Mapping Stellar Surfaces. II. An Interpretable Gaussian Process Model for Light Curves

Rodrigo Luger1,2,5, Daniel Foreman-Mackey1, and Christina Hedges3,4

1 Center for Computational Astrophysics, Flatiron Institute, New York, NY, USA; rluger@flatironinstitute.org
2 Virtual Planetary Laboratory, University of Washington, Seattle, WA, USA
3 Bay Area Environmental Research Institute, Moffett Field, CA, USA
4 NASA Ames Research Center, Moffett Field, CA, USA

Received 2021 February 9; revised 2021 March 30; accepted 2021 April 2; published 2021 August 27

Abstract

The use of Gaussian processes (GPs) as models for astronomical time series data sets has recently become almost ubiquitous, given their ease of use and flexibility. In particular, GPs excel at marginalization over the stellar signal when the variability due to starspots is treated as a nuisance, as in exoplanet transit modeling. However, these effective models are less useful in cases where the starspot signal is of primary interest, since it is not obvious how the parameters of the GP relate to physical parameters like the spot size, contrast, and latitudinal distribution. Instead, it is common practice to explicitly model the effect of individual starspots on the light curve and attempt to infer their properties via optimization or posterior inference. Unfortunately, this process is ill-posed and often computationally intractable when applied to stars with more than a few spots and/or to ensembles of many stars. Here we derive a closed-form expression for a GP that describes the light curve of a rotating, evolving stellar surface conditioned on a given distribution of starspot sizes, contrasts, and latitudes. We demonstrate that this model is correctly calibrated, allowing one to robustly infer physical parameters of interest from one or more light curves, including the typical spot radii and latitudes. Our GP has far-ranging implications for understanding the variability and magnetic activity of stars from light curves and radial velocity measurements, as well as for modeling correlated noise in exoplanet searches. Our implementation is efficient, user-friendly, and open-source, available in the package starry_process.

Unified Astronomy Thesaurus concepts: Analytical mathematics (38); Time series analysis (1916); Gaussian Processes regression (1930); Starspots (1572)

1. Introduction

Over the past two decades, Gaussian processes (GPs; Rasmussen & Williams 2005) have gained traction as a leading tool for modeling correlated signals in astronomical data sets. In particular, GPs are commonly used to model stellar variability in photometric time series (e.g., Brewer & Stello 2009; Aigrain et al. 2016; Luger et al. 2016; Foreman-Mackey et al. 2017; Angus et al. 2018) and radial velocity (RV) measurements (e.g., Rajpaul et al. 2015; Jones et al. 2017; Perger et al. 2021). The GPs are popular models for these applications because they allow marginalization over a stochastic noise process specified only by a kernel describing its autocorrelation structure. There are several popular open-source implementations that allow efficient evaluation of GPs, and these have been widely demonstrated to be useful effective models for the time series when the stochastic variability due to the star is primarily a nuisance (e.g., Ambikasaran et al. 2015; Foreman-Mackey et al. 2017; Gilbertson et al. 2020).

A major source of stellar variability in both light curves and RV data sets is the modulation induced by magnetically driven surface features like starspots rotating in and out of view. While GPs excel at marginalizing over stellar rotational variability, they have been less useful when the goal is to make inferences about the actual source of this variability, such as the properties of starspots and the magnetic processes that generate them. While it is straightforward to derive posterior constraints on the hyperparameters of an effective GP model for observations of a star, it is not clear what those constraints actually tell us about the stellar surface. Specifically, in all but a few restricted cases, there is no first-principles relationship between the descriptive parameters of a typical GP model (see Section 2.1) and the physical properties of the stellar surface that is being observed. For instance, it may be tempting to interpret the GP amplitude hyperparameter as some measure of the spot contrast or total number of spots or the GP timescale hyperparameter as the spot lifetime, but there are no guarantees these interpretations will hold in general. After all, the choice of kernel is quite often ad hoc, providing an effective—as opposed to interpretable—description of the physics. There are two important exceptions to this: asteroseismic studies, in which the GP hyperparameters can offer direct insight into the behavior of the complex pulsation modes and thus physical properties of the stellar interior (e.g., Brewer & Stello 2009; Foreman-Mackey et al. 2017), and stellar rotation period studies, in which the period hyperparameter $P$ can usually be associated with the rotation period of the star (e.g., Angus et al. 2018). For spot-induced variability, on the other hand, GPs are usually used when the variability itself is a nuisance parameter. For example, if the goal is to constrain the properties of a transiting exoplanet or search for a planetary signal in an RV data set, a GP might be used to remove (or, better yet, marginalize over) the stellar variability (e.g., Haywood et al. 2014; Rajpaul et al. 2015;
Luger et al. (2017b). In this case, the physics behind the variability is irrelevant, so an effective model of this sort may be sufficient.

However, understanding the properties of stellar surfaces, and starspots in particular, is a crucial step toward understanding stellar magnetism, which plays a fundamental part in stellar interior structure and evolution. Stellar magnetic fields control the spin-down of stars over time, on which the field of gyrochronology is founded (Barnes et al. 2001; Angus et al. 2019). They affect wave propagation in stellar interiors and must be properly understood to interpret asteroseismic measurements (e.g., Fuller et al. 2015). Strong magnetic fields are also likely the driving force behind chemical peculiarity in Ap/Bp stars (Turcotte 2003; Sikora et al. 2018), as well as radius inflation in M dwarfs (Gough & Tayler 1966; Ireland & Browning 2018). Stellar magnetohydrodynamics (MHD) is therefore an active area of research, with many open questions (e.g., Miesch & Toomre 2009). Because of the nonlinearity of the MHD equations and the vast range of scales on which magnetic processes operate, there is still significant theoretical uncertainty concerning how dynamos operate in stars of different masses, how magnetic fields affect stellar rotation, and how starspots form (Yadav et al. 2015; Weber & Browning 2016). Observational constraints on starspots and other magnetically controlled surface features are therefore extremely valuable to understanding various problems in stellar astrophysics.

Moreover, even when the stellar signal is considered a nuisance, a physically driven variability model may be a better choice than an effective model in some cases, particularly when the signal of interest is small compared to the systematics. A specific example of this is in transmission spectroscopy of transiting exoplanets, where the contribution from unocculted spots and faculae to the spectrum can be an order of magnitude larger than that of the planet atmosphere (Rackham et al. 2018). In this case, failure to explicitly model the effect of starspots can lead to spurious features in the planet spectrum. A similar situation arises in extreme-precision RV (EPRV) searches for planets, where the stellar signal can be orders of magnitude larger than the planetary signal. While effective models of variability have often been successful at disentangling the planetary and stellar contributions (e.g., Rajpaul et al. 2015), these models can struggle when the (a priori unknown) orbital period of the planet is close to an alias of the rotational period of the star (Vanderburg et al. 2016; Damasso et al. 2019; Robertson et al. 2020). In this case, a physically driven model of variability would likely perform better.

When the goal is to learn about the stellar surface, the common approach in the literature has not been to use GPs but rather to explicitly forward model the surface. Such a model allows one to compute a stellar light curve or spectral time series conditioned on certain surface properties, a procedure that must then be inverted in order to constrain the surface based on a data set. We discussed this approach for the rotational light curves of stars in Luger et al. (2021b, hereafter Paper I), where we argued that a unique solution to the surface map of the star is not possible without the use of aggressive (and often ad hoc) priors. The degeneracies at play make it effectively impossible for one to know the exact configuration of starspots and other features on the surface of a star from its rotational light curve alone.

However, it is hardly ever the case that this is actually our end goal. After all, physics can be used to predict properties of stellar surfaces at a fairly high level, i.e., typical spot sizes, active spot latitudes, or approximate timescales on which spots evolve (e.g., Schuessler et al. 1996; Solanki et al. 2006; Cantiello & Braithwaite 2019). We are hardly ever interested in the particular properties of a particular spot, as we would not really know what to do with that information. Instead, we often treat (whether explicitly or not) the properties of a starspot as a draw from some parent distribution controlling (say) the average and spread in the radii of the spots. The parameters controlling this distribution are the ones that we can predict with physics; they are therefore also the ones we are usually interested in.

Thus, if it were possible to derive robust posterior constraints on the properties of each of the spots on a star, we could then marginalize (integrate) over them to infer the properties describing the distribution of all of the spots as a whole. We could do this using the forward-model approach described above by modeling the properties of each of the spots and computing the corresponding light curves. Then, we could solve the “inverse” problem via a posterior sampling scheme, such as Markov Chain Monte Carlo (MCMC), while including a few hyperparameters controlling the distribution of those properties across all spots, i.e., a one-level hierarchical model. The marginal posteriors for the hyperparameters, then, would encode what we actually wish to know. In practice, however, the degeneracies and often extreme multimodality of the distributions of individual spot properties would make this quite hard (and expensive) to perform. If only we could use the elegant machinery of GPs to perform this marginalization for us.

In this paper, we derive an exact, closed-form expression for the Gaussian approximation to the marginal likelihood of a light curve conditioned on the statistical properties of starspots, which allows us define an interpretable GP for stellar light curves. Our GP analytically marginalizes over the degenerate and often unknowable distributions of properties of individual starspots, revealing the constraints imposed on the bulk spot properties without the need to explicitly model or sample over properties of individual spots. It inherits the speed, ease of use, and all other properties of traditionally used GPs, with the added benefit of direct physical interpretability of its hyperparameters.

While our GP can be used to model the light curves of individual stars, it is particularly useful for ensemble analyses of light curves of many similar stars. As we showed in Paper I, the joint information content of the light curves of many similar stars can be harnessed to constrain the statistical properties of the surfaces of those stars, even in the presence of degeneracies that preclude knowledge about the surfaces of individual stars. By “similar,” we do not mean stars that look similar but rather stars whose spot properties are drawn from the same parent distribution. The parameters of this parent distribution are the ones we can constrain; they are also usually the physically interesting ones, such as typical spot sizes or active latitudes and the variance in those quantities across the population. Ensembles may thus comprise light curves of stars in a narrow spectral type, metallicity, and rotation period bin, which we might reasonably expect to have statistically similar surfaces. We encourage readers to read Paper I to better understand this and other points regarding the information theory behind stellar rotational light curves.

The present paper is organized as follows. We present an overview of the derivation of the GP in Section 2 and a suite of
tests on synthetic data to show that the model is calibrated in Section 3. We discuss our results and the limitations of our model in Section 4 and present straightforward extensions of the GP, including its application to time-variable surfaces, in Section 5. In Section 6, we summarize our results and discuss topics we will address in future papers in this series.

Most of the math behind the algorithm is presented in the Appendices, followed by a series of supplementary figures (discussed in Section 3). Appendix A discusses the notation we adopt throughout the paper, and Table 2 lists the main symbols and variables, with links to their definitions. The algorithm developed in this paper is fully implemented in the `starry_process` code, which is available on GitHub and described in more detail in Luger et al. (2021a).

Finally, we note that all of the figures in this paper were autogenerated using the Azure Pipelines continuous integration service, which ensures that they are up to date with the latest version of the `starry_process` code. In particular, icons next to each of the figures link to the exact script used to generate them to ensure the reproducibility of our results. As in Paper I, the principal equations are accompanied by “unit tests”: `pytest`-compatible test scripts associated with the principal equations that pass (fail) if the equation is correct (wrong), in which case a clickable checkmark (×) is shown next to the equation label. In most cases, the validity of an equation is gauged by comparison to a numerical solution. Like the figure scripts, the equation unit tests are run on Azure Pipelines upon every commit of the code.

### 2. A GP for Starspots

In this section, we provide a brief overview of GPs (Section 2.1) and spherical harmonics (Section 2.2), followed by an outline of the derivation of our interpretable GP (Section 2.3). This derivation boils down to computing the mean and covariance of the stellar flux conditioned on certain physical properties of the star and its starspot distribution. In Sections 2.4 and 2.5, we derive useful extensions of the model. For convenience, we summarize the results of this entire section in Section 2.6. Most of the math is folded into the Appendices for readability; readers may want to refer to Appendix A in particular for a discussion of the notation and conventions we adopt.

#### 2.1. Brief Overview of GPs

Despite whatever mystique the words “Gaussian process” may evoke, a GP is nothing but a Gaussian distribution in many (formally infinite) dimensions. Specifically, it is a Gaussian distribution over functions spanning a continuous domain (in our case, the time domain). Similar to a multivariate Gaussian, which is described by a \((K \times 1)\) vector \(\mu\) characterizing the mean of the process and a \((K \times K)\) matrix \(\Sigma\) characterizing its covariance, a GP is fully specified by a mean function \(m(t)\) and a kernel function \(k(t, t')\). To say that a random vector-valued variable \(\mathbf{f}\) defined on a \((K \times 1)\) time array \(t\) is “distributed as a GP” means that we may write

\[
\mathbf{f} \sim \mathcal{N}(\mu, \Sigma),
\]

where the elements of the mean and covariance are given by \(\mu_i = m(t_i)\) and \(\Sigma_{ij} = k(t_i, t_j)\), respectively. Because of this relationship to multivariate Gaussians, GPs are easy to sample from. But, as we alluded to earlier, the real showstopper is the application of GPs to inference problems. Multivariate Gaussian distributions have a closed-form (marginal) likelihood function, so it is easy to compute the probability of one’s data conditioned on a given value of \(\mu\) and \(\Sigma\) (i.e., the “likelihood”; see Equation (14) below). This can, in turn, be maximized to infer the optimal values of the model parameters or used in a numerical sampling scheme to compute the probability of those parameters given the data (i.e., the “posterior”). Thanks to modern computer architectures, linear algebra packages, and GP algorithms, evaluating the GP likelihood may typically be done in a fraction of a second for a reasonably sized data set (i.e., \(K \lesssim 10^4\) data points).

Another big advantage of GPs is their flexibility. The GPs are often dubbed a class of “nonparametric” models, given that nowhere in the specification of the GP is there an explicit functional form for \(F\). Rather, a GP is a stochastic process whose draws can, in principle, take on any functional form, subject, however, to certain smoothness and correlation criteria of tunable strictness that are fully encoded in the covariance \(\Sigma\). In many applications, particularly when modeling stellar light curves, it is customary to restrict the problem by assuming that the process is stationary, such that we may write

\[
\Sigma_{ij} = k(t_i, t_j) = k(t_i - t_j) = k(\Delta t).
\]

A stationary process is one that is independent of phase (or, in this case, the actual value of the time \(t\)); rather, it depends only on the difference between the phases of two data points. The kernel of a stationary process is therefore a one-dimensional function, typically chosen from a set of standard functions with desirable smoothness and spectral properties.

The GP we derive in this paper is stationary and admits a representation as a one-dimensional kernel function. However, as we show in Section 2.5, the common practice of normalizing stellar light curves to their mean or median value breaks this stationarity. For this reason, it is more convenient to derive and present our GP covariance as a \((K \times K)\) matrix \(\Sigma\) and our GP mean as a \((K \times 1)\) vector \(\mu\) for arbitrary \(K\) instead of as a kernel and a mean function. Note, importantly, that these representations are equivalent given the definitions above.

#### 2.2. Spherical Harmonics

Before we dive into the computation of our GP, it is useful to introduce the spherical harmonics, a set of orthogonal functions on the surface of the sphere that we will use to describe the intensity field on the surface of a star (Figure 1). As we will see below, the spherical harmonics are a particularly convenient

\[\text{link to the exact script used to generate the figures in this paper were...}\]

\[\text{Cached data sets used in the figure generation are available on Zenodo:...}\]

\[\text{https://github.com/rodluger/starry_process}\]

\[\text{8 Cached data sets used in the figure generation are available on Zenodo:...}\]

\[\text{https://github.com/rodluger/starry_process}\]

\[\text{9 In this paper, we will use blackboard font (i.e., \(f\)) to denote random variables and serif font (i.e., \(f\)) to denote particular realizations of those variables. See Appendix A for a detailed explanation of our notation.}\]

\[\text{10 Given a one-dimensional array \texttt{mean} and a two-dimensional array \texttt{cov} in Python, sampling from the corresponding GP (if it exists) can be done in a single line of code by calling...}\]
basis on which to describe starspot distributions, as they will allow us to compute moments of the intensity distribution analytically. Of more immediate concern, Luger et al. (2019) showed that there is a linear relationship between the spherical harmonic expansion of a stellar surface and the total disk-integrated flux $f$ (i.e., the light curve) one would observe as the star rotates about a fixed axis. If the stellar surface intensity is described by a spherical harmonic coefficient vector $y$ (up to a certain degree $l_{\text{max}}$), the flux is given by

$$f = 1 + \mathcal{A}(l, P, u)y,$$

where $1$ is the ones vector and $\mathcal{A}$ is the starry design matrix, a purely linear operator that transforms from the spherical harmonic basis to the flux basis; it is a function of the stellar inclination $I$, the stellar rotation period $P$, and the stellar limb-darkening coefficients $u$, as well as the observation times (see Appendix B for details).

### 2.3. Computing the GP

Let $f = (f_0 f_1 \cdots f_{k-1})^T$ denote a random vector of $K$ flux measurements at times $t = (t_0 t_1 \cdots t_{k-1})^T$, defined in units such that a star with no spots on it will have unit flux. Conditioned on the stellar inclination $I$, the rotational period $P$, a set of limb-darkening coefficients $u$, and certain properties of the starspots $\theta$, (including their number, sizes, positions, and contrasts), we wish to compute the mean $\mu(l, P, u, \theta)$ and covariance $\Sigma(l, P, u, \theta)$ of $f$. Together, these specify a multidimensional Gaussian distribution, which we assume fully describes how our flux measurements are distributed:

$$f(l, P, u, \theta) \sim \mathcal{N}(\mu(l, P, u, \theta), \Sigma(l, P, u, \theta)).$$

As with any random variable, the mean and covariance may be computed from the expectation values of $f$ and $f^T$, respectively:

$$\mu(l, P, u, \theta) = E[f | l, P, u, \theta],$$

$$\Sigma(l, P, u, \theta) = E[f f^T | l, P, u, \theta] - \mu(l, P, u, \theta)\mu^T(l, P, u, \theta).$$

Given the linear relationship between flux and spherical harmonic coefficients (Equation (3)), we may write the mean and covariance of our GP as

$$\mu(l, P, u, \theta) = 1 + \mathcal{A}(l, P, u)\mu_\gamma(\theta),$$

$$\Sigma(l, P, u, \theta) = \mathcal{A}(l, P, u)\Sigma_\gamma(\theta)\mathcal{A}^T(l, P, u),$$

where

$$\mu_\gamma(\theta) = E[y | \theta]$$

and

$$\Sigma_\gamma(\theta) = E[y y^T | \theta] - \mu_\gamma(\theta)\mu_\gamma^T(\theta).$$

---

11 There are, of course, drawbacks to using this basis; in particular, the spherical harmonics are smooth, continuous functions that struggle to capture high-resolution features such as small starspots. We discuss this point at length in Section 4.1, where we show that our model is useful even when applied to stars with spots smaller than the effective resolution of the GP.

12 Note, importantly, that these are not the units we observe in. See Paper I and Section 2.5 below.

13 The true distribution of stellar light curves conditioned on $l, P, u,$ and $\theta$, is not Gaussian, so our assumption is formally wrong. But, as the saying goes, “all models are wrong; some are useful.” As we will show later, this turns out to be an extremely useful assumption.
are the mean and covariance of the distribution over spherical harmonic coefficient vectors \( y \). The bulk of the math in this paper (Appendix C) is devoted to computing the expectations in the expressions above, which are given by the integrals

\[
E[y | \theta] = \int y(x) p(x | \theta) dx
\]

and

\[
E[y y^T | \theta] = \int y(x)y^T(x) p(x | \theta) dx,
\]

where \( x \) is a random vector-valued variable corresponding to a particular distribution of features on the surface and \( p(x | \theta) \) is its probability density function (PDF). In Appendix C, we show that for suitable choices of \( \theta, y(x), \) and \( p(x | \theta) \), the integrals in the expressions above have closed-form solutions that may be evaluated quickly. While we present a few different ways of specifying \( \theta \), our default representation of the GP hyperparameters is

\[
\theta = (n \ c \ \mu_\phi \ \sigma_\phi \ r \ Y),
\]

where \( n \) is the number of starspots; \( c \) is their contrast (defined as the intensity difference between the spot and the background intensity as a fraction of the background intensity); \( \mu_\phi \) and \( \sigma_\phi \) are the mode and standard deviation of the spot latitude distribution, respectively; and \( r \) is the radius of the spots. For simplicity, the PDFs for the spot radius, spot contrast, and number of spots are chosen to be delta functions centered at \( r, c, \) and \( n \), respectively (Appendices C.1 and C.4), while the spot longitude is assumed to be uniformly distributed (Appendix C.3). Finally, the PDF for the latitude \( \phi \) of the spots is chosen to be a beta distribution in \( \cos \phi \) with (normalized) parameters \( a \) and \( b \), which have a one-to-one correspondence to the mode \( \mu_\phi \) and standard deviation \( \sigma_\phi \) of the distribution in \( \phi \) (Appendix C.2). This allows us to model starspot distributions with “active latitudes” of tunable width that are symmetric about the equator. The distribution is flexible enough to also model equatorial and isotropically distributed spots. Stars with multiple active latitudes can easily be modeled as a sum of GPs (Section 5.1). These choices for the spatial distribution of spots are based on the Sun, whose spots emerge in azimuthally symmetric belts at roughly the same latitude in both hemispheres, then migrate toward the equator over the course of the 11 yr cycle (Solanki et al. 2006).

In this paper, we assume that the parameters \( \theta \), described above are the physically interesting ones. That is, given a light curve \( f \) or an ensemble of \( M \) light curves of statistically similar stars \((f_0 \ f_1 \ \cdots \ f_{M-1})^T\), we wish to infer the statistical properties of the starspots, encoded in the entries of the vector \( \theta \). This is typically a tall order, since it requires marginalizing over all of the nuisance parameters, which include the nitty-gritty details of the size, contrast, and location of every spot (and, if \( M > 1 \), on every star in the ensemble). Fortunately, however, the GP we constructed does just that. Specifically, given the mean and covariance of the process, we are able to directly evaluate the log marginal likelihood of the \( m \)th data set conditioned on a specific value of \( \theta \), (as well as \( I, P, \) and \( u \)),

\[
\ln L_m(I, P, u, \theta) = -\frac{1}{2} r_m^{-1}(I, P, u, \theta) \cdot (\Sigma(I, P, u, \theta) + C_m)^{-1} r_m(I, P, u, \theta) - \frac{1}{2} \ln |\Sigma(I, P, u, \theta) + C_m| - K \ln (2\pi),
\]

where

\[
r_m(I, P, u, \theta) \equiv f_m - \mu(I, P, u, \theta)
\]

is the residual vector, \( C_m \) is the data covariance (which in most cases is a diagonal matrix whose entries are the squared uncertainty corresponding to each data point in the light curve), \( | \cdot | \) denotes the determinant, and \( K \) is the number of data points in each light curve. In an ensemble analysis, the joint marginal likelihood of all data sets is simply the product of the individual likelihoods, so in log space, we have

\[
\ln L(I, P, u, \theta) = \sum_m \ln L_m(I, P, u, \theta).
\]

The marginal likelihood may be interpreted as the probability of the data given the model. Typically, we are interested in the reverse: the probability of the model given the data, i.e., the posterior probability distribution. In later sections, we present a comprehensive suite of posterior inference exercises demonstrating that our GP model is correctly calibrated, allowing one to efficiently infer the statistical properties of starspots from light curves with minimal bias.

### 2.4. Marginalizing over Inclination

As we mentioned in the previous section, the equations for the mean and covariance of our GP (Equations (7) and (8), respectively) are conditioned on specific values of the stellar and spot properties. To obtain the posterior distribution for these parameters, we must typically resort to numerical sampling techniques, which often scale steeply with the number of parameters. It is therefore generally desirable to keep the total number of parameters small, especially when employing the GP in an ensemble setting. In such a setting, we might have light curves from \( M \) stars, all of which we believe to have similar spot properties (perhaps because they have similar spectral types and rotation periods, for example). The total number of parameters in our problem is therefore

\[
N = 4M + 5,
\]

since each of the stars will have their own set of four stellar properties (an inclination, a period, and usually two limb-darkening coefficients) but all will share the same five spot properties \( \theta \) (by assumption). For a reasonably sized ensemble of \( M = 100 \) stars, we would have to sample over \( N = 405 \) parameters. While large, this number is certainly not absurd, especially by modern standards. However, it does pose a problem when considering how complex the posterior

---

14 In Equation (14), we implicitly assume that all stars in the ensemble are observed at the same set of times \( t \). If this is not the case, the mean and covariance of the GP for each star must be computed from Equations (5) and (6) with the flux design matrix \( A(t_o) \) evaluated at the particular observation times \( t_o \).
is the random variable corresponding to the inclination. \( \Phi \) and \( \phi \) are given by Equations 19 and 20, respectively, and \( \Phi \) is the inclination-second-moment integral where we define the inclination first-moment integral

\[
e_i = \int \mathcal{A}(\mathbb{I}, P, u) E[y | \Theta] \, \rho(\mathbb{I}) \, d\mathbb{I}
\]

and the inclination second-moment integral

\[
E_i = \int \mathcal{A}(\mathbb{I}, P, u) E[y \, \gamma | \Theta]
\cdot \mathcal{A}(\mathbb{I}, P, u) \, \rho(\mathbb{I}) \, d\mathbb{I},
\]

and \( \mathbb{I} \) is the random variable corresponding to the inclination. The expectations inside the integrals in the expressions for \( e_i \) and \( E_i \) are given by Equations (11) and (12), respectively, and computed in Appendix C. If we are able to perform the integrals in those expressions, we can dramatically reduce the number of parameters in our ensemble problem. As we show in Section 2.4, if we assume that stellar inclinations are distributed isotropically, these integrals do in fact have closed-form solutions.

Finally, for future reference, it is useful to note that the mean of the GP is constant,

\[
\mu(P, u, \Theta) = (1 + e_i) \mathbf{1} = \mu \mathbf{1},
\]

since, by construction, our GP is longitudinally isotropic (see Appendix D.2).

### 2.5. Normalization Correction

In Paper I, we discussed a subtle but important point about stellar light curves: the common procedure of normalizing light curves to their mean or median level changes the covariance structure of the data, since it correlates all of the observations in a nontrivial way. When normalizing a light curve by the mean,\(^{15}\) the operation we perform is

\[
\tilde{f} = \frac{f}{\langle f \rangle},
\]

where \( \tilde{f} \) is the normalized unit-mean light curve, \( f \) is the measured light curve (in detector counts), and \( \langle f \rangle \) is the sample mean, i.e., the average value of a given star’s light curve (which we model as a sample from our GP). This may be close to but is in general different from the process mean, \( \mu(P, u, \Theta) \), since the mean of a draw from the GP is itself normally distributed with a variance that scales with the GP variance.\(^{16}\)

When modeling normalized light curves, we must correct our expression for the covariance matrix \( \Sigma \) of the GP. Computing the new covariance matrix is tricky, especially because the normalized process is not strictly Gaussian; the distribution of normalized light curves has heavy tails due to the fact that \( \tilde{f} \) diverges as the sample mean approaches zero. In fact, because of these tails, the covariance of the normalized process is formally infinite, since the probability of drawing a sample whose mean is arbitrarily close to zero is finite. If this is all starting to sound like a bad idea, that is because it is. A much safer approach is to resist the temptation to normalize the light curve and instead model the (unknown) amplitude of the variability as a multiplicative latent variable. However, this would require an extra parameter for every light curve, so the computational savings we achieved by marginalizing out the inclination would be gone. Fortunately, in practice, the variance of a stellar light curve is usually small compared to its mean; stellar variability amplitudes are typically at the level of a few percent or lower. When this is the case, the probability of drawing a GP sample whose mean is close to zero is extremely small, and we can make use of the approximate expression derived in R. Luger (2021, in preparation) for the covariance of a normalized GP.

\[
\tilde{\Sigma} \approx \frac{A}{\mu^2} \Sigma + z((A + B) (1 - q) \\
\times (1 - q) - A \, q \, q^T),
\]

where

\[
z \equiv \frac{\langle \Sigma \rangle}{\mu^2}
\]

is the ratio of the average element in \( \Sigma \) to the square of the mean of the GP. \( q \) is the ratio of the average of each row in \( \Sigma \).

\(^{15}\) In practice, the expressions derived here also work well for median-normalized light curves, since the distribution of the GP sample median is usually close to the distribution of the sample mean.

\(^{16}\) Importantly, the sample and process means will be different even in the absence of measurement error. In other words, the mean flux of a given star (i.e., the sample mean) will generally be different from the mean flux across all stars with similar surface properties (the process mean).
to the average element in $\Sigma$, and $A, B$ are order unity and zero scalars, respectively, given by the optimally truncated diverging series

$$A \equiv \sum_{i=0}^{i_{\text{max}}} \frac{(2i+1)!}{2^i i!} z^i,$$

$$B \equiv \sum_{i=0}^{i_{\text{max}}} \frac{2i(2i+1)!}{2^i i!} z^i,$$

where $i_{\text{max}}$ is the largest value for which the series coefficient at $i_{\text{max}}$ is smaller than the coefficient at $i_{\text{max}} - 1$. In the expressions above, it is assumed that the mean $\mu$ is constant, i.e., $\mu = \mu \mathbf{1}$. Since our GP is azimuthally isotropic (i.e., no preferred longitude), that is the case throughout this paper.

What Equation (24) allows us to do is effectively marginalize over the unknown normalization by modeling the normalized flux as a draw from a GP:

$$\tilde{f}(P, \mathbf{u}, \mathbf{\theta}) \sim \mathcal{N}(1, \tilde{\Sigma}(P, \mathbf{u}, \mathbf{\theta})).$$

(28)

This is appropriate as long as $z \ll 1$, for which the true distribution of $\tilde{f}$ is approximately Gaussian. In practice, we recommend employing this trick only for $z \lesssim 0.02$, for which the error in the approximation to the covariance is less than $10^{-6}$. In cases where the light-curve variability exceeds about 10%, we recommend modeling the multiplicative amplitude in each light curve as a latent variable, as discussed above.

Figure 2 shows an example of a covariance matrix normalized according to the procedure outlined above. The principal differences between the normalized and original covariances are an overall scaling and a small offset. However, the normalization also results in the process becoming nonstationary; the covariance between two points in a light curve is now slightly dependent on their phases.

### 2.6. Summary

As the computation of the GP relies on many interdependent equations scattered throughout the previous sections and the Appendices, it is useful to summarize the procedure for the case where we marginalize over the inclination (Section 2.4) and the light curves are normalized to their means (Section 2.5), which is likely to be the primary use case for our algorithm.

We model the mean-normalized flux $\hat{f}$ (Equation (23)) as a GP:

$$\tilde{f}(P, \mathbf{u}, \mathbf{\theta}) \sim \mathcal{N}(1, \tilde{\Sigma}(P, \mathbf{u}, \mathbf{\theta})).$$

(29)

The hyperparameters of the GP are the stellar rotation period $P$, the vector of limb-darkening coefficients $\mathbf{u}$, and the vector of parameters describing the spot distribution

$$\mathbf{\theta} = (n \ c \ \mu_0 \ \sigma_0 \ r)^\top,$$

(30)

consisting of the number of spots $n$, their contrast $c$ (the fractional intensity difference between the background and the spot), the mode $\mu_0$ and standard deviation $\sigma_0$ of the latitude distribution, and the radius of the spots $r$. The quantity $\tilde{\Sigma}$ is the covariance of the normalized process (Equation (24)), which is a straightforward correction to the true covariance of the process, accounting for changes in scale and phase introduced by the common process of normalizing light curves to a mean of unity. It depends on the true (constant) mean $\mu$ and true covariance $\Sigma$, given by Equations (18) and (19), respectively. Those expressions, in turn, depend on the inclination expectation integrals $e_I$ (Appendix D.2) and $E_I$ (Appendix D.3). Those, in turn, depend on the first and second moments of the distribution of the spherical harmonic coefficient vectors, $E[y | \mathbf{\theta}]$ and $E[y y^\top | \mathbf{\theta}]$, given by Equations (C8) and (C9), respectively. To compute those, we must evaluate four nested integrals (Equations (C10)–(C13) for the first moment and (C14)–(C17) for the second moment) corresponding to integrals over the radius, latitude, longitude, and contrast distributions. The computation of these integrals is discussed at length in Appendix C.
While lengthy (and quite tedious), all of the computations described above rely on equations whose solutions have a closed form.\footnote{The exception to this is the normalization correction (Section 2.5), which depends on a fast-to-evaluate series and thus adds negligible overhead to the computation.} Moreover, most of the terms in the expectation vectors and matrices may be computed recursively, and many may be precomputed, as they do not depend on user inputs. It is therefore possible to evaluate $\Sigma$ in an efficient manner. In the companion paper (Luger et al. 2021a), we discuss our implementation of the algorithm in a user-friendly Python package.

### 2.7. An Example

A concrete example of the GP derived above is presented in Figure 3, where we show random samples from the process evaluated up to spherical harmonic degree $l_{\text{max}} = 15$ and conditioned on different values of the hyperparameter vector $\theta$. Each column corresponds to a different random draw from the GP, while each row corresponds to a different value of $\theta$.

The images are intensity maps of the stellar surface seen in an equal-area Mollweide projection, in units such that a spotless star would have an intensity equal to 1 everywhere. Below them are the corresponding light curves (in units of parts per thousand deviation from the mean) over four rotation cycles, seen at inclinations varying from 15° (yellow) to 90° (dark blue), and assuming no limb darkening (i.e., $u = 0$). From top to bottom, the GP corresponds to a star with (a) small midlatitude spots, (b) small circumpolar spots, (c) large equatorial spots, and (d) small isotropic spots. See text for details.

**Figure 3.** Five random samples from the GP prior (columns) conditioned on four different values of $\theta$, (rows). The samples are shown on the surface of the star in a Mollweide projection (upper panels) alongside their corresponding light curves viewed over four rotation cycles at several different inclinations (lower panels). From top to bottom, the GP corresponds to a star with (a) small midlatitude spots, (b) small circumpolar spots, (c) large equatorial spots, and (d) small isotropic spots. See text for details.

\[
(\nu c \mu_o \sigma_s r) = (10.0 \ 0.10 \ 30.0 \ 5.00 \ 10.0)^	op, \quad (31a)
\]
\[
(10.0 \ 0.10 \ 60.0 \ 5.00 \ 10.0)^	op, \quad (31b)
\]
\[
(10.0 \ 0.05 \ 0.00 \ 5.00 \ 30.0)^	op, \quad (31c)
\]
\[
(20.0 \ 0.10 \ 0.00 \ 40.0 \ 10.0)^	op. \quad (31d)
\]

These correspond to (a) 10 spots of radius $10^\circ$ centered at $30^\circ \pm 5^\circ$ latitude with a contrast of 10%, (b) 10 spots of radius $10^\circ$ centered at $60^\circ \pm 5^\circ$ latitude with a contrast of 10%, (c) 10
spots of radius 30° centered at 0° ± 5° latitude with a contrast of 5%, and (d) 20 spots of radius 10° centered at 0° ± 40° latitude (a good approximation to a perfectly isotropic distribution; see Appendix C.2) with a contrast of 10%.

The surface maps in the figure show dark, compact features of roughly the expected size and contrast and at the expected latitudes. However, there are important differences between these maps and what we would obtain by procedurally adding discrete circular spots to a gridded stellar surface.

1. The spots are not circular. This is most evident in row (c), where some spots are distinctly asymmetric.

2. There is significant variance in contrast from one spot to another, even though our model implicitly treats \( c \) as constant. Within spots, the contrast is also not constant, even though (again) our model implicitly treats it as such.

3. There is ringing in the background. This is apparent to some extent in row (b), where there are small fluctuations in the brightness at low latitudes where no spots are present.

4. There are not exactly 10 (or 20) spots in those maps. This is most obvious in row (c), where only a few large distinct spots, plus maybe a few smaller ones, are visible.

5. There are bright spots in addition to dark spots. This may be the most glaring issue. We explicitly model spots as being dark, and yet there are (almost) just as many bright spots, particularly in row (d). While bright spots (such as plages) certainly exist in reality, we did not explicitly ask for them here.

While these may appear to be critical shortcomings of our model, it is important to keep in mind that a model consisting of discrete, circular, constant contrast spots is likely just as far (or perhaps even farther) from the truth. In fact, points (1) and (2) above suggest that our model is more flexible than the discrete spot model and thus (potentially) better suited to modeling real stellar surfaces. Points (3), (4), and (5), on the other hand, are more concerning, since they are due, respectively, to a truncation error in the spherical harmonic expansion, an intrinsic limitation of our Gaussian approximation, and the fact that Gaussian distributions are symmetric about the mean; a positive deviation is just as likely as a negative deviation of the same magnitude.\(^\text{18}\) However, as we have argued before, the true power of the GP is in its applicability to inference problems. In other words, while our GP has some undesirable features when used as a prior sampling distribution, the real test of the GP is when it is faced with data in an inference setting. As long as the data are sufficiently informative, it does not matter that the prior has finite support for unphysical configurations, as those will be confinedly rejected.

In Section 3 below, we exhaustively test the performance of our GP as an inference tool when used to model synthetic light curves. We will show that, despite the issues raised above, the GP is generally unbiased and correctly estimates the posterior variance when used to infer the spot properties \( \theta \).

But before we dive into calibration tests, it is worth pausing for a moment to take another look at Figure 3. While we focused on the shortcomings of the GP as a prior in the discussion above, it is important to appreciate that it even works in the first place. A GP, after all, is a nonparametric process describing a smooth and continuous function only via its covariance structure. The GP knows nothing about the existence of discrete spots, only how any two points on the surface are correlated. Because spherical harmonics are smooth functions with support over the entire sphere, the GP also does not know about features restricted to certain latitudes; in fact, in most applications of GPs to mapping problems in astronomy (such as in models of the cosmic microwave background, CMB; e.g., Wandelt 2012), the process is assumed to be isotropic, with no preferred angular direction. However, by prescribing the correct structure to the covariance matrix, we are able to approximately model compact spotlike features with given sizes and restricted to particular latitudes.

3. Calibration

3.1. Why We Need Calibration Tests

In the previous section, we derived a closed-form solution to the Gaussian approximation to the distribution of stellar surfaces (and their corresponding light curves) conditioned on a vector \( \theta \) of spot hyperparameters. As we mentioned, the real test of this GP is in how well it performs as a likelihood function for stellar light curves.

It is not immediately obvious that the GP approach should work, because the true marginal-likelihood function \( p(\tilde{y} | P, u, \theta) \) is certainly not Gaussian. To see why, let us generate \( 10^4 \) stellar surfaces sampled from the true distribution we are trying to model, that is, a surface with \( n = 5 \) discrete circular spots of fixed, uniform contrast \( c = 0.1 \) and radius \( r = 20^\circ \) at latitudes \( \mu_0 \pm \sigma_\phi = 30^\circ \pm 5^\circ \). Let us then expand each surface in spherical harmonics and visualize the distribution of coefficients \( y \). Figure 4 shows the joint distribution for five of the coefficients with the most non-Gaussian marginal distributions (selected by eye). Different slices through this distribution (in black) are skewed, strongly peaked, nonlinearly correlated, and even bimodal. Our approach in this paper is to model this distribution as a Gaussian (orange contours). While this may be a good approximation in certain regions of parameter space, it is certainly a poor approximation in others. In this section, we will show that, fortunately, the non-Gaussianity of the distribution is not generally an issue when doing inference with our GP, as the resulting posteriors are correctly calibrated.

3.2. Setup

We seek to demonstrate that our model is correctly calibrated by testing it on synthetic data, which we generate as follows. For each of the \( M \) synthetic light curves in a given ensemble, we create a rectangular \((150 \times 300)\) latitude–longitude grid of surface intensity values, all initialized to zero. We then add \( n \) spots to this grid, each of a fractional contrast \( c \) and radius \( r \) centered at latitude \( \phi \) and longitude \( \lambda \), by decreasing the intensity at all points within an angular distance \( r \) (measured along the surface of the sphere) by an amount \( c \). In order to compute the corresponding light curve, we expand the surface in spherical harmonics, although at a much higher degree \( l_{\text{max}}^{(0)} = 30 \) than the degree we will use in the inference step \( l_{\text{max}} = 15 \) to minimize potential ringing effects or other artifacts in the synthetic data. For reference, the chosen degree \( l_{\text{max}} = 30 \) is large enough to resolve features on the order of

\(^{18}\) Even still, the model favors dark spots over bright spots because the GP mean itself is lower than unity, in practice making positive deviations from unity less likely than negative deviations. This is why there are usually more dark spots than bright spots in the samples shown in the figure.
\[ 180^\circ/30 = 6^\circ \] across but small enough that the algorithm for computing the light curve is numerically stable.\(^{19}\) We compute the light curve at \( K \) points equally spaced over a baseline \( \Delta t \) using the starry algorithm (Appendix B), assuming an inclination \( \theta \), a rotational period \( P \), and limb-darkening coefficients \( u \). Finally, we divide the light curve by the mean and add Gaussian noise with a standard deviation \( \sigma_f \) to emulate photon noise.\(^{20}\) The default values/distributions of each of the

\(^{19}\) A more principled approach would perhaps be to generate light curves using a completely different model, such as by discretizing the surface at very high resolution and computing the flux via a weighted sum of the visible pixels. However, this would take orders of magnitude longer than the adopted approach and still be subject to artifacts due to the discretization scheme. We went to great lengths in Luger et al. (2019) to show that our flux computation from spherical harmonics is both accurate and precise, so we are confident that our synthetic light curves correctly represent the assumed spot distributions.

\(^{20}\) In theory, we should do this in the reverse order; we should add photon noise and then normalize the light curve to the mean, as that is the order in which those steps occur in reality. However, if we did that, we would have to normalize \( \sigma_f \) in our inference step, such that the variance of each of the normalized light curves in the ensemble would be different, requiring us to invert a different matrix for each light curve when computing the log likelihood (see Equation (14)). This would significantly increase the computational cost of our tests. Fortunately, in practice, the difference between these two approaches has a negligible effect on our results, so we opt for the faster of the two methods. Note, of course, that when applying our GP to real data, we will not have this choice.
parameters mentioned above are given in Table 1. Some, like the number of spots, their contrast, etc., are drawn from fiducial distributions, while others, like the photometric uncertainty, rotational period, and limb-darkening coefficients, are fixed across all M light curves in an ensemble. These fixed values are not realistic, but they greatly speed up the inference step, since they allow us to invert a single covariance matrix for all light curves when computing the log likelihood.

Figure 5 shows a synthetic data set generated from the default values listed in Table 1. While the light curves correspond to surfaces with the same statistical spot properties, they all look qualitatively different; the mapping of starspot properties to flux is nontrivial. In the inference step below, we assume that we observe only these 50 light curves (the figure only shows 49 of them), with no knowledge of the inclination of any individual star, and attempt to infer the spot properties.

### 3.3. Inference

We use our Python-based implementation of the GP (Luger et al. 2021a) to perform inference on the synthetic data sets. For simplicity, we assume we know the value of the period $P$, which is fixed at unity for all stars, as well as the value of the limb-darkening coefficients (fixed at zero for the default run). In practice, these will not be known exactly; we discuss this further in Section 5.5. Since we explicitly marginalize over the inclinations of individual stars, the only quantities we must constrain are the five parameters in the spot parameter vector $\vartheta$. (Equation (13)). We experimented with three different methods for doing posterior inference on our synthetic data sets: no-U-turn sampling, a variant of HMC (Duane et al. 1987; Hoffman & Gelman 2011); automatic differentiation variational inference (ADVI; Blei et al. 2016; Kucukelbir et al. 2016); and nested sampling (Skilling 2004, 2006). We obtained the best performance using the nested-sampling algorithm implemented in the dynasty package (Speagle 2020), so that is what we will use below.

Our sampling parameters are the number of spots $n$, their contrast $c$, their radius $r$, and the beta distribution parameters $a$ and $b$ describing their distribution in latitude. As we discuss in Appendix C.2, the parameters $a$ and $b$ are easier to sample in than the mode $\mu_\varphi$ and standard deviation $\sigma_\varphi$ provided we account for the Jacobian of the transformation (Equation (C62)) in our log probability function, which maps a uniform prior on $a$ and $b$ to a uniform prior on $\mu_\varphi$ and $\sigma_\varphi$.

We place uniform priors on all five quantities, with support in $1 \leq n \leq 50$, $0 < c \leq 1$, $10^{-3} \leq r \leq 30^\circ$, $0 \leq a \leq 1$, and $0 \leq b \leq 1$. Note that while $n$ formally represents an integer, its effect on the GP is a scaling of the covariance (see Equation (C9)); as such, it has support over all real numbers within the bounds listed above. We could restrict it to integer values, but this would make sampling quite tricky. Moreover, in practice, it is useful to allow for noninteger values to add some flexibility to the model; we discuss this in more detail in Section 4.4.

We use Equation (14) as our log-likelihood term, adding the log of the absolute value of Equation (C62) to enforce a uniform prior on $\mu_\varphi$ and $\sigma_\varphi$. As we mentioned above, the fact that $P$, $u$, and $\sigma_f$ are shared among all $M$ light curves means that $\Sigma + C_m$ is the same for all of them, greatly speeding up the likelihood evaluation, since we need only invert (or factorize) it a single time per sample. Our covariance is the covariance of the normalized process, given by Equation (24). Since we only consider light curves with variability limited to a few percent or less, the approximation for $\Sigma$ is always valid. Finally, we restrict our spherical harmonic expansion to $l_{\text{max}} = 15$ as a compromise between resolution, computational speed, and numerical stability (see Luger et al. 2021a).

We use the standard implementation of the nested sampler, dynasty.NestedSampler, with all arguments set to their default values (multi-ellipsoidal decomposition for bounds determination (Feroz et al. 2009), uniform sampling within the bounds, 500 live points, and no gradients) to perform our inference step. Convergence—defined as when the estimate of the remaining evidence $\ln Z$ drops below 0.5—is usually attained after 5000–10,000 samples and within a couple hours on a typical machine for most of the trials we perform.

Below, we describe several calibration runs: experiments where we generate an ensemble of light curves from synthetic stars with given properties (Section 3.2) and attempt to infer their statistical spot properties.

### 3.4. Default Run

The input parameters for the default run are shown in Table 1, and the corresponding light curves are shown in Figure 5. We run the nested sampler as described in the previous section and transform the posteriors in $\varphi$ and $\sigma$ via Equations (C53), (C57), and (C59). The results are shown in Figure 6, where we correctly infer all five parameters within two to three standard deviations. Posterior distributions for the spot radius $r$, central spot latitude $\mu_\varphi$, and spot contrast $c$ are fairly tight, while the distribution for the latitudinal scatter $\sigma_\varphi$ has wider tails, and the distribution for the number of spots $n$ is very poorly constrained. The latter, in particular, is degenerate with the spot contrast $c$; we discuss this at length in Paper I.

Figure 7 shows samples from the spot latitude posterior (hyper)distribution. Since the parameters $\mu_\varphi$ and $\sigma_\varphi$ characterize a distribution over spot latitudes, uncertainty in their values translates to uncertainty in the actual shape of the spot latitude distribution. Thus, the collection of blue curves in Figure 7 quantifies our knowledge of how spots are distributed on the surfaces of the stars in the data set. These distributions are

---

Luger, Foreman-Mackey, & Hedges

| Table 1                                      |
|---------------------------------------------|
| Default Parameters Used to Generate Synthetic Light Curves in the Calibration Tests |
| Symbol          | Description             | Value                     |
|-----------------|-------------------------|---------------------------|
| $n$             | Number of spots         | $\sim \mathcal{N}(20, 0^2)$ |
| $c$             | Spot contrast           | $\sim \mathcal{N}(0.05, 0^2)$ |
| $\phi$          | Spot latitude           | $\sim \mathcal{N}(30^\circ, 5^\circ)$ |
| $\lambda$       | Spot longitude          | $\sim [0^\circ, 360^\circ]$ |
| $r$             | Spot radius             | $\sim \mathcal{N}(15^\circ, 0^2)$ |
| $I$             | Stellar inclination     | $\sim \sin$               |
| $P$             | Rotational period       | 1 day                     |
| $u$             | Limb-darkening coefficients | $(0, 0^2)$                |
| $\sigma_f$      | Photometric uncertainty | $10^{-3}$                 |
| $K$             | Number of cadences per light curve | $10^3$               |
| $\Delta t$      | Time baseline           | 4 days                    |
| $M$             | Number of light curves in ensemble | 50                |
Figure 5. Synthetic data set generated by adding discrete spots to a stellar surface with parameters given by the default values listed in Table I. The surface maps for 49 synthetic stars are shown in a Mollweide equal-area projection above their corresponding light curves. All maps and light curves are plotted on the same scale. The cross on each map indicates the subobserver latitude assumed when generating the light curve. The blue curves correspond to the exact light curve, while the black dots are the observed light curve.
again consistent with the true distribution used to generate the
spots (orange curve) within less than two standard deviations. If the results in Figure 7 seem biased, recall from Figure 6 that the mean of the latitude distribution is consistent with the truth at 2σ–3σ. That is roughly the difference between the orange curve and the average of the blue curves. As we will see, inference with a larger ensemble (Figure 23) allows us to infer the mean latitude to within about 2°.

The results in Figures 6–8 are based on a single run, i.e., a single realization of the light-curve ensemble conditioned on the properties of Table 1. To properly gauge potential biases in our model, it is useful to perform the run under many different realizations of the data set. We therefore generate 100 ensembles of light curves in exactly the same way as above and perform inference on each of them. Figure 9 shows the marginal and joint posterior distributions for θ for all 100 trials. Posteriors for individual trials are shown as translucent colored curves in the top panel of each column (marginal posteriors) and as ellipses bounding the 1σ posterior level in the remaining panels (joint posteriors). The black curves in the top panel of each column show the marginal distributions of all samples across all trials, and the black contours in the remaining panels show the corresponding 1σ levels in the joint posteriors.

If our model is truly unbiased, in the limit of an infinite number of realizations of the data, the expectation value of the distribution of samples across all ensembles (the mean of the black curves in the marginal posterior plots) should coincide
with the true values (orange lines). This is approximately the case with the spot size $r$ posterior; on average, the posterior distributions are centered on the correct value. However, this is not the case for the spot latitude parameters $\mu_\phi$ and $\sigma_\phi$, for which our posterior means are biased high. While the modes of their posteriors are very close to the true values, the distributions have long tails extending to high latitudes and high variance, respectively.

The reason for this bias has to do with the normalization degeneracy discussed in Paper I; the total spotness of a stellar surface is not an observable in single-band photometry. In particular, this means that spots near the poles lie almost entirely in the null space. Applied to the problem at hand, this degeneracy makes it difficult to distinguish between stars with spots concentrated exclusively at midlatitudes (in this case, the truth) and stars with spots centered closer to the poles but with large latitudinal variance. The latter configuration leads to many spots close to the poles, whose effect on the light curve is negligible, and some spots at midlatitudes, whose effect on the light curve is similar to that of the former configuration. Thus, the data alone cannot be used to discriminate between these two scenarios, introducing the degeneracy we see in the posterior. In fact, it is clear that in the tails of the distribution, the mean spot latitude and the standard deviation of the spot latitudes are positively correlated. The bias we see is therefore not a shortcoming of the model but of the data itself. To get around this, we either need to impose stronger priors on $\mu_\phi$ and $\sigma_\phi$ (Section 4.4), observe in multiple wavelength bands (Paper I), or simply collect more data. As we will see below, the particular degeneracy described above is not perfect; for very large ($M \sim 1000$) ensembles, high-variance polar spots can confidently be ruled out.

The posteriors for the contrast $c$ and the number of spots $n$ are mostly unbiased. The contrast distribution has a bit of a tail; inspection of Figure 9 reveals that it too is positively correlated with the mean spot latitude and therefore suffers from the same degeneracy as above. And while the mean of the spot number distribution is roughly correct, the posterior is nearly unchanged across all runs and equally uninformative in all of them. This is yet another manifestation of the normalization degeneracy; the total number of spots is not an observable in single-band photometry (Paper I).

There is one final distribution that is instructive to consider: the distribution of errors on the inferred stellar inclination. Figure 10 shows a histogram of stellar inclination residuals (posterior mean minus true value) normalized to the posterior standard deviation for all stars across the 100 trials described above. For a correctly calibrated model, this distribution should equal the standard normal $\mathcal{N}(0, 1)$ in the limit of infinite trials. This, in fact, is roughly what we find (compare to the orange histogram in the figure). Our posterior has marginally heavier tails, meaning we tend to slightly underestimate the posterior variance, but in general, it is an excellent estimator of individual stellar inclinations.

Finally, Figure 11 shows the same posterior distributions as in Figure 9 but for 100 runs each with $M = 1000$ light curves. In addition to the constraints on all parameters (except the number of spots) being much tighter, the larger amount of data breaks the polar spot degeneracy discussed above. Given enough light curves, the model is capable of differentiating between concentrated midlatitude spots and high-latitude spots with large variance. Interestingly, the inferred radius appears to be biased high by a small amount. This is likely due to the fact that our prescription for generating the spots (Section 3.2) is different from how we actually model these spots. While we generate the spots as compact circular disks expanded at a high spherical harmonic degree, we model them as sigmoids (Appendix C.1) expanded at a significantly lower spherical harmonic degree. Some minor disagreement is therefore to be expected in the inferred radii.

### 3.5. Other Runs

In this section, we test the robustness of our model by changing one or more of the fiducial values listed in Table 1. Each of the runs below corresponds to a single realization of the ensemble data set, and the corresponding figures are presented in Appendix E.

Figures 17–20 show the results for different latitudinal distributions, keeping all other values in Table 1 the same. Specifically, Figure 17 corresponds to a run with midlatitude ($\mu_\phi = 45^\circ$ and $\sigma_\phi = 5^\circ$) spots, Figure 18 to a run with high-latitude ($\mu_\phi = 60^\circ$ and $\sigma_\phi = 5^\circ$) spots, Figure 19 to a run with equatorial ($\mu_\phi = 0^\circ$ and $\sigma_\phi = 5^\circ$) spots, and Figure 20 to a run with isotropically distributed spots ($\phi \sim \cos$). The results are largely consistent with those of the default run; in all cases, we infer the correct spot radius, the mean and standard deviation of the spot latitude, and the contrast distribution have $2\sigma$–$3\sigma$ confidence. The number of spots is equally unconstrained in all runs. In
Figure 18 and, to a lesser extent, Figure 17, the polar spot degeneracy discussed above is evident, particularly in the lower panels showing the latitudinal distribution of spots. Nevertheless, the distribution peaks near the correct latitude in both cases. Figures 19 and 20 are interesting because, while the true latitude distribution is unimodal, most of the posterior samples are not. In the equatorial case, the posterior peaks at very low (but nonzero) latitudes, and \( \sigma_\phi \) appears to be inconsistent with the true value at many standard deviations; however, recall that \( \sigma_\phi \) is a local approximation of the standard deviation of the PDF at the mode (Appendix C.2), which deviates from the true standard deviation (i.e., the square root of the variance, computed from the expectation of the second moment of the distribution) when the two modes are very closely spaced. In fact, the latitude PDF samples (lower panel in the figure) nearly span the true distribution, to the extent that our parameterization of the latitude distribution can approximate a zero-mean Gaussian. While the beta distribution in \( \cos \phi \) can be unimodal in \( \phi \) (see Equation (C51) and the first column of Figure 16), this happens only when \( \beta = 0 \), which occupies an infinitesimally thin hyperplane in parameter space. In practice, the majority of the posterior mass will be close to but not exactly at \( \beta = 0 \), leading to the bimodality in the figure. The same argument applies to Figure 20. In both cases, the posterior approximates the true distribution as best it can given the constraints of the adopted PDF.

Figure 21 tests the performance of the model on the light curves of stars with spots much smaller than the effective resolution of the GP. Our expansion to \( l_{\max} = 15 \) only allows us to model spots with radii \( r \gtrsim 10^4 \) (see Figure 15), so we place zero prior mass below this value. The figure shows the results of inference on a data set generated from spots with \( r = 3^9 \) (and an increased contrast \( c = 1 \) to enforce a comparable signal-to-noise ratio to the other trials). On the Sun, these would correspond to spots with diameters of about 70,000 km—typical of the larger spots during solar maximum. While the radius posterior is biased (as it must be, given our prior), the fact that it peaks at the lower bound of the prior suggests the presence of spots smaller than the model can capture. More importantly, however, the latitudinal parameters are inferred correctly and at fairly high precision; even though our model is biased against small spots, this does not affect inferences about their latitudes. On the other hand, the spot contrast is wrong by many standard deviations, since the model must compensate for the fact that the radii are biased high with a lower contrast to match the variability amplitude of the light curves.

Figures 22 and 23 show results for the default run but with extreme values of the number of light curves in the ensemble: \( M = 1 \) and 1000, respectively. These two figures underscore the power of ensemble analyses; a single light curve (Figure 22) is simply not informative enough about the properties of its spots. On the other hand, a very large ensemble can be extremely informative; the radius, latitude, and even the contrast are inferred correctly at high precision.

Figures 24–26 show results for the default run but with limb darkening. In all cases, we assume quadratic limb darkening with fiducial values \( u_1 = 0.5 \) and \( u_2 = 0.25 \) for all stars. From Figure 24, in which we assume we know the limb-darkening coefficients exactly, it is clear that the presence of limb darkening significantly degrades our ability to infer both the radii and latitudes of the spots. Limb darkening has a complicated effect on the mapping between surface features and disk-integrated flux, as it reveals information about the odd harmonics at the expense of introducing strong degeneracies with the even harmonics (Paper I). In practice, this leads to a higher uncertainty in the spot radii and latitudes relative to the same data set without limb darkening (Figure 6). Fortunately, this uncertainty can be dramatically reduced with more data, as evident in Figure 25, which shows the results of the same run but with \( M = 1000 \) light curves in the ensemble. The constraints on \( r, \mu_\phi \), and \( \sigma_\theta \) are now much tighter and in good agreement with the truth. While there is evidence for slight bias in the inferred radius, this is likely due to the difference between the model used to generate the spots and that used for inference, as discussed in the previous section. Finally, Figure 26 shows the results of inference on limb-darkened light curves under the (wrong) assumption that limb darkening is not present (\( u = 0 \)). Neglecting the effect of limb darkening can lead to biases in the spot radius and latitude parameters. While the model still favors midlatitude spots (at \(-45^\circ \) instead of \( 30^\circ \)), the constraints are deceptively tight and discrepant by many standard deviations. We discuss these points in more detail in Section 4.4.
The runs so far correspond to stars with many \( n = 20 \) spots, for which the resulting light curves are smooth due to the fact that many spots are in view at any given time. Figures 27 and 28 show what happens when the model is applied to stars with \( n = 2 \) and 1 spots, respectively. Despite large portions of the light curves being flat (and therefore extremely nonstationary) in these scenarios, the GP does surprisingly well, recovering the radii and latitude parameters within 2\( \sigma \)–3\( \sigma \) in both cases. Note that in order to preserve the same signal-to-noise ratio relative to the other runs, we gave the spots in Figure 27 a much higher contrast \( (c = 0.5) \). Even though the contrast is degenerate with the number of spots (which is very poorly constrained), the \( c \) posterior has a much heavier tail than in the other runs. Thus, in spite of the arguments in Luger et al. (2021b) about the difficulty in constraining \( c \) and \( n \) from single-band photometry, it is evident that the full covariance structure of the data encodes some information about the contrast and—to a much lesser extent—the number of spots. In Figure 28, we compensate for the smaller number of spots by increasing the spot radius to \( r = 45^\circ \pm 5^\circ \) instead, showing that the model can accurately model large spots, even in the presence of some (unmodeled) scatter in their sizes.

**Figure 9.** Similar to Figure 6 but showing the posterior distribution for 100 different synthetic data sets, all generated from the same default input parameters (Table 1). Each colored curve corresponds to a single run, with the corresponding 1\( \sigma \) contours shown as shaded ellipses for each pair of parameters. The black curves correspond to the distributions of all samples from each of the 100 runs, and the black contours again indicate 1\( \sigma \) contours in the joint posterior.

**Figure 10.** Distribution of stellar inclination residuals normalized to the posterior standard deviation for all 5000 stars across the 100 trials in Figure 9 (blue). The standard normal distribution is shown in orange for comparison. The inclination posteriors inferred with our GP are largely unbiased and have the expected variance.
In Figure 29, we add variance to all of the spot properties when generating the light curves; we add \( n = 20 \pm 3 \) spots to each star with radii \( r = 15^\circ \pm 3^\circ \), contrasts \( c = 0.05 \pm 0.01 \), and at latitudes \( \phi = 30^\circ \pm 5^\circ \). As before, we only explicitly account for the variance of the latitude distribution in our model. We correctly infer the latitude parameters and the contrast, but our radii appear to be biased high. This is likely due to the fact that larger spots have a bigger impact on the signal, so our inferred radius is a weighted average of all spot radii. In Appendix C.1, we derive an expression for the moment integrals of the spot size distribution assuming a uniform distribution between \( r - \Delta r \) and \( r + \Delta r \) (instead of a delta function at \( r \)), which can be used to compute the GP if one wishes to explicitly account for scatter in the spot sizes. We find that repeating the run shown in Figure 29 while explicitly sampling over the distribution in \( \Delta r \) shifts the posterior mass to lower radii, mitigating the bias described above.

Our final run is shown in Figure 30, in which we assume we know the true normalization of each light curve. That is, we assume that we can measure all light curves in units of the flux we would measure if the stars had no spots on them, and we do not normalize them (see Section 2.5). In practice, this would require knowledge of the brightness (or temperature) of the unspotted photosphere, which is not an observable in single-band photometry. This value can, in principle, be probed, however, in multiband photometry (e.g., Gully-Santiago et al. 2017; Guo et al. 2018), for which this run is extremely relevant. We again recover the radii and latitude parameters to within \( 2\sigma - 3\sigma \), but most importantly, we also infer the correct spot contrast and number of spots with fairly high precision. In particular, knowledge of the correct normalization breaks the \( c - n \) degeneracy. Photometric measurements in multiple bands (even just two) are therefore extremely useful when inferring spot properties. We discussed this point in Paper I.

4. Discussion

4.1. Small Spots

One of the biggest downsides of adopting a spherical harmonic representation of the stellar surface as the foundation of Figure 11. Same as Figure 9 but showing the posterior distributions of 100 realizations of ensembles with \( M = 1000 \) (instead of 50) light curves each. The constraints on most of the parameters are much tighter, and the polar spot degeneracy is gone.
of our GP is the inherent limitation it imposes on the resolution of surface features. In order to maximize computational efficiency and numerical stability, our default approach is to model the surface using an expansion up to degree $l_{\text{max}} = 15$, which can model features only as small as $\sim 180^\circ / 15 = 12^\circ$ across. Even on scales slightly larger than this, the presence of ringing can be seen (see panel (b) in Figure 3, where ringing is just barely noticeable in the equatorial region of the maps). The spherical harmonic basis consists of global modes, all of which contribute to the intensity everywhere on the surface. Localized features require constructive interference of modes inside and destructive interference of modes outside, often leading to a wavelike ringing pattern that gets worse as the size of the features gets smaller. Taken at face value, this might suggest that a different basis—such as the common choice of pixels on a grid, or perhaps a localized wavelet basis—would be better at modeling small spots. While this is probably true, it may be quite difficult to find closed-form expressions for the expectation integrals (Appendix C) that make the GP covariance evaluation tractable.

One option is to bypass the computation of the covariance on the surface of the star and write down an expression for the flux directly in terms of the properties of a starspot. Circular spots of uniform intensity can be modeled as spherical caps, which are almost identical. There is a slight difference in amplitude between the two cases; surfaces with dark spots have slightly higher light-curve amplitudes than those with bright spots.

Figure 12. Comparison of GP samples with dark (top) and bright (middle) spots alongside their corresponding light curves viewed at $I = 75^\circ$ (bottom). Without knowledge of the correct normalization (Section 2.5), it is very difficult to differentiate between the two from stellar light curves.

If we were to assume that all spots are circular, the flux by ignoring certain projection effects, such as the self-occultation of large spots by the limb of the star, and neglected variations in limb darkening within spots. Such a model could admit a closed-form solution to the GP covariance and may be better at capturing the effects of small spots at the expense of the ability to model larger spots.

In principle, small spots can be modeled under our current approach with negligible ringing by simply increasing the degree of the spherical harmonic expansion. As we discuss in (Luger et al. 2021a), however, the algorithm presented here becomes unstable for $I \gtrsim 15$, so doing so would require a reparameterization of the equations in Appendices C and D. Importantly, however, as we showed in Section 3, the current implementation of the GP is suitable to modeling light curves of stars with spots smaller than the limiting resolution of $\sim 10^\circ$. Consider Figure 21, which shows the results of doing inference on an ensemble of light curves of stars with small ($r = 3^\circ$) spots; these are comparable to some of the larger spots seen on the Sun with diameters of about 70,000 km. Even though our inferred radius and contrast are wrong, the fact that the radius posterior peaks at the lower prior bound of 10° is strongly suggestive of the presence of spots smaller than the resolution of the model. Moreover, the latitude mode and standard deviation posteriors are unbiased, and we correctly infer the presence of low-variance, midlatitude spots. With the above caveats in mind, our GP can therefore still be used to model stars with small spots.

4.2. Bright Spots

All of the calibration tests performed in the previous section assumed the stellar surface was dominated by dark spots. We can easily model the effect of bright spots by choosing negative values for the contrast parameter, i.e., $c < 0$. The top two panels of Figure 12 show five random samples from the GP with dark ($c = 0.1$) and bright ($c = -0.1$) spots; the random seed is the same for both panels, so the maps are identical in all other respects. While the surface maps can be easily distinguished by eye, the same is not true for the corresponding light-curve samples (bottom panel), which are almost identical. There is a slight difference in amplitude between the two cases; surfaces with dark spots have slightly higher light-curve amplitudes than...
surfaces with bright spots of the same contrast magnitude. However, the magnitude of the bright spots can be increased slightly to get a near-perfect match to the dark spot light curves, meaning it may be difficult (if not impossible) to tell the difference between dark and bright spots via the GP approach.

The reason for this degeneracy is rooted (once again) in the fundamental issue with photometry: we lack any information about the correct normalization of the light curve. Consider the dependence of the (unnormalized) GP covariance on the contrast: it enters in a single place, via Equation (C98), as $c^2$, meaning dark and bright spot models have exactly the same covariance. These models differ only in the mean of the unnormalized process, since that is proportional to $c$ (via Equation (C97)). However, as we argued in Paper I, the mean is not a direct observable. Instead, in single-band photometry, we are only sensitive to the ratio of the covariance to the square of the mean (see Equation (24)). From that equation, we can deduce that stars with dark spots (for which the light curve mean $\mu < 1$) will therefore have larger variance than stars with bright spots ($\mu > 1$), leading to the slight difference seen in the figure. However, since this is strictly a multiplicative factor affecting the covariance, it is degenerate with the two other properties that scale the covariance: the magnitude of the contrast and the total number of spots.23

We therefore conclude that single-band photometry is largely insensitive to the difference between bright and dark spots. However, it is important to bear in mind that the degeneracy described above exists only for the Gaussian approximation to the likelihood function. As we argued earlier, the true likelihood function is not a Gaussian; in particular, the true probability distribution has higher-order moments that we do not model here. These moments should, in principle, encode information about the sign of the spot contrast, but they may be very difficult to infer in practice. It may be possible to distinguish between dark and bright spots with traditional forward models of stellar surfaces, but (as we argued earlier) a statistically rigorous ensemble analysis of stellar light curves using such forward models is probably computationally intractable.

We can, however, skirt this degeneracy with observations in multiple bands, which can provide limited information about the correct normalization. Recently, Morris et al. (2018) used approximately coeval Kepler and Spitzer light curves of TRAPPIST-1 to argue that a bright spot model for the star is more consistent with the data. A detailed exploration of the effect of multiband photometry on the degeneracies of the mapping problem is deferred to a future paper in this series.

4.3. Comparison to Other Work

4.3.1. Synthetic Likelihoods, Random Fields, and Approximate Inference

The core idea behind the methodology presented in this paper—to compute the Gaussian approximation to an intractable multidimensional distribution in order to obtain a likelihood function for inference—is not new. Although the method likely goes by different names in different fields, it is a popular technique, particularly in the field of ecological population dynamics, where it is referred to as the synthetic likelihood (SL; Wood 2010) or Bayesian synthetic likelihood (BSL; Price et al. 2018) method. In many ecological systems, population growth is a chaotic process; observations of the size of a population over time can be dominated by steep spikes and drops in the population that occur due to sudden, random environmental pressure. While population growth can be forward modeled with ease, it is very difficult to use forward models to constrain basic growth parameters in an inference setting, since that requires marginalization over the extremely nonlinear noise processes. As a way around this, Wood (2010) introduced the SL method, in which, conditioned on a set of parameters of interest $\theta$, one computes the forward model $f(\theta)$ many times under different realizations of the noise and adopts the sample mean and covariance (usually of a summary statistic of the data) as the mean and covariance of the Gaussian likelihood function $p(f \mid \theta)$. Wood (2010) showed that, provided the number of forward-model samples is large enough, this “synthetic” likelihood allows one to infer the population growth parameters efficiently and without bias.

The method presented in this paper may be thought of as an SL method in the limit of an infinite number of forward-model samples. Unlike Wood (2010), whose method determines the mean and covariance of the distribution of some function of $f$ conditioned on $\theta$ by sampling, we are able to actually compute the mean and covariance of $f$ directly in closed form. While traditional SL methods are inherently noisy, our method employs the exact Gaussian approximation to the likelihood function.

Our GP is also closely related to techniques commonly employed in models of the CMB. In particular, it is a type of Gaussian random field (GRF) on the sphere, which is frequently used to model perturbations in the CMB (Wandelt 2012). In general, however, GRFs used in cosmology are isotropic; when expressed in the spherical harmonic basis, their covariance matrix is diagonal and admits a representation as a (one-dimensional) power spectrum. Our GP, in contrast, is anisotropic in the polar coordinate (i.e., the latitude) by construction.

23 There is also a small additive term in Equation (24), but this, too, depends only on the ratio of entries in the covariance matrix to the mean, so it is of little help in breaking the degeneracy.
4.3.2. Starspots and Stellar Variability

The methodology developed in this paper is closely related to that in Perger et al. (2021), who studied the effect of different starspot configurations on the autocorrelation and covariance of stellar RV measurements. The authors of that study compared the performance of various commonly used quasiperiodic kernels when applied to synthetic RV data sets, arguing that a new four-parameter quasiperiodic cosine kernel (QPC) can better capture the variability due to starspots. However, their study was empirical and related spot configurations to their effect on the covariance structure of the data primarily in a qualitative fashion. Their QPC kernel is a function of two interpretable hyperparameters (the rotation period and a spot timescale), as well as two amplitudes, which are not explicitly related to physical spot properties. Our GP, in contrast, is built from the ground up, such that all of its hyperparameters directly correspond to physical spot properties, allowing one to use it in starspot inference (not just marginalization) problems. While the methodology presented here applies to photometry, it is possible to extend it to model RV data sets as well; we discuss this in Section 5.

Recently, Morris (2020b) used Kepler, K2, and TESS light curves to derive a relationship between stellar age and spot coverage using an ensemble analysis similar to that proposed here. Because of the intractability of the marginal-likelihood function, that study used an ABC method to infer spot properties from a large ensemble of stars. Morris (2020b) developed a fast, approximate forward model for light curves of spotted stars (fleck; Morris 2020a), which they used to generate a large number of prior samples for different values of the spot radii, contrasts, and latitude distributions. For each collection of samples generated from a given set of hyperparameters, Morris (2020b) computed the distribution of the “smoothed amplitude,” the peak-to-trough difference of the (normalized, detrended) light curve. This distribution was then compared to the distribution of observed smoothed amplitude values among stellar clusters of different ages within an ABC algorithm, yielding approximate posterior distributions for the hyperparameters as a function of stellar age. While we believe the spot coverage results of that paper are predominantly driven by the prior (due to the strong degeneracy between the spot contrast and the number of spots; see Section 4.2 in Luger et al. 2021b), the ensemble analysis employed in that paper is nevertheless a powerful technique to infer spot properties. Our work builds on that of Morris (2020b) by deriving a closed-form solution to the likelihood function (as opposed to a sample-based likelihood-free inference algorithm) and harnessing the covariance structure of the data when doing inference (as opposed to relying solely on the amplitude of the data).

Finally, Basri & Shah (2020) recently presented a large suite of forward models of light curves of spotted stars, which they used to discuss the (complicated) dependence of various light-curve metrics on the physical spot parameters used to generate the data. They concluded that it is not possible to uniquely relate these metrics to the underlying starspot configuration. While we agree that this is the case for individual stars, our work stresses that it is possible to circumvent many of these degeneracies with ensemble analyses. Basri & Shah (2020) also concluded that it is not generally possible to uniquely disentangle differential rotation from spot evolution when their timescales are comparable, nor is it possible to confidently measure a rotation period when the evolution timescale is very short. However, their study relied on the effect these processes have on simple light-curve metrics, which are almost certainly not sufficient statistics of the data. Inference that takes into account the full covariance structure of the data while...
considering large ensembles of light curves could, in principle, break this degeneracy. While we do not explicitly model differential rotation in this paper, it will be the subject of a future paper in this series.

### 4.4. Caveats

We conclude our discussion with a list of several notes and caveats that should be kept in mind when using our algorithm and its Python implementation.

1. **The assumed latitude distribution is not Gaussian.** Because we require the first- and second-moment integrals (Equations (C63) and (C70)) to have closed-form solutions, there are restrictions on the PDF we can assume for the spot latitude. We find that a beta distribution in the cosine of the latitude is integrable in closed form and can be evaluated efficiently in terms of recursion relations. In many cases, particularly when \( \mu_\phi \lesssim 75^\circ \) and \( \sigma_\phi \lesssim 10^\circ \), the distribution in the spot latitude is close to a bimodal Gaussian with mean \( \pm \mu_\phi \) and standard deviation \( \sigma_\phi \) (see Figure 16). In general, however, \( \mu_\phi \) is formally equal to the mode (as opposed to the mean) of the distribution, and \( \sigma_\phi \) is the Laplace approximation to the local standard deviation at the mode.

2. **The number of spots \( n \) does not have to be an integer.** Samples from our GP prior will not generally have exactly \( n \) spots (see, e.g., Figure 3). This is due to the fact that our model is only an approximation to the true distribution of stellar surfaces conditioned on the spot properties. A corollary of this point is that \( n \) need not be an integer, which makes it easier in practice to sample over using modern inference techniques such as MCMC, HMC, ADVI, and nested sampling.

3. **Care should be taken when modeling large-amplitude light curves.** We discussed this point at length in Section 2.5. Modeling a light curve that has been normalized to its mean (or median) as a GP is conceptually a bad idea when the amplitude of variability is large compared to the mean. As a rule of thumb, if the amplitude of variability exceeds \( \sim 10\% \), we recommend not normalizing the light curve in this way and instead modeling the normalization amplitude as a latent variable.

4. **Keep in mind the polar spot degeneracy.** Even when modeling ensembles of light curves, there are still strong degeneracies at play (Paper I). In particular, spots centered on the poles are always in the null space, so it can be difficult in practice to rule out their presence. This can be seen in Figure 18, in which the model cannot distinguish between spots localized at \( 60^\circ \) and polar spots with high latitude variance. It may thus be advisable to adopt a prior that favors small values of \( \sigma_\phi \), such as the common inverse gamma prior for the variance. Alternatively, one could place an isotropic prior on the latitude (with density proportional to \( \cos \mu_\phi \)) to downweight very high latitude spots.

5. **Limb darkening matters.** The null space is extremely sensitive to limb darkening (Paper I). It is therefore extremely important to model it correctly; otherwise, there may be substantial bias in the inferred spot parameters. For stars with transiting exoplanets, it may be possible to infer the limb-darkening coefficients empirically, but in general, we recommend modeling them as latent variables with priors informed by theoretical models.

6. **Careful with the data.** In general, covariances can be very hard to estimate from noisy data. This makes it especially important to ensure one is correctly modeling the noise. When applying our GP to model real data, we recommend the usual inference practices of clipping outliers, modeling a latent white-noise (jitter) term, and modeling a small latent additive offset term to minimize the risk of bias in the posteriors of interest.

7. **Careful with the sample selection.** When performing an ensemble analysis of stellar variability, it is tempting to only analyze light curves that show variability in the first place. This is extremely dangerous, since the lack of variability could simply be due to low inclinations. It is extremely important to ensure that the sample selection step does not introduce bias. If there is reason to believe that there are two distinct populations within an ensemble—say, a population of active stars and a population of quiet (spotless) stars—we strongly recommend the use of a Gaussian mixture model.

### 5. Extensions

#### 5.1. Composite GPs

Thus far, we have assumed that spots are concentrated at a single latitude (above and below the equator). We baked this assumption directly into our choice of distribution function for the latitude (Appendix C.2), which has exactly two modes at \( \pm \mu_\phi \). However, it is possible (at least in principle) that certain stars could have two or more active latitudes, in which case our GP is not an appropriate description of the stellar surface.

Fortunately, Gaussian distributions (and thus also GPs) are closed under addition, meaning that the sum of two GPs is also a GP. We can thus construct more complex models for stellar variability by summing GPs with different spot hyperparameter vectors \( \theta \). The composite GP will then have a mean equal to the sum of the means of each GP and a covariance matrix equal to the sum of the covariance matrices of each GP. One possible application of this is to model stars with multiple active latitudes, as described above; an example of this is shown in Figure 13, where samples are drawn from a GP with small circumpolar spots and large equatorial spots (see caption for details). Since the composite GP inherits all of the properties of the standard GP, it can be used to do inference under more complex priors than those presented here.

It is also worth noting that this technique may be employed to model arbitrary distributions for parameters like the spot radius and the number of spots. In Appendix C.1, we present a formulation of our GP that admits a radius distribution half-width parameter \( \Delta r \); this generalizes our delta function distribution to a uniform distribution between \( r - \Delta r \) and \( r + \Delta r \). One may then compute the weighted sum of several GPs with half-widths \( \Delta r \) and central radii \( r \) to approximate any distribution of spot radii. Similarly, one may compute the weighted sum of several GPs with different values of the number of spots \( n \) to enforce any discrete distribution for that quantity. The reader should keep in mind that the cost of computing the GP covariance matrix will scale linearly with the number of GP components. In many cases, however, the computational
bottleneck is the covariance factorization step (Luger et al. 2021a), in which case, adding components to the model will result in negligible overhead.

5.2. Time Evolution

Another big limitation of the base algorithm is the implicit assumption that stellar surfaces are static. Our GP hyperparameters $\Theta$ describe the spatial configuration of starspots, but they say nothing about their evolution in time. We know from observations of the Sun and Kepler stars that temporal variability is extremely common; spots appear, disappear, and even migrate in latitude over time. While it may be possible to parameterize their evolution in a way that is general enough to capture all the ways in which they may change over time, such an approach is beyond the scope of the present paper. It is, however, straightforward to implement an uninformative temporal prior within the framework of our GP. To do this, we will make two simplifying assumptions.

1. The temporal process is stationary. This implies that there is no preferred time (or phase) and that the spatial covariance is the same at all points in time. Stars may still have active longitudes under this assumption, but there is no preferred longitude across all stars.

2. The temporal and spatial covariances are independent. This implies that the evolution of each spherical harmonic mode in time is independent of the evolution of any of the other modes in time.

There is some tension between these assumptions and our knowledge of how stellar surfaces evolve. Assumption (1) excludes surfaces whose total spottiness changes significantly or whose spots migrate in latitude, as both processes change the spatial covariance over time. Assumption (2) ignores the correlation between spherical harmonic modes due to the migration of spots, which requires the coherent evolution of many modes at once. These assumptions likely limit the ability of our GP to model light curves on very long baselines (i.e., on timescales of years) over which stellar activity cycles take place. However, given that the use case of our algorithm is likely to be the analysis of individual quarters of Kepler and individual sectors of TESS data, our assumptions are likely valid in most cases. Analyses of (say) all quarters of Kepler data could process each quarter at a time in a hierarchical framework in which hyperparameters like the mean spot latitude in each quarter are treated as functions of time.

The two assumptions listed above suggest a fairly straightforward form for the GP covariance in spherical harmonics and time,

$$\Sigma_{y'}^{(t)} = K \otimes \Sigma_y,$$

(32)

where $K$ is a $(K \times K)$ matrix describing the covariance among the $K$ points in time, $\Sigma_y$ is the $(N \times N)$ matrix describing the covariance among the $N \equiv (l_{\text{max}} + 1)^2$ spherical harmonic coefficients (Equation (10)), and $\otimes$ denotes the Kronecker product. The quantity $\Sigma_{y'}$ is the $(NK \times NK)$ temporal–spatial covariance whose coefficient at index $(Nk + n, Nk' + n')$ is the covariance between the spherical harmonic coefficients $y_n(l_k)$ and $y_{n'}(l_{k'})$, where the index $n$ is related to the spherical harmonic indices $l$ and $m$ via Equation (A3). Finally, because of our assumption of stationarity, the mean of the GP is still constant and equal to the mean of the static process.

The covariance matrix in Equation (32) can be sampled from to yield time-variable surface maps, or it can be transformed into flux space for sampling light curves or computing likelihoods. The covariance in flux space is given by

$$\Sigma^{(t)} = \mathcal{A}^t \Sigma_y^{(t)} \mathcal{A}^{t\top},$$

(33)

where

$$\mathcal{A}^t = \begin{pmatrix} a_0^t \\ a_1^t \\ \vdots \\ a_{K-1}^t \end{pmatrix}$$

(34)

is a design matrix constructed by staggering the rows of the standard design matrix $A$ (see Appendix B.1). The diagonal structure of $\mathcal{A}^t$ is due to the fact that each snapshot of the surface is only observed at a single phase.

In the temporally variable version of our GP, the covariance matrix in Equation (33) replaces the standard covariance $\Sigma$; it can similarly be modified (Section 2.5) to obtain the covariance of the normalized process. While $\Sigma^{(t)}$ is $(K \times K)$, it is computed from the contraction of a much larger $(NK \times NK)$ matrix, which is extremely inefficient to instantiate and operate on. Fortunately, it can be shown that

$$\Sigma^{(t)} = \mathcal{A}^t \Sigma_y^{(t)} \mathcal{A}^{t\top} = \mathcal{A}^t (K \otimes \Sigma_y) \mathcal{A}^{t\top} = \Sigma \otimes K,$$

(35)

that is, the flux covariance is just the elementwise $\otimes$ product of the GP covariance $\Sigma$ and temporal covariance $K$. This fact makes the temporal GP just as efficient to evaluate as the standard GP.

Armed with this algorithm for computing $\Sigma^{(t)}$, it only remains for us to decide on a structure for $K$. While this can, in principle, be any covariance matrix constructed from a stationary kernel, we recommend one of the common radial kernels, such as the exponential squared kernel

$$k_{E^2}(\Delta t) = \sigma^2 \exp\left(-\frac{\Delta t^2}{2\tau}\right)$$

(36)

or the Matérn-3/2 kernel

$$k_{M^3}(\Delta t) = \sigma^2\left(1 + \sqrt{3} \frac{\Delta t}{\tau}\right) \exp\left(-\sqrt{3} \frac{\Delta t}{\tau}\right)$$

(37)

with the variance parameter $\sigma^2$ set to unity (since the variance is already specified within $\Sigma$). In this case, the temporal covariance $K$ is a function of a single parameter: the timescale of the variability, $\tau$. Similar to the other hyperparameters of our GP, this parameter can be estimated in an inference setting. However, a detailed investigation of the ability of our temporal GP to accurately capture spot variability is beyond the scope of this paper and will be revisited in the future, along with an algorithm to explicitly model the effects of differential rotation on the covariance structure.
5.3. Marginalizing over Period and Limb Darkening

In Section 2.4, we discussed the value of marginalizing over the stellar inclination when computing the GP covariance (and thus the likelihood). When jointly analyzing the light curves of many stars (which we have repeatedly argued is the best way to infer their spot properties), it can be extremely useful to minimize the number of latent variables associated with individual stars by analytically marginalizing over them. This can dramatically reduce the number of free parameters, turning a difficult inference problem in (possibly) hundreds or thousands of dimensions into a much easier problem in a handful of dimensions. We showed how it is possible to marginalize away the dependence of our GP on the inclinations of individual stars, which is a huge step in this direction. However, our GP remains a function of two other quantities that will generally be different for different stars: the stellar rotation period $P$ and the limb-darkening coefficient vector $\mathbf{u}$. We argued that in some cases, one may be able to fix the period of each star at an estimate obtained in a preprocessing step (i.e., from a periodogram) and fix the limb-darkening coefficients at theoretical values, in which case the number of free parameters of the GP is equal to the size of the spot hyperparameter vector $\theta$, (five by default) and independent of the number of light curves in the ensemble. However, this procedure ignores any uncertainty in the period and limb-darkening coefficients, which could be significant; it is also subject to bias due to the fact that there may be systematic errors in theoretical models for the limb-darkening coefficients, particularly for low-mass stars (e.g., Kervella et al. 2017). A much better approach would be to analytically marginalize over these two quantities.

Consider Equation (B3) in Appendix B, from which the rows of the design matrix (which transforms vectors in the spherical harmonic basis to vectors representing the flux at a point in time) are computed. Marginalization over the inclination, which we demonstrate how to perform analytically in Appendix D, entails integrating over the term $\mathbf{R}_g(-I)$, the Wigner matrix that rotates the star by the inclination angle $I$ into the observer’s frame. Similarly, marginalization over the rotation period would entail integration over the term $\mathbf{R}_g\left(\frac{2\pi t}{P}\right)$, another Wigner matrix that rotates the star to the correct rotational phase at time $t = t_k$. Given an appropriate prior on the rotational angular frequency $\frac{2\pi}{P}$, it should be possible to follow the same procedure outlined in Appendix D to analytically compute the first two moments of the distribution of light curves marginalized over the rotation period.

Equation (B3) also makes clear the dependence of the GP on the limb-darkening coefficients, which enter via the limb-darkening operator $\mathbf{L}(\mathbf{u})$. The coefficients of this matrix are linear in the limb-darkening coefficients $u_1$ and $u_2$ (see Equation (B5)), so it should be possible to derive closed-form solutions to the relevant integrals over $\mathbf{L}(\mathbf{u})$ to yield the GP covariance marginalized over $\mathbf{u}$. It may be possible to compute the marginalized covariance even when parameterizing the quadratic limb-darkening coefficients in terms of the uncorrelated $q_1$ and $q_2$ parameters from Kipping (2013), which would allow us to incorporate hard constraints on positivity and strict limb darkening (as opposed to brightening) directly into the prior.

Given the complexity of the operations involved, performing these marginalizations is beyond the scope of the present paper. However, given the significant computational benefits of this marginalization, as well as the fact that the mapping problem is particularly sensitive to the limb-darkening coefficients (Paper I), marginalizing over $P$ and $\mathbf{u}$ will be the subject of an upcoming paper.

5.4. Modeling Transits and RV Data Sets

Finally, we would like to note that the GP developed in this paper is not limited to modeling rotational light curves of stars. We derived the covariance structure of spotted stellar surfaces in the spherical harmonic basis, which we then linearly transformed into flux space to obtain the light-curve GP presented above. However, our GP may be used to model any kind of observation that is linearly related to the spherical harmonic representation. If $\mathbf{A}$ is the linear operator that transforms the spherical harmonic representation $\mathbf{y}$ to the data vector $\mathbf{d}$ via $\mathbf{d} = \mathbf{A}\mathbf{y}$, then the mean and covariance of the GP model for $\mathbf{d}$ are given by

$$\mu_d = \mathbf{A} \mu_y,$$

and

$$\Sigma_d = \mathbf{A} \Sigma_y \mathbf{A}^T,$$

respectively, where $\mu_y$ and $\Sigma_y$ are the mean and covariance in the spherical harmonic basis derived in this paper. In Luger et al. (2019), we showed that occultation light curves are also linearly related to the spherical harmonic representation of the stellar surface, so it is straightforward to use our GP to model transits of planets across spotted stars, either to marginalize over the spot variability or to constrain the surface map of the star. In this context, $\mathbf{A}$ may be computed via the design_matrix() method of a Map instance of the starry package.

The formalism developed here may also be extended to model RV data sets, although this requires a bit more work. The instantaneous RV shift $v$ induced by a rotating spotted star may be approximated as

$$v = \frac{\int_{\delta} l V \, dS}{\int_{\delta} l \, dS},$$

where $l$ is the stellar intensity at a point on the surface, $V$ is the radial component of the rotational velocity vector at that point, and the integral is taken over the projected disk of the star. If we expand the surface intensity distribution in spherical harmonics, the integral in the denominator is just a classical starry integral, as it is just the disk-integrated intensity (i.e., the flux). The numerator may also be computed following the starry formalism, provided we weight the surface intensity representation by the velocity field $V$. In the case of rigid-body rotation, $V$ is exactly a dipole ($l = 1$) field, so the quantity $IV$...
can be expressed exactly as a product of spherical harmonics.\footnote{Higher-order effects such as differential rotation and convective blueshift can be easily modeled with a higher-degree expansion of $V$.} Since spherical harmonics are closed under multiplication, we may write $IV$ as a linear combination of spherical harmonics, meaning the integral in the numerator is also a starry integral. For some linear operators $A$ and $B$, we may therefore write
\begin{equation}
\psi = \frac{B \psi}{A \psi},
\end{equation}
where $\psi$ is the random variable representing the observed RV time series, $\gamma$ is the Gaussian random variable representing the spherical harmonic coefficients describing the stellar surface, and the division is performed elementwise. Since $\psi$ is the ratio of two Gaussian random variables, its distribution is not Gaussian. However, we can still compute the first two moments of the distribution of $\psi$ to derive a Gaussian approximation to it (similar to what we did in Section 2.5), which will yield the mean and covariance of the GP representation of the RV time series. Given the complexity of the operations described above, we defer this calculation (and the calibration of the resulting GP model) to future work.

6. Conclusions

This paper is the second in a series devoted to the development of statistically rigorous techniques to model stellar surfaces based on unresolved photometric and spectroscopic measurements. Here we presented a new GP model for stellar variability whose hyperparameters explicitly correspond to the physical properties of the stellar surface. Our GP allows one to efficiently compute the likelihood function for stellar light curves marginalized over nuisance parameters such as the specific sizes, positions, and contrasts of individual spots, which are generally unknowable due to the extreme degeneracies involved in the light-curve mapping problem. Our GP therefore makes it easy to do posterior inference on the real quantities of interest: parameters controlling the distribution of spot sizes, latitudes, and contrasts within a star and/or across many stars in an ensemble. Because our expression for the GP covariance has an exact, closed-form solution as a function of the spot parameters, it can be computed efficiently; a typical likelihood evaluation on a data set consisting of $K \sim 1000$ points takes between 30 and 60 ms on a modern laptop. Our algorithm is implemented in the open-source, user-friendly Python package starry\_process, which is pip-installable, available on GitHub (see footnote 7), and described in Luger et al. (2021a). The algorithm is implemented in a combination of C++ and Python, linked using the theano package. Because our GP covariance has an exact representation, so too do its derivatives. We therefore implement back-propagated derivatives with respect to all input parameters for out-of-the-box usage with gradient-based inference and optimization tools such as HMC, ADVI, and gradient-based nested sampling.

We devoted a large portion of this paper to testing the algorithm on a variety of synthetic data sets, showing that it is a well-calibrated and in most cases unbiased estimator for starspot properties. Below we list our main results.

1. Our GP works best for ensemble analyses. The light-curve mapping problem is extremely degenerate, as light curves contain a vanishingly small fraction of the total information about a stellar surface. However, the degenerate surface modes are a strong function of the observer’s viewing angle, so the light curves of stars seen at different inclinations constrain different components of the surface. We have shown that if we jointly analyze the light curves of many stars, we can break many of the degeneracies at play and uniquely infer the statistical properties of the spots across the ensemble. This type of analysis works best if the stars in the ensemble are statistically similar; i.e., the properties of their spots are all drawn from the same parent distribution, whose parameters we can constrain.

2. Typically, an ensemble of at least $M \sim 50$ light curves is needed to place meaningful constraints on starspot properties. This estimate is based on ensemble analyses of light curves with $K = 1000$ cadence each and a per-cadence precision of one part per thousand. Lower-quality, shorter-baseline observations or data sets contaminated by outliers will generally require larger values of $M$ for the same constraining power. The presence of strong limb darkening also degrades the information content of light curves, in which case an ensemble of hundreds or even 1000 light curves is recommended.

3. Our GP is in most cases an unbiased estimator for the spot radius and latitude distributions. We showed that our GP can accurately infer the angular size of spots and the mode and standard deviation of their distribution in latitude from stellar light curves. For the fiducial ensemble of $M = 50$ light curves described above, we are able to constrain the average spot radii to within a couple degrees and the average spot latitudes to within $5^\circ$.

4. Our GP can accurately infer stellar inclinations. We presented two versions of our GP: one conditioned on a specific value of the stellar inclination and one marginalized over inclination under an isotropic prior. In both cases, we find that we can accurately infer the inclinations of individual stars in an ensemble analysis (in the latter case, a simple postprocessing step can yield the inclination posterior distribution). While the inclination is not an observable for an individual stellar light curve, the population-level constraints on the spot properties achieved by the GP can break the degeneracies involving the inclination, allowing us to usually infer it to within about $10^\circ$ and without bias.

5. Our GP can be used to model small, Sun-like spots. The algorithm presented here is limited to a surface resolution of about $10^5$, corresponding to spots about an order of magnitude larger than typical sunspots. However, we have shown that when we apply our model to the light curves of stars with small ($r \sim 3^\circ$) spots, we can still infer their latitudinal distribution without bias, as well as the presence of spots below our resolution limit.

6. Our GP can be extended to model time-variable surfaces. The algorithm presented here was derived for static stellar surfaces, corresponding to perfectly periodic light curves. However, time variability can easily be modeled as the product of the starry\_process kernel and a kernel describing the covariance of the process in time, such as a
simple exponential squared or Matén-3/2 kernel. The hyperparameters of the temporal covariance are then strictly tied to the timescale on which the surface evolves. More complex temporal variability, such as that induced by differential rotation, will be the subject of a future paper in this series.

7. Our GP can be used in exoplanet transit modeling and extended to RV data sets. At its core, the starry_process GP defines a distribution over spherical harmonic representations of stellar surfaces. It can therefore be used as a physically interpretable prior when modeling transits of exoplanets across spotted stars, either to marginalize over the stellar inhomogeneity or to explicitly infer spot properties. When combined with the starry package, it can also be used as a prior on the stellar surface in Doppler imaging or tomography or even RV searches for exoplanets. The latter will be the subject of a future paper in this series.

The GP presented here has far-ranging applications for stellar light-curve studies. It serves as a drop-in replacement for commonly used GP kernels for stellar variability, which currently do not have physically interpretable parameters other than the rotation period and, in some cases, a spot evolution timescale. As such, it can be used to marginalize over stellar rotational variability signals in (say) transiting exoplanet searches, asteroseismic characterization of stars, RV searches, etc. It can also be used to learn about stellar surfaces directly: to infer the spot properties of main-sequence stars as a function of spectral type and age, differentiate between spot- and plage-dominated stellar surfaces, and better understand chemically peculiar massive stars and the spot properties of transiting exoplanet hosts for unbiased spectroscopic characterization of their atmospheres (to name a few).

The next papers in this series will focus on the following (in no particular order):

1. A more rigorous treatment of time variability. The paper will focus on modeling differential rotation, whose effect on the covariance of the process can be derived in a similar fashion to what we did here. This will enable direct inference about the differential rotation rates and spot evolution timescales of stars, processes whose effects on light curves are too similar for current methodology to reliably discern between them.

2. Explicit marginalization over the remaining stellar parameters. These include the stellar rotation period and limb-darkening coefficients, which can be marginalized over analytically under certain choices of prior. This will eliminate all per-star hyperparameters in the expression for the GP covariance, greatly speeding up the inference for large ensembles of stellar light curves.

3. Extension of this formalism to RV data sets. As we discussed in Section 5.4, it is possible to extend the methodology presented here to model the contribution of stellar surface variability to RV measurements, which can be used (for instance) to mitigate systematics in EPRV searches for exoplanets.

4. Extension of this formalism to Doppler imaging. As we will show in upcoming work, it is possible to derive an exact linear relationship between the wavelength-dependent spherical harmonic representation of a rotating star and its time-variable spectrum. This linearity makes it possible to adapt our GP formalism to the Doppler imaging problem, providing an efficient marginal-likelihood function for rigorous inference studies.

In keeping with other papers in the starry series, all figures in this paper are generated automatically from open-source scripts linked to in each of the captions ( ), and the principal equations link to associated unit tests that ensure the accuracy and reproducibility of the algorithm presented here ( ).

We would like to thank David W. Hogg, Michael Gully-Santiago, Adam Jermy, Megan Bedell, Will Farr, Dylan Simon, and the Astronomical Data Group at the Center for Computational Astrophysics for their help and many thought-provoking discussions that made this paper possible. R.L. is supported by a Flatiron Fellowship at the Flatiron Institute, a division of the Simons Foundation.

Appendix A
Notation

Unless otherwise noted, we adopt the following conventions throughout this paper: integers are represented by italic capital letters (i.e., K), scalars are represented by italic lowercase letters (i.e., x), column vectors are represented by bold lowercase letters (x), and matrices are represented by bold capital letters (X). In general, the elements of a vector x are denoted as x_i, and the elements of a matrix X are denoted as X_{ij}. Importantly, we make a distinction between quantities like X_{ij} and \textbf{X}_{ij}; the former is a scalar element of a matrix, while the latter is a matrix, which is itself a component of a higher-dimensional (in this case, four-dimensional) linear operator. Thus, lowercase bold symbols always represent vectors, and capital bold symbols always represent matrices.

We also make an explicit distinction between numerical quantities and random variables. The former are typeset in serif font (as above), while the latter are typeset in blackboard font. For example, the quantity \text{x} denotes a scalar random variable, while x denotes a particular realization of that variable. The same applies to vector-valued (x is a realization of \text{x}) and matrix-valued (X is a realization of \text{X}) random variables.

Much of the math in this paper involves vectors representing coefficients in the spherical harmonic basis, which are customarily indexed by two integers, l and m. We therefore make an exception to our indexing notation for quantities in the spherical harmonic basis; we use two indices to represent a scalar vector element, x_{lm}, and four indices to represent a scalar matrix element, X_{m,m'}. The upper indices correspond to the spherical harmonic degree, l \in [0, l_{\text{max}}] and l' \in [0, l_{\text{max}}], while the lower indices correspond to the spherical harmonic order, m \in [-l, l] and m' \in [-l', l']. Vector elements are arranged in order of increasing l and, within each l, in order of increasing m. For example, a vector x representing a quantity in the spherical harmonic basis up to degree l_{\text{max}} has components given by

\[ x = (x_0^0 \ x_{-1}^1 \ x_0^1 \ x_1^1 \ldots \ x_{l_{\text{max}}}^{l_{\text{max}}} \ldots \ x_{l_{\text{max}}}^{l_{\text{max}}} \ldots) \]
Table 2

List of Common Variables and Symbols Used Throughout This Paper

| Symbol | Description | References |
|--------|-------------|------------|
| I      | Vector of ones | … |
| ~      | Denotes a normalized vector-valued random variable | Section 2.5 |
| ⊙      | Elementwise product | Section 5.2 |
| ⊗      | Kronecker product | Section 5.2 |
| a      | GP hyperparameter: spot latitude shape parameter | Appendix C.2 |
| \( a^T \) | Row of the starry design matrix | Equation (B2) |
| A\(_1\) | Starry change-of-basis matrix | Appendix B |
| A      | Starry design matrix | Equation (B2) |
| \( \alpha \) | GP hyperparameter: spot latitude shape parameter | Appendix C.2 |
| b      | GP hyperparameter: spot latitude shape parameter | Appendix C.2 |
| \( \beta \) | GP hyperparameter: spot radius shape parameter | Appendix C.2 |
| c      | GP hyperparameter: spot contrast | Appendix C.4 |
| \( \epsilon \) | Spot contrast (random variable) | Appendix C.4 |
| \( \Gamma(\cdots) \) | Gamma function | … |
| \( D_u \) | Complex Wigner rotation matrix about an axis \( u \) | Appendix C.2.1 |
| \( \delta(\cdots) \) | Delta function | … |
| \( \delta_\theta \) | Kronecker delta | … |
| \( \Delta r \) | GP hyperparameter: spot radius spread | Appendix C.1 |
| \( E(\cdots) \) | Expected value | Equation (5) |
| \( \epsilon_l \) | First-moment integral of the inclination | Equation (20) |
| \( \epsilon_r \) | First-moment integral of the radius | Equation (C10) |
| \( \epsilon_\phi \) | First-moment integral of the latitude | Equation (C11) |
| \( \epsilon_\lambda \) | First-moment integral of the longitude | Equation (C12) |
| \( \epsilon_z \) | First-moment integral of the contrast | Equation (C13) |
| \( \epsilon_r \) | Second-moment integral of the inclination | Equation (21) |
| \( \epsilon_\theta \) | Second-moment integral of the radius | Equation (C14) |
| \( \epsilon_\phi \) | Second-moment integral of the latitude | Equation (C15) |
| \( \epsilon_\lambda \) | Second-moment integral of the longitude | Equation (C16) |
| \( \epsilon_z \) | Second-moment integral of the contrast | Equation (C17) |
| f      | Flux vector | Equation (3) |
| \( \mathcal{F}(\cdots) \) | Gaussian hypergeometric function | … |
| \( \beta_u \) | Vector of GP hyperparameters | Equation (13) |
| l      | Stellar inclination | … |
| \( \mathcal{I}(\Delta t) \) | Stellar inclination (random variable) | Section 2.4 |
| J      | Jacobian of the spot latitude transform | Equation (C62) |
| \( k(\Delta t) \) | GP kernel function | Equation (2) |
| K      | Number of points in light curve | … |
| \( K \) | Temporal covariance matrix | Section 5.2 |
| l      | Spherical harmonic degree | Appendix A |
| \( L \) | Limb-darkening operator | Appendix B.2 |
| \( \mathcal{L} \) | Likelihood function | Equation (14) |
| \( \lambda \) | Spot longitude (random variable) | Appendix C.3 |
| m      | Spherical harmonic order | Appendix A |
| M      | Number of light curves in ensemble | … |
| \( \mu \) | Flux GP mean | Equation (22) |
| \( \mu_s \) | Flux GP mean vector | Equation (7) |
| \( \phi_s \) | Spherical harmonic GP mean vector | Equation (9) |
| n      | GP hyperparameter: number of spots | Appendix C |
| m      | Number of spots contrast (random variable) | Appendix C |
| \( N(\mu, \sigma^2) \) | Normal distribution: mean \( \mu \), variance \( \sigma^2 \) | … |
| \( p(\cdots) \) | Probability, probability density | … |
| P      | Stellar rotation period | … |
| \( p^T \) | GP hyperparameter: spot radius | Appendix C.1 |
| \( R_u \) | Real Wigner rotation matrix about an axis \( u \) | Appendix C.2.1 |

Table 2 (Continued)

| Symbol | Description | References |
|--------|-------------|------------|
| s      | Spherical harmonic expansion of spot | Equation (C23) |
| \( \sigma_f \) | Photometric uncertainty | Section 3 |
| \( \sigma_p \) | GP hyperparameter: spot latitude standard deviation | Appendix C.2 |
| \( \Sigma \) | Flux GP covariance matrix | Equation (8) |
| \( \Sigma^{(0)} \) | Flux GP covariance matrix with temporal evolution | Equation (8) |
| \( \Sigma \) | Flux GP covariance matrix (normalized process) | Equation (24) |
| \( \Sigma^{(0)} \) | Spherical harmonic GP covariance matrix | Equation (10) |
| \( \Sigma^{(0)} \) | Spherical harmonic GP covariance with temporal evolution | Equation (10) |
| t      | Time | … |
| \( \tau \) | GP hyperparameter: timescale | Section 5.2 |
| u      | Limb-darkening coefficient vector | Appendix B.2 |
| \( u_1, u_2 \) | Linear and quadratic limb-darkening coefficients | Appendix B.2 |
| U      | Complex-to-real basis change operator | Appendix C.2.1 |
| \( \mathcal{U}(a, b) \) | Uniform distribution between \( a \) and \( b \) | … |
| \( \kappa \) | Random vector of spot properties | Equation (C3) |
| y      | Spherical harmonic coefficient vector | Equation (3) |
| \( \gamma \) | Spherical harmonic coefficient vector (random variable) | Equation (C4) |
| \( \mu_\phi \) | GP hyperparameter: spot latitude mode | Appendix C.2 |
| \( \phi \) | Spot latitude (random variable) | Appendix C.2 |
| z      | GP normalization number | Equation (25) |

while a matrix \( X \) in the same basis has components given by

\[
X = \begin{pmatrix}
X_{0,0} & X_{0,1} & X_{0,1} & X_{0,1} \\
X_{1,0} & X_{1,1} & X_{1,1} & X_{1,1} \\
X_{0,1} & X_{0,1} & X_{1,0} & X_{1,1} \\
X_{1,1} & X_{1,1} & X_{1,0} & X_{1,1}
\end{pmatrix}.
\]  

(A2)

For completeness, the element of a spherical harmonic vector \( \kappa \) with degree \( l \) and order \( m \) is at (flattened) index

\[
n = l^2 + l + m.
\]

(A3)

Conversely, the element at (flattened) index \( n \) has a degree and order

\[
l = \lfloor \sqrt{n} \rfloor
\]

and

\[
m = n - l^2 - l,
\]

(A4)

respectively. Note, finally, that our use of upper and lower indices is purely a notational convenience and should not be confused with exponentiation or a distinction between covariant and contravariant tensors. It also should not be confused with the notation used for the complex spherical harmonics, which also uses upper and lower indexing.

For reference, Table 2 lists the principal symbols, operators, and variables used throughout the paper, with links to the equations and/or sections in which they are presented.
Appendix B
Computing the Flux

B.1. Basic Expression

As we mentioned in Section 2.3, the flux \( f \) is a purely linear function of the spherical harmonic coefficient vector \( \mathbf{y} \):

\[
f = 1 + \mathbf{A} \mathbf{y}. \tag{B1}
\]

In this expression, we assume \( \mathbf{y} \) describes the surface intensity of the star at time \( t = 0 \) in a frame where \( \hat{\mathbf{x}} \) points to the right, \( \hat{\mathbf{y}} \) points up, and \( \hat{\mathbf{z}} \) points out of the page. Even though this is derived in detail in Luger et al. (2019), it is useful to expand on the computation of the design matrix \( \mathbf{A} \), which transforms from spherical harmonics to flux. Let \( \mathbf{a}_k^\top \) denote the \( k \)th row of \( \mathbf{A} \), such that

\[
\mathbf{A} = \begin{pmatrix}
\mathbf{a}_0^\top \\
\mathbf{a}_1^\top \\
\vdots \\
\mathbf{a}_{K-1}^\top 
\end{pmatrix}. \tag{B2}
\]

The row vector \( \mathbf{a}_k^\top \) encodes how the spherical harmonic coefficient vector projects onto the \( k \)th cadence in the flux time series and may be computed from

\[
\mathbf{a}_k^\top = \mathbf{r}^\top \mathbf{A}_1 \mathbf{R}_k(-l) \mathbf{R}_k \left( \frac{2\pi}{P} t_k \right) \mathbf{R}_k \left( \frac{\pi}{2} \right). \tag{B3}
\]

To understand the expression above, let us consider how each of its terms operate on the spherical harmonic representation \( \mathbf{y} \) from right to left. The quantity \( \mathbf{R}_k \) is a Wigner rotation matrix (described in detail in Appendix C.2.1), which in this case rotates the spherical harmonic representation of the star by an angle \( \pi/2 \) counterclockwise about \( \hat{\mathbf{x}} \) such that the north pole of the star points along \( \hat{\mathbf{z}} \). In this frame, we apply a second Wigner rotation matrix, \( \mathbf{R}_k \), to rotate the star about \( \hat{\mathbf{z}} \) counterclockwise (i.e., eastward) by an angle \( 2\pi t_k/P \), where \( P \) is the rotation period and \( t_k \) is the time at cadence \( t \). Next, we rotate the star by a clockwise angle of \( l \) about \( \hat{\mathbf{y}} \), where \( l \) is the stellar inclination (\( l = 0 \) corresponding to a pole-on view and \( l = \pi/2 \) corresponding to an edge-on view). With this last rotation, we are now in the observer’s frame.\footnote{In principle, one last rotation could be performed about \( \hat{\mathbf{z}} \) to orient the projected disk of the star on the plane of the sky; however, the disk-integrated flux is independent of the rotation angle along the plane of the sky (which we refer to as the obliquity), so this step is unnecessary.}

Following Luger et al. (2019), the next step is to project the representation of the star into a more convenient basis for performing the integration over the stellar disk. The change-of-basis matrix \( \mathbf{A}_1 \) (see Appendix B in Luger et al. 2019) projects the stellar map into the polynomial basis (Equation 7 in Luger et al. 2019), comprised of the sequence of monomials in Cartesian coordinates \( (1 x z y^2 x^2 z x y y^2 \cdots) \) where \( z = \sqrt{1 - x^2 - y^2} \) on the surface of the unit sphere. We can now compute the disk-integrated flux by integrating each of the terms in the basis over the unit disk, which is straightforward in the polynomial basis; the individual terms integrate to simple ratios of gamma functions. These are then assembled into the row vector \( \mathbf{r}^\top \), given by Equation (20) in Luger et al. (2019),

\[
\begin{pmatrix}
1 & -u_1 & 0 & \frac{m}{\sqrt{3}} \\
0 & 1 & -u_1 & 0 \\
\frac{m}{\sqrt{3}} & 0 & 1 & -u_1 \\
0 & 0 & 0 & 1 & -u_1 \\
0 & 0 & 0 & 0 & 1 & -u_1 \\
\end{pmatrix}. \tag{B5}
\]

The columns of \( \mathbf{L} \) are constructed from the coefficient vectors of each transformed spherical harmonic, which are in turn computed by multiplying each spherical harmonic by the spherical harmonic representation of the particular limb-darkening law.

In the presence of limb darkening, we may therefore replace our expression for the \( k \)th row of the flux design matrix \( \mathbf{A} \) (Equation (B3)) with

\[
\mathbf{a}_k^\top = \mathbf{r}^\top \mathbf{A}_1 \mathbf{L}(\mathbf{u}) \mathbf{R}_k(-l) \mathbf{R}_k \left( \frac{2\pi}{P} t_k \right) \mathbf{R}_k \left( \frac{\pi}{2} \right) \tag{B6}
\]

for a given limb-darkening coefficient vector \( \mathbf{u} \).
Appendix C
The Expectation Integrals

Our goal in this section is to find closed-form solutions to the first and second moments of the spherical harmonic representation of the stellar surface $\gamma$,

$$E[\gamma | \theta_i] = \int y(\mathbf{x}) p(\mathbf{x} | \theta_i) d\mathbf{x}, \quad (C1)$$

$$E[\gamma \gamma^T | \theta_i] = \int y(\mathbf{x}) y^T(\mathbf{x}) p(\mathbf{x} | \theta_i) d\mathbf{x}, \quad (C2)$$

which are linearly related to the mean and covariance of our GP (Section 2.3). Recall that $\mathbf{x}$ is a vector of parameters describing the exact configuration of features on the surface of a star, and $p(\mathbf{x} | \theta_i)$ is its PDF conditioned on hyperparameters $\theta_i$, which describe the distribution of the features on the surface of one or many stars. As we are specifically interested in modeling the effect of starspots on stellar light curves, we let

$$\mathbf{x} = (n \ c_0 \ \cdots \ c_{n-1} \ \lambda_0 \ \cdots \ \lambda_{n-1} \ \phi_0 \ \cdots \ \phi_{n-1} \ r_0 \ \cdots \ r_{n-1}) \quad (C3)$$

and

$$y(\mathbf{x}) = \sum_{i=0}^{n-1} c_i R_x(\lambda_i) R_x(\phi_i) s(\tau_i), \quad (C4)$$

where $n$ is the total number of spots, $c_i$ is the contrast of the $i$th spot, $\lambda_i$ is its longitude, $\phi_i$ is its latitude, and $\tau_i$ is its radius. The vector function $s(\tau_i)$ returns the spherical harmonic expansion of a negative unit brightness circular spot of radius $\tau_i$ at $\lambda = \phi = 0$, $R_x(\phi_i)$ is the Wigner matrix that rotates the expansion about $\hat{x}$ such that the spot is centered at a latitude $\phi_i$, and $R_y(\lambda_i)$ is the Wigner matrix that then rotates the expansion about $\hat{y}$ such that the spot is centered at a longitude $\lambda_i$; these three functions are detailed in the sections below. Equation (C4) thus provides a way of converting a random variable $\mathbf{x}$ describing the size, brightness, and position of spots to the corresponding representation in terms of spherical harmonics. Regarding this equation, two things should be noted. First, we define $y$ relative to a baseline of zero; i.e., a star with no spots on it will have $y = 0$ (which is why we add unity in the expression for the flux in Equation (3)). Second, and more importantly, we are not interested in any specific value of $y$; rather, we would like to know its expectation value under the probability distribution governing the different spot properties $\mathbf{x}$, i.e., $p(\mathbf{x} | \theta_i)$.

For simplicity, we assume that the total number of spots is fixed to a value $n$, i.e.,

$$p(n | \theta_i) = \delta(n - n), \quad (C5)$$

where $\delta$ is the delta function.26 We further assume that $p(\mathbf{x} | \theta_i)$ is separable in each of the four other spot properties and that all of the spots are drawn from the same distribution,

$$p(\mathbf{x} | \theta_i) = \prod_{i=0}^{n-1} p(c_i | \theta_i) \times p(\lambda_i | \theta_i) p(\phi_i | \theta_i) p(r_i | \theta_i), \quad (C6)$$

where

$$\theta_i = (n \ \theta_c \ \theta_\lambda \ \theta_\phi \ \theta_r) \quad (C7)$$

is the vector of hyperparameters describing the process and $\theta_c$, $\theta_\lambda$, $\theta_\phi$, and $\theta_r$ are yet to be specified. This allows us to rewrite the expectation integrals (Equations (C1) and (C2)) as

$$E[\gamma | \theta_i] = n \ e_c, \quad (C8)$$

and

$$E[\gamma \gamma^T | \theta_i] = n \ E_c, \quad (C9)$$

where we define the first-moment integrals

$$e_c \equiv \int s(\tau) p(\tau | \theta_i) d\tau, \quad (C10)$$

$$e_\phi \equiv \int R_x(\phi) e_c \ p(\phi | \theta_i) d\phi, \quad (C11)$$

$$e_\lambda \equiv \int R_y(\lambda) e_\phi \ p(\lambda | \theta_i) d\lambda, \quad (C12)$$

and

$$e_r \equiv -\int c \ e_\lambda \ p(c | \theta_i) dc \quad (C13)$$

and the second-moment integrals

$$E_r \equiv \int s(\tau) s^T(\tau) p(\tau | \theta_i) d\tau, \quad (C14)$$

26 When modeling a single star using the GP, this assumption is justified by definition. It is less justified when the GP is used to model an ensemble of stars, where each star may have a different total number of spots $n$. However, as we argue in the text, $n$ is extremely difficult to constrain from light curves, in particular because of how degenerate it is with the spot contrast. In practice, we find that assuming that all stars in the ensemble have the same number of spots $n$ leads to higher variance in the estimate of $n$, but it does not lead to noticeable bias in $n$ or any of the other hyperparameters.
\[ E_\delta = \int R_\delta (\phi) E_\delta R_\delta (\phi) p(\phi \mid \theta_\delta) d\phi, \]  
(C15)

\[ E_\phi = \int R_\phi (\lambda) E_\phi R_\phi (\lambda) p(\lambda \mid \theta_\phi) d\phi, \]  
(C16)

and

\[ E_c = \int c^2 E_c p(c \mid \theta_c) dc. \]  
(C17)

In Equations (C8) and (C9), we used the fact that both the mean and the variance of the sum of \( n \) independent, identically distributed random variables are equal to \( n \) times the individual mean and the variance, respectively.

We devote the remainder of this section to the computation of these eight integrals.

### C.1. The Radius Integrals

Below, we compute the first and second moments of the radius distribution \( (\epsilon_r, E_r) \) under a suitable spherical harmonic expansion \( s(r) \) of the spot profile and a suitable probability distribution function for the spot radius, \( p(r \mid \theta_r) \).

#### C.1.1. Spot Profile

We model the brightness \( b \) and angle \( \vartheta \) away from the center of a spot of negative unit intensity and radius \( s \) as

\[ b(r; \vartheta) = \frac{1}{1 + \exp \left( \frac{r - \vartheta}{s} \right)} - 1 \]  
(C18)

for some (constant) shape parameter \( s \). In the limit \( s \to 0 \), \( b \) approaches an inverted top-hat function with a half-width equal to \( r \), corresponding to a circular spot of uniform intensity. For \( s > 0 \), each half of \( b \) is a sigmoid with a half-width at half-maximum equal to \( r \). In our implementation of the algorithm, we choose \( s = 0.2 \), which is small compared to features of interest but not so small as to create numerical issues when computing model gradients (which would be undefined at the spot boundary if the spot profile were truly an inverted top hat).

Our goal now is to expand the function above in spherical harmonics. To that end, we note that in a frame where the spot is centered on \( \hat{z} \) (i.e., at polar angle \( \vartheta = 0 \)), the brightness profile is azimuthally symmetric, so the only nonzero coefficients in the spherical harmonic expansion are those with order \( m = 0 \). The corresponding spherical harmonics are simply proportional to the Legendre polynomials in \( \cos \vartheta \), so our task is simplified to finding the Legendre polynomial expansion of \( b \). Define a vector \( \theta \) of \( K \) equally spaced points between zero and \( \pi \), with coefficients given by

\[ \theta_k = \frac{k\pi}{K - 1}. \]  
(C19)

We wish to model the brightness evaluated at each \( \theta_k \) as a weighted combination of Legendre polynomials,

\[ B s_0(r) = b(r), \]  
(C20)

where \( b(r) \) is computed by evaluating Equation (C18) at each of the \( \theta_k \); \( B \) is a design matrix whose columns are the weighted Legendre polynomials,

\[ B_{k,l} = \sqrt{2l + 1} P_l(\cos \theta_k); \]  
(C21)

and \( s_0(r) \) are the coefficients of the expansion. These are related to the full vector of spherical harmonic coefficients describing the spot, \( s(r) \), by

\[ s_m^l(r) = s_0^l \delta_{m,0}, \]  
(C22)

or, in vector form,

\[ s(r) = \mathcal{I} s_0(r), \]  
(C23)

where \( \mathcal{I} \) is a rectangular \( (l_{\text{max}} + 1)^2 \times (l_{\text{max}} + 1) \) identity-like matrix with components

\[ \mathcal{I}_{m,l} = \delta_{m,l^2 + l + 1}, \]  
(C24)

and \( \delta \) is the Kronecker delta function. To find the coefficients \( s_0(r) \) (and hence \( s(r) \)), we solve the (linear) inverse problem,

\[ s_0 = \mathcal{B}^+ b(r), \]  
(C25)

where

\[ \mathcal{B}^+ = S (\mathcal{B}^\top \mathcal{B} + \epsilon I)^{-1} \mathcal{B}^\top \]  
(C26)

is the smoothed pseudo-inverse of \( \mathcal{B} \) with a small regularization parameter \( \epsilon \), \( I \) is the identity matrix, and \( S \) is a diagonal smoothing matrix with coefficients

\[ S_{k,l} = \exp \left[ -\frac{l(l + 1)}{2\xi^2} \right] \delta_{k,l} \]  
(C27)

for a smoothing strength \( \xi \). For \( \epsilon \to 0 \) and \( \xi \to \infty \), \( \mathcal{B}^+ \) is the exact pseudo-inverse of \( \mathcal{B} \). However, \( \epsilon > 0 \) is chosen for improved numerical stability, and \( \xi > 0 \) is chosen to mitigate the effect of ringing in the solution. In practice, we obtain good results with \( \epsilon \approx 10^{-9} \) and \( \xi \approx 15 \).

Figure 15 shows the intensity profile for spots of different radii expanded to a spherical harmonic degree \( l_{\text{max}} = 15 \). The average intensity within the spots is close to \( -1 \), and the half-widths at half-maximum equal to the spot radius, as expected. The effect of ringing due to the truncated spherical harmonic expansion is evident, although it is strongly suppressed compared to an expansion without the smoothing term (i.e., \( \xi = \infty \)). However, for \( r \lesssim 10^2 \), an expansion to \( l_{\text{max}} = 15 \) is insufficient to correctly model the spot, as can be seen from the \( r = 5^\circ \) profile (dashed curve). Expansions to a higher spherical harmonic degree allow one to model spots with radii smaller than \( 10^2 \), although at increased computational cost and potential numerical stability issues; we discuss this point at length in Section 4.1.

#### C.1.2. Probability Density Function

For simplicity, we will adopt a uniform probability distribution for the spot radius, characterized by a mean radius \( r \) and half-width \( \Delta r \),

\[ p(r \mid \theta_r) = \begin{cases} \frac{1}{2\Delta r} & r - \Delta r \leq r \leq r + \Delta r \\ 0 & \text{otherwise} \end{cases} \]  
(C28)

where the hyperparameters of the distribution are

\[ \theta_r = (r \Delta r \hat{r}). \]  
(C29)

As we argue in the text, in practice, it is often difficult to constrain the moments of the radius distribution above the first (the mean). It is therefore useful to also consider the limiting case of the radius distribution as \( \Delta r \to 0 \), in which case the
PDF becomes

\[ p(\mathbf{r} \mid \theta_r, \Delta r = 0) = \delta(\mathbf{r} - \mathbf{r}), \]  

where \( \delta \) is the delta function.

\subsection*{C.1.3. First Moment}

The first moment of the radius distribution is (Equation (C10))

\[ e_r \equiv \int s(\mathbf{r}) p(\mathbf{r} \mid \theta_r) \, d\mathbf{r} = \frac{1}{2\Delta r} \int_{r_0-\Delta r}^{r_0+\Delta r} s(\mathbf{r}) \, d\tau. \]  

(C31)

Using the equations from the previous section, its components may be written

\[ (e_r)_m^l = \frac{\delta_{m,0}}{2\Delta r} \int_{r-\Delta r}^{r+\Delta r} \tilde{s}_l^1(\mathbf{r}) \, d\mathbf{r} = \frac{\delta_{m,0}}{2\Delta r} \sum_{k=0}^{K-1} B_{0,k}^+ b_k(\mathbf{r}) \, d\tau = \frac{\delta_{m,0}}{2\Delta r} \sum_{k=0}^{K-1} B_{0,k}^+ c_k(\mathbf{r}, \Delta r), \]  

(C32)

where

\[ c_k(\mathbf{r}, \Delta r) = \frac{s}{2\Delta r} \ln \left( 1 + \chi_k^0(\mathbf{r}, \Delta r) \right) \]  

and

\[ \chi_k^0(\mathbf{r}, \Delta r) \equiv \exp \left( \frac{r - \Delta r - \theta_k}{s} \right) \]  

\[ \chi_k^1(\mathbf{r}, \Delta r) \equiv \exp \left( \frac{r + \Delta r - \theta_k}{s} \right), \]  

(C34)

In vector form, we may simply write

\[ e_r = \mathbb{I} B^+ c(r, \Delta r). \]  

(C35)

Note, finally, that in the limit \( \Delta r \to 0 \),

\[ \lim_{\Delta r \to 0} e_r = \mathbb{I} B^+ b(r). \]  

(C36)

\subsection*{C.1.4. Second Moment}

The second moment of the radius distribution is (Equation (C14))

\[ E_r \equiv \int s(\mathbf{r}) s^\top(\mathbf{r}) p(\mathbf{r} \mid \theta_r) \, d\mathbf{r} = \frac{1}{2\Delta r} \int_{r-\Delta r}^{r+\Delta r} s(\mathbf{r}) s^\top(\mathbf{r}) \, d\tau. \]  

(C37)

As before, its components may be written

\[ (E_r)_m^l = \frac{\delta_{m,0} \delta_{m',0}}{2\Delta r} \int_{r-\Delta r}^{r+\Delta r} \tilde{s}_m^1(\mathbf{r}) \tilde{s}_{m'}^1(\mathbf{r}) \, d\mathbf{r} = \frac{\delta_{m,0} \delta_{m',0}}{2\Delta r} \sum_{k=0}^{K-1} B_{k,k-1}^+ b_k(\mathbf{r}) b_{k'}(\mathbf{r}) \, d\tau = \frac{\delta_{m,0} \delta_{m',0}}{2\Delta r} \sum_{k=0}^{K-1} B_{k,k-1}^+ b_{k'}(\mathbf{r}) c_k(\mathbf{r}, \Delta r), \]  

(C38)

where

\[ C_{k,k'}(r, \Delta r) \equiv \frac{s}{2\Delta r} \begin{cases} \begin{align*} &\frac{\exp \left( \frac{r - \Delta r}{s} \right) \ln \left( \frac{1 + \chi_k^0}{1 + \chi_k^1} \right) - \frac{\ln \left( 1 + \chi_k^1 \right)}{1 + \chi_k^1} \end{align*} \end{cases} \quad k = k' \]  

(C39)

In vector form, this may be written

\[ E_r = \mathbb{I} B^+ C(r, \Delta r) B^+ \mathbb{T}. \]  

(C40)

Finally, in the limit \( \Delta r \to 0 \),

\[ \lim_{\Delta r \to 0} E_r = \mathbb{I} B^+ b(r) b^\top(r) B^+ \mathbb{T}. \]  

(C41)

\subsection*{C.2. The Latitude Integrals}

Our goal in this section is to compute the first and second moments of the latitude distribution \( e_\phi \) and \( E_\phi \), given by Equations (C11) and (C15), respectively. These involve integrals over the terms in the Wigner rotation matrix for spherical harmonics, which we discuss below.

\subsection*{C.2.1. Rotation Matrices}

The Wigner rotation matrix for real spherical harmonics up to degree \( l_{\text{max}} \) may be written as the block-diagonal matrix

\[ R = \begin{pmatrix} R^0 & & \\ & R^1 & \\ & & \ddots \end{pmatrix}. \]  

(C42)

where

\[ R^l = U^l \mathbf{D} U^l \]  

(C43)

is the Wigner rotation matrix for a single spherical harmonic degree,

\[ U^l = \frac{1}{\sqrt{2}} \begin{pmatrix} \begin{array}{c} i \\ 1 \end{array} & \begin{array}{c} i \\ -1 \end{array} \end{pmatrix} \]  

(C44)

describes the transformation from complex to real spherical harmonics, and \( \mathbf{D} \) is the Wigner matrix for complex spherical harmonics, whose terms are given by the expression

\[ D^l_{m,m'}(\alpha, \beta, \gamma) = \exp(-im')d^l_{m,m'}(\beta)\exp(-im\gamma), \]  

(C45)

where \( \alpha, \beta, \text{and} \gamma \) are the Euler angles describing the rotation in the \( \hat{z} - \hat{y} - \hat{z} \) convention and \( i \) is the imaginary unit (Collado et al. 1989). The terms of the \( \delta \)-matrix depend on powers of \( \sin(\beta/2) \) and \( \cos(\beta/2) \) (see Equation (C15) in Luger et al. 2019), but it is
convenient to use the half-angle formula to express these terms instead as
\[
d_{m,m'}^l(\beta) = \sum_{i=0}^{2l} \epsilon_{m,m',i}^l \text{sgn}(\sin \beta)^i (1 - \cos \beta)^{\frac{i}{2}} (1 + \cos \beta)^{\frac{3}{2}} \frac{m - m' - i}{},
\]
where
\[
\epsilon_{m,m',i}^l = \begin{cases} (1)^{-i} & \text{if } m - m' - i \text{ even} \\ 0 & \text{if } m - m' - i \text{ odd} \end{cases}
\]

\[
C\text{-}\text{2.2. Probability Density Function}
\]

The latitude integrals (Equations (C11) and (C15)) involve rotations by an angle \(\phi\) about \(\hat{\mathbf{x}}\), may be accomplished by choosing Euler angles \(\alpha = \frac{\pi}{2}, \beta = \phi,\) and \(\gamma = -\frac{\pi}{2}\), such that
\[
R^\phi_z(\phi) = U^l \hat{D}^\phi_z(\phi) U^l
\]
with
\[
D^\phi_z(\phi) = D^\phi_z\left(\frac{\pi}{2}, \phi, -\frac{\pi}{2}\right).
\]

From the expressions above, it is clear that all terms in \(R^\phi_z(\phi)\) are equal to (weighted) sums of powers of \((1 \pm \cos \phi)\). Since our goal is to compute the integrals of these terms multiplied by a PDF, it is convenient to model \(\cos \phi\) as a beta-distributed variable. As we will see, this choice will allow us to analytically compute the first two moments of the distribution of \(\phi\) conditioned on \(\theta_\phi\).

The beta distribution for \(\cos \phi\) has the hyperparameters \(\alpha\) and \(\beta\) (not to be confused with the Euler angles \(\alpha\) and \(\beta\)) and the PDF given by
\[
p(\cos \phi \mid \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (\cos \phi)^{\alpha-1} (1 - \cos \phi)^{\beta-1},
\]
where \(\Gamma\) is the gamma function. The implied distribution for \(\phi\) may be computed by a straightforward change of variable, \(\phi = \cos^{-1}\theta\).

For \(\phi \in \left[\frac{-\pi}{2}, \frac{\pi}{2}\right]\). Both \(\alpha\) and \(\beta\) are restricted to \((0, \infty)\). However, in practice, it is necessary to limit the values of these parameters to a finite range to ensure the numerical stability of the algorithm. It is also convenient to work with the log of these quantities because of their large dynamic range. We therefore introduce the modified parameters
\[
a = \frac{\ln \alpha - K_{00}}{K_{10} - K_{00}}
\]
\[
b = \frac{\ln \beta - K_{00}}{K_{11} - K_{10}}
\]
with inverse transform
\[
\alpha = \exp(K_{00} + (K_{10} - K_{00})a)
\]
\[
\beta = \exp(K_{10} + (K_{11} - K_{10})b),
\]
where the matrix
\[
K = \begin{pmatrix} 0 & 10 \\ \frac{1}{10} & 10 \end{pmatrix}
\]
defines the minimum and maximum values of \(\ln \alpha\) (top row) and \(\ln \beta\) (bottom row) we adopt in our implementation of the algorithm. The lower limits correspond to \(\alpha > 1\) and \(\beta > \frac{1}{2}\), which excludes distributions with unphysically sharp peaks at \(\phi = 90^\circ\). Both \(a\) and \(b\) are restricted to the domain \((0, 1)\) and together comprise the hyperparameter vector
\[
\theta_\phi = (a \ b \beta).
\]

Because of their trivial domain, these parameters are convenient to sample in when doing inference (provided we account for their implied prior on the spot latitudes; see below). However, \(a\) and \(b\) do not intuitively relate to physical quantities of interest. In many cases, it is more desirable to parameterize the latitude distribution in terms of a parameter \(\mu_\phi\) controlling the central latitude and a parameter \(\sigma_\phi\) controlling the dispersion in latitude among the spots. In this case, we may instead choose
\[
\theta_\phi = (\mu_\phi \ \sigma_\phi \beta).
\]

Over most of the parameter space in \(a\) and \(b\), the spot latitude distribution defined above is well approximated by a bimodal Gaussian. In particular, there exists a one-to-one relationship between \(a\) and \(b\) and the mean \(\mu_\phi\) and standard deviation \(\sigma_\phi\) of a normal approximation to the distribution. Moreover, we find that if we let \(\mu_\phi\) be the mode of the PDF and \(\sigma_\phi^2\) be a local approximation to the variance of the PDF, the relationship has a convenient closed form.

To compute \(\mu_\phi\), we differentiate Equation (C51) with respect to \(\phi\), set the expression equal to zero, and solve for \(\phi\) to obtain
\[
\mu_\phi = 2\tan^{-1}\left(\sqrt{2\alpha + \beta - 2 - \sqrt{4\alpha^2 - 8\alpha - 6\beta + 4\alpha\beta + \beta^2 + 5}}\right).
\]

To compute \(\sigma_\phi\), we note that the variance of a Gaussian distribution \(\varphi(\phi; \mu, \sigma^2)\) is the negative reciprocal of its curvature in log space:
\[
\sigma^2 = -\left(d^2 \ln \varphi(\phi; \mu, \sigma^2) / d\phi^2\right)^{-1}.
\]
We therefore twice differentiate the log of Equation (C51), negate it, take the reciprocal, and evaluate it at \(\phi = \mu_\phi\) to obtain...
a local approximation to the standard deviation of the distribution at the mode:

\[
\sigma_\phi = \frac{\sin \mu_\phi}{\sqrt{1 - \alpha + \beta + (\beta - 1) \cos \mu_\phi + \frac{\alpha - 1}{\cos \mu_\phi}}}. \tag{C59}
\]

For completeness, the inverse transform also has a closed form:

\[
\begin{align*}
\alpha &= \frac{2 + 4\sigma_\phi^2 + (3 + 8\sigma_\phi^2)\cos \mu_\phi + 2\cos(2\mu_\phi) + \cos(3\mu_\phi)}{16\sigma_\phi^4 \cos \left( \frac{\mu_\phi}{2} \right)^4}, \\
\beta &= \frac{\cos \mu_\phi + 2\sigma_\phi^2(3 + \cos(2\mu_\phi)) - \cos(3\mu_\phi)}{16\sigma_\phi^4 \cos \left( \frac{\mu_\phi}{2} \right)^4}. \tag{C60}
\end{align*}
\]

For reference, Figure 16 shows the latitude PDF and the corresponding Gaussian approximation for different values of \(\mu_\phi\) and \(\sigma_\phi\). The corresponding values of \(a\) and \(b\) are indicated within each panel. For \(\mu_\phi\) at intermediate latitudes and moderate values of \(\sigma_\phi\), the approximation is quite good. However, for \(\mu_\phi\) very close to the equator or poles, the curvature of the distribution changes significantly as a function of \(\phi\), so the variance is somewhat underestimated by the approximation; and for large \(\sigma_\phi\), the distribution becomes noticeably non-Gaussian.

The shaded panel at the lower left is a special case of the distribution \((\mu_\phi = 0^\circ, \sigma_\phi \approx 40^\circ)\); or equivalently, \(a \approx 0.06, b = 0\), which is approximately isotropic in latitude. In this panel, the orange curve instead corresponds to an isotropic (cosine) distribution in \(\phi\); note the excellent agreement. Thus, in addition to having closed-form moments, the beta distribution is quite flexible and, via the transforms outlined above, intuitive in how it affects the distribution of spots on the surface of a star.

In the following sections, we derive expressions for the moments of the distribution in terms of \(\alpha\) and \(\beta\), as this is somewhat more convenient; these can easily be transformed into expressions involving either \(\mu_\phi\) and \(\sigma_\phi\) or \(a\) and \(b\) via the equations above. The former parameterization is convenient when these properties are known and can be fixed; i.e., when using the GP as a restrictive prior for the light curve of a star whose spot distribution is already understood. However, for the purposes of posterior inference—that is, when trying to constrain the hyperparameters of the GP—we recommend sampling in the parameters \(a\) and \(b\), since their domains are trivial, with uncorrelated boundaries. Posterior constraints on these quantities may easily be transformed into constraints on \(\mu_\phi\) and \(\sigma_\phi\) via the equations above. Note, importantly, that this requires us to explicitly add a Jacobian term to the likelihood to account for the prior implied by sampling uniformly in \(a\) and \(b\) in the range \((0, 1)\). The Jacobian is given by \(\nabla\).
Adding the log of the absolute value of $J$ to the log likelihood corrects for the ad hoc prior on the latitude parameters introduced by our particular choice of parameterization, instead enforcing a uniform prior on the quantities $\mu$ and $\sigma$.

C.2.3. First Moment

Since the Wigner matrices are block diagonal, we may evaluate the moments of the distribution one spherical harmonic degree at a time. To that end, let us write the first-moment integral as

$$ e_\phi = \int R_{\frac{x}{2}}(\phi) e_{\phi} p(\phi | \theta_\phi) d\phi = \left(e_{\phi}^0 e_{\phi}^1 e_{\phi}^2 \ldots e_{\phi}^{l_{\max}}\right)' , $$

where

$$ e_{\phi}^l = \int R_{\frac{x}{2}}(\phi) e_{\phi}^l p(\phi | \theta_\phi) d\phi = U_l^{\rightarrow}p_{\phi,l} , $$

and we define

$$ p_{\phi,l} = \int D_{\frac{x}{2}}(\phi) e_{\phi}^l p(\phi | \theta_\phi) d\phi , $$

$$ \bar{e}_{\phi}^l = U^l e_{\phi}^l . $$

The integral $p_{\phi,l}$ defined above has a closed-form solution. To show this, we write the terms of $p_{\phi,l}$ as

$$ (p_{\phi,l})_m = \int \sum_{\mu=-l}^{l} (D_{\frac{x}{2}})_{m,\mu}(\phi) (\bar{e}_{\phi}^l)_\mu p(\phi | \theta_\phi) d\phi = \Gamma(\alpha + \beta) \sum_{\mu=-l}^{l} (\bar{e}_{\phi}^l)_\mu \times \exp \left[ \frac{it}{2} m - \mu \right] \sum_{i=0}^{2l} c_{m,p,i} (q_{\phi,l,i}(\theta_\phi) ,$$

where

$$ (q_{\phi,l,i}(\theta_\phi) = \frac{1}{2} \int_0^{\frac{\pi}{2}} \frac{\text{sgn}(\sin \phi) | \sin \phi| (\cos \phi)^{\alpha-1}}{(1 - \cos \phi)^{\alpha + \beta + i - 1}(1 + \cos \phi)^{\frac{\beta}{2}}} d\phi$$

$$ = \frac{1}{2} \int_0^{\frac{\pi}{2}} x^{\alpha-\frac{i}{2}}(1 - x)^{\frac{\beta}{2} - \frac{i}{2} - 1}(1 + x)^{\frac{\beta}{2}} d\phi \quad i \text{ even} ,$$

$$ = 0 \quad i \text{ odd} .$$

and in the last line, we make use of the transformation $x = \cos \phi$. The integral in the expression above has a closed-form solution in terms of the hypergeometric function $\text{F}_1$: 

$$ (q_{\phi,l,i}(\theta_\phi) = \left\{ \begin{array}{ll} \frac{\Gamma(\alpha + \beta + \frac{i}{2})}{\Gamma(\alpha + \beta - \frac{i}{2})} \text{F}_1 \left(-i, \alpha; l + \alpha + \beta - \frac{i}{2}; -1\right) & i \text{ even} \vphantom{\text{F}^2} \\
0 & i \text{ odd} . \end{array} \right. $$

In order to compute the integral $e_\phi$ (Equation (C11)), we must evaluate Equation (C69) for all $0 \leq l \leq l_{\max}$, $0 \leq i \leq 2l$, which can be done efficiently via upward recursion relations for both the gamma and hypergeometric functions.

C.2.4. Second Moment

Similarly to before, let us write the second-moment integral as

$$ E_\phi = \int R_{\frac{x}{2}}(\phi) E_{\phi} R_{\phi}^{\frac{3}{2}}(\phi) p(\phi | \theta_\phi) d\phi = \left(E_{\phi}^0 E_{\phi}^1 E_{\phi}^2 \ldots E_{\phi}^{l_{\max}}\right) , $$

where

$$ E_{\phi}^{l_{\max}} = \int R_{\frac{x}{2}}(\phi) E_{\phi}^{l_{\max}} R_{\phi}^{\frac{3}{2}}(\phi) p(\phi | \theta_\phi) d\phi = U_{l_{\max}}^{\rightarrow}E_{\phi}^{l_{\max}} U_{l_{\max}}^{\rightarrow} , $$

and we define

$$ E_{\phi}^{l,\prime} = \int D_{\frac{x}{2}}(\phi) E_{\phi}^{l,\prime} D_{\phi}^{\frac{3}{2}}(\phi) p(\phi | \theta_\phi) d\phi , $$

$$ E_{\phi}^{l,\prime} = U_l E_{\phi}^{l,\prime} U_l^{\rightarrow} . $$

As before, we may express the solution to the integral $P_{\phi,l}$ in closed form. Let us write the terms of $P_{\phi,l}$ as

$$(P_{\phi,l})_{m,m'} = \int \sum_{\mu=-l}^{l} \sum_{\mu'=-l}^{l'} (D_{\frac{x}{2}})_{m,\mu}(\phi) (E_{\phi}^{l,\prime})_{\mu,\mu'} \times (D_{\phi}^{\frac{3}{2}})_{m',\mu'}(\phi) p(\phi | \theta_\phi) d\phi$$

$$ = \Gamma(\alpha + \beta) \sum_{\mu=-l}^{l} (E_{\phi}^{l,\prime})_{\mu,\mu'} \exp \left[ \frac{it}{2} (m - \mu + m' - \mu') \right]$$

$$ \times \sum_{i=0}^{2l} c_{m,p,i}^{l'} (Q_{\phi,l}(\theta_\phi),$$

where, similarly to before,

$$(Q_{\phi,l})_{i,i')(\theta_\phi) = \frac{1}{2} \int_0^{\frac{\pi}{2}} \frac{\text{sgn}(\sin \phi)^{i+i'} | \sin \phi| (\cos \phi)^{\alpha-1}}{(1 - \cos \phi)^{i+i'+\frac{\beta}{2} - i}(1 + \cos \phi)^{\frac{\beta}{2}} d\phi$$

$$ = \frac{1}{2} \int_0^{\frac{\pi}{2}} x^{\alpha-\frac{i}{2}}(1 - x)^{i+i'-\frac{\beta}{2} - 1}(1 + x)^{\frac{\beta}{2}} d\phi \quad i \text{ even} \quad i \text{ odd} .$$

We may again express this integral in closed form:

$$ \left. \text{(C75)} \right\} \frac{1}{2} \int_0^{\frac{\pi}{2}} x^{\alpha-\frac{i}{2}}(1 - x)^{i+i'-\frac{\beta}{2} - 1}(1 + x)^{\frac{\beta}{2}} d\phi \quad i \text{ even} \quad i \text{ even} .$$
\[
(Q^{L'\nu}_{\nu})_{i,j}(\theta_0) = \frac{\left(\Gamma(\alpha)\Gamma(\alpha + \beta - \frac{i + i'}{2})\right)}{\Gamma(\alpha + \beta - \frac{i + i'}{2})} 2E_1\left(-\frac{i + i'}{2}, \alpha; I + I' + \alpha + \beta - \frac{i + i'}{2}, -1\right) \begin{cases} i + i' & \text{even} \\ i + i' & \text{odd}. \end{cases}
\]

As before, this integral may be evaluated recursively to efficiently compute all of the terms in \(E_{\phi}\).

### C.3. The Longitude Integrals

In this section, we will compute the first and second moments of the longitude distribution (\(e_\lambda\) and \(E_{\lambda}\), given by Equations (C12) and (C16), respectively). The math here is very similar to that in the previous section, as we are again dealing with integrals of Wigner matrices (Appendix C.2.1).

#### C.3.1. Probability Density Function

The longitude integrals (Equations (C12) and (C16)) involve rotations by an angle \(\lambda\) about \(y\), which may be accomplished by choosing Euler angles \(\alpha = 0, \beta = \lambda, \gamma = 0\), such that

\[
R_{\lambda} = U^{ij}D_{\lambda}^{ij}(\lambda)U^T
\]

with

\[
D_{\lambda}^{ij}(\lambda) = D^i(0, \lambda, 0).
\]

Since we expect the longitudinal distribution of features on the surfaces of stars to be (on average) isotropic, we will place a uniform prior on \(\lambda \in [-\pi, \pi)\):

\[
p(\lambda | \theta_0) = \begin{cases} \frac{1}{2\pi} & -\pi \leq \lambda < \pi \\ 0 & \text{otherwise}. \end{cases}
\]

We therefore have no hyperparameters controlling the longitudinal distribution, i.e.,

\[
\theta_\lambda = (\ ).
\]

#### C.3.2. First Moment

As before, we will solve for the terms of the moment integrals one spherical harmonic degree at a time:

\[
e_\lambda = \int R_{\lambda}(\lambda) e_\phi \ p(\lambda | \theta_0) \ d\lambda = (e_\lambda^0, e_\lambda^1, e_\lambda^2, \ldots, e_{\lambda_{\text{max}}}^\lambda, \ldots),
\]

where

\[
e_\lambda = \int R_{\lambda}(\lambda) e_\phi \ p(\lambda | \theta_0) \ d\lambda = U^{iT}p_\lambda^i,
\]

and we define

\[
p_\lambda^i = \int D_{\lambda}^{i,j}(\lambda) e_\phi \ p(\lambda | \theta_0) \ d\lambda,
\]

\[
ed_\phi^i \equiv U^ie_\phi^i.
\]

The integral \(p_\lambda^i\) defined above has a closed-form solution. To show this, we write the terms of \(p_\lambda^i\) as

\[
(p_\lambda^i)_{\mu\nu} = \int \sum_{\mu'=-\mu}^{\mu} D_{\lambda}^{i,j}(\lambda) (e_\phi^j)_{\mu'\nu'} \ p(\lambda | \theta_0) \ d\lambda = \sum_{\mu'=-\mu}^{\mu} (e_\phi^j)_{\mu'\nu'} \ \sum_{\nu'=-\nu}^{\nu} D_{\lambda}^{i,j}(\lambda) \ p(\lambda | \theta_0) \ d\lambda
\]

where

\[
(q_{\lambda}^j)_{\mu\nu} \equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{sgn}(\sin \lambda) \ y^{-\frac{1}{2}} (1 + \cos \lambda)^{\frac{\alpha}{2}} \ d\lambda
\]

whose terms may easily be computed by upward recursion. Since \(q_{\lambda}^j\) does not depend on any user inputs, it may be computed a single time as a preprocessing step for efficiency.

#### C.3.3. Second Moment

We write the second-moment integral as

\[
E_\lambda = \int R_{\lambda}^T(\lambda) E_{\phi} R_{\lambda}(\lambda) \ p(\lambda | \theta_0) \ d\lambda
\]

\[
= \begin{bmatrix} E_{\lambda}^{0,0} & E_{\lambda}^{0,1} & E_{\lambda}^{0,2} & \cdots & E_{\lambda}^{0,\lambda_{\text{max}}} \\ E_{\lambda}^{1,0} & E_{\lambda}^{1,1} & E_{\lambda}^{1,2} & \cdots & E_{\lambda}^{1,\lambda_{\text{max}}} \\ E_{\lambda}^{2,0} & E_{\lambda}^{2,1} & E_{\lambda}^{2,2} & \cdots & E_{\lambda}^{2,\lambda_{\text{max}}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E_{\lambda}^{\lambda_{\text{max}},0} & E_{\lambda}^{\lambda_{\text{max}},1} & E_{\lambda}^{\lambda_{\text{max}},2} & \cdots & E_{\lambda}^{\lambda_{\text{max}},\lambda_{\text{max}}} \end{bmatrix}
\]

where

\[
E_{\lambda}^{L\nu} = \int R_{\lambda}^T(\lambda) E_{\phi}^{L\nu} R_{\lambda}(\lambda) \ p(\lambda | \theta_0) \ d\lambda
\]

\[
= U^{T-1}p_{\lambda}^{L\nu} U^{T-1},
\]

and we define

\[
P_{\lambda}^{L\nu} = \int D_{\lambda}^{i,j}(\lambda) E_{\phi}^{L\nu} \ p(\lambda | \theta_0) \ d\lambda,
\]

\[
b_{\phi}^{L\nu} \equiv U^i b_{\phi}^{L\nu}.
\]

We then express the terms of \(P_{\lambda}^{L\nu}\) as

\[
(P_{\lambda}^{L\nu})_{\mu,\nu'} = \sum_{\mu'=-\mu}^{\mu} (D_{\lambda}^{L\nu})_{\mu',\nu'} \ E_{\phi}^{L\nu} \ p(\lambda | \theta_0) \ d\lambda
\]

\[
= \sum_{\mu'=-\mu}^{\mu} (D_{\lambda}^{L\nu})_{\mu',\nu'} \ (E_{\phi}^{L\nu})_{\mu'\nu'} \times \sum_{\nu'=-\nu}^{\nu} D_{\lambda}^{i,j}(\lambda) \ p(\lambda | \theta_0) \ d\lambda
\]

\[
= \sum_{\mu'=-\mu}^{\mu} (E_{\phi}^{L\nu})_{\mu'\nu'} \times \sum_{\nu'=-\nu}^{\nu} \sum_{\nu'=0}^{\nu} D_{\lambda}^{i,j}(\lambda) \ p(\lambda | \theta_0) \ d\lambda
\]

\[
= \sum_{\mu'=-\mu}^{\mu} (E_{\phi}^{L\nu})_{\mu'\nu'} \times \sum_{\nu'=-\nu}^{\nu} \sum_{\nu'=0}^{\nu} D_{\lambda}^{i,j}(\lambda) \ p(\lambda | \theta_0) \ d\lambda
\]
where \( \mathcal{Q}(\nu) \) is given by

\[
(\mathcal{Q}(\nu))_{ij} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \text{sgn}(\sin \lambda)^{i+j} \\
\times (1 - \cos \lambda)^{l+l'-i-j} (1 + \cos \lambda)^{i'-j} \, d\lambda
\]

\[
= \begin{cases} 
2^{i+j}(\lambda^{i+j}) (1+\lambda)^{l+i'} (1+\lambda')^{l+i} & i + i' \text{ even} \\\n0 & i + i' \text{ odd}
\end{cases}
\]

whence these terms may again be computed by upward recursion in a single preprocessing step.

\[\text{C.4. The Contrast Integrals}\]

The final integrals we must take in our computation of \( E[y \mid \theta]\) and \( \bar{E}[y' \mid \theta]\) are the integrals over the spot contrast distribution, \( \epsilon_c \) and \( \epsilon_c\). These are by far the easiest, since the spot contrast is a scalar multiplier of the spherical harmonic coefficient vector, so we can pull the terms \( \epsilon_c\) and \( \epsilon_c\) out of the integrals in Equations (C13) and (C17) to write

\[
\epsilon_c = -\epsilon_c \int c \, p(c \mid \theta) \, dc,
\]

\[
\epsilon_c = \epsilon_c \int c^2 \, p(c \mid \theta) \, dc.
\]

These integrals may be computed analytically for any choice of PDF \( p(c \mid \theta) \) with closed-form moments. However, in practice, it is quite difficult to constrain the spot contrast from light curves, let alone higher moments of its distribution; this is due largely to the fact that the contrast is extremely degenerate with the total number of spots (Paper I). In our implementation of the algorithm, we therefore choose the simplest possible probability distribution, a delta function

\[
p(c \mid \theta) = \delta(c - c),
\]

characterized by a single parameter, the contrast of the spots:

\[
\theta_c = (c \parallel). \tag{C96}
\]

The moment integrals are then trivial to evaluate:

\[
\epsilon_c = -c \, \epsilon_c, \tag{C97}
\]

\[
\epsilon_c = c^2 \, \epsilon_c, \tag{C98}
\]

\[\text{Appendix D}\]

\section*{Inclination}

In this section, we will compute the first- and second-moment integrals of the inclination distribution (Equations (20) and (21)), which allow us to compute the mean and covariance of the process that describes the flux marginalized over all values of the inclination (Equations (18) and (19)).

\[\text{D.1. Probability Density Function}\]

Similar to the latitude integrals, the process of inclining a star relative to the observer (see Equation (B3)) involves rotations by an angle \( -\varpi \) about \( \hat{x} \), which may be accomplished by choosing Euler angles \( \alpha = \pi/2 \), \( \beta = -\varpi \), and \( \gamma = -\pi/2 \), such that

\[
R_x^l(-\varpi) = U^{l,2} D_x^l(-\varpi) U^l
\]

with

\[
D_x^l(-\varpi) = D_l\left(\frac{\pi}{2}, -\varpi, -\frac{\pi}{2}\right).
\]

Since we expect an isotropic distribution of stellar rotation axes (absent prior constraints on individual stars), the prior probability density for the inclination \( I \) is simply

\[
p(I) = \sin I \tag{D3}
\]

for \( I \in [0, \pi/2] \).

\[\text{D.2. First Moment}\]

The expression for the first moment is

\[
\epsilon_l \equiv \int \mathcal{A}(I, P, \bm{u}) \, \epsilon_p(I) \, dI, \tag{D4}
\]

where

\[
\epsilon_p \equiv E[y \mid \theta] \tag{D5}
\]

is the first moment of the distribution over spherical harmonic coefficients (Equation (11)). We can use Equations (B2) and (B3) to express the element at index \( k \) (corresponding to the mean of the GP at time \( t = t_k \)) as

\[
(e^l)_k = \int a_k \mathcal{A}_1 \epsilon_p(I) \, dI \nonumber
\]

\[
= r^\top A_1 \int R_x(-\varpi) \epsilon_p(I) \, dI, \tag{D6}
\]

where we define

\[
(e^l)_k \equiv R_x(\frac{2\pi}{P}) R_z(\frac{\pi}{2}) \epsilon_p \tag{D7}
\]

as the expectation of \( y \) in the polar frame at time \( t = t_k \). At this point, it is convenient to invoke the fact that our GP is longitudinally isotropic; there is no preferred longitude on the surface of the star or, equivalently, no preferred phase in the light curve. The rotation about \( \hat{z} \) (i.e., the rotational axis of the star) therefore cannot change the expectation of \( y \), so

\[
(e^l)_k = (e^l)_0
\]

\[
= R_z(\frac{\pi}{2}) \epsilon_p
\]

\[
\equiv e^l. \tag{D8}
\]

We therefore have

\[
\epsilon_l \equiv \epsilon_l \mathbf{1}, \tag{D9}
\]

where

\[
\epsilon_l = r^\top A_1 \epsilon_p \tag{D10}
\]

and

\[
\epsilon_p = \int R_x(-\varpi) \epsilon_p(I) \, dI
\]

\[
= (e^l_0, e^l_1, e^l_2, \cdots, e^l_{2^n-1}) \tag{D11}
\]

is the expectation of \( y \) in the observer’s frame, and as before, we explicitly separate it out by spherical harmonic degree. As

\[27\]

In the presence of limb darkening, we must include the limb-darkening operator \( L(u) \) in Equation (D6); see Appendix B.2.
in Appendix C.2, we may write
\[
e_{i,j}' = \int R^T_{i,j}(\Xi) e_{i,j} p(\Xi) \, d\Xi
= U^{-1} p_{i,j}',
\]
and we define
\[
p_{i,j}' = \int D^T_{i,j}(\Xi) e_{i,j} p(\Xi) \, d\Xi,
\]
\[
e_{i,j}' = U' e_{i,j}'.
\]
The integral \( p_{i,j}' \) defined above has a closed-form solution. To show this, we write the terms of \( p_{i,j}' \) as
\[
(p_{i,j}')_m = \int \sum_{i=-l}^l (D^T_{i,j})_{m,i} (-1) (e_{i,j}')_m p(\Xi) \, d\Xi
= \sum_{i=-l}^l \epsilon_{i,m} (e_{i,j}')_m (q_{i,j}')_m,
\]
where
\[
(q_{i,j}')_m = \int_0^{\frac{\pi}{2}} \frac{(1 - \cos \Xi)^{\frac{3}{2} - i}}{(1 + \cos \Xi)^{\frac{3}{2} + i}} \sin \Xi \, d\Xi
= \frac{(-1)^i}{l - \frac{i}{2} + 1} \binom{1}{-\frac{i}{2}} x \binom{l}{\frac{i}{2} + 1},
\]
which may easily be computed recursively. As with the longitude integrals, the vector \( q_{i,j}' \) need only be computed a single time as a preprocessing step, as it does not depend on any user inputs.

\[
(E_{i,j}')_{k,l'} = \begin{pmatrix}
(E_{0,0}^0)_{k,l'} & (E_{0,1}^0)_{k,l'} & (E_{0,2}^0)_{k,l'} & \cdots & (E_{0,l_{\max}}^0)_{k,l'} \\
(E_{0,0}^1)_{k,l'} & (E_{0,1}^1)_{k,l'} & (E_{0,2}^1)_{k,l'} & \cdots & (E_{0,l_{\max}}^1)_{k,l'} \\
(E_{0,0}^2)_{k,l'} & (E_{0,1}^2)_{k,l'} & (E_{0,2}^2)_{k,l'} & \cdots & (E_{0,l_{\max}}^2)_{k,l'} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
(E_{0,0}^{l_{\max}})_{k,l'} & (E_{0,1}^{l_{\max}})_{k,l'} & (E_{0,2}^{l_{\max}})_{k,l'} & \cdots & (E_{0,l_{\max}}^{l_{\max}})_{k,l'}
\end{pmatrix}
\]

\[\text{D.3. Second Moment}\]

The expression for the second moment is
\[
E_f = \int \mathcal{A}(\Xi, P, u) E_y \mathcal{A}^T(I, P, u) p(\Xi) \, d\Xi,
\]
where
\[
E_y = E[y \ y^T | \theta,]
\]
is the second moment of the distribution over spherical harmonic coefficients (Equation (12)). We can use Equations (B2) and (B3) to express the element at index \( k, k' \) (corresponding to the covariance of the GP between times \( t = t_k \) and \( t' = t_{k'} \)) as
\[
(E_f)_{k,k'} = \int a_k^T (\Xi) E_y a_{k'}(\Xi) p(\Xi) \, d\Xi
= r^T A_k (E_y')_{k,k'} A_{k'}^T r,
\]
where
\[
(E_y')_{k,k'} = \int R_2(\Xi) (E_y')_{k,k'} R_2^T (\Xi) p(\Xi) \, d\Xi
\]
is the expectation of \( y \ y^T \) in the observer’s frame at times \( t = t_k \) and \( t' = t_{k'} \), and
\[
(E_y')_{k,k'} = R_{D_k} \left( \frac{2\pi}{P} t_k \right) R_{D_{k'}} \left( \frac{\pi}{2} \right)
\cdot E_y R_{D_k} \left( \frac{\pi}{2} \right) R_{D_{k'}} \left( \frac{2\pi}{P} t_{k'} \right)
\]
is the expectation of \( y \ y^T \) in the polar frame at times \( t = t_k \) and \( t' = t_{k'} \). The rest of the computation follows what we did in Appendix C.2, except that the number of operations required to compute \( E_f \) is a factor of \( K^2 \) larger than in the computation of expectations like \( E_o \) (Equation (C70)). That is because we must compute the integral of all terms of a matrix for each of the \( K^2 \) elements of \( E_f \). In Appendix D.4 below, we discuss strategies that can drastically improve the computational scaling of marginalizing over the inclination.

Let us write Equation (D20) in terms of its spherical harmonic components.

\[\text{(D22)\quad (E_y')_{k,k'} = \begin{pmatrix}
(E_{0,0}^{l_f})_{k,k'} & (E_{0,1}^{l_f})_{k,k'} & (E_{0,2}^{l_f})_{k,k'} & \cdots & (E_{0,l_{\max}}^{l_f})_{k,k'} \\
(E_{0,0}^{l_f})_{k,k'} & (E_{0,1}^{l_f})_{k,k'} & (E_{0,2}^{l_f})_{k,k'} & \cdots & (E_{0,l_{\max}}^{l_f})_{k,k'} \\
(E_{0,0}^{l_f})_{k,k'} & (E_{0,1}^{l_f})_{k,k'} & (E_{0,2}^{l_f})_{k,k'} & \cdots & (E_{0,l_{\max}}^{l_f})_{k,k'} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
(E_{0,0}^{l_f})_{k,k'} & (E_{0,1}^{l_f})_{k,k'} & (E_{0,2}^{l_f})_{k,k'} & \cdots & (E_{0,l_{\max}}^{l_f})_{k,k'}
\end{pmatrix}}\]

\[\text{(D23)\quad (E_{l_f})_{k,k'} = \int R_2(\Xi) (E_{l_f})_{k,k'} R_2^T (\Xi) p(\Xi) \, d\Xi
= (E_{l_f})_{k,k'} U^{-1} (P_{l_f})_{k,k'} U'^{-1} T,}\]
and we define
\[
(P_{l_f})_{k,k'} = \int D_2(\Xi) (E_{l_f})_{k,k'} D_2^T (\Xi) p(\Xi) \, d\Xi,
\]
\[
(\tilde{E}_{l_f})_{k,k'} = U' (E_{l_f})_{k,k'} U'^{-1} T.
\]
As before, we may express the solution to the integral in Equation (D24) in closed form. Let us write its terms as
\[ \begin{align*}
\langle (P_{l}^{m})_{h,k} \rangle_{m,m'} &= \int \sum_{\mu=-l}^{l} \sum_{\mu'=-l}^{l} (D_{l}^{m})_{\mu,\mu'} (-\bar{\mu}) \ p(\bar{\mu}) \ d\bar{\mu} \\
\times \ &\langle (E_{l}^{m'})_{h,k} \rangle_{\mu,\mu'} (D_{l'}^{m'})_{\mu',\mu''} (-\bar{\mu}) p(\bar{\mu}) \ d\bar{\mu} \\
= &\sum_{\mu=-l}^{l} \sum_{\mu'=-l}^{l} \langle (E_{l}^{m'})_{h,k} \rangle_{\mu,\mu'} \\
\times &\exp \left[ \frac{i\pi}{2} (m - \mu + m' - \mu') \right] \sum_{a=0}^{2l} \sum_{a'=0}^{2l'} e^{i l_{a} l_{a'}} r^{a} m_{m} m'_{m'} (Q_{l}^{l'})_{h,r},
\end{align*} \]  

where, similarly to before,✓

\[ (Q_{l}^{l'})_{h,r} \equiv \int_{0}^{2\pi} (-1)^{i+l'}(1 - \cos \bar{\mu})^{l+l'-i+l'} \times (1 + \cos \bar{\mu})^{i+l'} \sin \bar{\mu} \ d\bar{\mu} \]

\[ = (-1)^{i+l'} \int_{0}^{1} (1 - \chi)^{l+l'-i+l'} (1 + \chi)^{i+l'} \ d\chi \]

\[ = \frac{(-1)^{i+l'} \Gamma(l + l' - i + l')}{l + l' - i + l' + 1} y_{2} \left( \frac{1}{2} - \frac{i + l'}{2}; \frac{2 + l + l' - i + l'}{2}; -1 \right). \]  

which may again be computed recursively in a preprocessing step.

**D.4. Speeding Up the Computation**

The expressions in the previous section are a bit of a nightmare, particularly because of the dimensionality of some of the linear operators involved. The complexity of the expressions is due to the fact that the second moment of the spherical harmonic vector projected onto the sky (Equation (D20)) is time-dependent; it changes as the star rotates. Computing the second moment of the flux requires computing the outer product of this tensor with itself, leading to multi-indexed quantities like those in Equation (D26). In addition to being cumbersome to evaluate, the full second-moment matrix \( E_{l} \) (and hence the flux covariance matrix) is costly to compute. It is helpful that Equation (D27) does not depend on any user inputs and thus may be precomputed, but even so, we require evaluating the four nested sums in Equation (D20) \( \mathcal{O}(l_{\max}^{3}) \) times for each entry in the \( (K \times K) \) matrix \( E_{l} \).

Fortunately, the inner two sums in Equation (D20) do not depend on user inputs, so they may be precomputed, and Equation (D20) may be cast as a straightforward matrix dot product. In practice, we also find it helpful to take advantage of the phase independence (i.e., stationarity) of the covariance of our GP, as we did in Appendix D.2; any two entries \( (E_{l})_{h,k} \) and \( (E_{l})_{j,k} \) are the same if \( h - i = j - i \). If the data happen to be evenly sampled, such that the time difference between adjacent cadences is constant, then we need only compute the covariance at a total of \( K \) points (as opposed to \( K^{2} \)), as the covariance is a circulant matrix that is fully specified by a single vector of length \( K \).

In the more general case, where the data are not evenly sampled, we may still evaluate the covariance at a fixed number of points \( K' < K^{2} \) and approximate the full covariance matrix via interpolation. As long as the data are roughly evenly sampled, as is the case with Kepler or TESS light curves, this approximation leads to negligible error when \( K' \approx K \), affording the same \( \mathcal{O}(K) \) computational savings. Note that even in the case where the flux is normalized (see Section 2.5), the nonstationary correction to the covariance is applied after the step where we marginalize over the inclination, so this approach is still valid.

**Appendix E**

**Supplementary Figures**

Figures 17–30 below are referenced in Section 3.5, discussing additional calibration runs for our GP.
Figure 17. Same as Figures 6 and 7 but for midlatitude spots with $\mu_f = 45^\circ$. The radius and latitude parameters are again inferred correctly.
Figure 18. Same as Figure 17 but for high-latitude spots with $\mu_f = 60^\circ$. The radius and latitude parameters are again inferred correctly, although the model cannot rule out the presence of polar spots.
Figure 19. Same as Figure 17 but for equatorial spots with $\mu_\phi = 0^\circ$. Even though the model favors a bimodal distribution at low latitudes, the posterior strongly supports the presence of equatorial spots.
Figure 20. Same as Figure 17 but for isotropically distributed spots. The posterior accurately captures the cosine-like distribution of spot latitudes.
Figure 21. Same as Figures 6 and 7 but for high-contrast \( c = 1 \) small \( r = 3^\circ \) spots (significantly lower than the effective resolution of the model). The hatched regions in the posterior plots for the radius \( r < 10^\circ \) are excluded by the prior, since the model cannot capture features that small. Despite this, the spot latitude distribution is still inferred correctly, although the spot contrast is off by more than \( 10\sigma \).
Figure 22. Same as Figures 6 and 7 but for an inference based on a single light curve ($M = 1$). The constraints on all of the parameters are dramatically weaker.
Figure 23. Same as Figures 6 and 7 but for an inference based on 1000 light curves ($M = 1000$). The constraints on the radius and the latitude parameters are dramatically tighter.
Figure 24. Same as Figures 6 and 7 but for stars with (assumed known) limb-darkening coefficients $u_1 = 0.50$ and $u_2 = 0.25$. Limb darkening makes it much harder to infer the variance of the distribution of starspot latitudes.
Figure 25. Same as Figure 24 but for an ensemble consisting of $M = 1000$ light curves. For a sufficiently large ensemble, it is possible to correctly infer the spot radii and latitudes in the presence of limb darkening.
Figure 26. Same as Figure 24 but for $M = 500$ light curves and assuming no limb darkening when doing inference. Neglecting limb darkening leads to significant bias in the inferred spot radii and, to a lesser extent, the mean spot latitude.
Figure 27. Same as Figures 6 and 7 but for stars with $n = 2$ spots with high contrast $c = 0.5$. Our model correctly captures the increased contrast, but it is still strongly degenerate with the number of spots.
Figure 28. Same as Figures 6 and 7 but for a single \((n = 1)\) large \((r \sim \mathcal{N}(45°, 5°^2))\) spot on each star. Our model correctly infers the larger radius and infers the latitude within 3\(\sigma\), albeit with large uncertainty.
Figure 29. Same as Figures 6 and 7 but for stars with variance in their number of spots, $n \sim \mathcal{N}(20, 3^2)$; spot radii, $r \sim \mathcal{N}(15^\circ, 3^2)$; and the spot contrasts, $c \sim \mathcal{N}(0.05, 0.01^2)$. 
Figure 30. Same as Figures 6 and 7 but assuming that the light curves are not normalized and the true amplitude is known. Knowledge of the normalization breaks the $c - n$ degeneracy and allows us to infer the total number of spots.
References

Aigrain, S., Purviainen, H., & Pope, B. J. S. 2016, MNRAS, 459, 2408
Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O'Neil, M. 2015, ITPAM, 38, 252
Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., & Rajpaul, V. 2018, MNRAS, 474, 2094
Angus, R., Morton, T. D., Foreman-Mackey, D., et al. 2019, AJ, 158, 173
Barnes, S., Sofia, S., & Pinsonneault, M. 2001, ApJ, 548, 1071
Basri, G., & Shah, R. 2020, ApJ, 901, 14
Beaumont, M. A. 2019, AnRSA, 6, 379
Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. 2016, arXiv:1603.00670
Brewer, B. J., & Stello, D. 2009, MNRAS, 395, 2226
Cantiello, M., & Braithwaite, J. 2019, ApJ, 883, 106
Collado, J. R. A., Rico, J. F., López, R., Paniagua, M., & Ramírez, G. 1989, CoPhC, 52, 323
Damasso, M., Pinamonti, M., Scandariato, G., & Sozzetti, A. 2019, MNRAS, 489, 2555
Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. 1987, PhLB, 195, 216
Feroz, F., Hobson, M. P., & Bridges, M. 2009, MNRAS, 398, 1601
Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, AJ, 154, 220
Fuller, J., Cantiello, M., Stello, D., Garcia, R. A., & Bildsten, L. 2015, Sci, 350, 423
Gilbertson, C., Ford, E. B., Jones, D. E., & Stenning, D. C. 2020, ApJ, 905, 155
Gough, D. O., & Tayler, R. J. 1966, MNRAS, 133, 85
Gully-Santiago, M. A., Herczeg, G. J., Czekala, I., et al. 2017, ApJ, 836, 200
Guo, Z., Gully-Santiago, M., & Herczeg, G. J. 2018, ApJ, 868, 143
Haywood, R. D., Collier Cameron, A., Queloz, D., et al. 2014, MNRAS, 443, 2517
Hoffman, M. D., & Gelman, A. 2011, arXiv:1111.4246
Ireland, L. G., & Browning, M. K. 2018, ApJ, 856, 132
Jones, D. E., Stenning, D. C., Ford, E. B., et al. 2017, arXiv:1711.01318
Kervella, P., Bigot, L., Gallenne, A., & Thévenin, F. 2017, A&A, 597, A137
Kipping, D. M. 2013, MNRAS, 435, 2152
Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. 2016, arXiv:1603.00788
Luger, R., Agol, E., Foreman-Mackey, D., et al. 2019, AJ, 157, 64
Luger, R., Agol, E., Kruse, E., et al. 2016, AJ, 152, 100
Luger, R., Foreman-Mackey, D., & Hedges, C. 2021a, AJ, 162, 123
Luger, R., Foreman-Mackey, D., Hedges, C., & Hogg, D. W. 2021b, AJ, 162, 123
Luger, R., Lustig-Yaeger, J., & Agol, E. 2017a, ApJ, 851, 94
Luger, R., Sestovic, M., Kruse, E., et al. 2017b, NatAs, 1, 0129
Miesch, M. S., & Toomre, J. 2009, AnRFM, 41, 317
Morris, B. 2020a, JOSS, 5, 2103
Morris, B. M. 2020b, ApJ, 893, 67
Morris, B. M., Agol, E., Davenport, J. R. A., & Hawley, S. L. 2018, ApJ, 857, 39
Perger, M., Anglada-Escudé, G., Ribas, I., et al. 2021, A&A, 645, A58
Price, L. F., Drovandi, C. C., Lee, A., & Nott, D. J. 2018, J. Comput. Graph. Stat., 27, 1
Rackham, B. V., Apai, D., & Giampapa, M. S. 2018, ApJ, 853, 122
Rajpaul, V., Aigrain, S., Osborne, M. A., Reece, S., & Roberts, S. 2015, MNRAS, 452, 2269
Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning (Cambridge, MA: MIT Press)
Robertson, P., Stefansson, G., Mahadevan, S., et al. 2020, ApJ, 897, 125
Schuessler, M., Caligari, P., Ferriz-Mas, A., Solanki, S. K., & Stix, M. 1996, A&A, 314, 503
Sikora, J., Wade, G. A., Power, J., & Neiner, C. 2018, MNRAS, 483, 3127
Skilling, J. 2004, in AIP Conf. Ser. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th Int. Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. V. Toussaint (Melville, NY: AIP), 395
Skilling, J. 2006, Bayesian Analysis, 1, 833
Solanki, S. K., Inhester, B., & Schüssler, M. 2006, RPP, 69, 563
Speagle, J. S. 2020, MNRAS, 493, 3132
Turcotte, S. 2003, in ASP Conf. Ser. 305, Magnetic Fields in O, B and A Stars: Origin and Connection to Pulsation, Rotation and Mass Loss, ed. L. A. Balona, H. F. Henrichs, & R. Medupe (San Francisco, CA: ASP), 199
Vanderburg, A., Plavchan, P., Johnson, J. A., et al. 2016, MNRAS, 459, 3565
Wandelt, B. 2012, in Astrostatistical Challenges for the New Astronomy, ed. J. Hilbe (Berlin: Springer), 87
Weber, M. A., & Browning, M. K. 2016, ApJ, 827, 95
Wood, S. N. 2010, Natur, 466, 1102
Yadav, R. K., Gastine, T., Christensen, U. R., & Reiners, A. 2015, A&A, 573, A68