent with $e = 0$. MAP fits to the K-37 data are shown in Fig. 6.5 and Fig. 6.6.

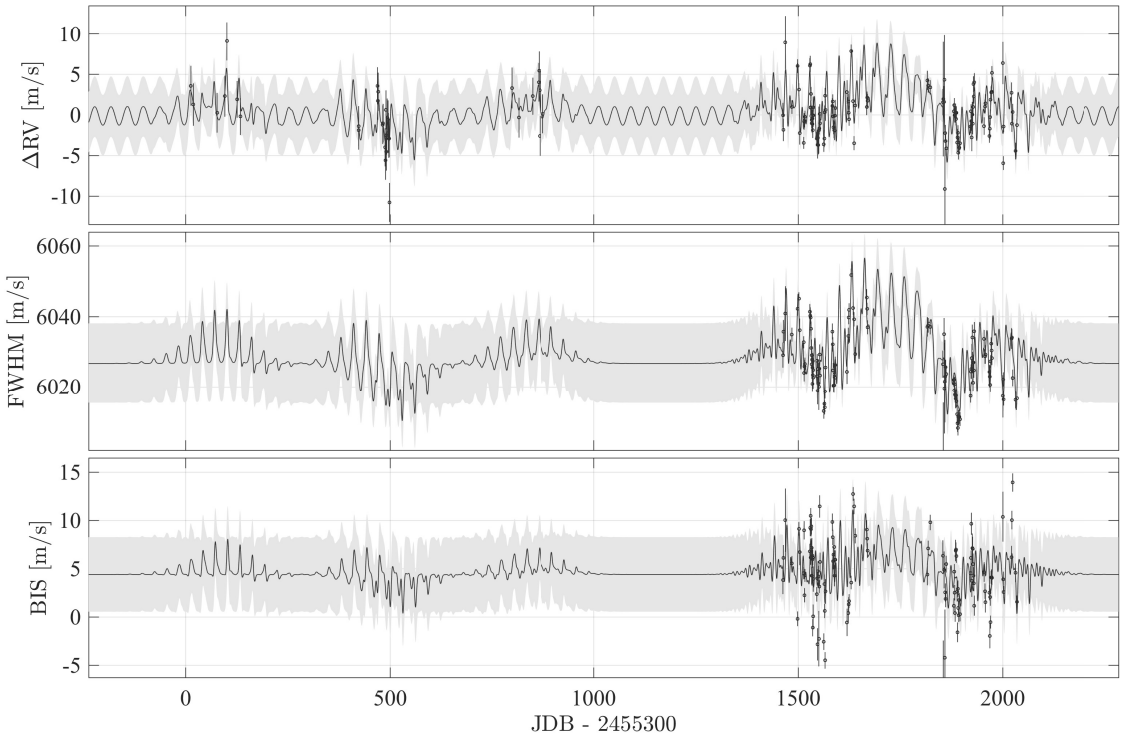


**Figure 6.5:** GP plus planet model MAP fit to the K-37 RVs and activity indicators. Black dots indicate observed data, with estimated errors; solid lines are model posterior means; and shaded regions denote $\pm\sigma$ posterior uncertainty.

The rms of the mean-subtracted RVs for K-37 was 3.06 m s$^{-1}$; after fitting the 0-planet model, the residual rms was 1.79 m s$^{-1}$; and for the 1-planet model, the residual rms was 1.68 m s$^{-1}$. For comparison, the rms of the estimated observational errors for the combined data set was 1.73 m s$^{-1}$, and our MAP model did not contain any added white noise.

We also investigated a number of other models, with between 2 and 5 planets (with orbital periods and periastron times either constrained to the values known for K-37b, c, d, or e; or with the orbital parameters entirely unconstrained). In all cases, we obtained model evidences of $\log \mathcal{Z} < -165$, indicating that none of these higher-order models was favoured over the single-planet (K-37d) model. Moreover, in the cases where we constrained the model orbital parameters to the values known for K-37b, c, and/or e, the planet masses we inferred were consistent with zero. Thus we detected no RV signatures for any of the other confirmed planets in the K-37 system.

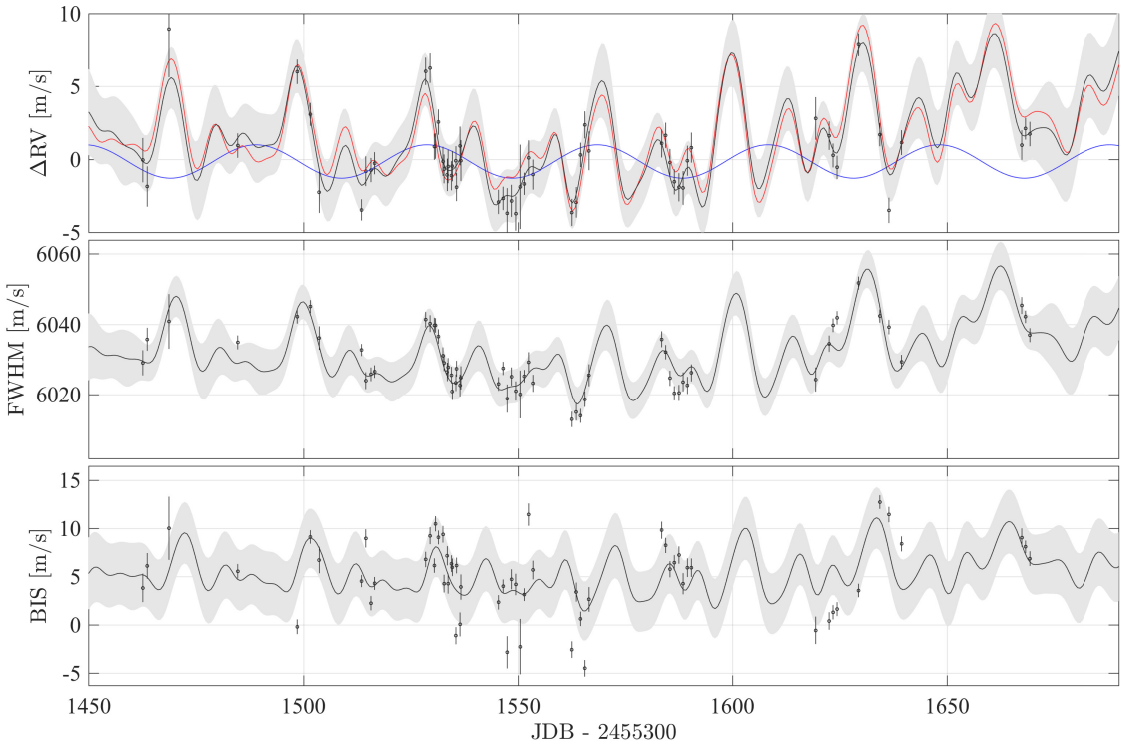


**Figure 6.6:** As for Fig. 6.5 but zoomed in, and in the case of the RVs, with the GP component (red) separated from the Keplerian component (blue) of the MAP model.

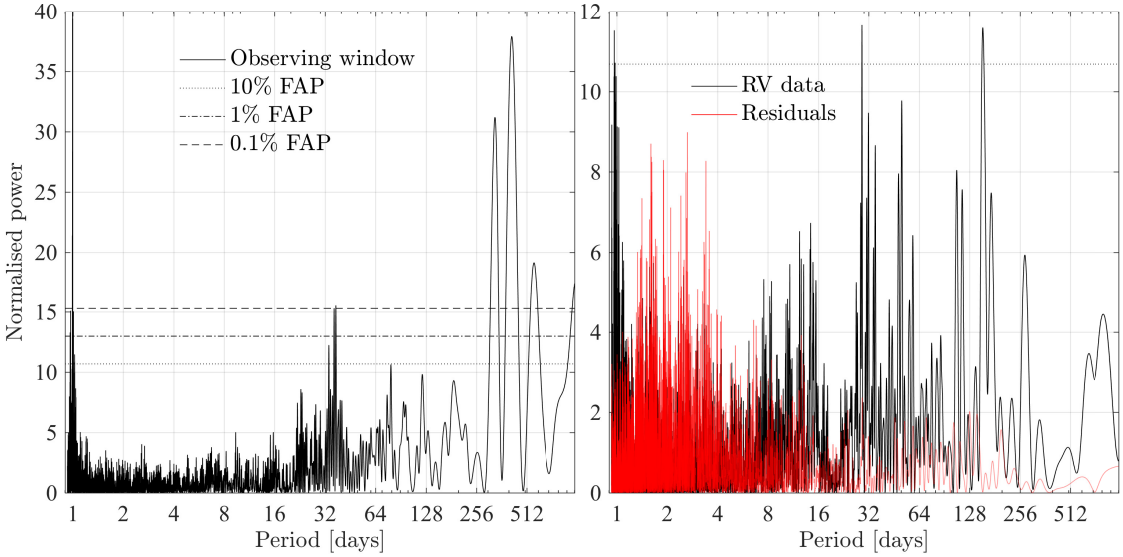


**Figure 6.7:** Normalised Lomb-Scargle periodogram of the observing window for the K-37 data (left panel), and of the RV residuals from the MAP GP plus 1-planet fit to the K-37 data (right panel). False alarm probabilities are plotted in both cases; note the different vertical scales in the two plots.

**Table 6.3:** As for Table 6.2 on pg. 171, but for our favoured (activity plus one planet) K-538b fit.

| Parameter | Description | Median | $\pm\sigma$ |
|---|---|---|---|
| $K_{\rm GP}$ | GP RV semi-amplitude | 1.55 m s$^{-1}$ | $\pm$0.58 m s$^{-1}$ |
| $P$ | GP period | 30.18 d | $\pm$0.13 d |
| $\lambda_{\rm p}$ | GP inv. harmonic complexity | 0.39 | $^{+0.05}_{-0.04}$ |
| $\lambda_{\rm e}$ | GP evolution time-scale | 250 d | $^{+51}_{-48}$ d |
| $\gamma$ | System velocity | $-$37.322 km s$^{-1}$ | $\pm$4 m s$^{-1}$ |
| $K_b$ | RV semi-amplitude | 1.35 m s$^{-1}$ | $\pm$0.27 m s$^{-1}$ |
| $P_b$ | Period | 81.737862 d | $\pm$8 s |
| $e_b$ | Eccentricity | 0.027 | $^{+0.022}_{-0.011}$ |
| $\omega_d$ | Periapsis longitude | 1.31$\pi$ | $^{+0.50}_{-1.04}\pi$ |
| $\Delta T_{\rm tr,b}$ | $\Delta$ Transit mid-point | $-$36 s | $^{+60}_{-53}$ s |
| $M_b$ | Mass | 8.9 M$_\oplus$ | $\pm$1.8 M$_\oplus$ |
| $\rho_b$ | Mean density | 4.23 g cm$^{-3}$ | $^{+1.10}_{-0.98}$ g cm$^{-3}$ |

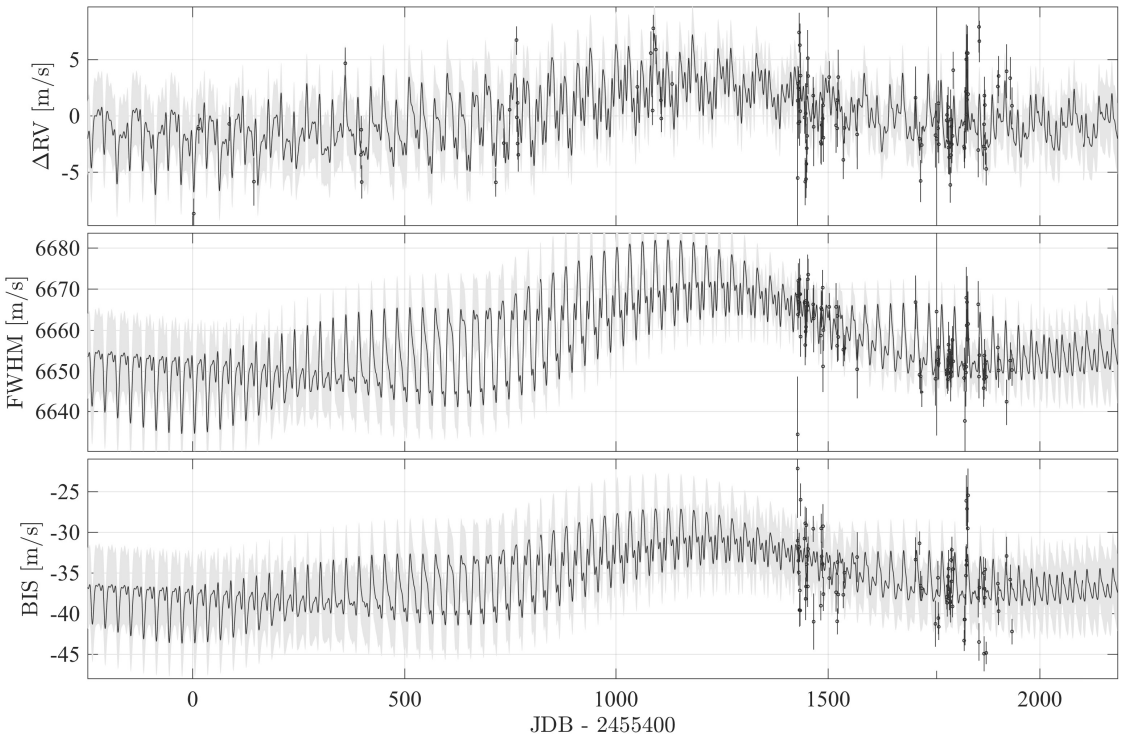


**Figure 6.8:** As for Fig. 6.5 on pg. 172, but for the K-538 data.

### Kepler-538

For the combined K-538 data set, we obtained $\log \mathcal{Z} = -201.79 \pm 0.01$ for a 0-planet model, and $\log \mathcal{Z} = -193.75 \pm 0.01$ for a 1-planet model. Thus, the 1-planet model was favoured by a factor of more than 1000 over the 0-planet

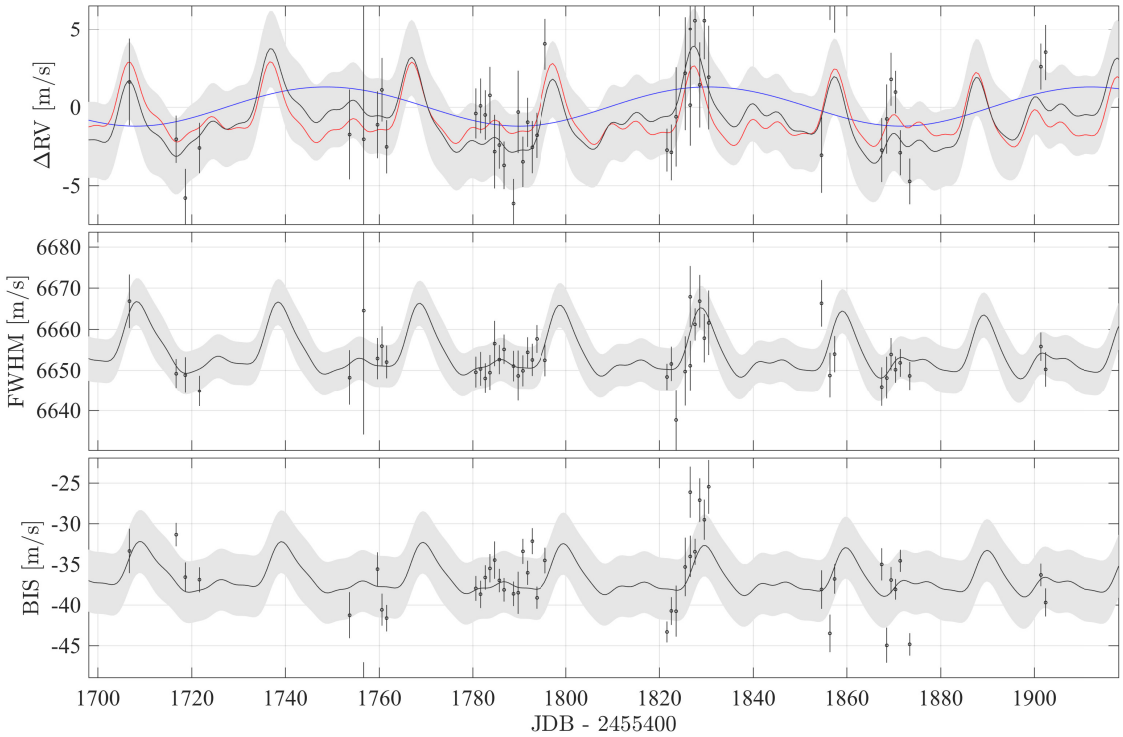


**Figure 6.9:** As for Fig. 6.6 on pg. 173, but for the K-538 data.

model, indicating overwhelming evidence in support of the more complex planetary model. Posterior distributions for parameters of interest from the 1-planet model are summarised in Table 6.3.

The rms of the mean-subtracted RVs for K-538 was 3.46 m $s^{-1}$; after fitting the 0-planet model, the residual rms was 2.30 m $s^{-1}$; and for the 1-planet model, the residual rms was 2.36 m $s^{-1}$. For comparison, the rms of the estimated observational errors in the combined data set was 2.36 m $s^{-1}$, and our MAP model did not contain any added white noise. Note that although the residual rms is larger for the 1-planet model, the complexity penalty for the planetary solution is much lower than when using the GP activity model to explain the observed RV variations, so the 1-planet model was strongly favoured in the model comparison test. Given the weaker correlation between the RVs and activity indicators observed in the K-538 data set (when compared to K-37), it is to be expected that the GP activity model will not do as good a job of explaining all the observed RV variations.

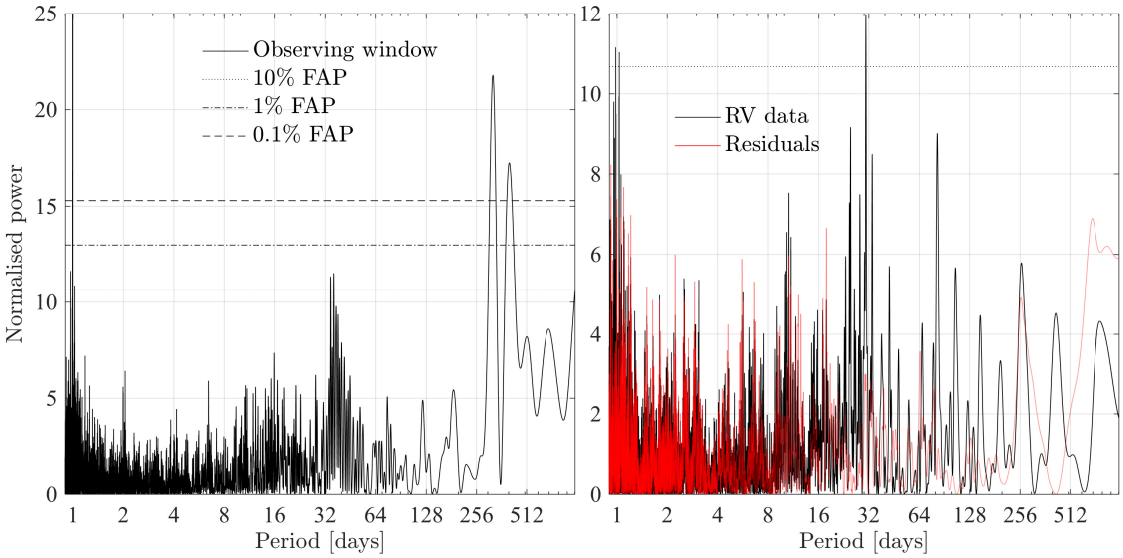


**Figure 6.10:** As for Fig. 6.7 on pg. 173, but for the K-538 data.

We also fitted to the data models containing up to 3 planets in total, i.e. K-538 b plus up two additional planets (with unconstrained orbital parameters for the additional planets). In all cases, we obtained model evidences $\log \mathcal{Z} < -200$, indicating that none of the higher-order models was favoured over the single-planet (K-538 b) model Thus we did not find strong RV evidence for further planets in the K-538 system.

### Additional support for the inferred masses

We performed a number of diagnostic tests, all of which lent support to our conclusions about the planets' masses. These tests are summarised below.

Firstly, we performed injection tests using Keplerian signals of similar properties to those of the planets we detected ($\Delta$RV $\sim 1$ m s$^{-1}$, periods on order of several weeks to months), and time sampling and noise/activity levels identical to the real data sets. The GP modelling framework led to reliable inferences about the periods and semi-amplitudes of these injected signals. This basic 'sanity check' confirmed that our modelling approach could be trusted to yield sensible conclusions about putative $\sim 1$ m s$^{-1}$ planetary signals in the real data sets.

Secondly, when not including the known transit ephemerides into the model priors, for both K-37d and K-538b the posterior distributions for the planets' predicted transit times nevertheless ended up tightly bracketing the known transit ephemerides. Conclusions about the planets' MAP RV semi-amplitudes (masses) also remained unchanged when not using the known transit ephemerides in the modelling, although discarding this information did broaden the posterior distributions for the planets' RV semi-amplitudes.

Thirdly, inferred planet parameters were insensitive to the exact treatment of noise and systematics. For example, even when allowing for a significant additive white noise component in our GP model, or when allowing a much larger activity signal amplitude, we ended up with consistent planet parameters.

Fourthly, when discarding the GP component of our model entirely – in other words, fitting only a Keplerian plus white noise, or a Keplerian plus white noise plus a linear component linking $\Delta$RV and FWHM observations – we obtained planet masses consistent (within $1\sigma$) with the masses obtained via the GP modelling approach. Given that the latter, GP-based approach placed much tighter constraints on the activity component of the $\Delta$RV observations, however, the resultant uncertainty in the inferred masses was smaller, i.e. the mass estimates were more precise. Moreover, the less complex, GP-free models were rejected on the basis of model evidences hundreds or thousands of times lower than the corresponding evidences for the GP plus planet models.

Finally, for both systems, the Bayesian evidence favouring the 1-planet model increased monotonically as we randomly and artificially increased the number of data points used in the modelling: behaviour consistent with expectations for genuine Keplerian signals.

### 6.2.4 Discussion and conclusions

In each data set, we made a robust detection of a planet, supported by multiple pieces of evidence including Bayes factors, consistency with transit information, insensitivity to treatment of systematics, and consistency with injection tests.

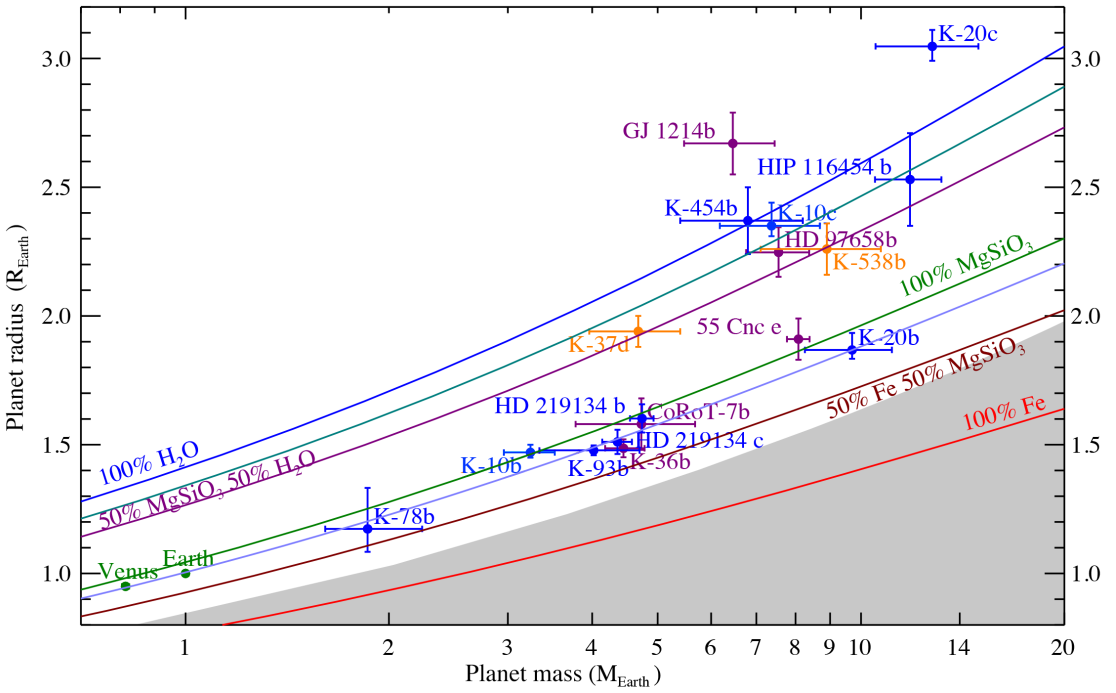


**Figure 6.11:** Mass-radius relation for planets smaller than 3.2 $R_\oplus$ and mass determinations better than 20% precision, as of June 2017. The details are as for Fig. 5.5 on pg. 152, but this plot adds the two planets (K-37d and K-538b) characterised in the analysis on the preceding pages; both new planets are shown in orange. The data for K-10b and K-10c are taken from the analysis in Chapter 5.

K-37d's radius of $1.92 \pm 0.02$ $R_\oplus$ and inferred mass of $4.68 \pm 0.72$ $M_\oplus$ points to a super-Earth with composition consistent with 50% $H_2O$ and 50% $MgSiO_3$. The non-detection of K-37e indicates an RV semi-amplitude $< 1$m $s^{-1}$; this suggests a mass lower than about 5 $M_\oplus$, assuming an inclination near 90°. K-538's radius of $2.26 \pm 0.1$ $R_\oplus$ and inferred mass of $4.68 \pm 0.72$ $M_\oplus$ points to a super-Earth with composition consistent with that of K-37d. See Fig. 6.11.

K-37d and K-538b thus join a relatively small number of confirmed planets with long-period orbits ($> 1$ month) and better than 20%-precision mass constraints. According to the NASA Exoplanet Archive, of June 2017, there are only 6 'small' (radii $\lesssim 3.2$ $R_\oplus$) confirmed exoplanets satisfying these criteria, and only about 10 such super-Earths (radii $< 10$ $R_\oplus$).

At the time of writing, analysis of both data sets is ongoing; evidence is emerging for a possible fifth planet in the K-37 system, i.e. K-37f, with period $148 \pm 1$ d.

*No empirical science is intended merely as a coherent fairy-tale. It is intended to consist of statements having application to the real world and believed because of their relation to that world.*

— Bertrand Russell, from the paper *Non-demonstrative inference* (Russell, 1959).

# 7

# A few other GP applications

## Contents



This, the penultimate chapter of this thesis, moves beyond the framework developed in Chapter 3 (and then applied in Chapters 4 to 6), and presents three further applications of GPs in the context of modelling stellar signals and detecting or studying exoplanets. Given space constraints, and that in most regards these applications remain 'works in progress,' each application is sketched only very briefly.

## 7.1 Characterising stellar rotation

### 7.1.1 Introduction

As noted in Section 2.3 of this thesis, spotted, rotating stars exhibit photometric and RV variability that is generally neither sinusoidal nor strictly periodic. The stochastic appearance, growth, persistence, and then decay of rotating active regions leads to quasi-periodic signals, governed by some overall stellar rotation period – or a distribution of periods, in the case of differential rotation.

The preceding several chapters of this thesis were concerned with identifying these stellar signals with a view to disentangling them from any exoplanet-induced variability present, so that the latter signals could be studied more accurately. Instead of trying to 'sweep away' these stellar signals, however, let us now concern ourselves with trying to characterise as accurately as possible the properties of the rotation giving rise to the stellar signals in the first place (as might be an essential part of a gyrochronological study: Barnes, 2007, 2010; Angus et al., 2015).

As discussed in Section 2.4 of this thesis, various 'standard' approaches are widely used to identify or model the (quasi)-periodic photometric or RV variability induced by active regions on rotating stars – including Fourier methods, GLS periodograms, ACF methods, wavelet methods, and spot modelling. Unfortunately, all of these approaches suffer from a number of fundamental drawbacks, e.g. assumption of an unrealistic (e.g. single stationary sinusoid) model, inability to deal with correlated noise or to account for observational uncertainties, vast and degenerate parameter spaces, non-probabilistic period estimates, and so on.

As before, then, these drawbacks lead us quite naturally to consider GPs as a principled, data-driven, and probabilistic alternative to the 'standard' techniques for characterising stellar rotation.

### 7.1.2 GP rotation period inference

For the same reasons outlined in Chapters 2 and 3 of this thesis, a GP with a QP (locally periodic) covariance kernel serves as a sensible prior for the functions that

could describe the RV or photometric signals of rotating, evolving active regions. The QP covariance kernel derived in Chapter 2, and used throughout the chapters that followed, was obtained by multiplying an SE kernel with a periodic one:

$$k_{\text{QP-SE}}(t_i, t_j) = h^2 \exp\left(-\frac{\sin^2\left(\frac{\pi(t_i - t_j)}{P}\right)}{2\lambda_{\text{p}}^2} - \frac{(t_i - t_j)^2}{2\lambda_{\text{e}}^2}\right); \tag{7.1}$$

Given some photometric or RV observations $\mathcal{D} = \{(t_i, y_i)\}_{i=1}^n = (\mathbf{t}, \mathbf{y})$, we can now follow the approach set out in Section 2.6.3 by assuming that some functional relationship $y = f(t)$ exists, and then placing a GP prior on $f$, i.e. $f \sim \mathcal{GP}(\rho, k_{\text{QP-SE}})$. Using the relations derived in Section 2.6.3, we can then evaluate the posterior distributions for various hyperparameters, in order to make inferences about the stellar rotation giving rise to any (quasi-)periodicity present in $\mathbf{y}$. In particular, the marginal posterior distribution for the hyperparameter $P$ will provide both an estimate for the stellar rotation period, as well as detailed information about the uncertainty in this estimate.

Using a GP with QP kernel to infer a stellar rotation period is not a new idea (Haywood et al., 2014; Rajpaul et al., 2015; Vanderburg et al., 2015; Littlefair et al., 2017). However, there does not appear to exist discussions in the literature on the subtleties involved in interpreting the hyperparameters of our QP kernel, nor much in the way of guidelines about sensible priors for these hyperparameters.[1] Examples abound of authors simply fixing the values of some hyperparameters, e.g. $\lambda_{\text{p}}$, to 'reasonable' values. The next section attempts to fill in some of these gaps in the literature.

### 7.1.3 Interpreting and constraining the hyperparameters

Because the hyperparameters defining $k_{\text{QP-SE}}$ (i.e., $h$, $P$, $\lambda_{\text{p}}$, and $\lambda_{\text{e}}$) serve to define the covariance structure of functions, rather than the shapes of the functions themselves, their interpretation is not always straightforward.

[1] As Bretthorst (1988) demonstrated, the incorporation of appropriate prior information, even if only qualitative in nature, can have an enormously beneficial effect on the *quantitative* information one can extract from a given set of observations. See also Appendix B.

### Signal variance, $h^2$

The scale hyperparameter $h^2$ defines the signal variance of functions generated by the $k_{\text{QP-SE}}$ covariance kernel; loosely, $h$ may be interpreted as an amplitude for these functions. When modelling the signals of spots on a rotating star, we expect $h$ to be related (amongst other factors) to the fraction of an observed stellar disc covered in spots.

For fixed values of $P$, $\lambda_{\text{p}}$, and $\lambda_{\text{e}}$, simple numerical experiments reveal that the relation between $h$ and the rms of typical mean-subtracted functions $f$ will take the form[2] $\log \text{rms}(f) = \log(|h|) - \alpha^2$, where $\alpha = \alpha(P, \lambda_{\text{p}}, \lambda_{\text{e}})$, $\alpha \in \mathbb{R}$; as $\lambda_{\text{p}} \to 0$, $\alpha^2 \to 0$, so that $\text{rms}(f) = |h|$. An analogous expression may be written down for the relationship between $h$ and the typical semi-amplitude of functions generated by $k_{\text{QP-SE}}$.

To see why $\alpha$ depends on the hyper-parameters, consider a set of observations with baseline $\Delta T$, in which we know a periodic signal is present. Provided $P \ll \Delta T$, the observations should cover several cycles of the periodic signal, so we don't expect severe under- or over-estimation of the signal rms due to inadequate coverage, and $P$ becomes largely irrelevant. However, $k_{\text{QP-SE}}$ decays monotonically with increasing values of $|t_i - t_j|$, and the rate of decay is governed by $1/\lambda_{\text{e}}$. Therefore, to achieve a given covariance value, the 'amplitude' $h$ has to be proportionately larger for larger values of $\lambda_{\text{e}}$ compared to smaller values of $\lambda_{\text{e}}$. Similar considerations apply to $\lambda_{\text{p}}$; moreover, one has to take into account 'interference' between the $\lambda_{\text{p}}$ and $\lambda_{\text{e}}$ terms inside the exponential function defining $k_{\text{QP-SE}}$.

More intuitively, one may reflect that very complex, non-sinusoidal variations should not necessarily be expected to have the same rms variation or semi-amplitude as sinusoidal variations (think, for example, of a function that oscillates several times near zero before undergoing one large excursion, and then repeats itself); the

[2] We could write down the same relationship without the logarithms. However, as there is an inherent scatter in the rms variation of randomly-generated functions, and as this scatter is proportional to the amplitude of the random functions, the relationship between $h$ and $\text{rms}(f)$ is most easily inspected in log space, especially when $h$ is allowed to span many orders of magnitude.

same considerations apply to a quasi-periodic function whose shape changes over time, regardless of the harmonic complexity of the function.

Given that $|h| = \mathrm{rms}(f) \cdot e^{\alpha^2} > \mathrm{rms}(f)$, one can use the rms variation in the observed data $\mathbf{y}$ to place a lower prior bound on $h$ (assuming all the variation is to be explained by the GP; straightforward adjustments can be made when this is not the case). For typical, non-pathological values of $\lambda_\mathrm{p}$ and $\lambda_\mathrm{e}$ (e.g. $0 < \lambda_\mathrm{p} < 10$, $1 < \lambda_\mathrm{e}/P < 10$; see below), one finds that $0 < \alpha^2 \lesssim 1$; this, in turn, allows upper prior bounds to be placed on $h$.

Thus we find that the interpretation of $h$ is not quite as straightforward as one might have hoped, although in practice it is usually safe to regard $h$ as an amplitude, to within a factor of about 2.

It seems that it should be possible to parametrise (even if only approximately) the relationship between $\alpha$ and the hyper-parameters; such an investigation is deferred to future work.

### Period, $P$

The hyper-parameter $P$ may generally be identified as a 'period' in the conventional sense. To wit, if a function $f(t)$ is strictly periodic with period $P$, as will be the case when $\lambda_\mathrm{e} \to \infty$, then $f(t + P) = f(t)\ \forall\ t \in \mathbb{R}$. For finite values of $\lambda_\mathrm{e}$, $f(t)$ will be quasi-periodic, in which case $f(t) \approx f(t + nP)\ \forall\ t \in \mathbb{R}$, $n \in \mathbb{Z} : |nP| \lesssim \lambda_e$. The only caveat (encoded in $|nP| \lesssim \lambda_e$) is that the evolution time-scale, $\lambda_e$, should be large enough that $P$ *can* be interpreted as a period, for if $\lambda_e$ is similar to or smaller than $P$, then functions will change too rapidly for even quasi-periodicity to be well-defined. See below for more on how $\lambda_\mathrm{e}$ should be constrained.

Given the relatively straightforward interpretation of $P$, prior constraints for $P$ should simply be informed by whatever knowledge one has of the (quasi)-periodicity inherent in the signal in question. In the case of modelling a spotted, rotating star, such prior constraints may be informed by the properties of stars in general (e.g., rotation periods of 5 minutes or 1000 yr could be ruled out for a main-sequence star), the properties of a given star (e.g., its age, spectral type, and activity level),

and/or preliminary analyses of the signal in question (e.g., an ACF analysis might point to an approximate period which a GP model could then refine).

**Harmonic complexity, $\boldsymbol{\lambda_p}$**

$\lambda_p$ seems to be the QP hyperparameter that is most often misinterpreted (e.g. as a 'smoothness' parameter), or simply set to a value claimed to be 'reasonable,' e.g. $\lambda_p = 2$. In fact, $\lambda_p$ determines the harmonic complexity of functions generated by generated $k_{\text{QP-SE}}$. As $\lambda_p \to \infty$, functions become sinusoidal; for smaller values, the functions have increasingly complex harmonic content.

One intuitive and non-parametric way of gaining insight into what this means is to consider the number of turning points and inflection points (quasi-)periodic functions have per cycle. This may be investigated numerically by generating a large number of random functions for fixed hyperparameter values, examining the typical numbers of stationary points and inflection points in these functions, repeating for different hyperparameter values, and so on. Such experiments reveal that provided $\lambda_e > P$, the expected number of turning points per period, $\langle N \rangle$, can be approximated by

$$\langle N \rangle \approx \frac{1}{2}\left(3\lambda_p^{-1} + 4\right), \tag{7.2}$$

while the expected number of inflection points per period, $\langle M \rangle$, can be approximated by

$$\langle M \rangle \approx 2\lambda_p^{-1} + 2. \tag{7.3}$$

Both of these approximations are good to within about 10%, and for larger values of $\lambda_p$, the scatter in $M$ and $N$ becomes smaller: $\sigma^2(N) \sim \lambda_p^{-3/4}$, and $\sigma^2(M) \sim \lambda_p^{-2/3}$.

We see that for large values of $\lambda_p$, functions with period $P$ will almost always have exactly two turning points per cycle: the simplest possibility for a differentiable periodic function.[3]

Prior upper bounds on $\lambda_p$ may be informed by the consideration that in practice it is usually not feasible to use a finite set of noisy data to distinguish between

[3]Consider $g(t) := f(t) - f(t + P/2)$, with $f(t)$ differentiable and periodic with period $P$. $g(t + P/2) = f(t + P/2) - f(t + P) = f(t + P/2) - f(t) = -g(t)$; thus, by the intermediate value theorem, $g(t)$ has at least one zero. But $g(t) = 0 \implies f(t) = f(t + P/2) = f(t + P)$; thus, by the mean value theorem, there exist points $\xi_1 \in (t, t + P/2), \xi_2 \in (t + P/2, t + P) : f'(\xi_1) = f'(\xi_2) = 0$.

the sinusoidal-like variations with $\lambda_{\rm p} \gtrsim 5$ and with $\nu \to \infty$. Lower bounds on $\lambda_{\rm p}$ may be informed by the consideration that, in most circumstances, the signals from complex configurations of active regions may be well approximated by around three or four Fourier terms (Jeffers and Keller, 2009), indicating a relatively low harmonic complexity consistent with $\lambda_{\rm p} \gtrsim 0.5$. This is easily verified by direct simulation, e.g. using the SOAP 2.0 code (Dumusque et al., 2014a).

### Evolution time-scale, $\boldsymbol{\lambda_{\rm e}}$

Intuition suggests that in order for functions to have clear (quasi-)periodicity, it is necessary to enforce $2\lambda_{\rm e} > P$; if $2\lambda_{\rm e} < P$, covariance decays by a factor of $e$ over time-scales shorter than a single period, in which case $f(t+P)$ would look significantly different to $f(t)$, and inference about periodicity becomes futile.

In fact, it is possible to derive analytically a more stringent constraint. If we require that the covariance kernel itself should have at least one local maximum (apart from at $\tau := t_i - t_j = 0$), we need to solve $\partial k/\partial\tau = 0$. It may be shown that this leads to the following constraint:

$$\lambda_{\rm e} > \sqrt{\frac{3P^2\lambda_p^2}{2\pi}}; \tag{7.4}$$

$\lambda_{\rm e} = \sqrt{3P^2\lambda_p^2/2\pi}$ is a critical point at which turning points disappear from $k_{\text{QP-SE}}$.[4] For $\lambda_{\rm p} \sim \mathcal{O}(1)$, equation 7.4 dictates $\lambda_{\rm e} \gtrsim \sqrt{2}P$.

Constraints on $\lambda_{\rm e}$ should, of course, also be informed as far as possible by astrophysical considerations. For example, if it is known that the typical persistence lifetime of active regions on a certain type of star is of order at least two rotation periods, this would suggest $\lambda_{\rm e} > 2P$. Upper bounds on $\lambda_{\rm e}$ may be informed by similar practical considerations as for $\lambda_{\rm p}$: given a finite set of noisy observations, models with $\lambda_{\rm e}/P \gg 10$ will probably be indistinguishable from those with $\lambda_{\rm e}/P \to \infty$.

[4] This result does not seem to appear in the literature.

### 7.1.4 Performance on a simple problem

In order to perform some simple tests of how a GP with QP kernel may be used for characterising stellar rotation – and especially to probe the impact of placing sensible priors on the hyperparameters – I generated 1000 synthetic light curves of spotted, rotating stars. The stellar rotation periods varied from 1 d to 50 d, and the stellar surfaces contained between 1 and 200 rotating dark spots or bright plages, each with a persistence lifetime of between 0.5 and 5 rotation periods. Each synthetic light curve was sampled with between 100 and 500 observations (spanning 1500 d, though including large gaps in the temporal coverage), each with a signal-to-noise ratio of between 1 and 50.

I then compared the accuracy of period inference using a GP (specifically, from the posterior distribution for the period hyper-parameter in $k_{\text{QP-SE}}$) with that from a GLS periodogram analysis (i.e. best-fitting sinusoidal period) and an ACF analysis (implementation as per McQuillan et al., 2014). Posterior sampling in the case of the GP was performed using the MultiNest nested sampling algorithm (Feroz and Hobson, 2008; Feroz et al., 2009, 2013).

I found that, even without any informative prior constraints on the GP hyper-parameters, periods could be inferred slightly more accurately than either GLS or ACF methods permitted, although $\sim 10\%$ of the inferred GP periods differed by more than $4\sigma$ (as estimated directly from the posterior) from the true rotation periods. When imposing the 'generic' constraints discussed in the previous section (e.g. $\lambda_{\text{e}} > \sqrt{3P^2\lambda_p^2/2\pi}$, $0.1 < \lambda_{\text{p}} < 10$, etc.), the results improved significantly. More than 90% of the inferred periods agreed with the true periods within $2\sigma$, while fewer than 1% of the periods diverged from the true periods by more than $3\sigma$. Moreover, the median absolute error of the GP periods was about an order-of-magnitude smaller than that of the GLS periods, and about half an order of magnitude smaller than that of the ACF periods.

Fig. 7.1 shows an example of a synthetic light curve, and the corresponding GP period posterior and GLS power spectrum. Fig. 7.2 compares the error distributions from the GP analysis and the GLS analysis.

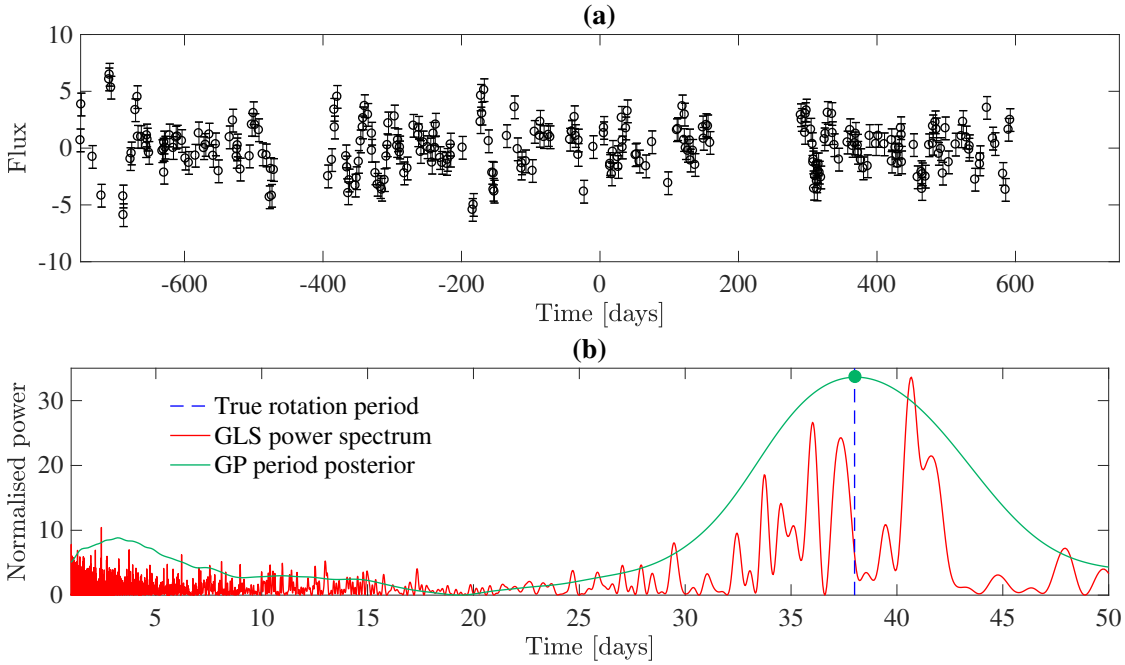


**Figure 7.1:** (a) One of the 1000 synthetic light curves of a spotted, rotating star. In this example the star has rotation period 38 d. (b) GLS power spectrum vs. GP period posterior distribution for this light curve (with the posterior renormalised to have the maximum value as the power spectrum). Note that the latter distribution is almost unimodal, and peaks very near to (within a few hours of) the true rotation period of 38 d. The GLS power spectrum, by contrast, is hugely multimodal, and achieves its peak value some 4 d away from the correct rotation period.

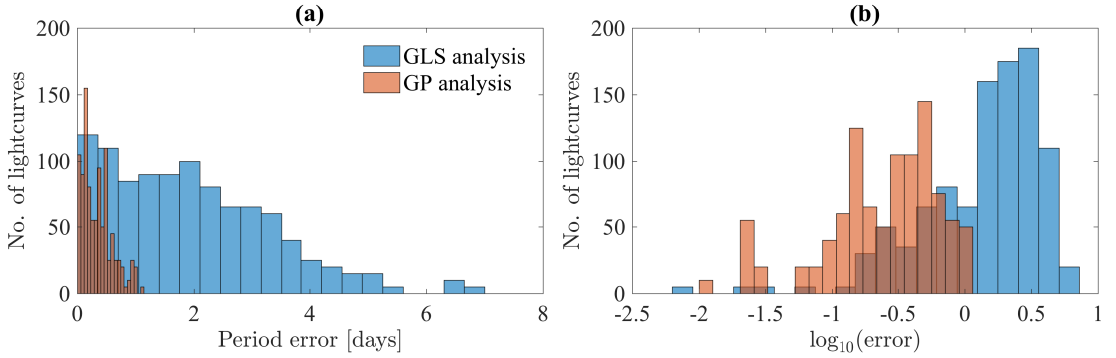


**Figure 7.2:** Distribution of absolute errors in the periods estimated by the GP analysis and the GLS analysis, scaled (a) linearly and (b) logarithmically. The median error from the GP analysis is an order of magnitude smaller than that from the GLS analysis.

Of course, the synthetic data set on which this test was based was somewhat simplistic, as was my implementation of the GLS and ACF analyses. Nevertheless, the test served at least as a 'sanity check' that demonstrated (i) that a standalone GP model can and indeed does handily outperform conventional models for the quasi-period signals from a spotted, rotating star, and moreover, (ii) that placing sensible priors on the GP hyperparameters can significantly improve inference about rotation periods.

### 7.1.5 Applications to real data

The paper *Inferring probabilistic stellar rotation periods using Gaussian processes* (Angus et al., 2018), of which I am a co-author, attempts to recover the rotation periods of several hundred, simulated *Kepler*-like light curves for solid-body rotators (with a range of stellar activity levels, spot lifetimes, inclination angles, etc.; see Aigrain et al., 2015) using three different methods: a GP method akin to the one described in Section 7.1.2 above; an ACF method; and a GLS method. The paper demonstrates that the GP method produces the most accurate rotation periods of all three methods, with significantly fewer outliers. The same paper then goes on to measure the rotation periods of over a thousand *Kepler* Object of Interest host stars, all of which had rotation periods previously measured by an ACF method (McQuillan et al., 2013a). Very good agreement with the ACF period catalogue is obtained, demonstrating that the GP method works well not only on synthetic but also on real photometric data. The paper argues that the extra computational expense of GP-based rotation period inference is justifiable, and that its advantages will be especially significant in cases of low SNR and/or uneven time sampling, and in the context of population inference.

Given that some of the results in Section 7.1.3 above were derived after the Angus et al. (2018) paper was written, there is clear scope for improvement to the accuracy and/or precision of the GP-extracted rotation periods published in the aforesaid paper.

A separate project to which I have made minor contributions (to be published as Esselstein et al., 2017, *in prep.*) attempts to use *K2* Campaign 5 photometry to measure rotation periods for Solar-age stars in the open cluster M67. Such measurements would enable calibration of the relationship between stellar age, mass and rotation rate, which in turn would provide a valuable test of stellar evolution models, and a tool to determine the ages of field stars. Given the challenges inherent in measuring the rotation periods of moderately active stars in a crowded field, a number of injection tests are performed to assess the performance of various period measurement techniques, including a GP method similar to the one in the Angus et al. (2018) paper.

### 7.1.6 Towards characterising differential rotation

Let us conclude the discussion on using GPs for characterising stellar rotation with a few embryonic remarks on how GPs might be used to identify or study *differential* stellar rotation. While fairly accurate recovery of an 'overall' rotation period is generally feasible using only full-disc light curves, the same cannot be said of differential rotation shear. In a blind exercise testing the recoverability of differential rotation signals from *Kepler* light curves, seven teams using a range of different techniques all failed to distinguish between differential and solid-body rotation under almost all circumstances (Aigrain et al., 2015).

A useful starting point, from the GP perspective, is to consider the result that if $f \sim \mathcal{GP}(\mu_1, k_1)$ and $g \sim \mathcal{GP}(\mu_2, k_2)$, then $f + g \sim \mathcal{GP}(\mu_1 + \mu_2, k_1 + k_2)$ (Rasmussen and Williams, 2006). Since we expect the covariance kernel $k_{\text{QP-SE}}$ to serve as a good model for the photometric signal of evolving active regions rotating at a single angular velocity, we could sum such QP covariance kernels, each with a different rotation period, to produce a model for evolving and differentially-rotating active regions.

In principle, each active region (or active latitude) should contribute a term to the summation. Unfortunately, we would not know *ab initio* how many terms to include, and the number could be very large. To make matters more tractable, we

could try to parametrise the period distribution. Suppose, say, that the probability of observing an active region with period $P$ takes the form

$$p(P) \propto \exp\left(\frac{-(P-\rho)^2}{2\delta^2}\right), \tag{7.5}$$

i.e. $P \sim \mathcal{N}(\rho, \delta)$. Thus there will be some 'overall' rotation period, $\rho$, although periods as short as about $\rho - 3\delta$ will be found at the stellar equator, or as long as about $\rho + 3\delta$ at the poles. Let us denote $k_{\text{SE}}(\tau) = h^2 \exp(-\tau^2/2\lambda_{\text{e}}^2)$, with $\tau = t_i - t_j$ as before; if we integrate over all possible stellar rotation periods, we find

$$k_{\text{RQP,1}} := \int_{-\infty}^{\infty} k_{\text{QP-SE}} \cdot p(P) dP \approx k_{\text{SE}} \int_{\rho-3\delta}^{\rho+3\delta} \exp\left(-\frac{\sin^2\left(\frac{\pi\tau}{P}\right)}{2\lambda_{\text{p}}^2} - \frac{(P-\rho)^2}{2\delta^2}\right) dP, \tag{7.6}$$

since $p(P) \approx 0$ when $|P - \rho| > 3\delta$. The 'RQP' in $k_{\text{RQP,1}}$ stands for 'rational quasi-periodic,' by analogy with the rational quadratic kernel defined in equation 3.28, since this new kernel is formed as a scale mixture (infinite sum) of QP kernels with different characteristic periods. Although the integral in equation 7.6 is not analytically tractable, numerical quadrature can be used to evaluate it quite accurately. For example, using Simpson's rule (based on quadratic interpolation of the integrand, with quartic error behaviour: Press, 2007), we find

$$k_{\text{RQP,1}} \approx \delta \cdot k_{\text{SE}} \left[ 4e^{-\sin^2\left(\frac{\pi\tau}{\rho}\right)/2\lambda_{\text{p}}^2} + e^{-\frac{9}{2}-\sin^2\left(\frac{\pi\tau}{\rho-3\delta}\right)/2\lambda_{\text{p}}^2} + e^{-\frac{9}{2}+\sin^2\left(\frac{\pi\tau}{\rho+3\delta}\right)/2\lambda_{\text{p}}^2} \right]; \tag{7.7}$$

it is straightforward to derive criteria under which the approximations inherent in the numerical quadrature will be justifiable. Thus we see that our new kernel, $k_{\text{RQP,1}}$ features just one extra hyperparameter compared to $k_{\text{QP-SE}}$ ($\delta$, which tells us about the spread of rotation periods), while the hyperparameter $\rho$ replaces the old hyperparameter $P$.

If we suppose instead that $P \sim \mathcal{U}(\rho - 3\delta, \rho + 3\delta)$ and then integrate over all possible periods to form a kernel $k_{\text{RQP,2}}$, Simpson's 3/8 rule (based on cubic interpolation of the integrand) leads us to the following approximation:

$$k_{\text{RQP,2}} \approx \frac{3\delta \cdot k_{\text{SE}}}{4} \sum_{i=0}^{3} \xi_i e^{-\sin^2\left(\frac{\pi\tau}{\rho-\delta(3-2i)}\right)/2\lambda_{\text{p}}^2}, \tag{7.8}$$

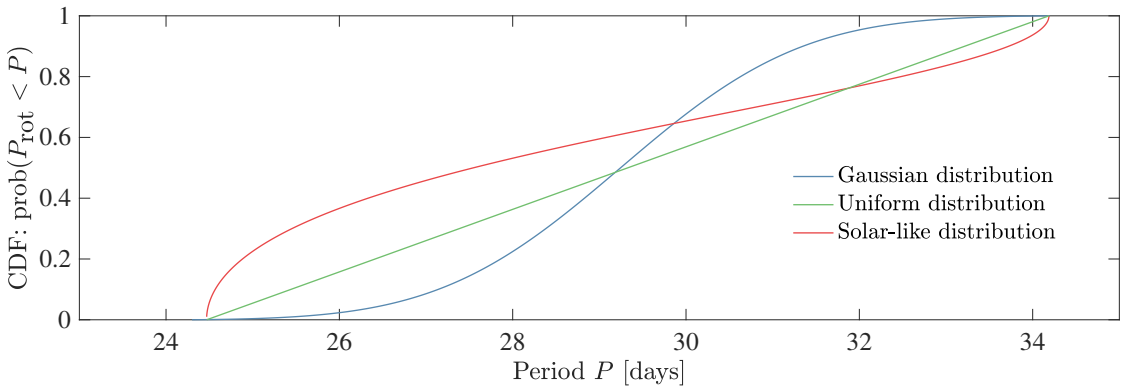


**Figure 7.3:** Three different stellar rotation period cumulative distribution functions (CDFs) considered in attempts to build a GP kernel to describe differential stellar rotation. The 'Solar-like' CDF is based on the expression in equation 7.9, with realistic values chosen for $\Phi_{1,2,3}$. The Gaussian and uniform distributions (see pp. 189–190) each have parameters $\rho$ and $\delta$ chosen to produce a reasonable match to the Solar-like distribution.

where $\xi_{0,3} = 1$ and $\xi_{1,2} = 3$.

Both $k_{\rm RQP,1}$ and $k_{\rm RQP,2}$ generally serve as good approximations to the integrals over period distributions on which they are based. However, given the assumed period distributions (Gaussian and uniform), these covariance kernels serve as little more than 'toy' models. Preliminary tests with a subset of the light curves from the Aigrain et al. (2015) data set suggest that degeneracies between the hyperparameters $\delta$ and $\lambda_{\rm p}$ confound efforts to use these new kernels to identify differential rotation unambiguously.

A more realistic, Solar-like differential rotation model would likely feature a quadratic relationship between angular velocity $\omega$ and $\sin^2\varphi$, where $\varphi$ is the stellar latitude (Schrijver and Zwaan, 2000), i.e. $\omega \sim \Phi_1 + \Phi_2 \sin^2\varphi + \Phi_3 \sin^4\varphi$. See Fig. 7.3. Integrating over all stellar latitudes, the kernel would take the form:

$$k_{\rm RQP,3} := k_{\rm SE} \int_{-\pi/2}^{\pi/2} \exp\left\{ -\frac{\sin^2\left[\pi\tau(\Phi_1 + \Phi_2 \sin^2\varphi + \Phi_3 \sin^4\varphi)\right]}{2\lambda_{\rm p}^2} \right\} \cos\varphi d\varphi; \qquad (7.9)$$

the factor of $\cos\varphi$ is included to take into account the decreasing projected area of active regions on higher latitudes.

Interestingly, applying Simpson's 3/8 rule to the integral in equation 7.9 reduces $k_{\rm RQP,3}$ to $k_{\rm QP\text{-}SE}$, with period re-defined in terms of $\Phi_1$, $\Phi_2$, and $\Phi_3$. This result suggests that, to first order at least, $k_{\rm QP\text{-}SE}$ may actually be a sufficient kernel

for describing differential rotation, and that more complex kernels constructed by summing over $k_{\text{QP-SE}}$ will not necessarily be better suited to the job.

In future, I aim to investigate how $k_{\text{QP-SE}}$ itself may be leveraged for characterising differential rotation. It seems plausible that photometric signals associated with differential rotation might favour smaller values of the hyperparameter $\lambda_{\text{p}}$, or be associated with broader posterior distributions for the hyperparameter $P$. Alternatively, given a sufficient quantity of high-SNR photometry (as available e.g. from *Kepler*), model comparison tests may permit some level of differentiation between signals best described with a single QP kernel vs. a combination of QP kernels with different periods.

## 7.2 Synthesising realistic observations

A GP application that has been mentioned almost only tangentially thus far in this thesis is using GPs to synthesise *realistic* synthetic observations.

In Chapter 4, a GP framework is first used to learn the properties of activity signals associated with $\alpha$ Cen B, and is thence used to generate realistic synthetic data for exposing the specific shortcomings of the model used to claim the existence of the planet $\alpha$ Cen Bb. As it turned out, the original model's failure to account for time-correlated nuisance signals (predominantly stellar in nature) present in the RV data led to the apparent false positive detection of $\alpha$ Cen Bb. In this case, tests performed with synthetic RV data comprising only a Keplerian signal plus white Gaussian noise (and perhaps a simplistic stellar activity signal) led to enormously optimistic conclusions about false alarm probabilities and planet detection prospects.

The more general message is the following. Real observations tend to be very 'messy': apart from containing (say) a few exoplanetary signals of interest, they'll typically also contain a variety of time-correlated stellar, instrumental, atmospheric, calibration, and other nuisance signals (e.g. Gilliland et al., 2011; Halverson et al., 2016). While it is easy to parametrise the exoplanetary contribution to the observed signal, most of the other time-correlated nuisance signals cannot be

easily parametrised or modelled. Uneven time sampling, interference between various signals, and limited observational accuracy and precision complicate matters further.

Therefore, while it is very easy to generate synthetic observations by adding i.i.d. white (serially uncorrelated) Gaussian noise to exoplanet signals from a parametric model, such synthetic observations will not be very realistic. For one, no matter how large the amplitude of the white noise, it will never interfere constructively or destructively with the exoplanet signals; it will simply 'dilute' the exoplanet signals. And yet the literature is replete with injection tests, detection probability tests, etc. based on precisely such simplistic synthetic data, generated via additive white Gaussian noise (AWGN) models. Even for the specific case of $\alpha$ Cen B, there are many published examples of planet-detectability studies where an AWGN model is assumed to be sufficient: see, e.g., Guedes et al. (2008), Endl et al. (2015), Bergmann et al. (2015). In fact, a recent study featuring a realistic, coloured noise model argued that around 1500 high-cadence RVs of $\alpha$ Cen B would be required to have a 90% chance of detecting an orbiting Earth-mass planet with 3.24 d period (Sulis et al., 2017) – more than three times as many as the 459 RVs originally used to make a 'detection' with a supposed FAP of 0.02%.

Greatly preferable to a simplistic AGWN model, of course, is to try to make synthetic observations precisely as 'messy' and correlated as the real observations. GPs provide a principled, data-driven, and non-parametric framework for doing this. Regardless of the origin of the nuisance signals (stellar, instrumental, telluric, etc.) in a given data set, GP regression allows their covariance properties to be learned; thence, infinitely many random, synthetic observations with the same covariance properties, and arbitrary sampling, may be generated. These synthetic observations may then be used to study detection limits, optimal observing strategies for planet detection or nuisance signal mitigation, or to identify artefacts introduced through sampling or through fitting a given model to observed data.

Chapter 5 of Rasmussen and Williams (2006) provides helpful examples of how appropriate GP covariance kernels may be constructed to capture and help interpret various behaviours that may be present in a data set, e.g. diurnal (or monthly,

seasonal, etc.) variations, smooth long-term trends, short-term irregularities, abrupt changes, measurement inaccuracies, and so on. Of course, certain types of nuisance signals may be modelled with simpler statistical tools than GPs, e.g. Poisson noise or cosmic ray strikes. Non-Gaussian outliers will also need to be modelled separately; see Appendix B.1.4 of this thesis for more on this subject.

Over the course of my doctoral studies, I have used GPs to synthesise realistic observations in the manner suggested above in a variety of contexts (apart from the published $\alpha$ Cen B study): e.g. studying RV detection limits for planets around AU Mic (A. M. Lagrange, 2016, *pers. comm.*, for a HARPS observing run); studying the RV detectability of a planet transiting Pr 0190, a very active star in the M44 cluster (L. Malavolta, 2016, *pers. comm.*, for a HARPS-N observing run); and devising an optimal observing schedule for constraining the properties of the planets in the Kepler-10 system (A. Mortier, 2017, *pers. comm.*, for a HARPS-N observing run that is ongoing at the time of writing).

## 7.3 Ultra-precise spectral extraction of RVs

### 7.3.1 Introduction

RV spectrographs have improved dramatically over the past few decades. Today, some state-of-the-art spectrographs have nominal precisions that reach below 10 cm $s^{-1}$: a factor of over a thousand better than the spectrographs of the 1970s. And yet, despite these instrumental advances, very little has changed in the way an individual RV is actually extracted from an observed spectrum – i.e., masked, weighted cross-correlation of the spectrum with that of an (essentially noise-free) template spectrum. See Section 2.2 and references therein for more details.

Superficially, this approach admits a few possible areas for improvement. First, there is the problem that the template spectrum (taken e.g. from a different star of similar spectral type) might be a good though never perfect match for any observed spectrum. There is also the concern that by masking off the continuum during cross-correlation, a small amount of genuine Doppler information might be lost, contained e.g. in very shallow absorption lines. Little weight ought to

be attached to these two concerns, however, as in many cases the performance of modern spectrographs appears very close to a fundamental limit for the case of a perfect template and a single spectrum corrupted only by photon noise (Connes, 1985; Bouchy et al., 2001; Pepe et al., 2014).

However, there are two other more subtle though potentially significant areas for improvement that do not seem to be discussed in the literature.

First is the assumption that there is nothing to be gained beyond comparing a single spectrum to a single (hopefully near-perfect) template. Yet what of the additional constraints that come from requiring that when computing RVs between any two *pairs* of observed spectra, results consistent with the template calculation should be obtained? For example, if 'spectrum 1' is found to have RV shift $\mathrm{RV}_1$ relative to a template, and 'spectrum 2' has an RV shift $\mathrm{RV}_2$ relative to that same template, an *additional* potential constraint comes from requiring that the RV shift of spectrum 2 computed relative to spectrum 1 (i.e. using spectrum 1 as the template) should equal $\mathrm{RV}_2 - \mathrm{RV}_1$. Given the potentially different wavelength axes of each observed spectrum and the template (see next point), one would not necessarily expect the error in $\mathrm{RV}_2 - \mathrm{RV}_1$ to fully determine the error in the RV obtained when cross-correlating spectrum 1 and spectrum 2 directly.

Second is the observation that an observed star's radial motion, combined with the Earth's own annual $\sim$ 30 km $\mathrm{s}^{-1}$ orbital motion (and perhaps also variable instrumental wavelength calibrations), will in practice mean that individual spectrograph pixels will correspond over time to slightly different photon emission wavelengths. Put differently, the same spectrograph will sample different parts (possibly at a sub-pixel level) of a stellar absorption spectrum over time. Given a large sample of spectra, then, such wavelength axis variations could be exploited to build what is effectively a noise-free template with much higher wavelength resolution than any template spectrum that could actually be observed with the spectrograph. This idea has strong parallels with multiple-frame 'super-resolution' techniques that successfully exploit sub-pixel level shifts in 2D images to generate

an image with higher resolution than any of the individual frames (Park et al., 2003; Farsiu et al., 2004).

I sketch below an outline of a GP-based approach that attempts to incorporate the above two considerations to deliver more precise RV extraction than would be possible when simply comparing a single observed spectrum to a perfect template.

### 7.3.2 A GP-based, template-free route to RVs

Suppose we have a set of $N$ spectra, $\{\mathcal{S}_1, \ldots, \mathcal{S}_N\}$, where the $i^{\text{th}}$ spectrum is specified by the set of triplets $\mathcal{S}_i = \{(\lambda_{ij}, f_{ij}, \sigma_{ij})\}_{j=1}^{M}$. Here, $M$ is the number of pixels in a single spectrum (for HARPS-N, 69 échelle orders of 4096 px each means $M \approx 280\,000$ px); $\lambda_{ij}$ are wavelengths, $f_{ij}$ fluxes, and $\sigma_{ij}$ errors in the fluxes, due e.g. to photon noise.

Numerical experiments reveal that simply cross-correlating observed spectra with each other, and thus attempting to extract RVs (perhaps after first 'unshifting' all spectra to build a master template spectrum), leads to very poor results. Presumably this is due, in large part, to the cross-correlations being extremely sensitive to noise in individual spectra, such that cross-correlation between any two individual spectra leads to an essentially meaningless RV estimate.

However, for each individual spectrum, we could try to build a model $f_i$ that aggregates information about pixel-level noise *and* covariances between nearby pixels to yield a lower-noise interpolant better suited to cross-correlation. After all, especially in deep spectral lines, nearby pixels should *not* be independent, and we could exploit this information when trying to figure out the best way to interpolate the noisy fluxes. Let us therefore place a GP prior on the function $f_i$, i.e. $f_i(\lambda) \sim \mathcal{GP}(\mu_i, k_i)$. The mean function $\mu_i(\lambda)$ will describe the continuum, while the covariance kernel will be used to build a model for the spectral absorption features.

Tests with a sample of HARPS-N spectra, with mean SNR at 550 nm ranging from 50 to 200, revealed that a Matérn kernel with $\nu = \frac{5}{2}$ did an excellent job of fitting absorption features across all wavelengths. Setting $\nu = \frac{3}{2}$ led to excessive flexibility (i.e. fitting of noise), while $\nu \geq \frac{7}{2}$ led to functions too smooth to fit sharp

absorption lines well. Thus, taking into account the observational noise, the kernel function would take the form (see equation 2.30)

$$k_i(\lambda_{ij}, \lambda_{ik}) = h^2 \left(1 + \frac{\tau}{\rho} + \frac{\tau^2}{3\rho^2}\right) \exp\left(-\frac{\tau}{\rho}\right) + \delta_{ik}\sigma_{ij}^2, \tag{7.10}$$

where $\tau := \sqrt{5(\lambda_{ij} - \lambda_{ik})^2}$. The scale hyperparameter $h$ is related to the typical depth of spectral absorption lines, while $\rho$ is related to the line shape and typical density of lines per wavelength; hyperparameters may be learned through maximum likelihood (see equation 2.38) or posterior exploration. Once the hyperparameters have been learned, we can compute the GP posterior (predictive) distribution: $g_i := f_i(\lambda)|\mathcal{S}_i \sim \mathcal{GP}(\mu_i^*, k_i^*)$, where $\mu_i^*$ and $k_i^*$ are to be computed using the relations in equations 2.27 and 2.28.

Apart from being more amenable to cross-correlation than the noisy, 'raw' spectrum $\mathcal{S}_i$, $g_i$ may now be evaluated at arbitrary wavelengths that extrapolate or interpolate $\lambda_{ij}$, as will be necessary when shifting spectra in wavelength or velocity space (e.g. when performing barycentric velocity corrections, cross-correlation calculations, etc.). Such evaluations will be accompanied by principled flux uncertainty estimates that can then be propagated throughout subsequent inference.

In practice, it turns out to be convenient to model each order in a given spectrum *separately*, for two reasons. Firstly, while modelling (say) $M = 4096$ fluxes with a GP is quite straightforward computationally, modelling $M > 200,000$ fluxes per spectrum is neither feasible, given the $\sim \mathcal{O}(M^3)$ scaling of basic GP regression, nor desirable, given that we expect little or no covariance between widely-spaced spectral lines. Secondly, given that typical spectral features and continuum behaviour may be expected to vary significantly across orders, it will be preferable to allow for a different set of hyperparameters for each order, rather than forcing a single set of hyperparameters to model a single spectrum globally. Therefore, let us denote by $g_i^{(l)}(\lambda)$ our GP posterior for the $l^{\text{th}}$ order of the $i^{\text{th}}$ spectrum.

### Comparing all GP models

Let us now recall that the rest wavelengths $\lambda$ at which we measure stellar flux on Earth to be a function[5] of stellar RV relative to Earth, $v$, i.e. $\lambda = \lambda(v)$.

We can now consider a pair of GP posteriors (i.e. model spectra), say $g_i^{(l)}(\lambda(v))$ and $g_j^{(l)}(\lambda(v+\delta))$, and establish which velocity shift $\delta$ will maximise the overlap between the two models. The value of $\delta$ leading to the best match corresponds to the stellar RV shift between the two spectra:

$$\mathrm{RV}_{ij}^{(l)} = \underset{|\delta|<c}{\arg\max} \int_{-c}^{+c} g_i^{(l)}(\lambda(v))\, g_j^{(l)}(\lambda(v+\delta))dv; \tag{7.11}$$

in practice, the arg max of the integral may be found via a global optimisation routine. Given that each model $g_i^{(l)}(\lambda(v))$ is accompanied by a posterior uncertainty, this may be propagated via a Monte Carlo approach to an uncertainty in $\mathrm{RV}_{ij}^{(l)}$, i.e. $\sigma\left(\mathrm{RV}_{ij}^{(l)}\right)$.

If we thus cross-correlate every GP posterior with every other GP posterior, we can build an $N \times N$ matrix of RV shifts for each order. Given the anti-symmetry $\mathrm{RV}_{ij}^{(l)} + \mathrm{RV}_{ji}^{(l)} = 0$, and that $\mathrm{RV}_{ii}^{(l)} = 0$, only $N(N-1)/2$ of these RV shifts actually need to be computed.

### Aside: measuring a signal via pairwise differences

As a brief aside, let us consider how any function can be expressed as a sum of pairwise differences. Suppose we have a function $y = f(t)$, with its values measured at $N$ points, so that we have a data set $\{(t_i, y_i, \sigma_i)\}_{i=1}^N$. Here, $y_i$ is the value of the function relative to some origin (zero), and $\sigma_i$ is the measurement error in each $y_i$.

Suppose now that rather than making a single absolute measurement of $y_i$ relative to a fixed zero point, we are instead able to measure the $N(N-1)/2$ pairwise differences between all values, so that our data would take the form $\{(t_{ij}, y_{ij}, \sigma_{ij})\}_{i,j=1}^N$, where $y_{ij} = y_i - y_j$. It turns out that we can then reconstruct the individual $y_i$ values, relative to an origin defined by the average[6] of all $y_i$ values:

$$y_i = \frac{1}{N}\sum_{j=1}^N y_i = \frac{1}{N}\sum_{j=1}^N (y_i - y_j) + \frac{1}{N}\sum_{j=1}^N y_j = \frac{1}{N}\sum_{j=1}^N y_{ij} + \gamma \tag{7.12}$$

[5]The functional relationship between stellar Doppler velocity, photon emission wavelength, and observed photon wavelength is given in equation 2.6.

[6]To account for observation errors, we should really be weighting each $y_i$ by $\sigma_i^{-2}$.

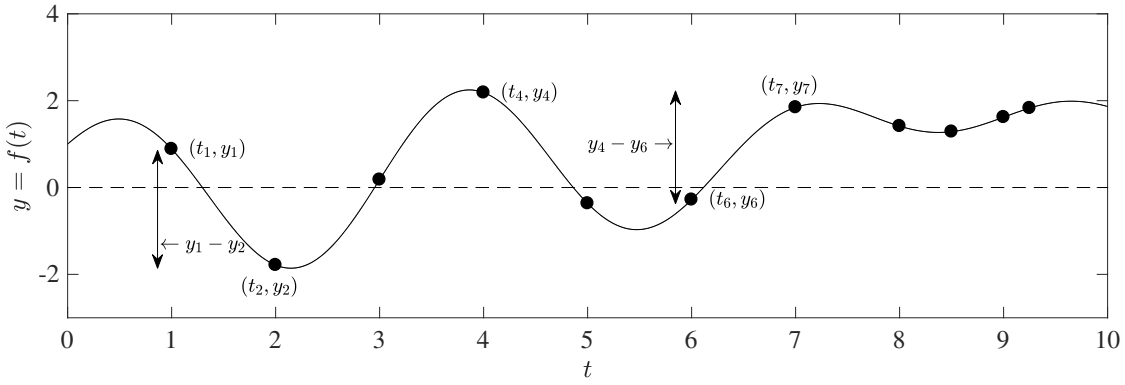


**Figure 7.4:** Illustration of an arbitrary function with function values measured relative to a fixed origin (dashed line) or via a sum of pairwise differences (see text on pg. 198).

where $\gamma := \frac{1}{N}\sum_{j=1}^{N} y_j$ is our new origin; note that $\gamma$ is the same for all $y_i$. If we are not interested in this offset, we may as well set $\gamma = 0$.

We could of course choose to interpret $f(t)$ as a time-dependent stellar RV signal, and $y_{ij}$ as pairwise differences between RVs measured at discrete times. Regardless of the complexity of $f(t)$ – it might comprise binary stellar, instrumental, exoplanetary, and other contributions – we could reconstruct it, up to an additive constant, by averaging over $y_{ij}$. Why bother? The advantage of this approach is that working with $\sim N^2$ rather than $N$ measurements leads to a $\sim \sqrt{N}$ improvement in the SNR of the reconstructed signal, assuming the $\sigma_{ij}$ are independent (Hassan and Sabieh Anwar, 2010).[7]

### Aggregating the pairwise RV shifts

We now turn to the problem of aggregating the pairwise RV shifts to estimate a single RV shift for each spectrum.

Following the idea in the preceding aside, we can average over the pairwise RV shifts in each column (or row) to estimate an RV shift for the $i^{\text{th}}$ spectrum, relative to some (arbitrary) offset. We compute a variance-weighted mean as follows:

$$\mathrm{RV}_i^{(l)} = \sum_{j=1}^{N} \mathrm{RV}_{ij}^{(l)} / \sigma\left(\mathrm{RV}_{ij}^{(l)}\right)^2 \Big/ \sum_{j=1}^{N} 1/\sigma\left(\mathrm{RV}_{ij}^{(l)}\right)^2 + \gamma, \tag{7.13}$$

[7] This is an important assumption. In practice, when dealing with RV shifts extracted from model spectra, sizeable barycentric velocity corrections, and even sub-pixel level interpolations and extrapolations associated with different stellar radial velocities, seem sufficient to ensure that there will be a component of independent noise in each $y_{ij}$ estimate.

where $\gamma$ is a constant for all $i$, and

$$\sigma\left(\mathrm{RV}_i^{(l)}\right) = \sqrt{1 \Big/ \sum_{j=1}^{N} \sigma\left(\mathrm{RV}_{ji}^{(l)}\right)^2}. \tag{7.14}$$

For convenience, we can set $\gamma = 0$, as in the earlier example. We can now also compute a variance-weighted mean across all $L$ orders to obtain a final estimate[8] of the stellar RV of the $i^{\mathrm{th}}$ spectrum:

$$\mathrm{RV}_i = \sum_{l=1}^{L} \mathrm{RV}_i^{(l)} / \sigma\left(\mathrm{RV}_i^{(l)}\right)^2 \Big/ \sum_{l=1}^{L} 1/\sigma\left(\mathrm{RV}_i^{(l)}\right)^2, \tag{7.15}$$

with

$$\sigma\left(\mathrm{RV}_i\right) = \sqrt{1 \Big/ \sum_{l=1}^{L} \sigma\left(\mathrm{RV}_i^{(l)}\right)^2}. \tag{7.16}$$

Given the iterative variance-weighted averaging over noisy realisations of the same RV shifts, we expect the noise power to decrease as $\sim 1/\sqrt{N}$, where $N$ is the number of spectra included in the modelling, and also as $\sim 1/\sqrt{L}$, where $L$ is the number of échelle orders incorporated. Another significant though less obvious advantage of the method is that by building a continuous model for each spectrum, and exploiting pixel-level covariances when interpolating and extrapolating observed fluxes,[9] an implicit 'master spectrum' is constructed with higher resolution than would be possible with any real template spectrum observed with the same spectrograph.

### 7.3.3 Illustration of the method's performance

I implemented the method described above in a couple hundred lines of straightforward MATLAB code. The details of the implementation will be presented in a forthcoming paper.

To ascertain that the method actually works, at least in principle, I took a single HARPS-N spectrum of the bright G5V star HD 127334 (V magnitude $m_V = 6.5$)

[8] In practice it might be desirable to exclude those orders known to be most strongly contaminated by activity-sensitive features (e.g. the Calcium II H & K lines).

[9] Consider that the Earth's own orbital motion will lead to Doppler shifts of observed photons, such that the same spectrograph pixels at different times of the year will correspond to different stellar emission wavelengths, all else being equal. Thus it becomes possible to learn the 'true' stellar spectrum at far more wavelengths than there are pixels on the detector.

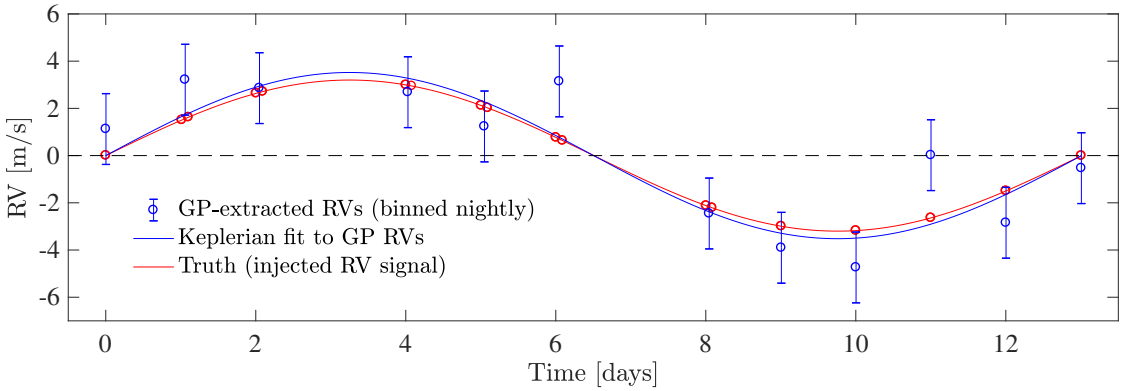


**Figure 7.5:** GP-based extraction of a 3.2 m s$^{-1}$ sinusoidal signal from 35 Doppler-shifted spectra (SNR = 200), using just one échelle order, i.e. < 2% of the full spectra. For visual clarity, the 35 extracted RVs and estimated errors have been binned to one point per night. The sine amplitude is recovered as $3.5 \pm 0.7$ m s$^{-1}$.

and created 35 copies of it, and assigned to each a 'time stamp' corresponding to 35 real HARPS-N spectra of HD 127334, taken over 13 nights in June 2015. To each copy of the original spectrum I added random realisations of photon noise at a level consistent with the real spectra (SNR $\sim$ 200), and then Doppler-shifted the spectra using a sinusoidal signal with amplitude 3.2 m s$^{-1}$ and period 13 d. I also added an appropriate Doppler contribution from the Earth's orbital motion over the observation baseline. Thus I ended up with 35 spectra from which to extract a synthetic RV signals; the advantage of this approach was that my spectra were confounded only by photon noise, and not by any time-variable instrumental signals (e.g. CCD cold plate temperature oscillations).

I found that even using only a single échelle order (order $l = 50$ out of $L = 69$ available, i.e. using < 2% of the available data), I was able to recover the injected RV signal with a median precision of about 1.5 m s$^{-1}$ per observation. See Fig. 7.5. When using all 69 orders, the median error per measurement was 25 cm s$^{-1}$: in line with the predicted $\sqrt{L}$ improvement (actually slightly poorer than the prediction, given the above-average SNR of the $l = 50$ order used in the initial calculation). Though remarkably small, the estimated $\sim$ 25 cm s$^{-1}$ errors were consistent with the RV scatter on nights with 2 or 3 spectra available – the method does not 'know' which RVs correspond to observations closely spaced in time.

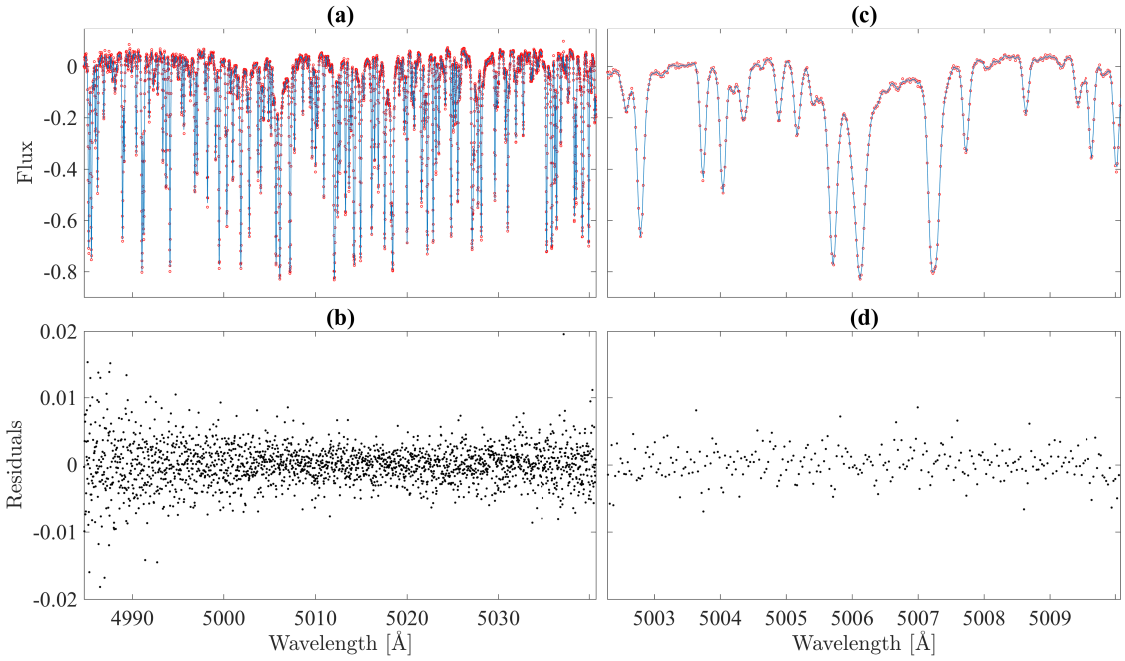


**Figure 7.6:** (a) Echelle order $l = 36$ of a single HARPS-N spectrum of HD 127334, plus GP model fit. The red circles are renormalised fluxes, and the blue solid line is the GP posterior mean, with the shaded blue region around it (nearly invisible at this scale) representing $\pm\sigma$ posterior uncertainty. (b) Residuals from the GP fit. Panels (c) and (d) are as for panels (a) and (b), but zoomed in to a narrower wavelength range from the same order.

Thus satisfied that the method works in principle, I turned to two real data sets: (i) the 35 HARPS-N spectra of HD 127334, spanning 13 nights, as described above; and (ii) 148 HARPS-N spectra of Kepler-10 (K-10, $m_V = 11.2$), spanning 17 months. For both data sets I also had available RVs derived from the spectra via the HARPS-N DRS pipeline, for comparison with RVs extracted using my GP method; in the case of K-10, the DRS RVs were the same ones used in the main analysis in Chapter 5.

An example of a GP fit to a single échelle order from a single spectrum is shown in Fig. 7.6 for the case of HD 127334, and in in Fig. 7.7 for the case of K-10. Qualitatively, at least, we see that the GP serves as a remarkably good model for the observed spectra. The residuals from the GP fits appear white, and have rms scatter consistent with photon noise estimates for the respective orders.

I found that when using only a subset of the available spectra, and/or a subset of the échelle orders within the available spectra, the median error of extracted

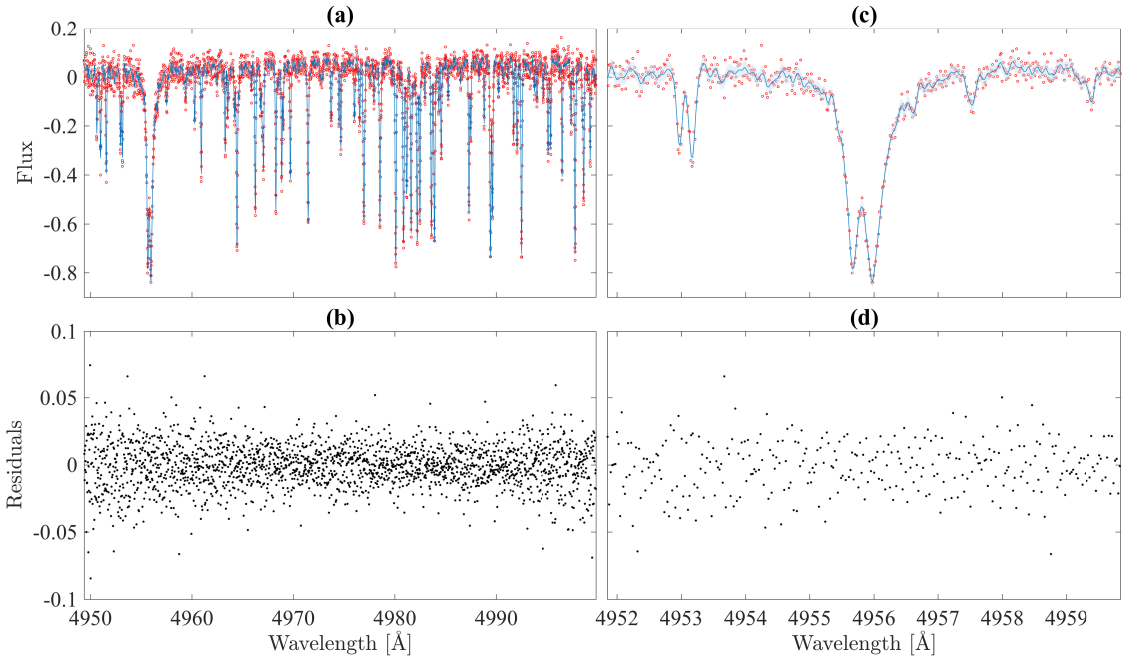


**Figure 7.7:** As for Fig. 7.6, but now for order $l = 35$ of a single HARPS-N spectrum of K-10. Note the lower SNR of the K-10 spectrum compared to the HD 127334 spectrum.

RVs did indeed scale like $N^{-1/2}$ and $L^{-1/2}$, where $N$ and $L$ are the total number of spectra and échelle orders, respectively, used in the modelling. See Fig. 7.8. This was a significant finding because the error estimates are bounded below (amongst other things) by the scatter in RVs computed from independent orders, and in RVs computed from cross-correlation between different pairs of spectra. As such, the estimated errors should not be over-optimistic.

I found that when using all available HD 127334 spectra and échelle orders, the median error of the 35 GP-extracted RVs was 36 cm s$^{-1}$, and the rms variation 1.03 m s$^{-1}$. The HARPS-N DRS values were 43 cm s$^{-1}$ and 1.17 m s$^{-1}$, respectively. In the case of K-10, the median error of the 148 GP-extracted RVs was 1.52 m s$^{-1}$, and the rms variation 4.32 m s$^{-1}$; the DRS values were 1.73 m s$^{-1}$ and 4.72 m s$^{-1}$, respectively. In both cases, then, the GP approach led to a 15–20% improvement in RV precision, and a 10–15% reduction in final rms variation, compared to the DRS. The time series of the actual RVs extracted by both approaches are shown in Fig. 7.9. Note, in particular, that the GP and DRS RVs are almost always consistent with each other within $1\sigma$. Moreover, as noted in Section 5.5 of this thesis, modelling

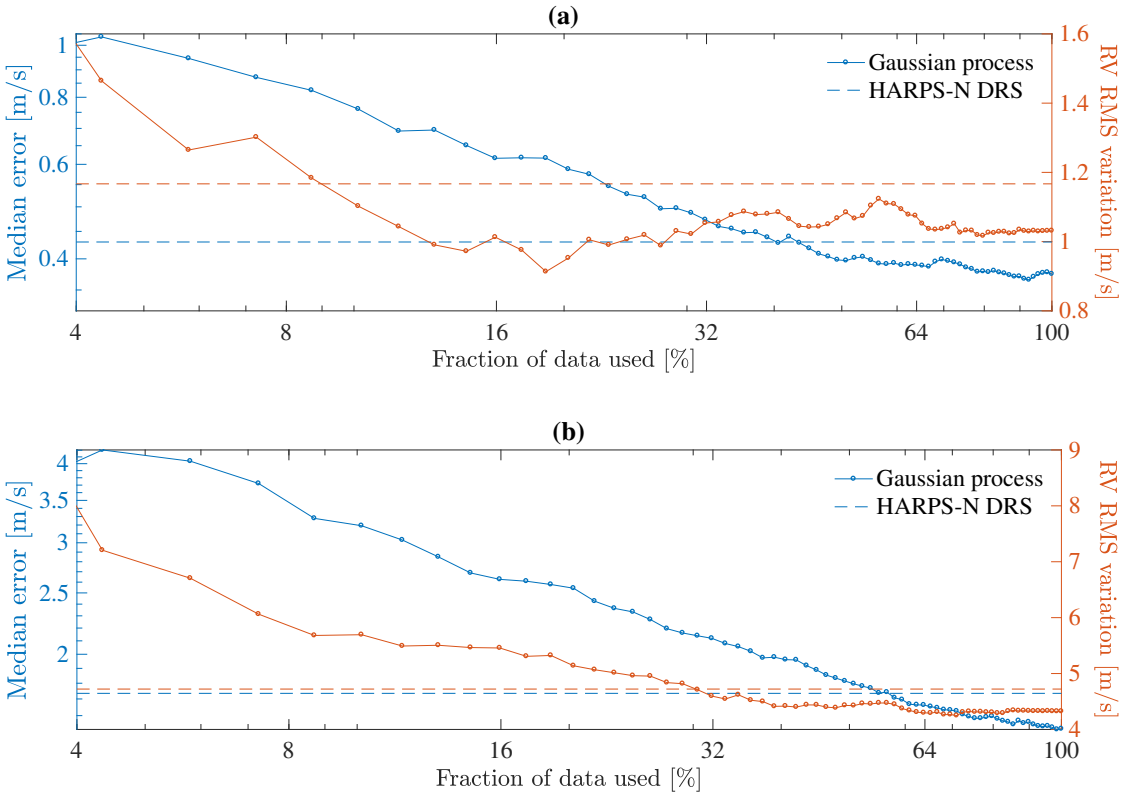


**Figure 7.8:** (a) GP-extracted RV median errors and rms variation vs. the fraction of available data used, where 100% corresponds to all 35 spectra and all 69 échelle orders for HD 127334. Note the log-linear decrease in median error with quantity of data modelled. The dashed lines indicate the median error and rms variation of the full set of RVs as extracted by the HARPS-N DRS. (b) As for the preceding panel, but this time for the case of K-10, with up to 148 spectra and 69 échelle orders modelled.

the K-10 RVs extracted via the GP approach led to planet parameters for K-10b and K-10c consistent with those inferred from the DRS RVs.

At face value, a 10–20% precision improvement might seem modest, yet in the quest to find and accurately characterise the smallest exoplanets, any incremental improvement is valuable. Indeed, this performance seems nothing short of remarkable, given the conceptual and computational simplicity of the GP approach: there is no template spectrum, no binary mask, nor even any stellar astrophysical priors in the modelling. Moreover, the results presented here come from a somewhat crude, 'first attempt' implementation spanning only a couple hundred lines of code.

A current disadvantage of the GP method that should be mentioned is computational complexity. On a modern laptop, fitting a GP model to a single order from a

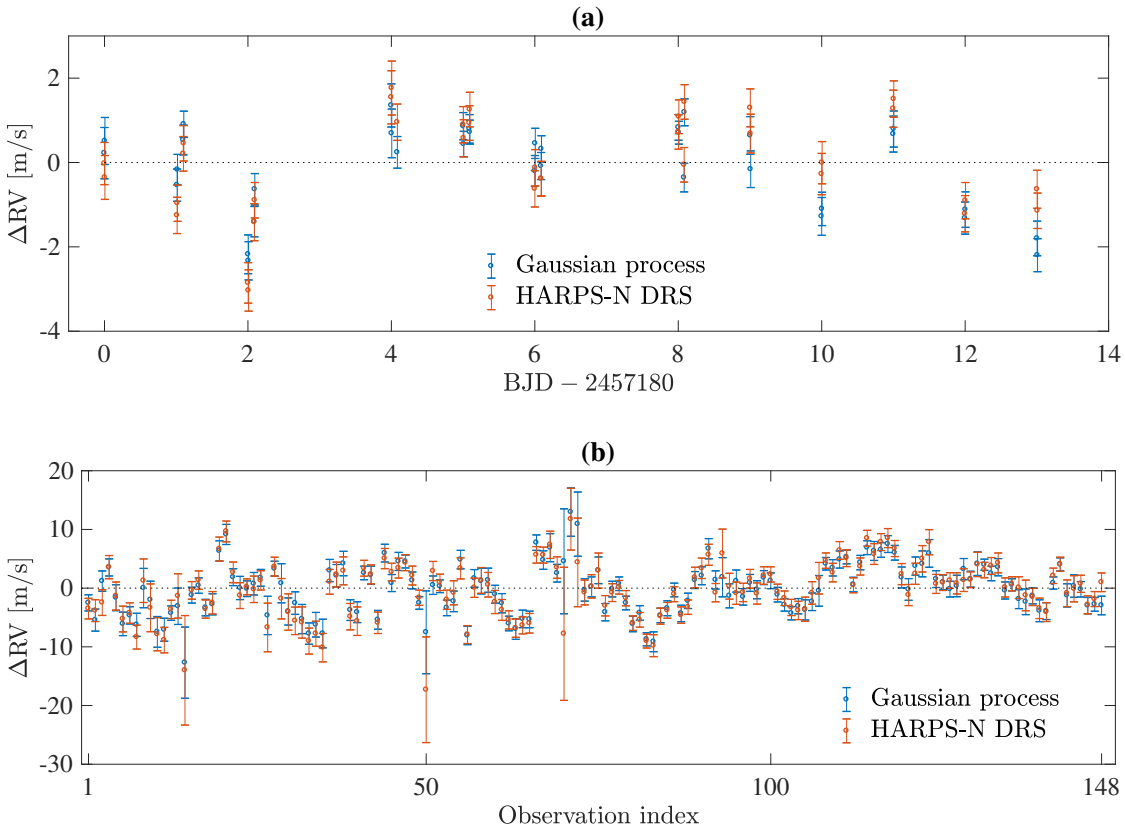


**Figure 7.9:** (a) 35 mean-subtracted HD 127334 RVs extracted by the GP method, plotted in blue along with estimated errors, with the HARPS-N DRS RVs and errors in orange. (b) As for the preceding panel, but for the case of 148 K-10 RVs. Because of the long observing baseline and uneven temporal coverage, the K-10 RVs are plotted a function of observation sequence number (i.e. 1 to 148) to aid visual inspection of RV variability.

single HARPS-N spectrum (i.e. 4096 pixels) takes a matter of seconds. Doing this for 69 orders across 148 different spectra (in the K-10 case) takes a matter of hours, and the subsequent cross-correlation (and associated Doppler-shifting and re-evaluation) between all pairs of GP models can take many further hours. In short, the extra precision does not come cheaply. However, there are many straightforward ways in which the basic method could be sped up. For example, once the hyperparameters for a given order are learnt for a few spectra, these could be used as priors for subsequent spectra (assuming the absorption lines in a given order don't change drastically over time). Given what should be very small covariance between widely-spaced absorption lines, sparse matrix inversion techniques and other approximate-GP approaches could be used to speed up the modelling significantly.

### 7.3.4 Future considerations

I have sketched a basic (though admittedly quite computationally-intensive) scheme for extracting RVs by direct modelling of observed spectra, without requiring cross-correlation with a masked template spectrum. Despite the theoretical advantages and already-excellent performance of the method applied to real data, a number of levels of additional sophistication may be built on top of the basic scheme.

First, instead of the sequential approach to extracting RVs (i.e. modelling chunks of spectra with separate GPs, then cross-correlating them, averaging RVs, etc.), all spectra could be modelled jointly, with the relative velocity shifts of all spectra included as free parameters in the modelling. At present, such an approach would be wholly unfeasible on account of its enormous computational requirements, i.e. modelling millions to tens of millions of observations simultaneously. With suitable approximations (e.g. sparse GP methods), however, it might be rendered tractable.

Secondly, (micro-)telluric lines may be identified and modelled separately from the rest of the spectrum: these lines should not shift with the rest of the stellar spectrum (Cunha et al., 2014). Early GP-based attempts at identifying and removing such lines using HARPS-N Solar spectra (Dumusque et al., 2015) seem very promising, with the incidental possibility of building very high-quality telluric spectra as a by-product. Similarly, it might be possible to identify and model certain instrumental contributions to time-correlated changes in observed spectra.

Thirdly, activity-sensitive lines may be identified either *ab initio* or through the modelling process itself. After all, an ensemble of model spectra should contain more information about stellar activity than a handful of CCF activity indicators. By decreasing the weights attached to lines or wavelength ranges identified as activity-sensitive, the SNR of any RV signals *not* due to stellar activity (e.g. due to exoplanets) could be boosted.

Finally, separate from these improvements to the ways in which stellar RV signals could be extracted, very high-resolution template stellar spectra could be built by 'unshifting' and stacking all model spectra, taking into account the different emission wavelengths sampled by each spectrum. While this does not necessarily

have relevance to extracting RVs or subsequent attempts to find exoplanetary signals in those RVs, such spectra may well be of interest to stellar astrophysicists.

Even in its current, 'basic' implementation, however, there are already specific cases in which it seems the GP method should have be able to offer particularly dramatic advantages over to the standard, template cross-correlation approach. Such cases would include RV extraction (i) from low-resolution spectra, where the 'super-resolution' concept could be strongly leveraged; (ii) from stellar spectra for which good templates are not readily available, e.g. brown dwarf infrared spectra; and (iii) in cases where very many RVs are available, so that the $\sim 1/\sqrt{N}$-precision of the method could be best exploited.

*I regret only one thing, which is that the days are so short and that they pass so quickly. Once never notices what has been done; one can only see what remains to be done.*

— Marie Skłodowska Curie, in an 1894 letter to her brother, as quoted in *Madame Curie: ihr Leben und Wirken* (Curie, 1938).

# 8 Conclusions

## Contents



## 8.1 Review and conclusions

In this thesis, I have leveraged GPs to develop tools for studying smaller planets around a wider variety of stars than has previously been possible, and for constraining more accurately and precisely the properties of known planets.

In Chapter 3, I developed a sophisticated and principled Bayesian framework, based on advanced application of GPs, for modelling RV time series jointly with ancillary activity-sensitive proxies, thus allowing activity signals to be constrained and disentangled from genuine planetary signals. I showed that this framework succeeds even in cases where existing techniques would fail to detect planets, e.g. the case of a weak planetary signal with period identical to its host star's rotation period.

In Chapter 4, in the first significant application of the GP framework, I demonstrated that $\alpha$ Cen Bb – until 2016, thought to be the closest exoplanet to Earth, and also the lowest minimum-mass exoplanet around a Sun-like star – was,

in fact, an astrophysical false positive. I showed that the original 'detection' likely arose from a combination of inadequate modelling of stellar activity signals, and a failure to understand the impact of the observing window on statistical inference.

Next, in Chapter 5, I used the framework to re-characterise the well-studied Kepler-10 system, thereby resolving a mystery involving the mass of planet Kepler-10c. I showed that the original discrepancies between the Kepler-10c masses inferred from two different RV spectrographs (Keck-HIRES and HARPS-N) likely arose from a failure to model correlated nuisance signals present in the RVs, as well as a failure to account for the impact of the observing window, as in the $\alpha$ Cen B case. My analysis incidentally strengthened existing evidence for a third planet in the Kepler-10 system. I also provided a few recommendations for future observing strategies, with a view to obtaining optimal coverage of planets' orbital phases.

In Chapter 6, I used my framework to help detect and characterise various exoplanets, *viz.* HD 175607b, Kepler-37d, and Kepler-538b. HD 175607b is a non-transiting planet whose orbital period overlaps with the apparent rotation period of its host star; using the GP framework allowed the stellar signal to be identified, leading to a secure detection of the planet. Kepler-37d and Kepler-538b, on the other hand, are both transiting exoplanets, and I employed my GP framework to assist the characterisation of these planets in the face of significant activity signals from their host stars. Both stars thus characterised contributed to the relatively small number of long-period super-Earths with accurately constrained masses.

Lastly, in Chapter 7, I presented in outline form three promising applications of GPs in the context of modelling stellar signals and studying exoplanets, *viz.* GPs for enhanced characterisation of stellar rotation; for generating realistic synthetic observations, and modelling in a systematic way the effects of an observing window function; and for ultra-precise extraction of RV shifts directly from observed spectra, without requiring template cross-correlation.

Perhaps the overarching conclusion of this thesis is that when trying to detect or characterise exoplanetary signals in noisy observations, significant benefits may be derived from honing the tools one uses to analyse the observations. (By 'tools' I mean

the physical models, the machinery of statistical inference, and so on.) Conversely, using inappropriate tools may lead to sub-optimal or even wholly misleading results.

Another way of phrasing this would be to say that what one can learn about a given of observations depends critically on the questions one asks. If one insists on computing a Lomb-Scargle periodogram for some RV data, say, one will always be answering a question about the degree to which a *single stationary sinusoid* suffices to describe the data, allowing for white Gaussian noise. Yet if the data show evidence for multiple frequencies or complex, non-sinusoidal behaviour, the answers to this question will be misleading, and more realistic questions (models) will be needed if one is to obtain realistic answers. As Jaynes put it: 'a false premise built into a model that is never questioned cannot be removed by any amount of new data' (Jaynes, 2003).

I argued that a GP with suitably-chosen covariance structure, e.g. a quasi-periodic covariance, represents a sensible and physically-motivated prior for the stochastic RV signals one would expect to be generated by evolving active regions on rotating stars. Such GP priors obviate the use of *ad hoc* parametric activity models, which may be overly-simplistic, or unrealistically flexible. By using a GP to model stellar activity signals *jointly* with planetary signals, one can avoid unwittingly 'mutilating' a data set through a theoretically-unsound process of iterative filtering (e.g. pre-whitening). Moreover, by modelling activity signals across multiple time series (i.e. RVs and all available activity-sensitive time series), one can place the tightest constraints on the activity signals, which is greatly preferable to throwing away a great deal of relevant information on the grounds that it might be difficult to incorporate into the modelling. Finally, by doing all of this within the framework of Bayesian inference, the relative evidence for competing models (with different numbers of planets, activity components of different complexities, etc.) may be computed in a principled way. This is the approach that I advocated and employed in this thesis, and I believe my results thus far bear testimony to its merits.

Of course there is an enormous deal to be said for advances in instrumentation, but recent history (e.g. the Kepler mission: Gilliland et al. 2011; and indeed also

some of the cases studied in this thesis) shows that careful modelling of stellar nuisance signals is the *sine qua non* for unlocking the full potential even of existing planet-hunting instruments, let alone next-generation instruments.

A theme that featured strongly albeit unexpectedly in Chapters 4 and 5 of this thesis was the extent to which discrete sampling can confound inference about a continuous signal. In principle, discrete sampling should be largely irrelevant if one is using the 'correct' model for a signal, which would automatically account for interference between the various possible components (e.g. multiple planets plus stellar activity) of the signal, regardless of when it happened to be sampled. Yet in practice, of course, one can never be sure that one *is* using the correct model, so rigorous model comparison tests become essential, with a close study of an observing window function serving as a useful adjunct. With the benefit of hindsight, one could contend that if a more realistic activity model were used *ab initio* in the $\alpha$ Cen B and the Kepler-10 cases, the apparent false positive planet detection and planet mass discrepancy, respectively, would have been avoided at the outset, and it would not have been necessary to give a second thought to sampling. Nevertheless, in both cases, a *post hoc* study of the observing window function did prove useful for diagnosing probable shortcomings in existing models.

Though my focus in this thesis was primarily on RV data, the tools and insights I developed have applicability to other forms of time series data as well. For example, the GP framework from Chapter 3 can accommodate photometric data quite naturally, as demonstrated with the Gliese 15 A modelling in Section 3.4.3. GPs for studying stellar rotation, as in Section 7.1, finds natural application on stellar photometric data (light curves). Finally, the insights about model comparison and sampling apply to any form of time series data, e.g. astrometric, photometric, etc.

## 8.2 Future work

The work in this thesis lends itself naturally to a number of possible extensions, a few of which I've outlined below.

First, while the GP framework presented in Chapter 3 has already borne significant fruit, I would like to investigate ways in which it could be improved and extended. (I do of course also intend to continue applying the framework to new and interesting data sets.) Even incremental improvements will be invaluable if the framework is to be used to detect the weakest possible planetary signals. For example, more realistic relationships between RVs and ancillary, activity-sensitive time series might be incorporated into the framework. It might also be worth adding a component to the framework to account for possible nuisance signals of instrumental rather than stellar origin.

Secondly, I'd like to develop a new, sophisticated 'quick look' pipeline for use at telescopes, in order to optimise the scheduling of different targets within upcoming RV and photometric surveys. Using some of the tools and insights developed in this thesis, the pipeline will keep track of the parameter space of 'allowed' models (with and without planets) for various stars, sequentially updating it after each observation, and selecting the next star to observe, to maximise the chances of establishing which targets host planets in a specific mass and orbital period range. Stellar activity will be modelled jointly with possible planetary signals, and the scheduling will also take into account achieving optimal phase coverage for planet candidates (as in Chapter 5). Compared to traditional survey scheduling, involving significant *ad-hoc* human decision-making, such an approach should increase the efficiency of surveys as a whole, and will also make it much easier to model their sensitivities, and thus to convert survey results into planet incidence estimates. This type of optimised scheduling will be particularly powerful for dedicated surveys using most of the time on a given telescope, e.g. the Terra Hunting Experiment on HARPS3 (Thompson et al., 2016).

Thirdly, I'd like to keep investigating ways in which GPs might be better employed to characterise stellar rotation and other (quasi)periodic astrophysical phenomena. As noted in Chapter 7, the work on stellar rotation is being led by collaborators, though it is clear to me that much remains to be done on the GP front, e.g. trying to derive an appropriate kernel for identifying differential stellar rotation.

Fourthly, I'd like investigate 'upgrading' some of the statistical tools I've employed in this thesis, e.g. leveraging Bayesian quadrature as a more principled and hopefully more expeditious approach to $n$-planet model comparison (which always entails evaluating difficult integrals with high dimensionality) than Monte Carlo approaches.

Finally, and perhaps most significantly, I'd like to develop to maturity the GP-based approach to extracting RVs from observed spectra, outlined in the latter part of Chapter 7. This work would have the potential to make a significant and lasting impact on the exoplanet field, with possible future adoption by almost all astronomers working with RV data.

## 8.3 Closing remarks

In years to come, I expect I might revisit this thesis and wish I had done things a little differently. Presumably, some of the models I have espoused here will end up being supplanted by better (or at least more complete) ones. The analytical techniques I today dare to call 'state of the art' will sooner or later grow to seem crude, as the assumptions I now hope are reasonable may eventually come to be regarded as untenable. (Certainly I'll regret not starting to write the whole thing a good deal earlier.) Yet such doubts can have no bearing on the fact that exoplanetary science will continue to move forward ever more swiftly and inexorably. Innumerable new worlds *will* be discovered and illuminated in exquisite, heretofore-unimagined detail, and each such world will provide further challenge to humanity's conceits and imagined self-importance. If nothing else, then, I am content to reflect that this thesis has embodied a few small yet fruitful contributions to contemporary exoplanetary science, and that with the passage of time, it might just help pave ways to further frontiers of knowledge.

# Appendices

*Mathematics is the queen of the sciences [...] She often condescends to render service to astronomy and other natural sciences, but in all relations, she is entitled to first rank.*

— Carl Friedrich Gauss, as quoted in *Gauss zum Gedächtnis* (von Waltershausen, 1856).

# A

# Covariance between observations and derivative observations

## Contents



Some of the mathematical expressions given below already appear, albeit without further explanation, in Chapter 3 of this thesis. The aim of this appendix is to provide some further details to allow these covariance kernels and their derivatives to be implemented computationally.

Suppose we have a GP, $G$, characterised by a covariance function $\gamma$. Then the covariance between an observation of $G$ at time $t_i$, and an observation of its derivative $\dot{G}$ at time $t_j$, is given by

$$\gamma^{(G,dG)}(t_i, t_j) = \frac{\partial}{\partial t}\gamma^{(G,G)}(t, t_j)\Big|_{t=t_i}, \tag{A.1}$$

where $\gamma^{(G,G)}(t_i, t_j)$ is used to denote the covariance between (non-derivative) observations of $G$ at times $t_i$ and $t_j$. Similarly, the covariance between two observations of $\dot{G}$ at times $t_i$ and $t_j$ is

$$\gamma^{(dG,dG)}(t_i, t_j) = \frac{\partial}{\partial t'}\frac{\partial}{\partial t}\,\gamma^{(G,G)}(t, t')\Big|_{t=t_i}\Big|_{t'=t_j}. \tag{A.2}$$

For convenience, we present below the relevant expressions for $\gamma^{(G,dG)}$, $\gamma^{(dG,dG)}$ etc. for the two covariance functions considered in some detail in Chapter 3. The above relations are, however, valid for any covariance function $\gamma$; as such, the two examples below serve as an illustration of how the relevant expressions can be derived for any covariance function.

## A.1 Squared-exponential covariance

Using our 'generalised squared-exponential' covariance function, the expression for the covariance between two observations of a process $G$ at times $t$ and $t'$ is:

$$\gamma_{\mathrm{SE}}^{(G,G)}(t,t') = \sum_{i=1}^{N} h_i^2 \exp\left[-\frac{(t-t')^2}{2\lambda_i^2}\right]; \tag{A.3}$$

the $h_i$'s control the relative amplitude of the $N \geq 1$ components with evolutionary time-scales $\lambda_i$. For the case $N = 1$, this reduces to the standard square-exponential covariance function.

Following equation (A.1), the covariance between an observation of $G$ at time $t$ and an observation of $\dot{G}$ at time $t'$ is given by

$$\gamma_{\mathrm{SE}}^{(G,dG)}(t,t') = -\sum_{i=1}^{N} h_i^2 \frac{(t-t')}{\lambda_i^2} \exp\left[-\frac{(t-t')^2}{2\lambda_i^2}\right], \tag{A.4}$$

and, by equation (A.2), the covariance between an observation of $G$ at time $t'$ and an observation of $\dot{G}$ at time $t$ is given by

$$\gamma_{\mathrm{SE}}^{(G,dG)}(t',t) = -\gamma_{\mathrm{SE}}^{(G,dG)}(t,t'). \tag{A.5}$$

The covariance is symmetric, by definition, from which it follows that $\gamma_{\mathrm{SE}}^{(G,dG)}(t',t) = \gamma_{\mathrm{SE}}^{(dG,G)}(t,t')$. Finally, the covariance between two observations of $\dot{G}$ at times $t$ and $t'$ is given by

$$\gamma_{\mathrm{SE}}^{(dG,dG)}(t,t') = \sum_{i=1}^{N} h_i^2 \left[\frac{1}{\lambda_i^2} - \frac{(t-t')^2}{\lambda_i^4}\right] \exp\left[-\frac{(t-t')^2}{2\lambda_i^2}\right]. \tag{A.6}$$

## A.2 Quasi-periodic covariance

Using a quasi-periodic covariance function formed from a squared exponential kernel, the expression for the covariance between two observations of a process $G$ at times $t$ and $t'$ is:

$$\gamma_{\mathrm{QP}}^{(G,G)}(t,t') = \exp\left\{-\frac{\sin^2\left[\pi(t-t')/P\right]}{2\lambda_{\mathrm{p}}^2} - \frac{(t-t')^2}{2\lambda_{\mathrm{e}}^2}\right\}, \tag{A.7}$$

where $P$ and $\lambda_{\mathrm{p}}$ correspond to the period and length scale of the periodic component of the variations, and $\lambda_{\mathrm{e}}$ is an evolutionary time-scale. While $\lambda_{\mathrm{e}}$ has units of time, $\lambda_{\mathrm{p}}$ is dimensionless, as it is relative to $P$.

Defining $\phi := 2\pi(t-t')/P$, we then obtain the expressions for the covariance between derivative and non-derivative observations:

$$\gamma_{\mathrm{QP}}^{(G,dG)}(t,t') = g_{\mathrm{qp}}(t,t')\left[-\frac{\pi\sin\phi}{2P\lambda_{\mathrm{p}}^2} - \frac{t-t'}{\lambda_{\mathrm{e}}^2}\right], \tag{A.8}$$

and

$$\gamma_{\mathrm{QP}}^{(G,dG)}(t',t) = -\gamma_{\mathrm{QP}}^{(G,dG)}(t,t'); \tag{A.9}$$

by symmetry of the covariance, $\gamma_{\mathrm{QP}}^{(G,dG)}(t',t) = \gamma_{\mathrm{QP}}^{(dG,G)}(t,t')$. Finally, the covariance between two observations of $\dot{G}$ at times $t$ and $t'$ is given by

$$\gamma_{\mathrm{QP}}^{(dG,dG)}(t,t') = \gamma_{\mathrm{QP}}^{(G,G)}(t,t') \times$$
$$\left[-\frac{\pi^2\sin^2\phi}{4P^2\lambda_{\mathrm{p}}^4} + \frac{\pi^2\cos\phi}{P^2\lambda_{\mathrm{p}}^2} - \frac{\phi\sin\phi}{2\lambda_{\mathrm{p}}^2\lambda_{\mathrm{e}}^2} - \frac{(t-t')^2}{\lambda_{\mathrm{e}}^4} + \frac{1}{\lambda_{\mathrm{e}}^2}\right]. \tag{A.10}$$

*If you ask naïve but relevant questions, then almost immediately the person doesn't know the answer, if he is an honest man.*

— Richard Feynman, in a 1963 lecture published in the posthumous collection *The Meaning of It All: Thoughts of a Citizen Scientist* (Feynman, 1998).

# B

# GPs: a few frequently asked questions

## Contents



This appendix is a compilation of questions I am frequently asked about GPs in general (Section B.1), and about GPs specifically for the joint modelling of stellar activity and exoplanet signals (Section B.2), together with my attempts at answers to these questions. I have relegated these questions and answers to an appendix to avoid cluttering the corresponding material in the main body of my thesis. Nevertheless, I have found that contemplating these questions has helped to clarify a few of my ideas about GPs, and about probability theory more broadly. I hope that this appendix may prove similarly useful to the reader.

# B.1 GPs in general

## B.1.1 How are GPs relevant to modelling functions whose shapes are not Gaussian?

It seems to be a somewhat common misconception that the 'Gaussian' in 'Gaussian process' refers in some way to the shapes of the functions that may be drawn from a GP, as though the functions themselves should be expected to take the form of some sort of 'bell curve'. This is not the case.

Recall that a GP is a joint distribution over infinitely many random variables, any finite collection of which must have a joint multivariate Gaussian distribution. In general, each random variable may have its own mean value (and variance, and covariance with the other random variables), so that there are no *a priori* restrictions on the actual values these random variables can take; therefore, there are no *a priori* restrictions on the sorts of functions that may be generated by a GP.

The restriction imposed by Gaussianity is related instead to the *uncertainty* in random variables, and therefore to the *uncertainty* in the functions generated by a GP. The Gaussianity tells us, for example, that significantly less probability be assigned to random functions that stray very far from the process's mean function, compared to functions close to its mean function.[1] Gaussianity also determines (via the process's covariance structure) the way in which knowing the value of a function at one point will shrink our uncertainty about the value of this function at other points.

Consider, for example, functions drawn from the QP covariance kernel defined in equation 2.37, a few examples of which are plotted in Fig. 2.15(c). Clearly these functions do not in any way resemble a Gaussian distribution. In this representation, the only hint of Gaussianity might come the way in which a large number of random, quasi-periodic functions will be scattered around the mean function.

[1] Using a more heavy-tailed distribution such as a Student's-$t$ process would render functions far from the mean function more probable; see section B.1.4.

### B.1.2 Doesn't 'Gaussian' imply uncorrelated signals?

No. It is often assumed that Gaussian noise necessarily refers to white (uncorrelated) noise, yet neither property implies the other. Gaussianity refers to the probability distribution with respect to a value, e.g. the probability of a signal falling within any particular range of amplitudes, while the term 'white' refers to the way the signal power is distributed (i.e., independently) over time or among frequencies. See also Section B.1.1. We can therefore find Gaussian white noise, but also Poisson, Cauchy, etc. white noise. We can also find coloured Gaussian noise – for example, serially-correlated red noise, the sort of signal produced by Brownian motion.

The simplifying assumption of additive *white* Gaussian noise (AWGN) is sometimes a reasonable one when modelling real-world signals corrupted by noise, and leads to especially simple mathematical models. Nevertheless, nuisance signals associated with stellar activity processes, for example, are manifestly correlated over time (displaying such properties as quasi-periodicity, some degree of smoothness, etc.). Thus it becomes necessary to adopt something more sophisticated than an AWGN model – a natural extension to which comes in the form of a GP model.

### B.1.3 Is it not restrictive to assume that there must exist covariance between observations?

No. If a signal is uncorrelated, it is white noise. A GP model can trivially accommodate white Gaussian noise via the covariance kernel $k(x_i, x_j) = \sigma_i^2 \delta_{ij}$. It is usually also extremely straightforward to account for other forms of white noise (e.g. Poisson noise) in data modelling.

The real power of GPs in the context of regression, however, lies in their ability to model *correlated* signals (i.e. signals that are not white noise) that cannot easily be parametrised.[2] The signal in question might be difficult to parametrise because of a stochastic component, or because of imperfect knowledge of the process giving rise to the signal, such that any parametric model we could build would

[2]Of course, whether these correlated signals are referred to as 'correlated noise' or '(correlated) nuisance signals' is somewhat context-dependent. Stellar activity signals might be noise to an exoplanet hunter, but an interesting signal to a stellar astrophysicist.

necessarily be predicated on *ad hoc* assumptions, and in practice would likely be either overly simplistic or underdetermined.

While GPs give us a principled way to model correlated signals that we do not know how to parametrise, an easier yet far more dangerous approach to modelling is to pretend that signals containing genuine correlation can be treated simply as white noise (Morales, 1999; Lund et al., 2006; Smith et al., 2007). Unfortunately, such approaches continue to find widespread use in exoplanetary science. There are countless studies, for example, in which an additive Gaussian 'jitter' term is injudiciously used to soak up all variation that cannot be explained via an *ad hoc* parametric model. The danger inherent in such approaches is discussed in Section 2.5 of this thesis.

### B.1.4 Is the assumption of Gaussian noise reasonable?

Often, but not always.

In probability theory, the central limit theorem establishes that when independent random variables are added, their sum tends towards a Gaussian distribution, even if the original variables are not normally distributed.[3] In astronomical experiments, measurement errors usually contain contributions from many independent sources (photon noise, thermo-mechanical noise, calibration errors, telescope and detector effects, atmospheric effects, etc.), so that the final noise distribution may be expected to be approximately Gaussian.

Furthermore, from an information theoretic approach, the maximum entropy principle may be used to show that in the absence of detailed knowledge of the effective noise distribution (other than that it has a finite variance), a Gaussian distribution would be the most conservative choice, i.e., maximally non-committal about missing information (Gregory, 2005; Sivia and Skilling, 2006).

Nevertheless, situations may arise in which Gaussian noise becomes contaminated by a few non-Gaussian outliers – or where the noise distribution is, for some reason,

[3] In more precise terms, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a finite expectation value and variance, will be approximately normally distributed, regardless of the underlying distribution (Fischer, 2011).

altogether non-Gaussian. In the former case, a number of techniques exist for detecting and mitigating the effects of outliers (see e.g. Rousseeuw and Leroy, 2005; Stegle et al., 2008; Kim and Ghahramani, 2008). In the latter case, it is possible to generalise GPs (most directly) to Student-$t$ processes.

In addition to a mean function and a covariance function, a Student-$t$ process (TP) requires a parameter $\nu$ to specify how *light tailed* the process is. A GP may be seen as the $\nu \to \infty$ limiting case, i.e. the TP with the lightest possible tails. As is the case with a GP, a TP may be used to place a non-parametric prior over functions, and somewhat remarkably, closed-form expressions exist also for the marginal likelihood and predictive distribution of a TP. While TPs are more algebraically complex than GPs, they offer most of the benefits of GPs, and the advantage of being able to cope directly with heavy-tailed noise behaviour. See Vanhatalo et al. (2009), Jylänki et al. (2011) and Shah et al. (2014) for more details on TPs.

### B.1.5 Are GPs not overly flexible, 'black box' models with little grounding in physics?

In principle, GPs *can* be used to construct models that will fit arbitrary data sets arbitrarily well. They can be to interpolate data, for example, although there are certainly much simpler means (e.g. spline interpolation) for achieving such ends.

In this thesis, however, GPs are used as a tool for performing principled Bayesian inference about signals that cannot easily be parametrised (e.g. stellar activity signals; spectra from a wide variety of stars), but about which we nevertheless have significant prior physical knowledge. This knowledge is encoded into the choice of GP covariance function, the covariance hyper-parameter priors, the choice of mean function, the mean parameter priors, etc. Additionally, overly-complex models are automatically penalised, while those which are most compatible with both our prior physical knowledge and observed data are favoured. Thus the prior GP model is grounded in physics, and the posterior GP model, in turn, is used to make inferences about physical processes we have observed.

In the framework used in Chapter 3 of this thesis, for example, the GP models incorporate physics ranging from the expected forms of signals from evolving spots on a rotating star with long-term magnetic cycles, to the Keplerian signals of exoplanets orbiting the star, to correlations expected between the radial velocity variations of the star and activity-induced changes in the shape of its spectral lines, and more. Clearly, then, the model is strongly grounded in physics. While the model cannot and does not incorporate *all* relevant physics, it also makes relatively few simplifying or restrictive assumptions; it might therefore be seen as being honest in a quantitive way about the physics it does not incorporate.[4]

### B.1.6 Does the use of priors not confer GPs a misleading advantage over other models?

No. Any model masquerading as 'prior-free' must nevertheless make certain assumptions about the system under investigation, and implicitly imposes (potentially arbitrary, unjustified) priors of some sort.

If our initial state of knowledge about a parameter (or model, distribution, etc.) is indeed complete ignorance, we should take care to construct a non-informative prior that is a consistent reflection of this ignorance. For example, given a parameter $\theta$ that could potentially scale many orders of magnitude (e.g. an RV semi-amplitude, or a planet's orbital period), a log uniform prior ($p(\theta) \propto 1/\theta$) would be a more appropriate choice than a uniform prior ($p(\theta) = \text{constant}$), as the latter prior, perhaps somewhat counter-intuitively, is strongly biased towards larger values of $\theta$, while the former prior is scale invariant, and assigns equal probability per decade. Conversely, inference will always benefit from the principled incorporation of any informative prior knowledge we might possess. This applies to inference in general, and not just to GPs.

Jaynes (2003) gives a comprehensive account of the demonstrable superiority of Bayesian inference methods over supposedly prior-free frequentist approaches; of the

[4] As an antithesis to the question about GPs being overly flexible, it is worth asking: for a given problem, might a given *ad hoc* model be too restrictive or simplistic? What about the infinitely many other *ad hoc* models that could model the same data as well, or better? Which of those models is most 'grounded in physics'?

inevitability of priors within a Bayesian framework; and of principled approaches (e.g. maximum entropy methods) for constructing 'objective' prior distributions.

## B.2 Exoplanets and stellar activity modelling

### B.2.1 How do you know that a GP will not 'absorb' some of the signal of an exoplanet?

Using the framework in Chapter 3 of this thesis, A GP would not be expected to 'absorb' some of the signal of an exoplanet for the following reasons.[5]

1. The same function (or simple affine transformations of it) is forced to model activity variability in multiple time series, e.g. $\Delta$RV, BIS, FWHM, $\log R'_{\rm HK}$, etc. While a planet would contribute a Keplerian signal to the $\Delta$RV time series, its signal would not be present in any of the other, activity-sensitive time series. Therefore, the only case in which a GP model could, in principle, absorb RV variability of planetary origin would be where nearly identical variability happened to be simultaneously present in the activity-sensitives time series.
2. Even if a planet's orbital period happened to be identical to its host star's rotation period, stellar activity-induced variability in (say) a BIS or $\log R'_{\rm HK}$ time series would not generally be expected to have the same phase, Keplerian shape, and coherence over many rotation periods as would the planet's RV signature. Therefore, once again, the GP model would not be expected to mistake a planetary signal for an activity signal.
3. Finally, even if confronted with a stellar activity signal identical to all respects to a planetary signal, a key point is that a GP model with QP kernel has significantly more flexibility than a Keplerian model. The latter model enforces strict periodicity, and RV signals of a very specific form, whereas the GP model can describe essentially any smooth function exhibiting locally-periodic behaviour. Within a Bayesian inferential framework, then, a simpler Keplerian model should

[5] As an antithesis to the question of GPs potentially absorbing some of the signal of an exoplanet, it is worth asking: for a given problem, might an overly-simplistic activity model force the planetary component(s) of the model to inflate to account for some of the residual activity variability?

always be favoured over a vastly more complex GP model, assuming of course the former suffices to describe a given signal. Given the complexity penalty associated with a GP model, it would only be used to explain variability that cannot be ascribed to a Keplerian function.

The above discussion applies to the case of a single planet plus stellar activity. In the case of multiple planets, it is conceivable that a GP could artificially absorb planetary variability *if* an inadequate model for the planets were used, e.g. if there were signals from three planets present but only a single Keplerian model were used. Nevertheless, in such cases, Bayesian model comparison tests should strongly favour (say) a '3 planets plus activity' model over a 'one planet plus activity' model.

### B.2.2 Will the use of a GP lead to more false-positive or false-negative exoplanet detections?

There is no part of the GP modelling framework in Chapter 3 that would increase the inherent resemblance of a given RV nuisance signal (stellar, instrumental, or otherwise) to a Keplerian signal, and so the use of GPs should not in any way increase the rate of false-positive planet detections. On the contrary, while simplistic parametric models might fit activity signals imperfectly (allowing some of their variability to be absorbed by a Keplerian model), a GP model should do a better job of accounting for activity signals, and *decrease* the rate of false positives.

Conversely, as discussed in Section B.2.1, there are various reasons why the GP modelling framework in Chapter 3 would not be expected to mistake any part of planetary signal for an activity signal, and so the framework should not increase the possibility of false-negative planet detections. On the contrary, once again, the superior activity modelling should facilitate cleaner 'extraction' of planetary signals, and a lower false negative rate. Of course, GPs are not 'magic bullets.' There will always be a noise floor below which all planetary signals will be undetectable, regardless of the sophistication of the modelling technique. Yet the use of the GP framework should, at least, permit one to detect and characterise planets closer to that noise floor than more simplistic techniques might allow.

*Next o'er his books his eyes began to roll,*
*In pleasing memory of all he stole;*
*How here he sipp'd, how there he plunder'd snug,*
*And suck'd all o'er like an industrious bug.*

— Alexander Pope, from 'The Dunciad', Book I (1743).

# Bibliography


Aerts, C., Christensen-Dalsgaard, J., and Kurtz, D. W. (2010). *Asteroseismology*. Springer.

Aigrain, S. et al. (2015). Testing the recovery of stellar rotation signals from Kepler light curves using a blind hare-and-hounds exercise. *MNRAS*, 450(3):3211–3226.

Aigrain, S., Pont, F., and Zucker, S. (2012). A simple method to estimate radial velocity variations due to stellar activity using photometry. *MNRAS*, 419:3147–3158.

Allen, C. W. (1976). *Astrophysical Quantities*. Athlone Press.

Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., and O'Neil, M. (2014). Fast Direct Methods for Gaussian Processes. *ArXiv e-prints*, arXiv:1403.6015.

Anglada-Escudé, G. et al. (2016). A terrestrial planet candidate in a temperate orbit around Proxima Centauri. *Nature*, 536:437–440.

Angus, R., Aigrain, S., Foreman-Mackey, D., and McQuillan, A. (2015). Calibrating gyrochronology using Kepler asteroseismic targets. *MNRAS*, 450:1787–1798.

Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., and Rajpaul, V. (2018). Inferring probabilistic stellar rotation periods using gaussian processes. *MNRAS*, 474(2):2094–2108.

Antonini, M., Barlaud, M., Mathieu, P., and Daubechies, I. (1992). Image coding using wavelet transform. *IEEE Transactions on Image Processing*, 1:205–220.

Artigau, É. et al. (2014). SPIRou: the near-infrared spectropolarimeter/high-precision velocimeter for the Canada-France-Hawaii telescope. In *Ground-based and Airborne Instrumentation for Astronomy V*, volume 9147 of *Proc. SPIE*, page 914715.

Babcock, H. W. (1953). The Solar Magnetograph. *ApJ*, 118:387.

Bailes, M. et al. (2011). Transformation of a Star into a Planet in a Millisecond Pulsar Binary. *Science*, 333:1717.

Baluev, R. V. (2013). Detecting multiple periodicities in observational data with the multifrequency periodogram - I. Analytic assessment of the statistical significance. *MNRAS*, 436:807–818.

Baranne, A., Mayor, M., and Poncet, J. L. (1979). CORAVEL - A new tool for radial velocity measurements. *Vistas in Astronomy*, 23:279–316.

Baranne, A., Queloz, D., Mayor, M., Adrianzyk, G., Knispel, G., Kohler, D., Lacroix, D., Meunier, J.-P., Rimbaud, G., and Vin, A. (1996). ELODIE: A spectrograph for accurate radial velocity measurements. *A&AS*, 119:373–390.

Barclay, T., Endl, M., Huber, D., Foreman-Mackey, D., Cochran, W. D., MacQueen, P. J., Rowe, J. F., and Quintana, E. V. (2014). Radial Velocity Observations and Light Curve Noise Modeling Confirm That Kepler-91b is a Giant Planet Orbiting a Giant Star. *ArXiv e-prints*, arXiv:1408.3149.

Barclay, T. et al. (2013). A sub-Mercury-sized exoplanet. *Nature*, 494:452–454.

Barnes, S. A. (2007). Ages for Illustrative Field Stars Using Gyrochronology: Viability, Limitations, and Errors. *ApJ*, 669:1167–1189.

Barnes, S. A. (2010). A Simple Nonlinear Model for the Rotation of Main-sequence Cool Stars. I. Introduction, Implications for Gyrochronology, and Color-Period Diagrams. *ApJ*, 722:222–234.

Basri, G., Borucki, W. J., and Koch, D. (2005). The Kepler Mission: A wide-field transit search for terrestrial planets [review article]. *New A Rev.*, 49:478–485.

Batalha, N. M. et al. (2011). Kepler's First Rocky Planet: Kepler-10b. *ApJ*, 729:27.

Beaulieu, J.-P. et al. (2006). Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing. *Nature*, 439:437–440.

Berger, J. O. and Pericchi, L. R. (1996). The Intrinsic Bayes Factor for Model Selection and Prediction. *Journal of the American Statistical Association*, 91(433):109–122.

Bergmann, C., Endl, M., Hearnshaw, J. B., Wittenmyer, R. A., and Wright, D. J. (2015). Searching for Earth-mass planets around $\alpha$ Centauri: precise radial velocities from contaminated spectra. *International Journal of Astrobiology*, 14:173–176.

Bettex, A. W. (1965). *The Discovery of Nature*. Simon and Schuster.

Bhattacharya, R. N. and Waymire, E. C. (2009). *Stochastic Processes with Applications*. SIAM.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Information Science and Statistics Series. Springer New York.

Boisse, I., Bonfils, X., and Santos, N. C. (2012). SOAP. A tool for the fast computation of photometry and radial velocity induced by stellar spots. *A&A*, 545:A109.

Boisse, I., Bouchy, F., Hébrard, G., Bonfils, X., Santos, N., and Vauclair, S. (2011). Disentangling between stellar activity and planetary signals. *A&A*, 528:A4.

Boisse, I., Moutou, C., Vidal-Madjar, A., Bouchy, F., Pont, F., Hébrard, G., Bonfils, X., Croll, B., Delfosse, X., Desort, M., and et al. (2009). Stellar activity of planetary host star HD 189 733. *A&A*, 495:959–966.

Bolmont, E., Raymond, S. N., Leconte, J., and Matt, S. P. (2012). Effect of the stellar spin history on the tidal evolution of close-in planets. *A&A*, 544:A124.

Bond, I. A. et al. (2004). OGLE 2003-BLG-235/MOA 2003-BLG-53: A Planetary Microlensing Event. *ApJ*, 606:L155–L158.

Bonfils, X. et al. (2007). The HARPS search for southern extra-solar planets. X. A $m \sin i = 11 M_{\oplus}$ planet around the nearby spotted M dwarf GJ 674. *A&A*, 474:293–299.

Borucki, W. J. (2016). KEPLER Mission: development and overview. *Reports on Progress in Physics*, 79(3):036901.

Boss, A. P. et al. (2007). Working Group on Extrasolar Planets. *Transactions of the International Astronomical Union, Series A*, 26:183–186.

Bouchy, F. and Carrier, F. (2001). P-mode observations on $\alpha$ Cen A. *A&A*, 374:L5–L8.

Bouchy, F., Pepe, F., and Queloz, D. (2001). Fundamental photon noise limit to radial velocity measurements. *A&A*, 374:733–739.

Bouchy, F., Pont, F., Melo, C., Santos, N. C., Mayor, M., Queloz, D., and Udry, S. (2005). Doppler follow-up of OGLE transiting companions in the Galactic bulge. *A&A*, 431:1105–1121.

Bressan, A., Marigo, P., Girardi, L., Salasnich, B., Dal Cero, C., Rubele, S., and Nanni, A. (2012). PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code. *MNRAS*, 427:127–145.

Bretthorst, G. L. (1988). *Bayesian Spectrum Analysis and Parameter Estimation*, volume 48. Springer-Verlag.

Brinkworth, C. S., Marsh, T. R., Morales-Rueda, L., Maxted, P. F. L., Burleigh, M. R., and Good, S. A. (2005). Rotational period of WD 1953-011- a magnetic white dwarf with a star-spot. *MNRAS*, 357:333–337.

Brown, D. J. A., Collier Cameron, A., Hall, C., Hebb, L., and Smalley, B. (2011). Are falling planets spinning up their host stars? *MNRAS*, 415:605–618.

Buchhave, L. A. and Latham, D. W. (2015). The Metallicities of Stars with and without Transiting Planets. *ApJ*, 808:187.

Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.

Cabrera, J., Csizmadia, S., Lehmann, H., Dvorak, R., Gandolfi, D., Rauer, H., Erikson, A., Dreyer, C., Eigmüller, P., and Hatzes, A. (2014). The Planetary System to KIC 11442793: A Compact Analogue to the Solar System. *ApJ*, 781:18.

Campbell, B. and Walker, G. A. H. (1979). Precision radial velocities with an absorption cell. *PASP*, 91:540–545.

Campbell, B., Walker, G. A. H., and Yang, S. (1988). A search for substellar companions to Solar-type stars. *ApJ*, 331:902–921.

Carolo, E. et al. (2014). A vigorous activity cycle mimicking a planetary system in HD 200466. *ArXiv e-prints*, arXiv:1406.0697.

Carroll, B. W. and Ostlie, D. A. (2013). *An Introduction to Modern Astrophysics*. Pearson, 2nd edition.

Cegla, H. M., Watson, C. A., Marsh, T. R., Shelyag, S., Moulds, V., Littlefair, S., Mathioudakis, M., Pollacco, D., and Bonfils, X. (2012). Stellar jitter from

variable gravitational redshift: implications for radial velocity confirmation of habitable exoplanets. *MNRAS*, 421:L54–L58.

Chalupka, K., Williams, C. K. I., and Murray, I. (2012). A Framework for Evaluating Approximation Methods for Gaussian Process Regression. *ArXiv e-prints*, arXiv:1205.6326.

Charbonneau, D., Brown, T. M., Latham, D. W., and Mayor, M. (2000). Detection of Planetary Transits Across a Sun-like Star. *ApJ*, 529:L45–L48.

Charbonneau, D., Brown, T. M., Noyes, R. W., and Gilliland, R. L. (2002). Detection of an Extrasolar Planet Atmosphere. *ApJ*, 568:377–384.

Charpinet, S. et al. (2011). A compact system of small planets around a former red-giant star. *Nature*, 480:496–499.

Chauvin, G., Lagrange, A.-M., Dumas, C., Zuckerman, B., Mouillet, D., Song, I., Beuzit, J.-L., and Lowrance, P. (2004). A giant planet candidate near a young brown dwarf. Direct VLT/NACO observations using IR wavefront sensing. *A&A*, 425:L29–L32.

Christensen-Dalsgaard, J. (2004). Physics of Solar-like oscillations. *Sol. Phys.*, 220:137–168.

Cochran, W. D., Hatzes, A. P., and Hancock, T. J. (1991). Constraints on the companion object to HD 114762. *ApJ*, 380:L35–L38.

Connes, P. (1985). Absolute astronomical accelerometry. *Ap&SS*, 110:211–255.

Cosentino, R. et al. (2014). HARPS-N @ TNG, two year harvesting data: performances and results. In *Ground-based and Airborne Instrumentation for Astronomy V*, volume 9147 of *Proc. SPIE*, page 91478C.

Cowan, N. B. and Fujii, Y. (2017). Mapping Exoplanets. *ArXiv e-prints*, arXiv:1704.07832.

Cunha, D., Santos, N. C., Figueira, P., Santerne, A., Bertaux, J. L., and Lovis, C. (2014). Impact of micro-telluric lines on precise radial velocities and its correction. *A&A*, 568:A35.

Curie, E. (1938). *Madame Curie : ihr Leben und Wirken*. Th. Knaur, Berlin.

da Silva, L., Girardi, L., Pasquini, L., Setiawan, J., von der Lühe, O., de Medeiros, J. R., Hatzes, A., Döllinger, M. P., and Weiss, A. (2006). Basic physical parameters of a selected sample of evolved stars. *A&A*, 458:609–623.

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36:961–1005.

Dawson, R. I. and Fabrycky, D. C. (2010). Radial Velocity Planets De-aliased: A New, Short Period for Super-Earth 55 Cnc e. *ApJ*, 722(1):937.

Deacon, N. R., Schlieder, J. E., and Murphy, S. J. (2016). A nearby young M dwarf with a wide, possibly planetary-mass companion. *MNRAS*, 457:3191–3199.

Deck, K. M., Agol, E., Holman, M. J., and Nesvorný, D. (2014). TTVFast: An Efficient and Accurate Code for Transit Timing Inversion Problems. *ApJ*, 787(2):132.

Demory, B.-O. et al. (2016). A map of the large day-night temperature gradient of a super-Earth exoplanet. *Nature*, 532:207–209.

Dempsey, R. C., Bopp, B. W., Strassmeier, K. G., Granados, A. F., Henry, G. W., and Hall, D. S. (1992). Line profile asymmetries in chromospherically active stars. *ApJ*, 392:187–200.

Desidera, S., Gratton, R. G., Endl, M., Claudi, R. U., Cosentino, R., Barbieri, M., Bonanno, G., Lucatello, S., Martinez Fiorenzano, A. F., Marzari, F., and Scuderi, S. (2004). No planet around HD 219542 B. *A&A*, 420:L27–L30.

Desort, M., Lagrange, A.-M., Galland, F., Udry, S., and Mayor, M. (2007). Search for exoplanets with the radial-velocity technique: quantitative diagnostics of stellar activity. *A&A*, 473:983–993.

Dobbs-Dixon, I., Lin, D. N. C., and Mardling, R. A. (2004). Spin-Orbit Evolution of Short-Period Planets. *ApJ*, 610:464–476.

Dravins, D. (1987). Stellar granulation. I - The observability of stellar photospheric convection. *A&A*, 172:200–224.

Dumusque, X. (2012). *Mitigating stellar signals in the quest for other Earths*. PhD thesis, University of Geneva.

Dumusque, X., Boisse, I., and Santos, N. C. (2014a). SOAP 2.0: A Tool to Estimate the Photometric and Radial Velocity Variations Induced by Stellar Spots and Plages. *ApJ*, 796(2):132.

Dumusque, X., Borsa, F., Damasso, M., Díaz, R. F., Gregory, P. C., Hara, N. C., Hatzes, A., Rajpaul, V., et al. (2017). Radial-velocity fitting challenge. II. First results of the analysis of the data set. *A&A*, 598:A133.

Dumusque, X. et al. (2011a). The HARPS search for southern extra-solar planets. XXX. Planetary systems around stars with Solar-like magnetic cycles and short-term activity variation. *A&A*, 535:A55.

Dumusque, X. et al. (2012). An Earth-mass planet orbiting $\alpha$ Centauri B. *Nature*, 491:207–211.

Dumusque, X. et al. (2014b). The Kepler-10 Planetary System Revisited by HARPS-N: A Hot Rocky World and a Solid Neptune-Mass Planet. *ApJ*, 789:154.

Dumusque, X. et al. (2015). HARPS-N Observes the Sun as a Star. *ApJ*, 814:L21.

Dumusque, X., Santos, N. C., Udry, S., Lovis, C., and Bonfils, X. (2011b). Planetary detection limits taking into account stellar noise. II. Effect of stellar spot groups on radial-velocities. *A&A*, 527:A82.

Dumusque, X., Udry, S., Lovis, C., Santos, N. C., and Monteiro, M. J. P. F. G. (2011c). Planetary detection limits taking into account stellar noise. I. Observational strategies to reduce stellar oscillation and granulation effects. *A&A*, 525:A140.

Edelson, R. A. and Krolik, J. H. (1988). The discrete correlation function - A new method for analyzing unevenly sampled variability data. *ApJ*, 333:646–659.

Eisenstein, G. (1847). *Mathematische Abhandlungen. Besonders aus dem Gebiete der höhern Arithmetik und der elliptischen Functionen*. G. Reimer.

Endl, M., Bergmann, C., Hearnshaw, J., Barnes, S. I., Wittenmyer, R. A., Ramm, D.,

Kilmartin, P., Gunn, F., and Brogt, E. (2015). The Mt John University Observatory search for Earth-mass planets in the habitable zone of $\alpha$ Centauri. *International Journal of Astrobiology*, 14:305–312.

Fabrycky, D. C. (2010). *Non-Keplerian Dynamics of Exoplanets*, pages 217–238. The University of Arizona Press.

Farsiu, S., Robinson, M. D., Elad, M., and Milanfar, P. (2004). Fast and Robust Multiframe Super Resolution. *IEEE Transactions on Image Processing*, 13:1327–1344.

Fellgett, P. (1955). A Proposal for a Radial Velocity Photometer. *Optica Acta*, 2:9–16.

Feroz, F. and Hobson, M. P. (2008). Multimodal nested sampling: an efficient and robust alternative to Markov Chain Monte Carlo methods for astronomical data analyses. *MNRAS*, 384:449–463.

Feroz, F. and Hobson, M. P. (2014). Bayesian analysis of radial velocity data of GJ667C with correlated noise: evidence for only two planets. *MNRAS*, 437:3540–3549.

Feroz, F., Hobson, M. P., and Bridges, M. (2009). MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics. *MNRAS*, 398:1601–1614.

Feroz, F., Hobson, M. P., Cameron, E., and Pettitt, A. N. (2013). Importance Nested Sampling and the MultiNest Algorithm. *ArXiv e-prints*, arXiv:1306.2144.

Feynman, R. P. (1998). *The Meaning of it All: Thoughts of a Citizen Scientist*. Penguin.

Figueira, P., Santos, N. C., Pepe, F., Lovis, C., and Nardetto, N. (2013). Line-profile variations in radial-velocity measurements. Two alternative indicators for planetary searches. *A&A*, 557:A93.

Fischer, D. A. and Valenti, J. (2005). The Planet-Metallicity Correlation. *ApJ*, 622:1102–1117.

Fischer, H. (2011). *A History of the Central Limit Theorem*. Springer.

Flammarion, C. (1862). *La pluralité des mondes habités*. Didier et Cie., Paris.

Foreman-Mackey, D. (2016). corner.py: Scatterplot matrices in Python. *The Journal of Open Source Software*, 2016.

Foreman-Mackey, D., Agol, E., Angus, R., and Ambikasaran, S. (2017). Fast and scalable Gaussian process modeling with applications to astronomical time series. *ArXiv e-prints*, arXiv:1703.09710.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306–312.

French, A. P. (2001). *Vibrations and Waves*. CBS Publishers & Distributors.

Fressin, F. et al. (2011). Kepler-10 c: a 2.2 Earth Radius Transiting Planet in a Multiple System. *ApJS*, 197:5.

Gandolfi, D. et al. (2010). Transiting exoplanets from the CoRoT space mission. XIV. CoRoT-11b: a transiting massive "hot-Jupiter" in a prograde orbit around a rapidly rotating F-type star. *A&A*, 524:A55.

García, R. A. et al. (2014). Rotation and magnetism of Kepler pulsating Solar-like stars. Towards asteroseismically calibrated age-rotation relations. *A&A*, 572:A34.

Gibbs, M. N. (1998). *Bayesian Gaussian processes for regression and classification*. PhD thesis, University of Cambridge.

Gibson, N. P., Aigrain, S., Roberts, S., Evans, T. M., Osborne, M., and Pont, F. (2012). A Gaussian process framework for modelling instrumental systematics: application to transmission spectroscopy. *MNRAS*, 419(3):2683–2694.

Gilliland, R. L. et al. (2011). Kepler Mission Stellar and Instrument Noise Properties. *ApJS*, 197:6.

Gillon, M. et al. (2017). Seven temperate terrestrial planets around the nearby ultracool dwarf star TRAPPIST-1. *Nature*, 542:456–460.

Glebocki, R. and Gnacinski, P. (2005). VizieR Online Data Catalog: Catalog of Stellar Rotational Velocities (Glebocki+ 2005). *VizieR Online Data Catalog*, 3244.

Gomes da Silva, J., Santos, N. C., Bonfils, X., Delfosse, X., Forveille, T., Udry, S., Dumusque, X., and Lovis, C. (2012). Long-term magnetic activity of a sample of M-dwarf stars from the HARPS program . II. Activity and radial velocity. *A&A*, 541:A9.

Gonzalez, G. (1997). The stellar metallicity-giant planet connection. *MNRAS*, 285:403–412.

Gregory, P. (2005). *Bayesian Logical Data Analysis for the Physical Sciences*, volume 10. Cambridge University Press Cambridge, UK.

Gregory, P. C. (2016). An apodized Kepler periodogram for separating planetary and stellar activity signals. *MNRAS*, 458:2604–2633.

Griffin, R. F. (1967). A Photoelectric Radial-Velocity Spectrometer. *ApJ*, 148:465.

Grunblatt, S. K., Howard, A. W., and Haywood, R. D. (2015). Measuring the Mass of Kepler-78b Using a Gaussian Process Model. *ArXiv e-prints*, arXiv:1501.00369.

Guedes, J. M., Rivera, E. J., Davis, E., Laughlin, G., Quintana, E. V., and Fischer, D. A. (2008). Formation and Detectability of Terrestrial Planets around $\alpha$ Centauri B. *ApJ*, 679:1582–1587.

Hadden, S. and Lithwick, Y. (2014). Densities and Eccentricities of 139 Kepler Planets from Transit Time Variations. *The Astrophysical Journal*, 787(1):80.

Hale, G. E. (1908). On the Probable Existence of a Magnetic Field in Sun-Spots. *ApJ*, 28:315.

Hall, J. C. (2008). Stellar Chromospheric Activity. *Living Reviews in Solar Physics*, 5:2.

Halverson, S. et al. (2016). A comprehensive radial velocity error budget for next generation Doppler spectrometers. In *Ground-based and Airborne Instrumentation for Astronomy VI*, volume 9908 of *Proc. SPIE*, page 99086P.

Handley, W. J., Hobson, M. P., and Lasenby, A. N. (2015). POLYCHORD: next-generation nested sampling. *MNRAS*, 453:4384–4398.

Hara, N. C., Boué, G., Laskar, J., and Correia, A. C. M. (2017). Radial velocity data analysis with compressed sensing techniques. *MNRAS*, 464:1220–1246.

Hassan, U. and Sabieh Anwar, M. (2010). Reducing noise by repetition: introduction to signal averaging. *European Journal of Physics*, 31:453–465.

Hatzes, A. P. (1996). Simulations of Stellar Radial Velocity and Spectral Line Bisector Variations: I. Nonradial Pulsations. *PASP*, 108:839.

Hatzes, A. P. (2012). Astronomy: Meet our closest neighbour. *Nature*, 491:200–201.

Hatzes, A. P. (2013). The Radial Velocity Detection of Earth-mass Planets in the Presence of Activity Noise: The Case of $\alpha$ Centauri Bb. *ApJ*, 770:133.

Hatzes, A. P. and Cochran, W. D. (1993). Long-period radial velocity variations in three K giants. *ApJ*, 413:339–348.

Hatzes, A. P., Dvorak, R., Wuchterl, G., Guterman, P., Hartmann, M., Fridlund, M., Gandolfi, D., Guenther, E., and Pätzold, M. (2010). An investigation into the radial velocity variations of CoRoT-7. *A&A*, 520:A93.

Hatzes, A. P. et al. (2006). Confirmation of the planet hypothesis for the long-period radial velocity variations of $\beta$ Geminorum. *A&A*, 457:335–341.

Hatzes, A. P. et al. (2011). The Mass of CoRoT-7b. *ApJ*, 743:75.

Haywood, R., Cameron, A. C., Queloz, D., Barros, S., Deleuil, M., Fares, R., Gillon, M., Hatzes, A., Lanza, A., Lovis, C., et al. (2014). Disentangling planetary orbits from stellar activity in radial-velocity surveys. *International Journal of Astrobiology*, 13(02):155–157.

Haywood, R. D. (2015). *Hide and Seek: Radial-Velocity Searches for Planets around Active Stars*. PhD thesis, University of St Andrews.

Haywood, R. D. et al. (2014). Planets and Stellar Activity: Hide and Seek in the CoRoT-7 system. In Booth, M., Matthews, B. C., and Graham, J. R., editors, *IAU Symposium*, volume 299 of *IAU Symposium*, pages 237–240.

Hébrard, G. et al. (2008). Misaligned spin-orbit in the XO-3 planetary system? *A&A*, 488:763–770.

Heintz, W. D. (1978). Reexamination of suspected unresolved binaries. *ApJ*, 220:931–934.

Henry, G. W., Marcy, G. W., Butler, R. P., and Vogt, S. S. (2000). A Transiting "51 Peg-like" Planet. *ApJ*, 529:L41–L44.

Hershey, J. L. (1973). Astrometric analysis of the field of AC +65 6955 from plates taken with the Sproul 24-inch refractor. *AJ*, 78:421–425.

Högbom, J. A. (1974). Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *A&AS*, 15:417.

Holczer, T., Mazeh, T., Nachmani, G., Jontof-Hutter, D., Ford, E. B., Fabrycky, D., Ragozzine, D., Kane, M., and Steffen, J. H. (2016). Transit Timing Observations from Kepler. IX. Catalog of the Full Long-cadence Data Set. *ApJS*, 225:9.

Hou, F. (2014). *Bayesian Statistics using Stellar Radial Velocity Data*. PhD thesis, New York University.

Howard, A. W. et al. (2013). A rocky composition for an Earth-sized exoplanet. *Nature*, 503:381–384.

Howard, A. W. et al. (2014). The NASA-UC-UH ETA-Earth Program. IV. A Low-mass Planet Orbiting an M Dwarf 3.6 PC from Earth. *ApJ*, 794(1):51.

Ingrosso, G., Novati, S. C., de Paolis, F., Jetzer, P., Nucita, A., and Zakharov, A. (2011). Search for exoplanets in M31 with pixel-lensing and the PA-99-N2 event revisited. *General Relativity and Gravitation*, 43:1047–1060.

Isaacson, H. and Fischer, D. (2010). Chromospheric Activity and Jitter Measurements for 2630 Stars on the California Planet Search. *ApJ*, 725:875–885.

Jacob, W. S. (1855). On certain Anomalies presented by the Binary Star 70 Ophiuchi. *MNRAS*, 15:228.

Jaynes, E. T. (2003). *Probability Theory*. Cambridge University Press.

Jeffers, S. V. and Keller, C. U. (2009). An analytical model to demonstrate the reliability of reconstructed 'active longitudes'. In Stempels, E., editor, *15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*, volume 1094 of *American Institute of Physics Conference Series*, pages 664–667.

Jenkins, J. S., Ramsey, L. W., Jones, H. R. A., Pavlenko, Y., Gallardo, J., Barnes, J. R., and Pinfield, D. J. (2009). Rotational Velocities for M Dwarfs. *ApJ*, 704:975–988.

Jenkins, J. S. and Tuomi, M. (2014). The Curious Case of HD 41248. A Pair of Static Signals Buried Behind Red Noise. *ApJ*, 794:110.

Jenkins, J. S., Tuomi, M., Brasser, R., Ivanyuk, O., and Murgas, F. (2013). Two Super-Earths Orbiting the Solar Analog HD 41248 on the Edge of a 7:5 Mean Motion Resonance. *ApJ*, 771:41.

Johnson, J. A., Aller, K. M., Howard, A. W., and Crepp, J. R. (2010). Giant Planet Occurrence in the Stellar Mass-Metallicity Plane. *PASP*, 122:905.

Jones, H. R. A., Butler, R. P., Tinney, C. G., Marcy, G. W., Carter, B. D., Penny, A. J., McCarthy, C., and Bailey, J. (2006). High-eccentricity planets from the Anglo-Australian Planet Search. *MNRAS*, 369:249–256.

Jylänki, P., Vanhatalo, J., and Vehtari, A. (2011). Robust Gaussian Process Regression with a Student-$t$ Likelihood. *Journal of Machine Learning Research*, 12(11):3227 – 3257.

Keller, C. U., Schüssler, M., Vögler, A., and Zakharov, V. (2004). On the Origin of Solar Faculae. *ApJ*, 607:L59–L62.

Kim, H.-C. and Ghahramani, Z. (2008). *Outlier Robust Gaussian Process Classification*, pages 896–905. Springer.

Kipping, D. M. (2012). An analytic model for rotational modulations in the photometry of spotted stars. *MNRAS*, 427:2487–2511.

Kipping, D. M., Schmitt, A. R., Huang, X., Torres, G., Nesvorný, D., Buchhave, L. A.,

Hartman, J., and Bakos, G. Á. (2015). The Hunt for Exomoons with Kepler (HEK): V. A Survey of 41 Planetary Candidates for Exomoons. *ApJ*, 813:14.

Koch, D. G. et al. (2010). Kepler Mission Design, Realized Photometric Performance, and Early Science. *ApJ*, 713:L79–L86.

Kochukhov, O. (2016). Doppler and Zeeman Doppler Imaging of Stars. In Rozelot, J.-P. and Neiner, C., editors, *Lecture Notes in Physics, Berlin Springer Verlag*, volume 914 of *Lecture Notes in Physics, Berlin Springer Verlag*, page 177.

Konacki, M., Torres, G., Sasselov, D. D., and Jha, S. (2003). High-Resolution Spectroscopic Follow-up of OGLE Planetary Transit Candidates in the Galactic Bulge: Two Possible Jupiter-Mass Planets and Two Blends. *ApJ*, 597:1076–1091.

Kuss, M. (2006). *Gaussian Process Models for Robust Regression, Classification, and Reinforcement Learning*. PhD thesis, Technische Universität Darmstadt.

Lang, K. R. (1974). *Astrophysical formulae: A compendium for the physicist and astrophysicist*. Springer-Verlag.

Lanza, A. F., Damiani, C., and Gandolfi, D. (2011). Constraining tidal dissipation in F-type main-sequence stars: the case of CoRoT-11. *A&A*, 529:A50.

Lanza, A. F., Das Chagas, M. L., and De Medeiros, J. R. (2014). Measuring stellar differential rotation with high-precision space-borne photometry. *A&A*, 564:A50.

Lanza, A. F. et al. (2010). Photospheric activity, rotation, and radial velocity variations of the planet-hosting star CoRoT-7. *A&A*, 520:A53.

Lanza, A. F., Rodonò, M., Mazzola, L., and Messina, S. (2001). Long-term starspot evolution, activity cycle and orbital period variation of SZ Piscium. *A&A*, 376:1011–1030.

Latham, D. W., Stefanik, R. P., Mazeh, T., Mayor, M., and Burki, G. (1989a). The unseen companion of HD 114762 - A probable brown dwarf. *Nature*, 339:38–40.

Latham, D. W., Stefanik, R. P., Mazeh, T., and Torres, G. (1989b). Spectroscopic Searches for Low-Mass Companions of Stars. In *Bulletin of the American Astronomical Society*, volume 21 of *BAAS*, page 1224.

Lázaro-Gredilla, M., Quiñonero Candela, J., Rasmussen, C. E., and Figueiras-Vidal, A. R. (2010). Sparse Spectrum Gaussian Process Regression. *J. Mach. Learn. Res.*, 11:1865–1881.

Lee, B.-C., Han, I., Park, M.-G., Mkrtichian, D. E., Hatzes, A. P., and Kim, K.-M. (2014). Planetary companions in K giants $\beta$ Cancri, $\mu$ Leonis, and $\beta$ Ursae Minoris. *A&A*, 566:A67.

Leighton, R. B., Noyes, R. W., and Simon, G. W. (1962). Velocity Fields in the Solar Atmosphere. I. Preliminary Report. *ApJ*, 135:474.

Lin, D. N. C., Bodenheimer, P., and Richardson, D. C. (1996). Orbital migration of the planetary companion of 51 Pegasi to its present location. *Nature*, 380:606–607.

Lindegren, L. and Dravins, D. (2003). The fundamental definition of "radial velocity". *A&A*, 401:1185–1201.

Linsky, J. L. and Avrett, E. H. (1970). The Solar H and K Lines. *PASP*, 82:169.

Littlefair, S. P., Burningham, B., and Helling, C. (2017). Robust detection of quasi-periodic variability: a HAWK-I mini survey of late-T dwarfs. *MNRAS*, 466:4250–4258.

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Ap&SS*, 39:447–462.

Lovis, C. et al. (2011). 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.

Lovis, C. and Fischer, D. (2011). *Radial Velocity Techniques for Exoplanets*, chapter 3, pages 27–53. The University of Arizona Press.

Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W.-L., and Nichols, T. E. (2006). Non-white noise in fMRI: Does modelling have an impact? *NeuroImage*, 29(1):54 – 66.

MacKay, D. J. (1998). Introduction to Gaussian Processes. *NATO ASI Series F Computer and Systems Sciences*, 168:133–166.

Maldonado, J., Martínez-Arnáiz, R. M., Eiroa, C., Montes, D., and Montesinos, B. (2010). A spectroscopy study of nearby late-type stars, possible members of stellar kinematic groups. *A&A*, 521:A12.

Mamajek, E. (2016). Cumulative Number of Exoplanets Discoveries Versus Time. `https://figshare.com/articles/Cumulative_Number_of_Exoplanets_Discoveries_Versus_Time/4057704`.

Mamajek, E. E. and Hillenbrand, L. A. (2008). Improved Age Estimation for Solar-Type Dwarfs Using Activity-Rotation Diagnostics. *ApJ*, 687:1264–1293.

Marcy, G. W. and Butler, R. P. (1992). Precision radial velocities with an iodine absorption cell. *PASP*, 104:270–277.

Marcy, G. W. et al. (2014). Masses, Radii, and Orbits of Small Kepler Planets: The Transition from Gaseous to Rocky Planets. *ApJS*, 210:20.

Marsh, T. R. and Horne, K. (1988). Images of accretion discs. II - Doppler tomography. *MNRAS*, 235:269–286.

Mathur, S. et al. (2011). Granulation in Red Giants: Observations by the Kepler Mission and Three-dimensional Convection Simulations. *ApJ*, 741:119.

Mayor, M. et al. (2003). Setting New Standards with HARPS. *The Messenger*, 114:20–24.

Mayor, M. and Queloz, D. (1995). A Jupiter-mass companion to a Solar-type star. *Nature*, 378:355–359.

McQuillan, A., Aigrain, S., and Mazeh, T. (2013a). Measuring the rotation period distribution of field M dwarfs with Kepler. *MNRAS*, 432:1203–1216.

McQuillan, A., Mazeh, T., and Aigrain, S. (2013b). Stellar Rotation Periods of the Kepler Objects of Interest: A Dearth of Close-in Planets around Fast Rotators. *ApJ*, 775:L11.

McQuillan, A., Mazeh, T., and Aigrain, S. (2014). Rotation Periods of 34,030 Kepler Main-sequence Stars: The Full Autocorrelation Sample. *ApJS*, 211:24.

Meunier, N., Desort, M., and Lagrange, A.-M. (2010). Using the Sun to estimate Earth-like planets detection capabilities. II. Impact of plages. *A&A*, 512:A39.

Meunier, N. and Lagrange, A. (2013). Using the Sun to study the impact of stellar activity on exoplanet detectability. *Astronomische Nachrichten*, 334(1-2):141–144.

Michel, E. et al. (2008). CoRoT Measures Solar-Like Oscillations and Granulation in Stars Hotter Than the Sun. *Science*, 322:558–.

Middelkoop, F. (1982). Magnetic structure in cool stars. IV - Rotation and CA II H and K emission of main-sequence stars. *A&A*, 107:31–35.

Morales, J. M. (1999). Viability in a pink environment: why "white noise" models can be dangerous. *Ecology Letters*, 2(4):228–232.

Mordasini, C., Alibert, Y., Benz, W., Klahr, H., and Henning, T. (2012). Extrasolar planet population synthesis. IV. Correlations with disk metallicity, mass, and lifetime. *A&A*, 541:A97.

Mordasini, C., Alibert, Y., Benz, W., and Naef, D. (2009). Extrasolar planet population synthesis. II. Statistical comparison with observations. *A&A*, 501:1161–1184.

Mortier, A., Faria, J. P., Santos, N. C., Rajpaul, V., et al. (2016). The HARPS search for southern extra-solar planets. XXXIX. HD 175607, the most metal-poor G dwarf with an orbiting sub-Neptune. *A&A*, 585:A135.

Mortier, A., Santos, N. C., Sousa, S., Israelian, G., Mayor, M., and Udry, S. (2013). On the functional form of the metallicity-giant planet correlation. *A&A*, 551:A112.

Mortier, A., Santos, N. C., Sozzetti, A., Mayor, M., Latham, D., Bonfils, X., and Udry, S. (2012). The frequency of giant planets around metal-poor stars. *A&A*, 543:A45.

Morton, T. D., Bryson, S. T., Coughlin, J. L., Rowe, J. F., Ravichandran, G., Petigura, E. A., Haas, M. R., and Batalha, N. M. (2016). False Positive Probabilities for all Kepler Objects of Interest: 1284 Newly Validated Planets and 428 Likely False Positives. *ApJ*, 822:86.

Motalebi, F. et al. (2015). The HARPS-N Rocky Planet Search. I. HD 219134 b: A transiting rocky planet in a multi-planet system at 6.5 pc from the Sun. *A&A*, 584:A72.

Moulton, F. R. (1899). The limits of temporary stability of satellite motion, with an application to the question of the existence of an unseen body in the binary system 70 Ophiuchi. *AJ*, 20:33–37.

Nardetto, N., Mourard, D., Kervella, P., Mathias, P., Mérand, A., and Bersier, D. (2006). High resolution spectroscopy for Cepheids distance determination. I. Line asymmetry. *A&A*, 453:309–319.

Nielsen, M. B., Gizon, L., Schunker, H., and Karoff, C. (2013). Rotation periods of 12 000 main-sequence Kepler stars: Dependence on stellar spectral type and comparison with $v \sin i$ observations. *A&A*, 557:L10.

Noyes, R. W., Hartmann, L. W., Baliunas, S. L., Duncan, D. K., and Vaughan, A. H. (1984a). Rotation, convection, and magnetic activity in lower main-sequence stars. *ApJ*, 279:763–777.

Noyes, R. W., Weiss, N. O., and Vaughan, A. H. (1984b). The relation between stellar rotation rate and activity cycle periods. *ApJ*, 287:769–773.

Osborne, M. A. (2010). *Bayesian Gaussian Processes for Sequential Prediction, Optimisation and Quadrature*. PhD thesis, University of Oxford.

O'Toole, S. J., Tinney, C. G., and Jones, H. R. A. (2008). The impact of stellar oscillations on doppler velocity planet searches. *MNRAS*, 386:516–520.

Park, S. C., Park, M. K., and Kang, M. G. (2003). Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20:21–36.

Pasquini, L. et al. (2008). CODEX: the high-resolution visual spectrograph for the E-ELT. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7014 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 70141I–70141I.

Pepe, F., Ehrenreich, D., and Meyer, M. R. (2014). Instrumentation for the detection and characterization of exoplanets. *Nature*, 513:358–366.

Pepe, F. et al. (2002a). HARPS: ESO's coming planet searcher. Chasing exoplanets with the La Silla 3.6-m telescope. *The Messenger*, 110:9–14.

Pepe, F. et al. (2011a). Hunting for the lowest-mass exoplanets. In Sozzetti, A., Lattanzi, M. G., and Boss, A. P., editors, *IAU Symposium*, volume 276, pages 13–24.

Pepe, F. et al. (2013). An Earth-sized planet with an Earth-like density. *Nature*, 503:377–380.

Pepe, F., Lovis, C., Ségransan, D., Benz, W., Bouchy, F., Dumusque, X., Mayor, M., Queloz, D., Santos, N. C., and Udry, S. (2011b). The HARPS search for Earth-like planets in the habitable zone. I. Very low-mass planets around HD 20794, HD 85512, and HD 192310. *A&A*, 534:A58.

Pepe, F., Mayor, M., Galland, F., Naef, D., Queloz, D., Santos, N. C., Udry, S., and Burnet, M. (2002b). The CORALIE survey for southern extra-solar planets VII. Two short-period Saturnian companions to HD 108147 and HD 168746. *A&A*, 388:632–638.

Pepe, F. A. et al. (2010). ESPRESSO: the Echelle spectrograph for rocky exoplanets and stable spectroscopic observations. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7735.

Perryman, M. (2011). *The Exoplanet Handbook*. Cambridge University Press.

Pollack, J. B., Hubickyj, O., Bodenheimer, P., Lissauer, J. J., Podolak, M., and Greenzweig, Y. (1996). Formation of the Giant Planets by Concurrent Accretion of Solids and Gas. *Icarus*, 124:62–85.

Pourbaix, D. et al. (2002). Constraining the difference in convective blueshift between the components of alpha Centauri with precise radial velocities. *A&A*, 386:280–285.

Press, W. H. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press.

Queloz, D. (1995). Echelle Spectroscopy with a CCD at Low Signal-To-Noise Ratio. In

Philip, A. G. D., Janes, K., and Upgren, A. R., editors, *New Developments in Array Technology and Applications*, volume 167 of *IAU Symposium*, page 221.

Queloz, D., Bouchy, F., Moutou, C., Hatzes, A., Hébrard, G., Alonso, R., Auvergne, M., Baglin, A., Barbieri, M., Barge, P., and et al. (2009). The CoRoT-7 planetary system: two orbiting super-Earths. *A&A*, 506:303–319.

Queloz, D., Henry, G. W., Sivan, J. P., Baliunas, S. L., Beuzit, J. L., Donahue, R. A., Mayor, M., Naef, D., Perrier, C., and Udry, S. (2001). No planet for HD 166435. *A&A*, 379:279–287.

Quirrenbach, A. et al. (2014). CARMENES instrument overview. In *Ground-based and Airborne Instrumentation for Astronomy V*, volume 9147 of *Proc. SPIE*, page 91471F.

Raftery, A. E. (1999). Bayes Factors and BIC: Comment on Weakliem. Technical report, University of Washington.

Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., and Roberts, S. (2015). A Gaussian process framework for modelling stellar activity signals in radial velocity data. *MNRAS*, 452(3):2269–2291.

Rajpaul, V., Aigrain, S., and Roberts, S. (2016). Ghost in the time series: no planet for Alpha Cen B. *MNRAS*, 456:L6–L10.

Rajpaul, V., Buchhave, L. A., and Aigrain, S. (2017). Pinning down the mass of Kepler-10c: the importance of sampling and model comparison. *MNRAS*, 471:L125–L130.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.

Reece, S., Ghosh, S., Rogers, A., Roberts, S., and Jennings, N. R. (2014). Efficient State-space Inference of Periodic Latent Force Models. *J. Mach. Learn. Res.*, 15(1):2337–2397.

Reece, S. and Roberts, S. (2010). An introduction to Gaussian processes for the Kalman filter expert. In *Proceedings of the 13th International Conference on Information Fusion*, pages 1–9. IEEE.

Rieutord, M. and Rincon, F. (2010). The Sun's Supergranulation. *Living Reviews in Solar Physics*, 7:2.

Roberts, S., Osborne, M., Ebden, M., Reece, S., Gibson, N., and Aigrain, S. (2013). Gaussian processes for time-series modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1984):20110550.

Robertson, P. and Mahadevan, S. (2014). Disentangling Planets and Stellar Activity for Gliese 667C. *ApJ*, 793(2):L24.

Rousseeuw, P. J. and Leroy, A. M. (2005). *Robust regression and outlier detection*, volume 589. John Wiley & Sons.

Rowe, J. F. (2014). Validation of Kepler's Multiple Planet Candidates. III. Light Curve Analysis and Announcement of Hundreds of New Multi-planet Systems. *ApJ*, 784:45.

Russell, B. (1959). Non-demonstrative inference. *The Centennial Review of Arts & Science*, 3(3):237–257.

Russell, H. N. (1906). On the light variations of asteroids and satellites. *ApJ*, 24:1–18.

Sahlmann, J., Lazorenko, P. F., Ségransan, D., Martín, E. L., Queloz, D., Mayor, M., and Udry, S. (2013). Astrometric orbit of a low-mass companion to an ultracool dwarf. *A&A*, 556:A133.

Sahu, K. C. et al. (2006). Transiting extrasolar planetary candidates in the Galactic bulge. *Nature*, 443:534–540.

Santerne, A., Díaz, R. F., Almenara, J.-M., Bouchy, F., Deleuil, M., Figueira, P., Hébrard, G., Moutou, C., Rodionov, S., and Santos, N. C. (2015). PASTIS: Bayesian extrasolar planet validation - II. Constraining exoplanet blend scenarios using spectroscopic diagnoses. *MNRAS*, 451:2337–2351.

Santos, N. C. et al. (2004). The HARPS survey for southern extra-solar planets. II. A 14 Earth-masses exoplanet around $\mu$ Arae. *A&A*, 426:L19–L23.

Santos, N. C. et al. (2014). The HARPS search for southern extra-solar planets. XXXV. The interesting case of HD 41248: stellar activity, no planets? *A&A*, 566:A35.

Santos, N. C., Israelian, G., and Mayor, M. (2001). The metal-rich nature of stars with planets. *A&A*, 373:1019–1031.

Santos, N. C., Mayor, M., Naef, D., Pepe, F., Queloz, D., Udry, S., Burnet, M., Clausen, J. V., Helt, B. E., Olsen, E. H., and Pritchard, J. D. (2002). The CORALIE survey for southern extra-solar planets. IX. A 1.3-day period brown dwarf disguised as a planet. *A&A*, 392:215–229.

Sarkka, S., Solin, A., and Hartikainen, J. (2013). Spatio-temporal Learning via Infinite-Dimensional Bayesian Filtering and Smoothing: A Look at Gaussian Process Regression Through Kalman Filtering. *Signal Processing Magazine, IEEE*, 30(4):51–61.

Sato, B. et al. (2012). Substellar Companions to Seven Evolved Intermediate-Mass Stars. *PASJ*, 64:135.

Scargle, J. D. (1982). Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *ApJ*, 263:835–853.

Scargle, J. D. (1989). Studies in astronomical time series analysis. III - Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data. *ApJ*, 343:874–887.

Schmitt, D. (1987). An alpha-omega-dynamo with an alpha-effect due to magnetostrophic waves. *A&A*, 174:281–287.

Schrijver, C. J. (2002). Solar spots as prototypes for stellar spots. *Astronomische Nachrichten*, 323:157–164.

Schrijver, C. J. and Zwaan, C. (2000). *Solar and Stellar Magnetic Activity*. Cambridge University Press.

Schuster, A. (1898). On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism (Journal of Geophysical Research)*, 3:13.

Schwamb, M. E. et al. (2013). Planet Hunters: A Transiting Circumbinary Planet in a Quadruple Star System. *ApJ*, 768:127.

Seager, S., editor (2010). *Exoplanets (Space Science Series)*. University of Arizona Press.

Seager, S. and Deming, D. (2010). Exoplanet Atmospheres. *ARA&A*, 48:631–672.

See, T. J. J. (1895). Perturbations in the motion of the double star 70 Ophiuchi = T, 2272. *AJ*, 15:180–180.

Ségransan, D. et al. (2011). The HARPS search for southern extra-solar planets. XXIX. Four new planets in orbit around the moderatly active dwarfs HD 63765, HD 104067, HD 125595, and HIP 70849. *A&A*, 535:A54.

Setiawan, J., Hatzes, A. P., von der Lühe, O., Pasquini, L., Naef, D., da Silva, L., Udry, S., Queloz, D., and Girardi, L. (2003). Evidence of a sub-stellar companion around HD 47536. *A&A*, 398:L19–L23.

Shah, A., Wilson, A., and Ghahramani, Z. (2014). Student-$t$ Processes as Alternatives to Gaussian Processes. In Kaski, S. and Corander, J., editors, *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pages 877–885, Reykjavik, Iceland. PMLR.

Sivia, D. and Skilling, J. (2006). *Data Analysis: A Bayesian Tutorial*. Oxford University Press, 2 edition.

Skumanich, A. (1972). Time Scales for Ca II Emission Decay, Rotational Braking, and Lithium Depletion. *ApJ*, 171:565.

Smith, A. T., Singh, K. D., and Balsters, J. H. (2007). A comment on the severity of the effects of non-white noise in fMRI time-series. *NeuroImage*, 36(2):282 – 288.

Snellen, I. A. G., de Kok, R. J., de Mooij, E. J. W., and Albrecht, S. (2010). The orbital motion, absolute mass and high-altitude winds of exoplanet HD 209458b. *Nature*, 465:1049–1051.

Snelson, E. L. (2008). *Flexible and efficient Gaussian process models for machine learning*. PhD thesis, University College London.

Soter, S. (2006). What Is A Planet? In *AAS/Division for Planetary Sciences Meeting Abstracts #38*, volume 38 of *Bulletin of the American Astronomical Society*, page 1295.

Sousa, S. G., Santos, N. C., Israelian, G., Lovis, C., Mayor, M., Silva, P. B., and Udry, S. (2011a). Spectroscopic characterization of a sample of metal-poor Solar-type stars from the HARPS planet search program. Precise spectroscopic parameters and mass estimation. *A&A*, 526:A99.

Sousa, S. G., Santos, N. C., Israelian, G., Mayor, M., and Udry, S. (2011b). Spectroscopic stellar parameters for 582 FGK stars in the HARPS volume-limited sample. Revising the metallicity-planet correlation. *A&A*, 533:A141.

Stegle, O., Fallert, S., MacKay, D., and Brage, S. (2008). Gaussian Process Robust Regression for Noisy Heart Rate Data. *IEEE Transactions on Biomedical Engineering*, 55(9):2143 – 2151.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, New York.

Strand, K. A. (1943). 61 Cygni as a Triple System. *PASP*, 55:29–32.

Strassmeier, K. G. (2009). Starspots. *A&A Rev.*, 17:251–308.

Struve, O. (1952). Proposal for a project of high-precision stellar radial velocity work. *The Observatory*, 72:199–200.

Sulis, S., Mary, D., and Bigot, L. (2017). A Study of Periodograms Standardized Using Training Datasets and Application to Exoplanet Detection. *IEEE Transactions on Signal Processing*, 65:2136–2150.

Sumi, T. et al. (2011). Unbound or distant planetary mass population detected by gravitational microlensing. *Nature*, 473:349–352.

Tassoul, J.-L. (2000). *Stellar Rotation*. Cambridge University Press.

Thompson, S. J. et al. (2016). HARPS3 for a roboticized Isaac Newton Telescope. In *Ground-based and Airborne Instrumentation for Astronomy VI*, volume 9908 of *Proc. SPIE*, page 99086F.

Tipping, M. E., Faul, A., Avenue, J. J. T., and Avenue, J. J. T. (2003). Fast Marginal Likelihood Maximisation for Sparse Bayesian Models. In *Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics*, pages 3–6.

Tobias, S. M. (2002). The Solar dynamo. *Philosophical Transactions of the Royal Society of London Series A*, 360:2741–2756.

Todorov, K., Luhman, K. L., and McLeod, K. K. (2010). Discovery of a Planetary-mass Companion to a Brown Dwarf in Taurus. *ApJ*, 714:L84–L88.

Toner, C. G. and Gray, D. F. (1988). The starpatch on the G8 dwarf XI Bootis A. *ApJ*, 334:1008–1020.

Torres, G., Andersen, J., and Giménez, A. (2010). Accurate masses and radii of normal stars: modern results and applications. *A&A Rev.*, 18:67–126.

Traub, W. A. and Oppenheimer, B. R. (2010). *Direct Imaging of Exoplanets*, pages 111–156. University of Arizona Press.

Triaud, A. (2016). Exoplanets: Migration of giants. *Nature*, 537:496–497.

Tuomi, M., Anglada-Escude, G., Jenkins, J. S., and Jones, H. R. A. (2014). Filtering out activity-related variations from radial velocities in a search for low-mass planets. *ArXiv e-prints*, arXiv:1405.2016.

Tuomi, M. et al. (2013). Signals embedded in the radial velocity noise. Periodic variations in the $\tau$ Ceti velocities. *A&A*, 551:A79.

Udalski, A., Zebrun, K., Szymanski, M., Kubiak, M., Soszynski, I., Szewczyk, O., Wyrzykowski, L., and Pietrzynski, G. (2002). The Optical Gravitational Lensing Experiment. Search for Planetary and Low- Luminosity Object Transits in the Galactic Disk. Results of 2001 Campaign – Supplement. *Acta Astron.*, 52:115–128.

Udry, S. and Santos, N. C. (2007). Statistical Properties of Exoplanets. *ARA&A*, 45:397–439.

van de Kamp, P. (1963). Astrometric study of Barnard's star from plates taken with the 24-inch Sproul refractor. *AJ*, 68:515–521.

Van Eylen, V. and Albrecht, S. (2015). Eccentricity from Transit Photometry: Small Planets in Kepler Multi-planet Systems Have Low Eccentricities. *ApJ*, 808:126.

van Leeuwen, F. (2007). Validation of the new Hipparcos reduction. *A&A*, 474:653–664.

Vanderburg, A. et al. (2015). Characterizing K2 Planet Discoveries: A Super-Earth Transiting the Bright K Dwarf HIP 116454. *ApJ*, 800:59.

Vanhatalo, J., Jylänki, P., and Vehtari, A. (2009). Gaussian process regression with Student-$t$ likelihood. In Bengio, Y., Schuurmans, D., Lafferty, J. D., Williams, C. K. I., and Culotta, A., editors, *Advances in Neural Information Processing Systems 22*, pages 1910–1918. Curran Associates, Inc.

Vidal-Madjar, A., Lecavelier des Etangs, A., Désert, J.-M., Ballester, G. E., Ferlet, R., Hébrard, G., and Mayor, M. (2003). An extended upper atmosphere around the extrasolar planet HD 209458b. *Nature*, 422:143–146.

von Waltershausen, W. S. (1856). *Gauss zum Gedächtnis*. S. Hirzel, Leipzig.

Walker, G. A. H. (2012). The first high-precision radial velocity search for extra-solar planets. *New A Rev.*, 56:9–15.

Watson, F. T., Fletcher, L., and Marshall, S. (2011). Evolution of sunspot properties during Solar cycle 23. *A&A*, 533:A14.

Weiss, L. M. et al. (2016). Revised Masses and Densities of the Planets around Kepler-10. *ApJ*, 819:83.

Weiss, L. M. and Marcy, G. W. (2014). The Mass-Radius Relation for 65 Exoplanets Smaller than 4 Earth Radii. *ApJ*, 783:L6.

Wolszczan, A. (1994). Confirmation of Earth-Mass Planets Orbiting the Millisecond Pulsar PSR B1257+12. *Science*, 264:538–542.

Wolszczan, A. and Frail, D. A. (1992). A planetary system around the millisecond pulsar PSR1257+12. *Nature*, 355:145–147.

Zechmeister, M. and Kürster, M. (2009). The generalised Lomb-Scargle periodogram. A new formalism for the floating-mean and Keplerian periodograms. *A&A*, 496:577–584.

Zeng, L. and Sasselov, D. (2013). A Detailed Model Grid for Solid Planets from 0.1 through 100 Earth Masses. *PASP*, 125:227.

Zhou, G. et al. (2017). HAT-P-67b: An Extremely Low Density Saturn Transiting an F-subgiant Confirmed via Doppler Tomography. *AJ*, 153:211.