Inferring probabilistic stellar rotation periods using Gaussian processes

Ruth Angus[1], Timothy Morton[2], Suzanne Aigrain[3], Daniel Foreman-Mackey[5,4] & Vinesh Rajpaul[3]

ABSTRACT

The light curves of spotted, rotating stars often vary in a non-sinusoidal and quasi-periodic fashion—spots move on the stellar surface and have finite lifetimes, causing stellar flux variations to slowly shift in phase. A strictly periodic sinusoid therefore cannot accurately model a rotationally modulated stellar light curve. Physical models of the stellar surface also have many drawbacks preventing effective inference, such as highly degenerate or high-dimensional parameter spaces. In this work, we introduce an appropriate *effective* model: a Gaussian Process (GP) with a quasi-periodic covariance kernel function. Using this highly flexible model, we are able to sample from the posterior Probability Density Function (PDF) of the periodic parameter, marginalising over the other kernel hyperparameters using a Markov Chain Monte Carlo approach. To test the effectiveness of this method, we infer rotation periods from 333 simulated stellar light curves, demonstrating that our results are more accurate than both a sine-fitting periodogram and an AutoCorrelation Function (ACF) method. We also demonstrate that it works effectively on real data, by inferring rotation periods for 1102 *Kepler* stars. Because this method delivers posteriors PDFs it will enable hierarchical studies involving stellar rotation, particularly those involving population modelling, such as inferring stellar ages, obliquities in exoplanet systems, or characterising star-planet interactions.

---

[1]Simons Fellow, Department of Astronomy, Columbia University, NY, NY, RuthAngus@gmail.com

[2]Department of Astrophysical Sciences, Princeton University, Princeton, NJ

[3]Subdepartment of Astrophysics, University of Oxford, UK

[5]Department of Astronomy, University of Washington, Seattle, WA

[4]Sagan Fellow

1. Introduction

The brightness of a spotted, rotating star often varies in a non-sinusoidal and Quasi-Periodic (QP) manner, due to active regions on its surface which rotate in and out of view. Complicated surface spot patterns produce non-sinusoidal variations, and the finite lifetimes of these active regions and differential rotation on the stellar surface produce quasi-periodicity (Dumusque et al. 2012). A strictly periodic sinusoid is therefore not necessarily a good model choice for these time-series. A physically realistic model of the stellar surface would, ideally, perfectly capture the complexity of shapes within stellar light curves as well as the quasi-periodic nature, allowing for extremely precise probabilistic period recovery when conditioned on the data. However, such physical models require many free parameters in order to accurately represent a stellar surface, and some of these parameters are extremely degenerate (*e.g.* Russell 1906; Jeffers and Keller 2009; Kipping 2012). 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. Considering that stars may have hundreds of spots, the number of free parameters in such a model quickly becomes unwieldy, especially to explore its posterior Probability Density Function (PDF). Simplified spot models, such as the one described in Lanza et al. (2014) where only two spots are modelled, have produced successful results; however, such relatively inflexible models sacrifice precision.

Standard methods to measure rotation periods include detecting peaks in a Lomb-Scargle (Lomb 1976; Scargle 1982) (LS) periodogram (e.g. Reinhold et al. 2013), Auto-Correlation Functions (ACFs), (McQuillan et al. 2013b) and wavelet transforms (García et al. 2014). The precisions of the LS periodogram and wavelet methods are limited by the suitability of the model choice: a sinusoid for the LS periodogram, and a choice of mother wavelet, assumed to describe [VR: spelling] the data over a range of transpositions (see, *e.g.* Carter and Winn 2010), for the wavelet method. In contrast, the ACF method is much better suited to signals that are non-sinusoidal. In fact, as long as the signal is approximately periodic the ACF will display a peak at the rotation period, no matter its shape. A drawback of the ACF method, however, is that it requires data to be evenly-spaced[6], which is not exactly the case with *Kepler* light curves (although in many cases it can be approximated as uniformly sampled) and not nearly the case for most ground-based measurements, such as the future Large Synoptic Survey Telescope (*LSST*). An ACF is also an operation performed on the data rather than a generative model of the data, and so is not inherently probabilistic. This means that the effects of the observational uncertainties

[6]Edelson and Krolik (1988) describe a method for computing ACFs for unevenly-spaced data.

cannot be formally propagated to constraints on the rotation period.

In this work, we introduce an *effective* model for rotationally modulated stellar light curves which captures the salient behaviour but is not physically motivated—although some parameters may indeed be *interpreted* as physical ones. An ideal effective model should have a small number of non-degenerate parameters and be flexible enough to perfectly capture non-sinusoidal and QP behaviour. A Gaussian process (GP) model fulfills these requirements. We thus use a GP as the generative model at the core of a method to probabilistically infer accurate and precise stellar rotation periods. This enables us to estimate the posterior PDF of the rotation period, thereby producing a justified estimate of its uncertainty while simultaneously marginalizing over parameters which model correlated noise.

GPs are commonly used in the machine learning community and increasingly in other scientific fields such as biology, geophysics and cosmology. More recently, GPs have been used in the stellar and exoplanet fields within astronomy, to capture stellar variability or instrumental systematics (see *e.g.* Gibson et al. 2012; Haywood et al. 2014; Dawson et al. 2014; Barclay et al. 2015; Haywood 2015; Evans et al. 2015; Rajpaul et al. 2015, 2016; Aigrain et al. 2016). They are useful in regression problems involving any stochastic process, specifically when the probability distribution for the process is a multi-variate Gaussian. If the probability of obtaining a dataset is a Gaussian in N dimensions, where N is the number of data points, a GP can describe that dataset. An in-depth introduction to Gaussian processes is provided in Rasmussen and Williams (2005).

GP models parameterise the covariance between data points by means of a parametric kernel function that models the autocorrelation behaviour of the signal and defines the form of the covariance matrix. As a qualitative demonstration, we present the time-series in Figure 1: the *Kepler* light curve of KIC 5809890. This is a relatively active star that rotates once every $\sim$ 30.5 days, with stochastic variability typical of *Kepler* FGK stars. Clearly, data points in this light curve are correlated. Points close together in time are tightly correlated, and points more widely separated are loosely correlated. A GP models this variation in correlation with the separation between data points; that is, it models the *covariance structure* rather than the data directly. This lends GPs their flexibility—they can model any time series with a similar covariance structure. In addition, a very simple function can usually capture the covariance structure of a light curve, whereas modelling the time series itself might require much more complexity. Figure 1 demonstrates how a GP model fits the light curve of KIC 5809890.

A range of kernel functions can describe stellar variability. For example, the simplest and most commonly used kernel function [VR: it's true that it's the most commonly used kernel, but not true that it's the 'simplest' – a constant or linear kernel is simpler; an

exponential is as 'simple' in terms of number of hyper-parameters; etc. There are of course other motivations for using SE kernel; see e.g. Rasmussen & Williams pg. 14, although these details are probably not relevant to this paper. Maybe it's the simplest stationary kernel function that admits infinitely-differentiable function draws, although I stand to be corrected on that.], the 'Squared Exponential' (SE), defined as follows, could adequately fit the KIC 5809890 light curve:

$$k_{i,j} = A \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right). \tag{1}$$

Here $A > 0$ is the amplitude of covariance [VR: either say $A > 0$ or use A^2 in the definition of this kernel; need to ensure positive values so QP covariance matrix has nonnegative eigenvalues i.e. is positive semi-definite.], l is the length scale of exponential decay, and $x_i - x_j$ is the separation between data points. The SE kernel function has the advantage of being very simple, with just two parameters, A and l. If l is large, two data points far apart in x will be tightly correlated, and if small they will be loosely correlated. Another property of the SE kernel function is that it produces functions that are infinitely differentiable, making it possible to model a data set and its derivatives simultaneously. However, The SE kernel function does not well describe the covariance in stellar light curves, nor is it *useful* for the problem of rotation period inference because it does not capture periodic behaviour. Inferring rotation periods thus requires a periodic kernel function. For this reason, we use the 'Quasi-Periodic' kernel. Rasmussen and Williams (2005) model QP variability in CO_2 concentration on the summit of the Mauna Loa volcano in Hawaii (data from Keeling and Whorf 2004) using a kernel which is the product of a periodic and a SE kernel: the QP kernel. This kernel is defined as

$$k_{i,j} = A \exp\left[-\frac{(x_i - x_j)^2}{2l^2} - \Gamma^2 \sin^2\left(\frac{\pi(x_i - x_j)}{P}\right)\right]. \tag{2}$$

[VR: there was a term missing from the above definition, viz. $(x_i - x_j)$ inside the sin(). Without it, the sin() term reduces to a constant. Also, I changed 2π to π: if you define it with 2π then the period will differ from the physical period by a factor of 2, given the squared term modifying the sine wave: $\sin^2(\frac{\theta}{2}) = \frac{1}{2} - \frac{1}{2}\cos\theta$. This can be checked by plotting some covariance kernels with large l and noting that peaks appear with expecting spacing P only if you remove the factor of 2 that originally appeared in the definition.] It is the product of the SE kernel function, which describes the overall covariance decay, and an exponentiated, squared, sinusoidal kernel function that describes the periodic covariance structure. P can be interpreted as the rotation period of the star, and Γ controls the amplitude of the $\sin^2$ term. If Γ is very large, only points almost exactly one period away are tightly correlated and points that are slightly more or less than one period away are very loosely correlated. If Γ is small, points separated by one period are tightly correlated, and points separated by

slightly more or less are still highly correlated, although less so. [VR: behaviour for large vs. small values of Γ was described backwards; i.e. correct explanation is that large Γ leads to only points ∼ one period apart being tightly correlated, with weak correlations for other separations. Another way of expressing this would be to say that large values of Γ lead to periodic variations with increasingly complex harmonic content. Or in non-technical terms: more bumps and wiggles per function repetition.] This kernel function allows two data points that are separated in time by one rotation period to be tightly correlated, while also allowing points separated by half a period to be weakly correlated. We also use an extra parameter σ which is an additional white noise term added to the diagonal elements of the covariance matrix. [VR: I think it might be clearer if you just included this white noise term directly in the definition of the kernel, especially for people who might not be familiar with GPs and for whom it won't necessarily be obvious how the white noise term connects to the rest of the kernel. I.e. I'd suggest you simply defined the kernel as below.]

$$k_{i,j} = A \exp\left[-\frac{(x_i - x_j)^2}{2l^2} - \Gamma^2 \sin^2\left(\frac{\pi(x_i - x_j)}{P}\right)\right] + \sigma^2 \delta_{ij}. \quad (3)$$

This can be interpreted to represent underestimation of observational uncertainties — if the uncertainties reported on the data are too small, it will be non-zero — or it can capture any remaining "jitter" or residuals that are not captured by the effective GP model. We use this QP kernel function to produce the GP model that fits the *Kepler* light curve in in Figure 1.

To infer a stellar rotation period P from a light curve, we fit this QP-kernel GP model to the data. As with any model-fitting exercise, the likelihood of the model could be maximized to find the maximum-likelihood value for P. In this study, however, we explore the full posterior PDFs using a Markov Chain Monte Carlo (MCMC) procedure. While this approach comes at a computational cost, such posterior exploration importantly provides a justified uncertainty estimate.

This paper is laid out as follows. In §3.1 we describe the simulated light curves used to test the three rotation period recovery methods (ACF, LS periodogram and GP). The GP method is described in §2 and the ACF and LS periodogram methods are described in §3.3. In §4 we apply our method to real *Kepler* data, and the results are discussed in §5. [VR: the order doesn't make sense. As things stand, you've laid it out as §3.1,§2, then §3.3. Presumably should be §2,§3.1, §3.3 – and §3.2 is missing].

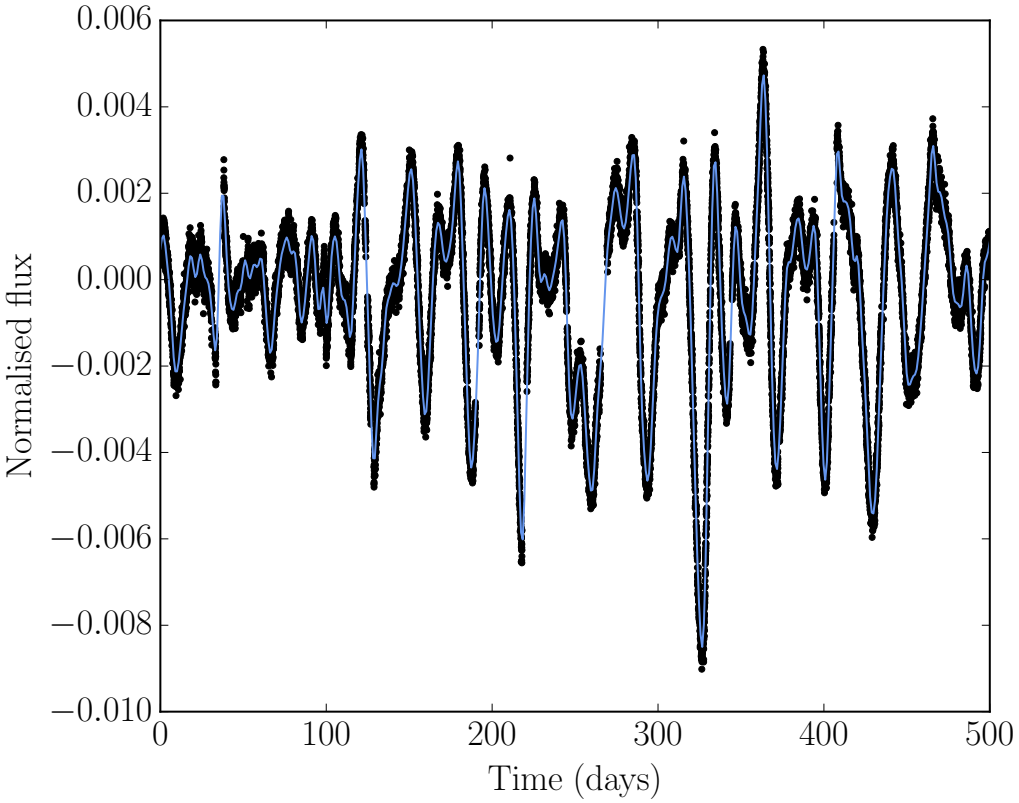

Fig. 1.— Light curve of KIC 5809890, an active star with a rotation period of ~ 30.5 days. The blue line shows a fit to the data using a Gaussian process model with a QP covariance kernel function.

2. GP Rotation Period Inference

In order to recover a stellar rotation period from a light curve using a quasi-periodic Gaussian process (QP-GP), we sample the following posterior PDF:

$$p(\boldsymbol{\theta} \,|\, y) \propto \mathcal{L}(y \,|\, \boldsymbol{\theta}) p(\boldsymbol{\theta}), \tag{4}$$

where y are the light curve flux data, $\boldsymbol{\theta}$ are the hyperparameters of the kernel described in Equation 2, $\mathcal{L}$ is the QP-GP likelihood function, and $p(\boldsymbol{\theta})$ is the prior on the hyperparameters. Sampling this posterior presents several challenges:

- The likelihood evaluation [VR: spelling] is computationally expensive;
- The GP model is very flexible, sometimes at the expense of reliable recovery of the period parameter; and
- The posterior may often be multimodal.

This Section discusses how we address these challenges through the implementation details of the likelihood (Section 2.1), priors (Section 2.2), and sampling method (Section 2.3).

2.1. Likelihood

The GP likelihood is similar to the simple Gaussian likelihood function used for optimisation problems where the uncertainties are Gaussian and uncorrelated. The latter can be written

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{n=1}^{N} \frac{(y_n - \mu)^2}{\sigma_n^2} - \frac{N}{2} \ln(2\pi\sigma_n^2), \tag{5}$$

where y_n are the data, μ is the mean model and σ_n are the Gaussian uncertainties [VR: spelling] on the data. VR: the above expression is incorrect. It should read:

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{n=1}^{N} \frac{(y_n - \mu)^2}{\sigma_n^2} - \frac{1}{2} \sum_{n=1}^{N} \ln(2\pi\sigma_n^2), \tag{6}$$

or equivalently

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{n=1}^{N} \left\{ \frac{(y_n - \mu)^2}{\sigma_n^2} + \ln(2\pi\sigma_n^2) \right\}. \tag{7}$$

The problem with the expression as written above is that either the second term must be included in the summation, assuming $\sigma_n \neq \sigma$, in which case the factor of N should not

appear in the second term; or if $\sigma_n = \sigma \; \forall n$ (which now no longer represents the general case), then the factor of N is correct, but the subscript n should be removed from σ_n^2.

The equivalent equation in matrix notation is

$$\ln \mathcal{L} = -\frac{1}{2}\mathbf{r^T C^{-1} r} - \ln |\mathbf{C}| + \text{constant}, \tag{8}$$

(VR: why not just write $\frac{N}{2} \log 2\pi$ instead of 'constant'?) where $\mathbf{r}$ is the vector of residuals and $\mathbf{C}$ is the covariance matrix,

$$\mathbf{C} = \begin{pmatrix} \sigma_1^2 & \sigma_{2,1} & \cdots & \sigma_{N,1} \\ \sigma_{1,2} & \sigma_2^2 & \cdots & \sigma_{N,2} \\ & & \vdots & \\ \sigma_{1,N} & \sigma_{2,N} & \cdots & \sigma_N^2 \end{pmatrix} \tag{9}$$

In the case where the uncertainties are uncorrelated, the noise is 'white', (which is a frequent assumption made by astronomers and is sometimes justified) and the off-diagonal elements of the covariance matrix are zero. However, in the case where there is evidence for correlated 'noise'[7], as in the case of *Kepler* light curves, those off-diagonal elements are non-zero. With GP regression, a covariance matrix generated by the kernel function $\mathbf{K}$ replaces $\mathbf{C}$ in the above equation (for our purposes, the QP kernel of Equation 2).

Evaluating this likelihood [VR: spelling] for a large number of points can be computationally expensive. For example, evaluating $\mathcal{L}$ for an entire *Kepler* lightcurve (∼40,000 points) takes about ∼5 s— too slow to perform inference on large numbers of light curves[8]. The matrix operations necessary to evaluate the GP likelihood[9] scale as $N \ln(N)^2$, where N is the number of data points in the light curve.

We accelerate the likelihood calculation using two complementary strategies: subsampling the data and splitting the light curve into independent sections. To subsample *Kepler* data, for example, we randomly select 1/30th of the points in the full light curve (an average

[7]In our case the 'noise' is actually the model! Incidentally, this approach is the reverse of the regression techniques usually employed by astronomers. In most problems in astronomy one tries to infer the parameters that describe the mean model and, if correlated noise is present, to marginalise over that noise. Here, the parameters describing the correlated noise are what we are interested in and our mean model is simply a straight line at $y = 0$.

[8]All computational times cited in this section are based on evaluations on a single core of a 2015 Macbook Pro, 3.1 GHz Intel Core i7.

[9]These operations use the fast matrix solver HODLR (Ambikasaran et al. 2014), implemented in the `george` (Foreman-Mackey et al. 2014) python package.

of ∼1.5 points per day). This decreases the likelihood evaluation time by a factor of about 50, down to about 100 ms. We then split the light curve into equal-sized chunks containing approximately 300 points per section (corresponding to about 200 days), and evaluate the log-likelihood as the sum of the log-likelihoods of the individual sections (all using the same parameters $\boldsymbol{\theta}$). This reduces computation time because the section-based likelihood evaluation scales as $mn\ln(n)^2$, where n is the number of data points per section and m is the number of sections. This method further reduces computation time for a typical light curve (subsampled by a factor of 30) by about a factor of two, to about 50 ms.

2.2. Priors

2.2.1. Non-period Hyperparameters

The flexibility of this GP model allows for posterior multimodality and "over-fitting"–like behavior. For example, if l is small, the non-periodic factor in the covariance kernel may dominate, allowing for a good fit to the data without requiring any periodic covariance structure—even if clear periodic structure is present. Additionally, if Γ is too large, the GP model becomes extremely flexible and can fit the data without varying the period.

Managing this flexibility to focus on reliably retrieving the period parameter requires imposing priors on the non-period GP parameters. While there are no strictly *a priori* reasons to believe that any of the hyperparameters should take values in any specific range, we have found in this work that reasonable priors for any particular dataset can be determined by experimentation. We note this is a somewhat subjective art, with a goal of not allowing the GP to become so flexible that it does not require periodicity to fit the data even in the presence of clear periodicity. In particular, we find it necessary to avoid large values of A and Γ, and small values of l; though the exact details of what works well may differ among datasets (see **??** and **??** for two different examples of hyperparameter priors).

VR: in my opinion, this is the weakest part of the paper. I don't agree that there are no *a priori* reasons to believe that any of the hyperparameters should take values in any specific range, and the phrase 'somewhat subjective art' seems likely to set alarm bells ringing for referees/readers. Going through the hyperparameters in turn, the following things come to mind.

- Amplitude, A: their is a closed form solution for the maximum-likelihood value of A (solve $\partial_A \mathcal{L} = 0$). Steve Roberts alluded to this in his 2012 tutorial paper and I know I've worked through the mathematics myself – offhand I can't remember the

solution, but it's a fairly straightforward looking thing involving the precision matrix like $A \sim \mathbf{y}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{y}$ with some scaling constant. I'll see if I can dig up my workings and/or find a reference. I think this means that if you're anyway going to use a flat prior on A, your MAP solution will end up corresponding to the ML solution. In which case you could speed things up a good deal by just setting A to its optimal value *ab initio*, and not bothering to explore the full posterior PDF. More importantly though, if for some reason the MAP value you use doesn't correspond more or less to the ML value (e.g. if something went wrong with the MCMC convergence) – in other words, if you end up using a non-optimal value for A – I've found you can sometimes get rather bizarre behaviour. For example, when I fit data with an obvious period, sometimes I end up recovering a very wrong period if A is also very wrong e.g. much too large. Such problems can be avoided quite straightforwardly by constraining A to sensible values.

- Harmonic complexity, Γ: for large values of l ($l \gg P$), Γ can be related to the typical number of stationary/non-stationary points of inflection, per period, found in functions drawn from the corresponding GP. This behaviour will be independent of the exact value of l. For 'intermediate' values of l ($l \gtrsim P$) this can still be done but the behaviour will be sensitive to the value of l. And small values of l e.g. $l < P$ are probably not relevant – see below. If useful I could make a plot or two illustrating this. Anyway, the point is that depending on the processes giving rise to observed stellar signals, you may not expect to require certain types of functions to model observed signals. In typical simulations e.g. of arbitrary numbers of spots on rotating stars, I think you're unlikely to end up with more than say three or four turning points per period in the disk-integrated photometric signal (as opposed to say 20; of course 2 turning points is the minimum for periodic functions). Such simulations, crude though they may be, can be used to come up with physically-motivated priors for Γ.

- Evolution time-scale, l: it's sensible to enforce something like $l > P$ if we desire functions that have at least some fairly clear periodicity, as intuitively $l < P$ would mean the functions are allowed to evolve significantly over time scales shorter than one period. So a function $f(t + P)$ would look signficantly different to $f(t)$, in which case any claims of periodicity become dubious. Actually, you can come with a more rigorous criterion (which also takes into account dependencies on Γ) than this – see below.

- Suzanne and I have shown that provided

$$\frac{P^2}{l^2\Gamma^2} < \frac{4\pi}{3}, \tag{10}$$

the QP covariance function will have at least one inflection point other than at $|x_i - x_j| = 0$ (the derivation is straightforward though a bit long). So $\frac{P^2}{l^2\Gamma^2} = \frac{4\pi}{3}$ corresponds to a critical point at which secondary maxima disappear, and the behaviour of the kernel becomes dominated by the SE term rather than the periodic term. For $\Gamma \sim \mathcal{O}(1)$, this translates into $l \gtrsim P/2$. In practice satisfying this criterion doesn't absolutely guarantee that function draws will have clear evidence of periodicity, but it does give you a quantitative boundary between obviously periodic behaviour and not-so-obviously periodic behaviour (if you're trying to fit rotation periods then obviously you don't want to use an SE-like kernel to do so) and provides some justification for a more stringent criterion such as $l > P$, or $l > 2P$ (ensure that functions evolve significantly over time-scales longer than 2 periods), etc.

The upshot of all of this is that I don't think constraining hyperparameters needs to be some sort of dark art or ad hockery, and it'd do the underlying method a disservice by suggesting as much. As I see it, there are at least two ways to proceed here.

1. Provide as much detail on constraining the hyperparameters in this paper as we possibly can, including quantitative justification, illustrations of qualitatively-different behaviours for different regions of the hyperparameter space, etc. Disadvantages: if we do this it'd be necessary to re-run your fits with the new priors; and the paper would be come a good deal longer, with the focus possibly shifted away from the essential idea of using a GP to infer rotation periods.

2. Alternatively, at least allude to some of these approaches to constraining hyperparameters (instead of saying that it's a subjective art) but defer a focused discussion on how to do that to a short, separate paper/research note. As I've already done much of the work for such a study, I'd be happy to take the lead on that at some point later this year. Advantage here would be your paper remains more focused and you don't need to re-run any of your fits. Indeed I imagine that if you check your MAP values for A, l, etc., they'd usually be consistent with any rigorously-motivated priors (since you get generally sensible results). Perhaps we could show in the follow-up paper how using the physically-motivated priors could lead to improved period inference (at least I hope that would be the case!). Any differential rotation tests could also be shifted to the other paper – see my later remarks on this.

2.2.2. Period

While using an uninformed prior on period (e.g., log-flat) often suffices, we also develop a method to construct an informed period prior, based on the autocorrelation function (ACF). The ACF has proven to be very useful for measuring stellar rotation periods (McQuillan et al. 2012, 2013b, 2014); however, the method has several shortcomings, most notably the inability to deliver uncertainties, but also the necessity of several heuristic choices, such as a timescale on which to smooth the ACF, how to define a peak, whether the first or second peak gets selected, and what constitutes a secure detection. While this paper presents a rotation period inference method that avoids these shortcomings, it seems prudent to still use information available from the ACF. We thus use the ACF to define a prior on period, which can help the posterior sampling converge on the true period.

We do not attempt to decide which single peak in the ACF best represents the true rotation period, but rather we identify several *candidate* periods and define a weighting scheme in order to create a noncommital, though useful, multimodal prior. While this procedure does not avoid heuristic choices, the fact that its goal is simply to create a *prior* for a probabilistically [VR: spelling] justified inference scheme rather than to identify a single correct period softens the potential impact of these choices.

As another innovation beyond what ACF methods in the literature have presented, we also bandpass filter the light curves (using a 5th order Butterworth filter, as implemented in `SciPy`) before calculating the autocorrelation. This suppresses power on timescales shorter than a chosen minimum period $P_{\rm min}$ and longer than a chosen maximum $P_{\rm max}$, producing a cleaner autocorrelation signal than an unfiltered light curve.

We use the following procedure to construct a prior for rotation period given a light curve:

1. For each value of P_i, where $i = \{1, 3, 5, 10, 30, 50, 100\}$ d, we apply a bandpass filter to the light curve using $P_{\rm min} = 0.1$ d and $P_{\rm max} = P_i$. We then calculate the ACF of the filtered light curve out to a maximum lag of $2P_i$ and smooth it with a boxcar filter of width $P_i/10$.

2. We identify the time lag corresponding to the first peak of each of these ACFs, as well as the first peak's trough-to-peak height, creating a set of candidate periods T_i and heights h_i.

3. We assign a quality metric Q_i to each of these candidate periods, as follows. First, we

model the ACF as a damped oscillator with fixed period T_i:

$$y = Ae^{-t/\tau} \cos \frac{2\pi t}{T_i}, \tag{11}$$

where and t is the lag time, and find the best-fitting parameters A_i and τ_i by a non-linear least squares minimization procedure. We then define the following heuristic quality metric:

$$Q_i = \left(\frac{\tau_i}{T_i}\right)\left(\frac{N_i h_i}{R_i}\right), \tag{12}$$

where h_i is the height of the ACF peak at T_i, N_i is the length of the lag vector in the ACF (directly proportional to the maximum allowed period P_i), and R_i is the sum of squared residuals between the damped oscillator model and the actual ACF data. The idea behind this quality metric to give a candidate period a higher score if

(a) it has many regular sinusoidal peaks, such that the decay time τ_i is long compared to the oscillation period T_i,

(b) the ACF peak height is high, and

(c) the damped oscillator model is a good fit (in a χ^2 sense) to the ACF, with extra bonus for being a good fit over more points (larger N_i, or longer P_i).

4. Given this set of candidate periods T_i and quality metrics Q_i, we finally construct a multimodal prior on the P parameter of the GP model as a weighted mixture of Gaussians:

$$p(\ln P) = \frac{\sum_i Q_i \left(0.9\mathcal{N}(\ln T_i, 0.2) + 0.05\mathcal{N}(\ln(T_i/2), 0.2) + 0.05\mathcal{N}(\ln(2T_i), 0.2)\right)}{\sum_i Q_i}. \tag{13}$$

That is, in addition to taking the candidate periods themselves as mixture components, we also mix in twice and half each candidate period at a lower level, which compensates for the possibility that the first peak in the ACF may actually represent half or twice the actual rotation period. The period width of 0.2 in log space (corresponding to roughly 20% uncertainty) is again a heuristic choice, balancing a healthy specificity with the desire to not have the results of the inference overly determined by the ACF prior.

Incidentally, while we use the procedure described here to create a prior on P which we use while inferring the parameters of the quasi-periodic GP model, this same procedure may also be used in the service of a rotation-period estimating procedure all on its own, perhaps being even more robust and accurate than the traditional ACF method. We leave exploration of this possibility to future work.

2.3. Sampling

To sample the posterior in a way that is sensitive to potential multimodality, we use the `emcee3`[10] MCMC sampler. `emcee3` is the successor to the `emcee` project (Foreman-Mackey et al. 2013) that includes a suite of ensemble MCMC proposals that can be combined to efficiently sample more distributions than the stretch move (Goodman, J. and Weare 2010) in `emcee`. For this project, we use a weighted mixture of three proposals. First, we include a proposal based on the `kombine` package[11] (Farr & Farr, in prep.) where a kernel density estimate (KDE) of the density represented by the complementary ensemble is used as the proposal for the other walkers. The other two proposals are a "Differential Evolution (DE) MCMC" proposal (Ter Braak 2006; Nelson et al. 2014) and the "snooker" extension of DE (ter Braak and Vrugt 2008).

We initialize 500 walkers with random samples from the prior and use a weighted mixture of the KDE, DE, and snooker proposals with weights of 0.4, 0.4, and 0.2 respectively. We run 50 steps of the sampler at a time, checking for convergence after each iteration, up to a maximum of 50 iterations. We declare convergence if the total effective chain length is at least 8× the maximum autocorrelation time. When convergence is achieved, we discard the first two autocorrelation lengths in the chain as a burn-in, and randomly choose 5000 samples as representative of the posterior. This fitting process takes several hours for a typical simulated light curve, though in some cases it can take 12 hours or longer to converge.

3. Performance and Comparison to Literature: Simulated Data

In order to benchmark this new rotation period recovery method, we apply it to a set of simulated light curves and compare to the performance of established literature methods. Section 3.1 describes the simulated data we use; Section 3.2 demonstrates the performance of the QP-GP method; and Section 3.3 compares to the performance of the Lomb-Scargle periodogram and autocorrelation function methods.

---

[10] https://github.com/dfm/emcee3

[11] https://github.com/bfarr/kombine

3.1. Simulated light curves

We take our test data set from the Aigrain et al. (2015) 'hare and hounds' rotation period recovery experiment. These light curves result from placing dark, circular spots with slowly evolving size on the surface of bright, rotating spheres, ignoring limb-darkening effects. Aigrain et al. (2015) simulated one thousand such light curves to test the ability of participating teams to recover both the stellar rotation periods and the rotational shear (the amplitude of surface differential rotation). However, in this work, in order to focus on demonstrating reliable period recovery, we select only the 333 light curves without differential rotation, as differential rotation may produce additional scatter in the measured rotation periods. In future we intend to test to what extent we can recover differential rotation using the GP method and will then use the full set of 1000 light curves. [VR: a few months ago I did a bunch of tests with kernels I constructed specifically to account for differential rotation. I need to dig up that work but I remember the clear conclusion at which I arrived was that there seems to be an unbreakable degeneracy between any extra hyperparameter(s) to account for differential evolution, and the standard QP hyperparameters. This might change if enforcing $l \gg P$; I can't remember. Anyway, I mention this now just as a little warning that we might not be able to count on the GP approach yielding many useful insights as far as differential rotation goes.]

Each of these light curve simulations uses a real *Kepler* long-cadence time array: one data point every thirty minutes over a four year duration. 90% of the rotation periods of the simulations come from a log-uniform distribution between 10 and 50 days, and 10% from a log-uniform distribution between 1 and 10 days. Figure 2 shows the distribution of solid-body rotation periods. The simulations also have a range of stellar inclination angles, activity levels, spot lifetimes and more (see Table 3.1). In order to preserve *Kepler* noise properties, Aigrain et al. (2015) add real *Kepler* light curves with no obvious astrophysical variability to the theoretical rotational modulated light curves. Figure 3 shows an example of a simulated light curve with a period of 20.8days.

3.2. Method Performance

We apply the QP-GP inference method described in Section 2 to each of these 333 simulated light curves. As discussed in Section 2.2.1, reliable inference requires defining a useful set of priors on the non-period hyperparameters. For this simulation dataset, we determined these by first running the method using very broad priors on all the non-period parameters (log-flat between -20 and 20) and then inspecting the distribution of their posteriors for those cases that successfully recovered the true period. We also experimented with constraining

Table 1: Ranges and distributions of parameters used to simulate light curves in Aigrain et al. (2015)

Parameter	Range	Distribution
Rotation period, P_{rot}	10 - 50 days (90%)	log uniform
	1 - 10 days (10%)	log uniform
Activity cycle length	1 - 10 years	log uniform
Inclination	0 - 90°	Uniform in $\sin^2 i$
Decay timescale	(1 - 10) $\times P_{rot}$	log uniform

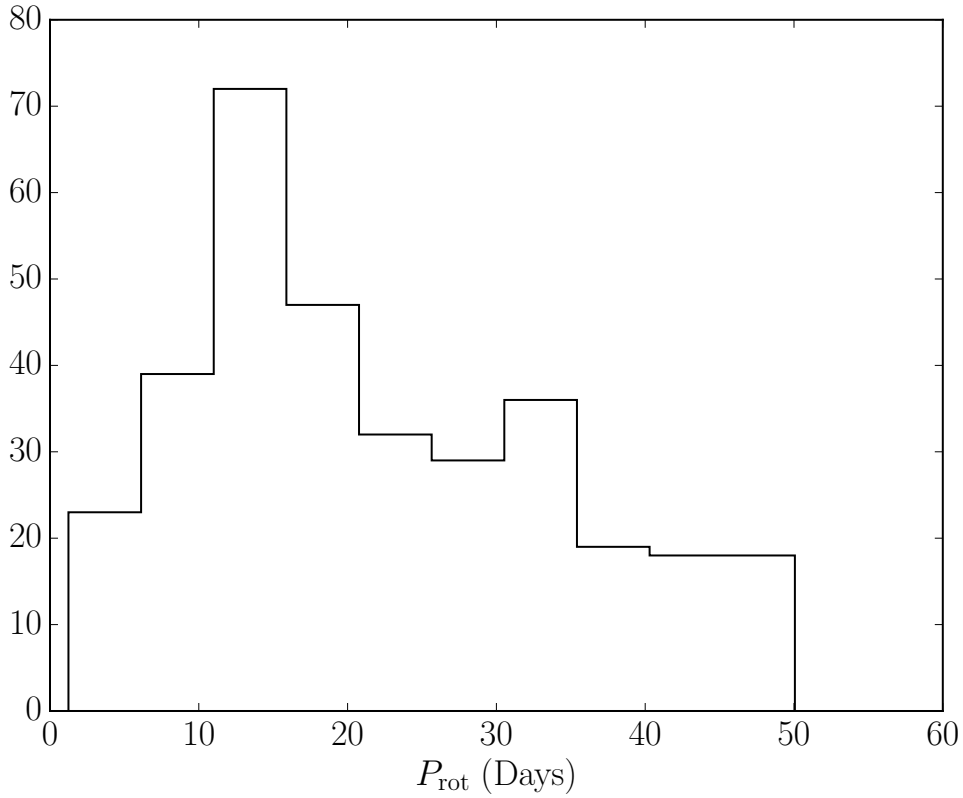

Fig. 2.— A histogram of the rotation periods used to generate the 333 simulated light curves in Aigrain et al. (2015).

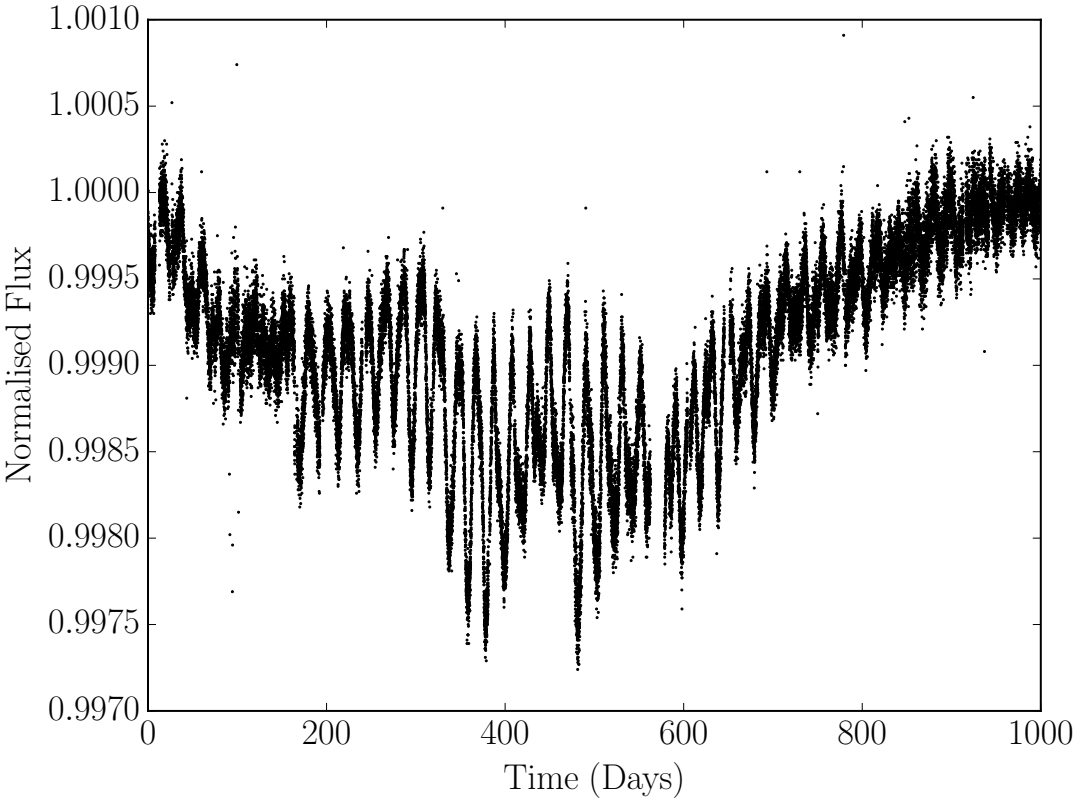

Fig. 3.— An example simulated light curve. This 'star' has a rotation period of 20.8 days.

the allowed ranges of the parameters after discovering that some regions of parameter space (such as large values of A and Γ and small values of l) tended to allow fits that ignored the desired periodicity. We list the final priors and bounds this process led us to adopt in Table 3.2. We note that we do not have any specific quantitative justification for the use of these priors, except that they produce good results for this particular test set of light curves. We thus caution that other data sets might have different noise properties, potentially requiring modifications to these priors [VR: might want to reword or remove these remarks; see my earlier comments] (e.g., see Section 4). For the period prior, we tried two different methods: an uninformed (log-flat) prior between 0.5 d and 100 d, and an ACF-informed prior (Section 2.2.2).

Figures 4 and 5 summarize our results compared to the injected 'true' stellar rotation periods, for the uninformed and ACF-informed priors on period, respectively. To assess the performance of the QP-GP and other period recovery methods, we compute Median Absolute Deviations (MADs) of the results, relative to the input periods. We also compute the Median Relative Deviation (MRD), as a percentage. These metrics are presented for the three different methods tested in this paper in table 3.3.2. The informative and uninformative prior versions of the GP method have MRDs of 1.85% and 2.45% respectively. The marginal posterior distributions of the QP kernel hyperparameters, for the example simulated light curve in figure 3, are shown in figure 6.

3.3. Comparison with literature methods

3.3.1. ACF

We measure an ACF-based period for each light curve, following the method of McQuillan et al. (2013a). After calculating the ACF for a light curve, we smooth it by convolving with a Gaussian filter ($\sigma = 9$ d), and select the rotation period as the lag-time of the highest

Table 2: Priors and bounds on the natural logarithms of the GP model parameters.

Parameter	Prior	Bounds
$\ln A$	$\mathcal{N}(-13, 5.7)$	(-20, 0)
$\ln l$	$\mathcal{N}(7.2, 1.2)$	(2, 20)
$\ln \Gamma$	$\mathcal{N}(-2.3, 1.4)$	(-10, 3)
$\ln \sigma$	$\mathcal{N}(-17, 5)$	(-20, 0)
$\ln P$	Uniform / ACF-based	$(\ln 0.5, \ln 100)$

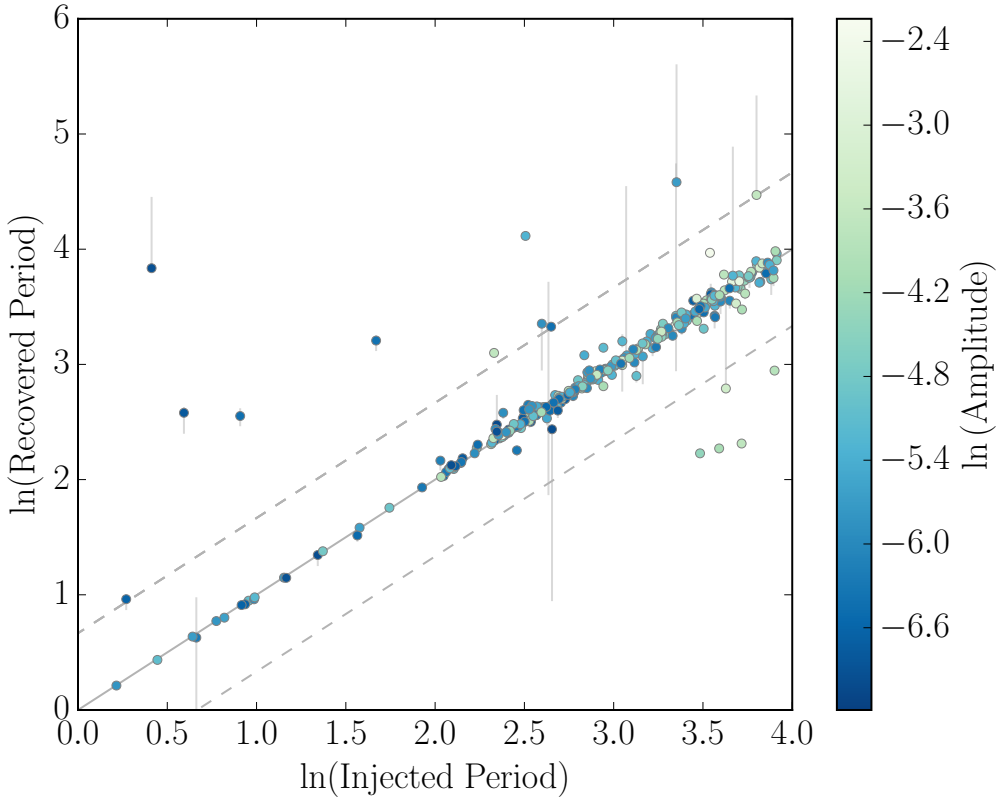

Fig. 4.— The 'true' rotation periods used to generate 333 simulated light curves vs the rotation periods measured using the GP technique. Since the posterior PDFs of rotation periods are often non-Gaussian, the points plotted here are the highest likelihood samples. The uncertainties are the 16th and 84th percentiles. In many cases, the uncertainties are under-estimated. The ACF-informed prior on rotation period used to generate these results is described in the appendix. The MAD of these results is 0.38 in log-space, *i.e.* 1.46 days.

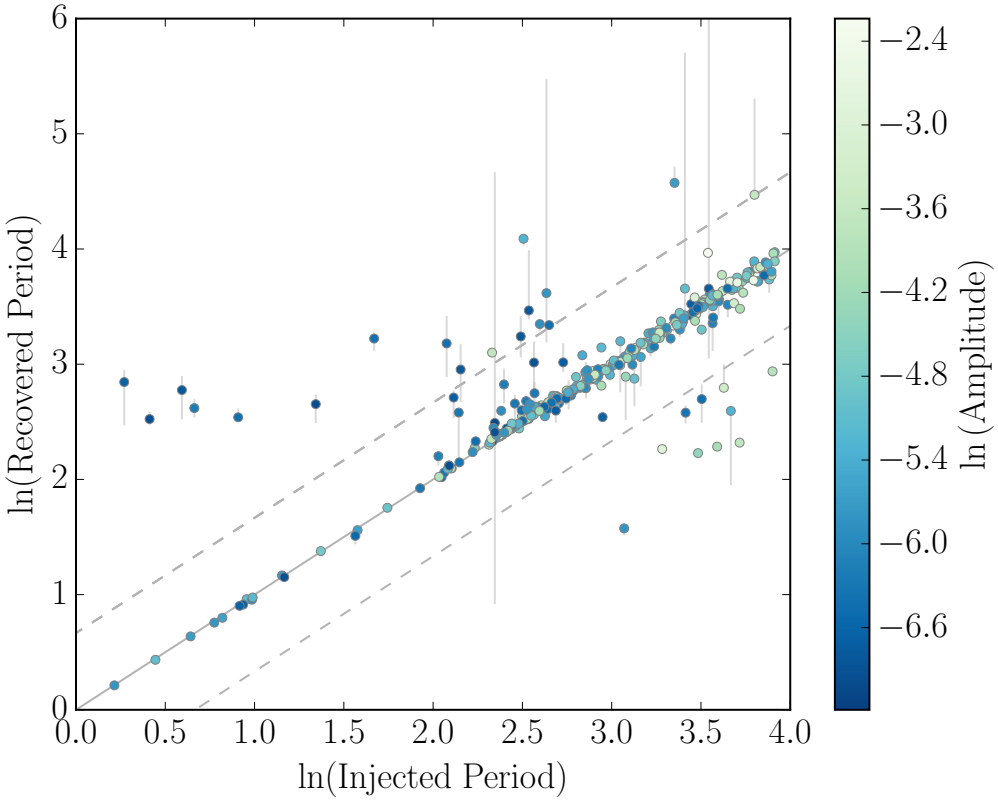

Fig. 5.— The 'true' rotation periods used to generate 333 simulated light curves vs the rotation periods measured using the GP technique with no . Since the posterior PDFs of rotation periods are often non-Gaussian, the points plotted here are the highest likelihood samples. The uncertainties are the 16th and 84th percentiles. In many cases, the uncertainties are under-estimated. An uninformative prior, flat in the natural log of the rotation period between 0.5 and 100 days was used to generate these results. The MAD of these results is 0.48 in log-space, *i.e.* 1.61 days.

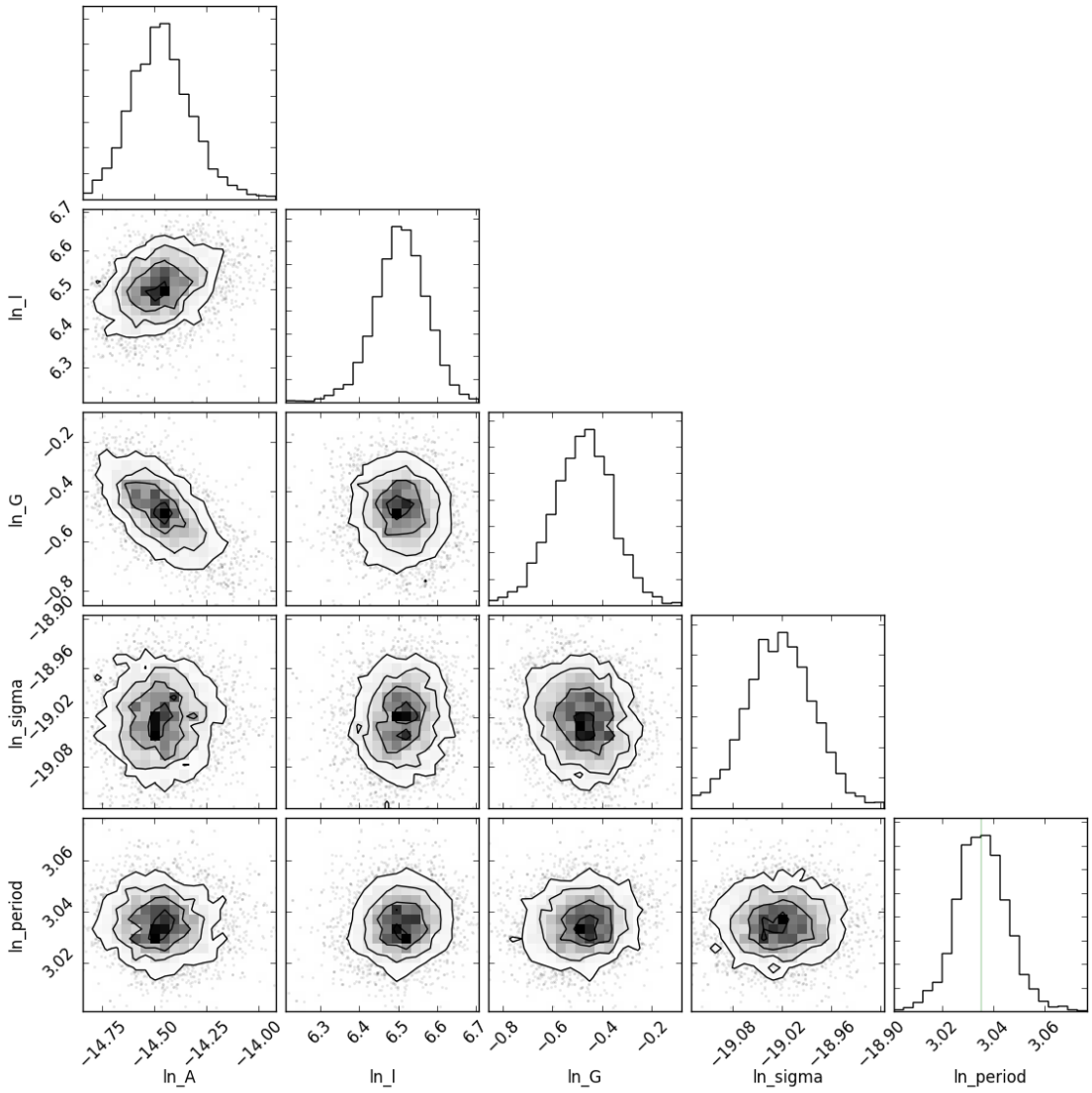

Fig. 6.— Marginal posterior PDFs of the QP GP model parameters, fit to the light curve in figure 3. The green line in the period panel shows the injected period. [VR: might it be possible to make the green line darker? It took me a little while to find it.] This figure was made using `corner.py` (Foreman-Mackey 2016).

peak in the ACF less than 100 d. This is not always the first peak—the second can be larger than the first if two active regions are at or near opposite longitudes on the surface of the star, producing a light curve with two dips per rotation period. We restrict our search to periods less than 100 d because instrumental noise and light-curve systematics significantly distort *Kepler* light curves on longer timescales, rendering longer periods realistically unrecoverable. Figure 3 contains an example ACF of the light curve in Figure 7.

The ACF method has proven extremely useful for measuring rotation periods. The catalogue of rotation periods of *Kepler* stars provided in McQuillan et al. (2013a) has been widely used by the community and has provided ground-breaking results for stellar and exoplanetary science. The method also performed well in the Aigrain et al. (2015) recovery experiment, producing a large number of accurate rotation period measurements (see, e.g., their Figure 8). Another advantage is its fast implementation speed. However, because the ACF method is non-probabilistic, ACF-estimated rotation period uncertainties are poorly defined—a clear disadvantage.

We apply the ACF method to the sample of 333 simulated light curves. Figure 8 shows ACF-measured versus true rotation periods, with the $2n$ and $\frac{1}{2}n$ harmonic lines as dashed lines. The MRD of the ACF-recovered periods is 3.65 % (see table 3.3.2 for a side-by-side comparison with the other methods). The injected and recovered rotation periods generally agree well, though with a few drastic over- or underestimates. Additionally, of the points clustered close to the 1:1 line, more fall slightly below than above it, *i.e.* rotation periods are preferentially underestimated. This stems from how peak positions in the ACF are measured. ACFs of stellar light curves have similar functional forms to the QP kernel function in Equation 2: periodic functions added to decaying exponentials. In such functions the peak positions can be shifted towards the left (towards shorter periods), because the decaying exponential raises the left side of each peak more than the right. It is possible to model this effect; however, standard practice simply measures the peak position without taking it into account. We adopt this standard practice here to faithfully compare our new method to that used in the literature.

3.3.2. LS periodogram

For each simulated light curve, we compute a LS periodogram [12] over a grid of 10,000 periods, evenly spaced in frequency, between 1 and 100 days. We adopt the period of

[12]LS periodograms were calculated using the gatspy Python module: https://github.com/astroML/gatspy/tree/master/gatspy/periodic.

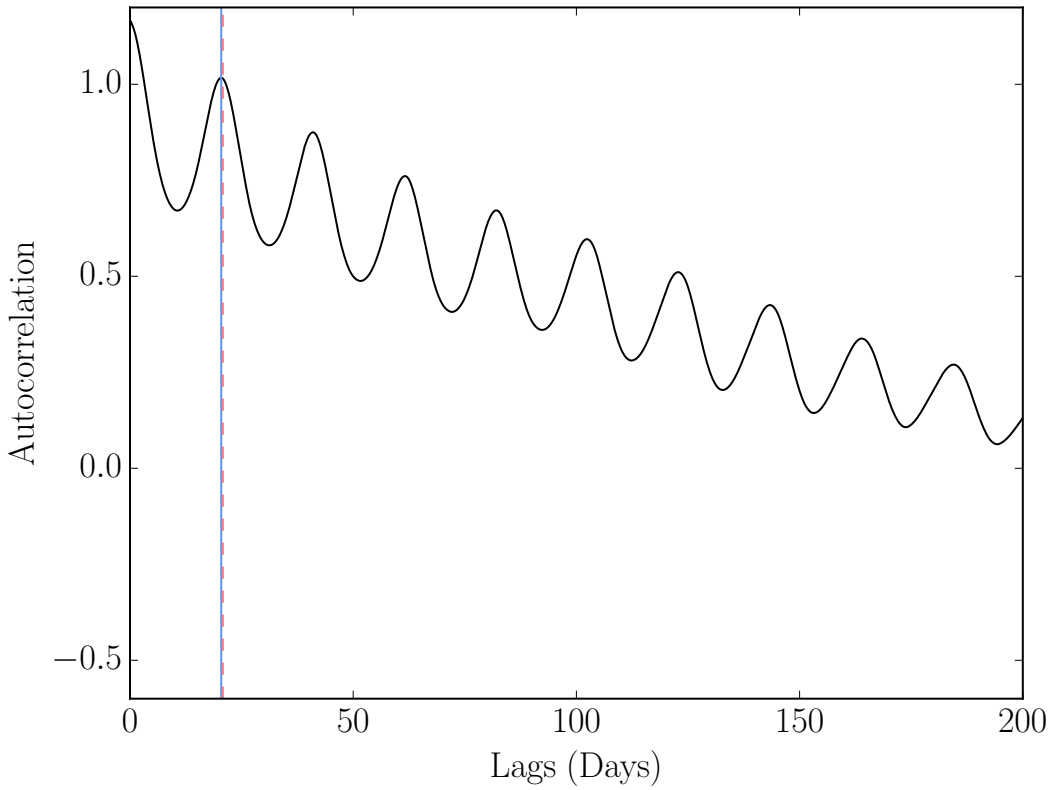

Fig. 7.— An autocorrelation function of the simulated light curve shown in figure 3. The vertical blue line shows the period measured using the ACF method (20.4 days) and the pink dashed line shows the period that was used to simulate the light curve (20.8 days).

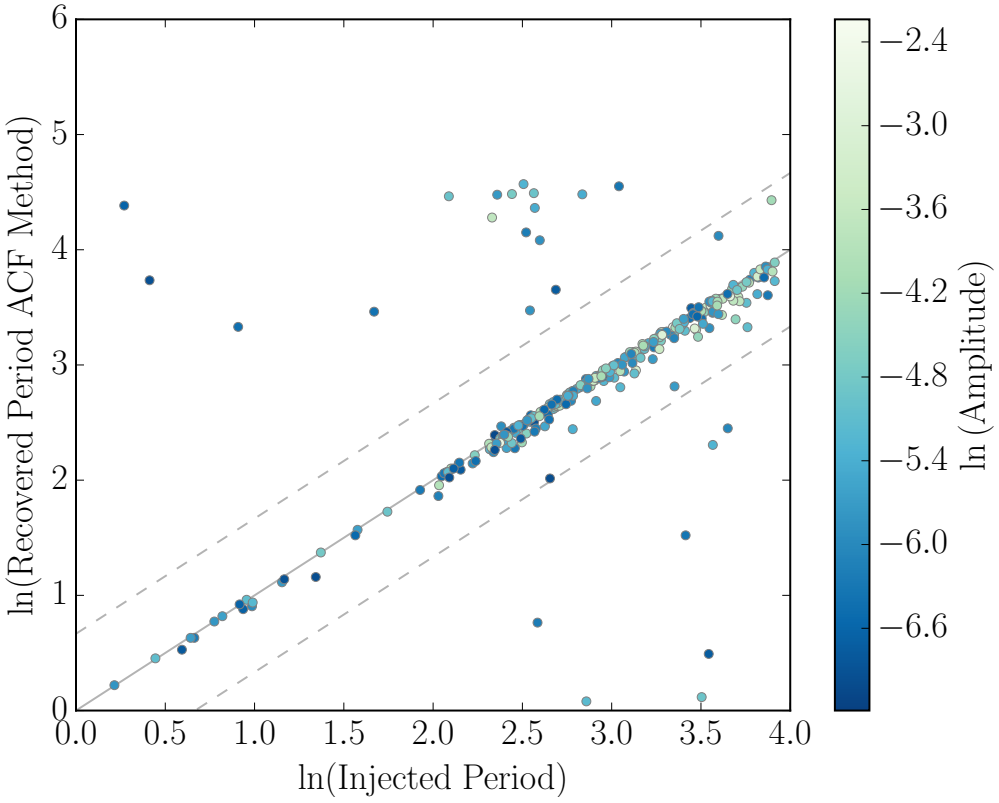

Fig. 8.— The 'true' rotation periods used to generate 333 simulated light curves vs the rotation periods measured using the ACF technique. Several light curves have over-estimated rotation periods and some are drastically underestimated.

the highest peak in the periodogram as the rotation period. Figure 9 shows the resulting recovered rotation periods as a function of true period. The MRD of the periodogram-recovered periods is 4.32% (see table 3.3.2 for a side-by-side comparison with the other methods). This method drastically overestimates many rotation periods, because long-term trends unrelated to rotation can cause excess power at long periods—a problem further exacerbated by instrumental noise. I don't quite understand this—what do we mean here by "instrumental noise" as distinct from long-term trends?

Table 3: Median absolute (MAD) and median relative (MRD) deviations for the ACF, LS periodogram and GP period recovery methods.

Method	MAD	MRD
ACF	0.60 days	3.65%
LS periodogram	0.98 days	4.32%
GP (no prior)	0.48 days	2.45%
GP (acf prior)	0.38 days	1.85%

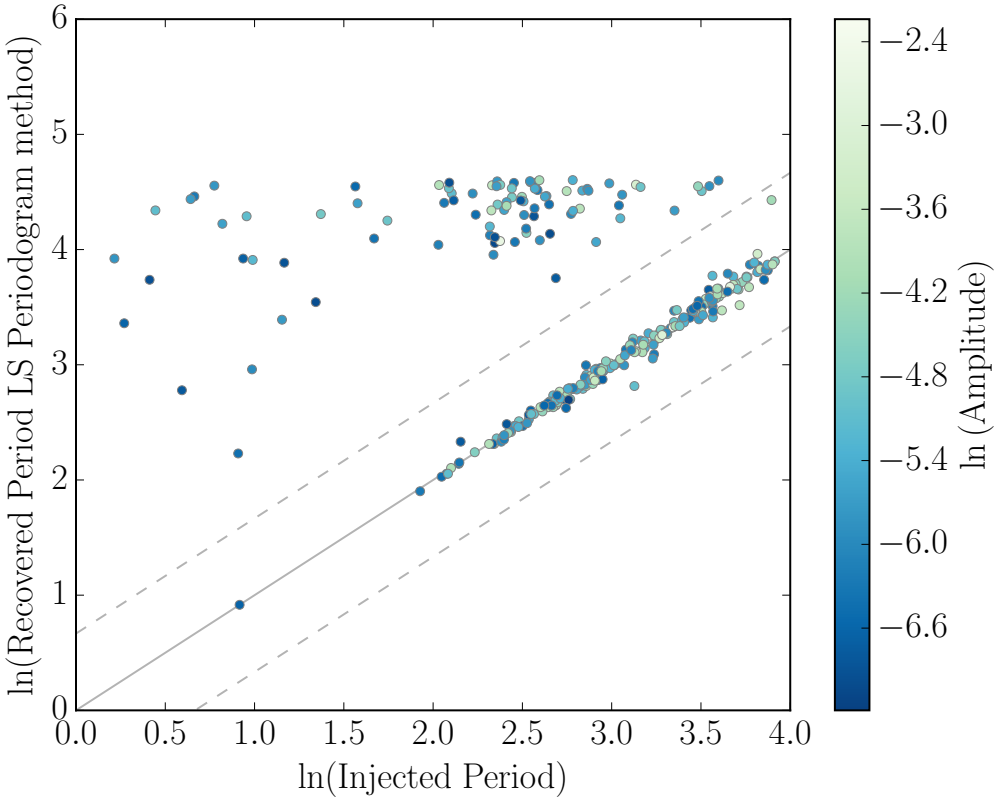

Fig. 9.— The 'true' rotation periods used to generate 333 simulated light curves vs the rotation periods measured using a LS periodogram technique. In many cases a large peak at a long period was present in the periodogram, producing a significant over-estimate of the period.

4. Real *Kepler* data

In order to test our rotation period inference method on real data, we apply it to a set of 1102 *Kepler* Object of Interest (KOI) host stars, 275 of which had rotation periods previously measured by the ACF method by McQuillan et al. (2013a). We use the pipeline-corrected flux (`pdcsap_flux` column in the *Kepler* light curve table), median-normalized and unit-subtracted, and mask out all known transiting planet candidate signals. As with the simulated light curves, we randomly subsample each light curve by a factor of 30 and split it into segments of about 300 points for the purposes of evaluating the likelihood. We also follow the same MCMC fitting procedure as with the simulated data, using the ACF-based prior as before.

Initially, we also use the same priors on the hyperparameters for the KOIs as for the simulated light curves (Table 3.2). However, we found that $\ln l$ and $\ln \Gamma$ tended toward slightly different values than the simulations. We also found that the allowed hyperparameter range that we used for the simulations was too large for the KOI population, as maybe $\sim$15% of the fits tended toward corners in the hyperparameter space, resulting in poor period measurements. As a result, after this initial test, we subsequently adjusted the priors and re-fit all the KOIs. The final priors and parameter ranges that we used are in Table 4.

Figure 10 compares the periods inferred with the GP method to the ACF-based periods from McQuillan et al. (2013a) for the 275 overlapping KOIs. This comparison shows generally very good agreement, with only a few exceptions, demonstrating that this method works not only on simulated data, but also on real data—with the caveat that for any particular data set, some care is needed regarding the setting the priors and ranges for the GP hyperparameters. Notably, below $\ln P$ of about 2.5, the GP method recovers periods systematically slightly larger than McQuillan et al. (2013a)—a likely consequence of correcting for the ACF peak measurement bias discussed in Section 3.3.1.

Table 4: Priors and bounds on the natural logarithms of the GP model parameters, for *Kepler* light curves

Parameter	Prior	Bounds
$\ln A$	$\mathcal{N}(-13, 5.7)$	(-20, 0)
$\ln l$	$\mathcal{N}(5.0, 1.2)$	(2, 8)
$\ln \Gamma$	$\mathcal{N}(1.9, 1.4)$	(0, 3)
$\ln \sigma$	$\mathcal{N}(-17, 5)$	(-20, 0)
$\ln P$	ACF-based	$(\ln 0.5, \ln 100)$

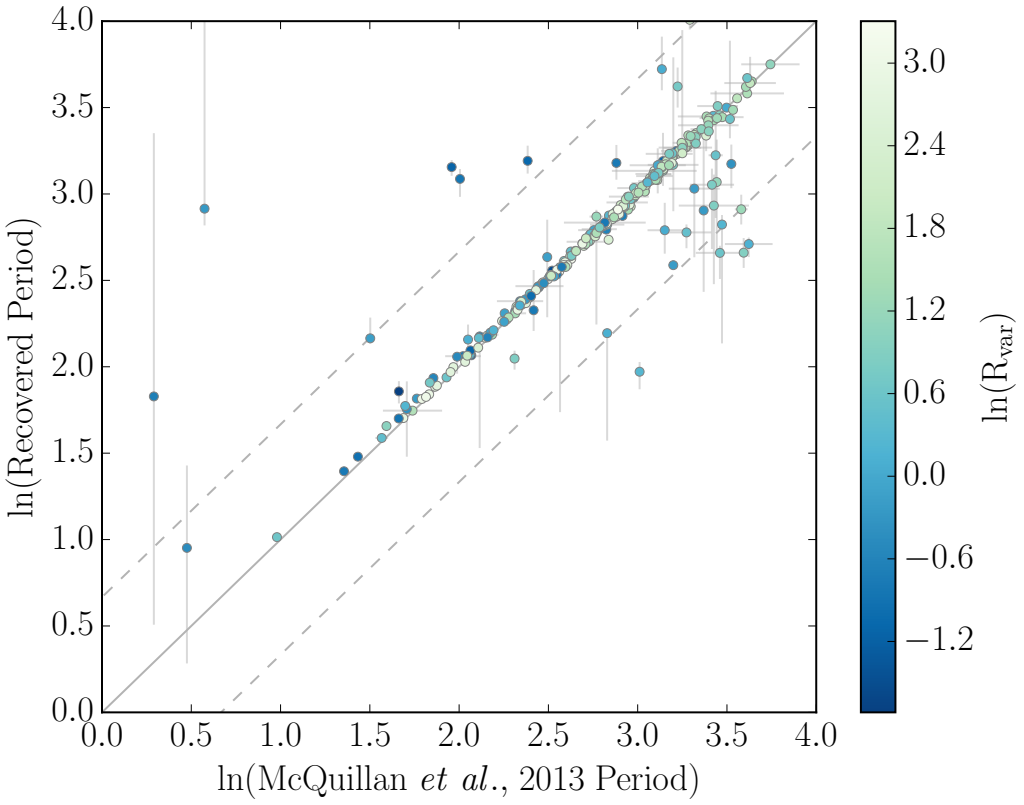

Fig. 10.— A comparison of our GP rotation period measurements to those of McQuillan et al. (2013a). The data points are coloured by the range of variability measured by McQuillan et al. (2013a), defined as the interval between the 5th and 95th percentiles of normalized flux per period bin in millimagnitudes. The results show good agreement.

5. Discussion

The uncertainties produced by the GP method are more representative than those produced by the ACF and periodogram methods because the posterior PDF of the period is explored using MCMC. Whilst they are an improvement, unfortunately these uncertainties are still underestimated in the majority cases. Of the rotation periods recovered from the simulated light curves, only 25% of the measured periods lie within 1 σ of the true period. 50% lie within 2 σ and 66% within 3 σ. What are these numbers for the uninformed-prior experiment? The largest outlier is 114 σ away from the true value. In the majority of these cases the uncertainties are underestimated due to the multi-modal nature of the period posterior PDFs, which makes them difficult to sample. Is there any evidence of this? Doesn't sound convincing to me. I would think the multi-modality would only be relevant if the posterior is closer to the 1/2x or 2x period, but this isn't most of them. The underestimated uncertainties can also be attributed to the model choice. Although a GP model is more appropriate than a sinusoid, it is still an imperfect, effective model. A perfect physical model of the star would still produce better results with more representative uncertainties. Not sure saying "a perfect model would give better results" is worth saying... VR: I agree this is probably not worth saying.

The QP kernel function represents a simplistic effective model of a stellar light curve. It adequately describes the data, captures that all-important periodic quality [VR: I think this is a bit vague...a sinusoidal model would capture 'periodic quality'. Maybe say something like: 'it manages to describe a wide range of quasi-periodic behavior'.] and is relatively simple, with only a few hyperparameters. It also satisfies the requirement to produce positive semi-definite covariance matrices. [VR: I don't think it's worth mentioning the PSD thing. It's not as though this is some advantage the QP kernel has over any other kernel function.] Whilst the QP kernel function evidently captures the periodic qualities of light curves adequately, it is still a somewhat arbitrary choice [VR: repetitive...maybe just say 'Nevertheless, it is still a somewhat arbitrary choice.']. Another valid choice would be a squared cosine function multiplied by a squared exponential,

$$k_{i,j} = A \exp\left(-\frac{(x_i - x_j)^2}{2l^2}\right) \cos\left(\frac{\pi(x_i - x_j)}{P}\right) \tag{14}$$

VR: missing $(x_i - x_j)$ term added above, and factor of 2 removed. This function produces a positive semi-definite matrix and has the P parameter of interest. The main practical difference between this cosine and the QP function is that the cosine function allows negative covariances and the QP function does not. Is is realistic to allow negative covariances? In practice, the ACFs of *Kepler* light curves often go negative. However, many stars have two active regions on opposite hemispheres that produce two brightness dips per rotation. If the

covariance is forced to be negative for two data points that are separated by half a rotation period, those light curves with two peaks per rotation period may not be well modelled. In other words, we do not want to force anti-correlation of points that are $\frac{1}{2}$ a period apart. It would be very worthwhile to test this assumption and this alternative kernel function in future. If CPU time were not limited there may be some benefit to performing formal model comparison with different kernel functions. However, the evidence integral is an ambitious calculation for models with likelihood functions that take milliseconds to compute, let alone those involving GPs with light curves containing thousands of data points which can take minutes.

Not all *Kepler* light curves show evidence of stellar rotation. In some cases perhaps the star has few or no active regions, it is rotating pole-on, or it rotates so slowly that the *Kepler* data detrending pipeline removes any signal. In other cases there may be another source of variability present in the light curve, generating a false period detection. These sources may be physical: *e.g.* binary star interactions, intra-pixel contamination from other astrophysical objects, pulsating variable stars, asteroseismic oscillations in giants and even stellar activity cycles. For some of these there is little we can do save attempt to eliminate these astrophysical false positives via alternative methods, *e.g.* apply colour cuts to avoid giant contamination. For some however, like the variable stars, they may have distinctive hyperparameters that identify them, for example long coherence timescales. Testing this is beyond the scope of this paper but may be an interesting follow-up study. As well as astrophysical contamination, there may also be instrumental sources of contaminating variability: *e.g.* temperature variations or pointing shifts of the *Kepler* spacecraft. These are unlikely to be periodic and, again, may produce unusual combinations of hyperparameters. This is something that we hope to test in future. In addition to this, there are several other aspects of the GP method we are continuing to develop, listed as follows:.

- To perform model selection with different kernel functions. Again, these tests have not been performed due to the cost of calculating the fully marginalised likelihood with GPs. This calculation may be prohibitively expensive, however simpler model selection tests can be performed. For example, the relative precision of rotation periods recovered using alternative kernel functions could be tested. [VR: this is something I've started exploring using MultiNest. Computation times seem quite manageable.]

- To design and implement a physically motivated kernel function. We have only explored the physical interpretation of one parameter in our kernel function, P. However, the other parameters may also be related to some physical processes. For example, the overall timescale for covariance fall-off, l may be related to spot lifetimes. A star with long spot lifetimes will show little variation in the overall shape and amplitude of its

light curve between rotations and l will be large. In contrast, the light curve of a star with short spot lifetimes may display non-repeating patterns and amplitudes that vary rapidly between rotations. In this case l will be small. The Γ parameter is related to the number of zero crossings within one rotation period: when Γ is small there are many zero crossings and vice versa. Since the number of zero crossings per rotation period is related to the number of active regions on the surface of the star, this parameter may also be of physical interest. In addition, instead of interpreting the parameters of the QP kernel function used here, it may be possible to design an entirely new kernel function, based on the physical processes that drive the light curve variability. This idea is being explored by another member of my research group.

- To attempt to detect differential rotation. We did not test our code on the light curves simulated with differential rotation in Aigrain et al. (2015) since we were only interested in recovering the most precise measurements of rotation period possible. In future we intend to investigate the possibility of recovering differential rotation by searching for close double peaks in the posterior PDFs of stars' rotation periods.

- [VR: for the two points above, you might want to reword slightly depending on the extent to which you want to try to (i) incorporate my suggestions re prior constraints etc. etc. vs. (ii) deferring this to the future paper, but at least alluding to the possibilities and results. For example I think one needn't try to 'design' a physically-motivated kernel – rather it seems not too difficult just to come up with physically-motivated priors for the QP kernel parameters. I've tried to design a physically motivated kernel to capture differential rotation but as noted earlier, that came with its own slew of degeneracies and other problems...more on this another time.]

- To build in a noise model for *Kepler* data. Another huge advantage of the GP method is, because it is a *generative* model of the data, the rotation period signal can be modelled at the same time as systematic noise features. One can then marginalise over the parameters of the noise model. This approach would be extremely advantageous for *Kepler* data since long-term trends are often removed by the *Kepler* detrending pipeline. Marginalising over the noise model at the same time as inferring the parameters of interest will insure that the periodic signal is preserved.

6. Conclusion

We attempted to recover the rotation periods of 333 simulated *Kepler*-like light curves for solid-body rotators (Aigrain et al. 2015) using three different methods: a new Gaus-

sian process method, a Lomb-Scargle periodogram method and an autocorrelation function method. We demonstrate that the GP method produces the most accurate rotation periods of the three techniques, providing a large improvement over the LS periodogram method and a moderate one over the ACF method. We also find that the standard version of the ACF method, most commonly implemented in the literature, often slightly under-predicts the rotation period due to a subtle feature of its peak detection algorithm. In addition, we compared the rotation periods of *Kepler* objects of interest measured using the GP method to those measured previously by McQuillan et al. (2013a). The good agreement between the two sets of results demonstrates that the GP method works well on real *Kepler* data.

Unlike the ACF and LS periodogram methods, the GP method provides posterior PDF samples which can be used to estimate rotation periods uncertainties. Although an improvement on competing methods, these uncertainties are still approximate because in many cases the posterior PDFs of the parameters are multi-modal. If it were possible to run an MCMC sampler for an infinite number of steps, the entire posterior PDF would eventually be sampled, and a more accurate uncertainty would result. Unfortunately, due to realistic computational resources and the limitations of MCMC sampling techniques for multi-modal posteriors, we are not always able to map out the entire posterior PDFs. I don't think this is the problem; I think the fits that we judge to have converged have actually converged.

In addition, although the GP model used here is clearly a good one, it is still only an effective model, not an accurate physical model. It can only capture the posterior PDF of the periodic component of the covariance matrix, not the actual rotation period of a physical star. We find that only a quarter of the reported uncertainties are accurate.

The main aim of this work is to develop a probabilistic rotation period inference method. Probabilistic periods are necessary for hierarchical Bayesian inference, particularly when performing population analysis. This new GP method is capable of generating a catalogue of probabilistic rotation periods which could, for example, reveal the period distribution (and therefore potentially the age distribution) of stars in the Milky Way. There is still room for improvement since the uncertainty estimates are not yet truly representative and it is still only an 'effective' model, not a physical one. However, it is probabilistic and provides more accurate periods than alternative methods. We therefore argue that the GP method is the best method for rotation period inference currently available.

This research was funded by the Simons Foundation and the Leverhulme Trust. TDM is supported by NASA grant NNX14AE11G, and acknowledges the hospitality of the Institute for Advanced Study, where part of this work was completed. V. R. thanks Merton College and the National Research Foundation of South Africa for financial support. Some of the data presented in this paper were obtained from the Mikulski Archive for Space Telescopes

(MAST). STScI is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS5-26555. Support for MAST for non-HST data is provided by the NASA Office of Space Science via grant NNX09AF08G and by other grants and contracts. This paper includes data collected by the Kepler mission. Funding for the Kepler mission is provided by the NASA Science Mission directorate.

REFERENCES

S. Aigrain, J. Llama, T. Ceillier, M. L. d. Chagas, J. R. A. Davenport, R. A. García, K. L. Hay, A. F. Lanza, A. McQuillan, T. Mazeh, J. R. de Medeiros, M. B. Nielsen, and T. Reinhold. Testing the recovery of stellar rotation signals from Kepler light curves using a blind hare-and-hounds exercise. *MNRAS*, 450:3211–3226, July 2015. doi: 10.1093/mnras/stv853.

S. Aigrain, H. Parviainen, and B. J. S. Pope. K2SC: Flexible systematics correction and detrending of K2 light curves using Gaussian Process regression. *MNRAS*, April 2016. doi: 10.1093/mnras/stw706.

S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. W. Hogg, and M. O'Neil. Fast Direct Methods for Gaussian Processes. *ArXiv e-prints*, March 2014.

T. Barclay, M. Endl, D. Huber, D. Foreman-Mackey, W. D. Cochran, P. J. MacQueen, J. F. Rowe, and E. V. Quintana. Radial Velocity Observations and Light Curve Noise Modeling Confirm that Kepler-91b is a Giant Planet Orbiting a Giant Star. *ApJ*, 800:46, February 2015. doi: 10.1088/0004-637X/800/1/46.

J. A. Carter and J. N. Winn. Parameter Estimation from Time-Series Data with Correlated Errors: A Wavelet-Based Method and its Application to Transit Light Curves. Astrophysics Source Code Library, October 2010.

R. I. Dawson, J. A. Johnson, D. C. Fabrycky, D. Foreman-Mackey, R. A. Murray-Clay, L. A. Buchhave, P. A. Cargile, K. I. Clubb, B. J. Fulton, L. Hebb, A. W. Howard, D. Huber, A. Shporer, and J. A. Valenti. Large Eccentricity, Low Mutual Inclination: The Three-dimensional Architecture of a Hierarchical System of Giant Planets. *ApJ*, 791:89, August 2014. doi: 10.1088/0004-637X/791/2/89.

X. Dumusque, F. Pepe, C. Lovis, et al. An Earth-mass planet orbiting α Centauri B. *Nature*, 491:207–211, November 2012. doi: 10.1038/nature11572.

R. A. Edelson and J. H. Krolik. The discrete correlation function - A new method for analyzing unevenly sampled variability data. *ApJ*, 333:646–659, October 1988. doi: 10.1086/166773.

T. M. Evans, S. Aigrain, N. Gibson, J. K. Barstow, D. S. Amundsen, P. Tremblin, and P. Mourier. A uniform analysis of HD 209458b Spitzer/IRAC light curves with Gaussian process models. *MNRAS*, 451:680–694, July 2015. doi: 10.1093/mnras/stv910.

D. Foreman-Mackey, D. W. Hogg, D. Lang, et al. emcee: The MCMC Hammer. *PASP*, 125: 306–312, March 2013. doi: 10.1086/670067.

D. Foreman-Mackey, S. Hoyer, J. Bernhard, and R. Angus. Fast Gaussian Processes for regression. October 2014. doi: 10.5281/zenodo.11989. URL http://dx.doi.org/10.5281/zenodo.11989.

Daniel Foreman-Mackey. corner.py: Scatterplot matrices in python. *The Journal of Open Source Software*, 24, 2016. doi: 10.21105/joss.00024. URL http://dx.doi.org/10.5281/zenodo.45906.

R. A. García, T. Ceillier, D. Salabert, S. Mathur, J. L. van Saders, M. Pinsonneault, J. Ballot, P. G. Beck, S. Bloemen, T. L. Campante, G. R. Davies, J.-D. do Nascimento, Jr., S. Mathis, T. S. Metcalfe, M. B. Nielsen, J. C. Suárez, W. J. Chaplin, A. Jiménez, and C. Karoff. Rotation and magnetism of Kepler pulsating solar-like stars. Towards asteroseismically calibrated age-rotation relations. *A&A*, 572:A34, December 2014. doi: 10.1051/0004-6361/201423888.

N. P. Gibson, S. Aigrain, S. Roberts, T. M. Evans, M. Osborne, and F. Pont. A Gaussian process framework for modelling instrumental systematics: application to transmission spectroscopy. *MNRAS*, 419:2683–2694, January 2012. doi: 10.1111/j.1365-2966.2011.19915.x.

Goodman, J. and J. Weare. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5:1, 2010.

R. D. Haywood. *Hide and Seek: Radial-Velocity Searches for Planets around Active Stars*. PhD thesis, University of St Andrews, November 2015.

R. D. Haywood, A. Collier Cameron, D. Queloz, S. C. C. Barros, M. Deleuil, R. Fares, M. Gillon, A. F. Lanza, C. Lovis, C. Moutou, F. Pepe, D. Pollacco, A. Santerne, D. Ségransan, and Y. C. Unruh. Planets and stellar activity: hide and seek in the CoRoT-7 system. *MNRAS*, 443:2517–2531, September 2014. doi: 10.1093/mnras/stu1320.

S. V. Jeffers and C. U. Keller. An analytical model to demonstrate the reliability of reconstructed 'active longitudes'. In E. Stempels, editor, *15th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun*, volume 1094 of *American Institute of Physics Conference Series*, pages 664–667, February 2009. doi: 10.1063/1.3099201.

D. Keeling, C. and P. Whorf, T. Atmospheric CO2 from Continuous Air Samples at Mauna Loa Observatory, Hawaii, U.S.A. *Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory*, October 2004.

D. M. Kipping. An analytic model for rotational modulations in the photometry of spotted stars. *MNRAS*, 427:2487–2511, December 2012. doi: 10.1111/j.1365-2966.2012.22124.x.

A. F. Lanza, M. L. Das Chagas, and J. R. De Medeiros. Measuring stellar differential rotation with high-precision space-borne photometry. *A&A*, 564:A50, April 2014. doi: 10.1051/0004-6361/201323172.

N. R. Lomb. Least-squares frequency analysis of unequally spaced data. *Ap&SS*, 39:447–462, February 1976. doi: 10.1007/BF00648343.

A. McQuillan, S. Aigrain, and S. Roberts. Statistics of Stellar Variability in Kepler Data with ARC Systematics Removal. In E. Griffin, R. Hanisch, and R. Seaman, editors, *IAU Symposium*, volume 285 of *IAU Symposium*, pages 364–365, April 2012. doi: 10.1017/S1743921312001081.

A. McQuillan, S. Aigrain, and T. Mazeh. Measuring the rotation period distribution of field M dwarfs with Kepler. *MNRAS*, 432:1203–1216, June 2013a. doi: 10.1093/mnras/stt536.

A. McQuillan, T. Mazeh, and S. Aigrain. Stellar Rotation Periods of the Kepler Objects of Interest: A Dearth of Close-in Planets around Fast Rotators. *ApJ*, 775:L11, September 2013b. doi: 10.1088/2041-8205/775/1/L11.

A. McQuillan, T. Mazeh, and S. Aigrain. Rotation Periods of 34,030 Kepler Main-sequence Stars: The Full Autocorrelation Sample. *ApJS*, 211:24, April 2014. doi: 10.1088/0067-0049/211/2/24.

B. Nelson, E. B. Ford, and M. J. Payne. RUN DMC: An Efficient, Parallel Code for Analyzing Radial Velocity Observations Using N-body Integrations and Differential Evolution Markov Chain Monte Carlo. *ApJS*, 210:11, January 2014. doi: 10.1088/0067-0049/210/1/11.

V. Rajpaul, S. Aigrain, M. A. Osborne, S. Reece, and S. Roberts. A Gaussian process framework for modelling stellar activity signals in radial velocity data. *MNRAS*, 452: 2269–2291, September 2015. doi: 10.1093/mnras/stv1428.

V. Rajpaul, S. Aigrain, and S. Roberts. Ghost in the time series: no planet for Alpha Cen B. *MNRAS*, 456:L6–L10, February 2016. doi: 10.1093/mnrasl/slv164.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.

T. Reinhold, A. Reiners, and G. Basri. Rotation and differential rotation of active Kepler stars. *A&A*, 560:A4, December 2013. doi: 10.1051/0004-6361/201321970.

H. N. Russell. On the light variations of asteroids and satellites. *ApJ*, 24:1–18, July 1906. doi: 10.1086/141361.

J. D. Scargle. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *ApJ*, 263:835–853, December 1982. doi: 10.1086/160554.

Cajo JF Ter Braak. A markov chain monte carlo version of the genetic algorithm differential evolution: easy bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3):239–249, 2006.

Cajo JF ter Braak and Jasper A Vrugt. Differential evolution markov chain with snooker updater and fewer chains. *Statistics and Computing*, 18(4):435–446, 2008.