

## Interpretable Gaussian Processes for Starspots

RODRIGO LUGER,<sup>1,2,\*</sup> DANIEL FOREMAN-MACKEY,<sup>1</sup> AND CHRISTINA HEDGES<sup>3,4</sup>

<sup>1</sup>*Center for Computational Astrophysics, Flatiron Institute, New York, NY*

<sup>2</sup>*Virtual Planetary Laboratory, University of Washington, Seattle, WA*

<sup>3</sup>*Bay Area Environmental Research Institute, P.O. Box 25, Moffett Field, CA 94035, USA*

<sup>4</sup>*NASA Ames Research Center, Moffett Field, CA*

### ABSTRACT

The use of Gaussian processes (GPs) in stellar light curve modeling has recently become almost ubiquitous, given their ease of use and flexibility. GPs excel in particular at marginalization of the stellar signal in cases where the variability due to starspots rotating in and out of view is treated as a nuisance, such as in exoplanet transit modeling. However, in cases where the star spot signal is of primary interest, GPs have thus far been less useful: it is often difficult to relate their hyperparameters to the physical properties of interest, such as the size, contrast, and latitudinal distribution of the spots. 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 degenerate, ill-posed, and often computationally intractable when applied to stars with more than a few spots and/or to ensembles of many light curves. In this paper, we derive a closed-form expression for the mean and covariance of a Gaussian process model that describes the light curve of a rotating, evolving stellar surface conditioned on a given distribution of star spot sizes, contrasts, and latitudes. We show that our GP is unbiased, allowing one to correctly infer physical parameters of interest from one or more stellar light curves, including the typical radii and the mean and variance of the latitude distribution of starspots. Our GP has far-ranging implications for understanding the variability and magnetic activity of stars from both light curves and radial velocity (RV) measurements, as well as for robustly modeling correlated noise in both transiting and RV exoplanet searches. Our implementation is efficient, user-friendly, and open source, available as the Python package `starry_process`. 

*Keywords:* analytical mathematics — time series analysis — Gaussian processes regression — starspots

*Features:* open-source figures ; equation unit tests: 18 passed , 0 failed 

[rluger@flatironinstitute.org](mailto:rluger@flatironinstitute.org)

\* Flatiron Fellow

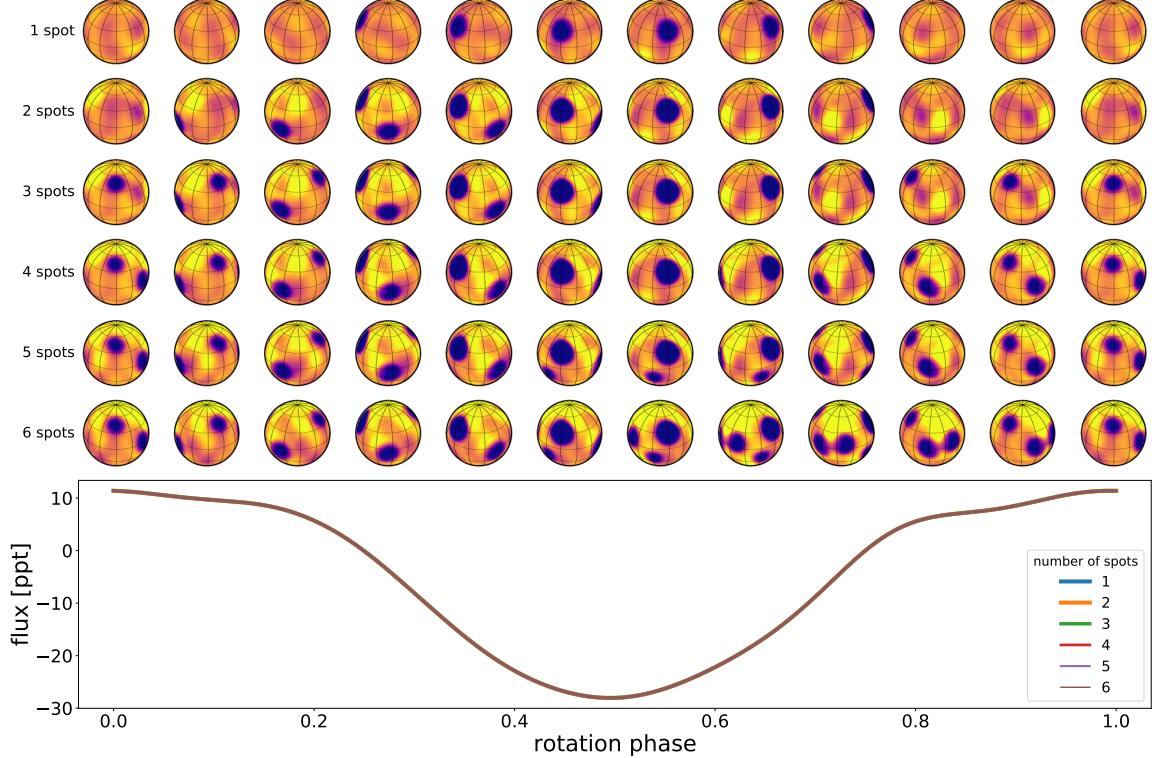
## 1. INTRODUCTION

### 1.1. Stellar surface modeling

The advent of space-based precision photometry with missions such as *Kepler* (Borucki et al. 2010) and *TESS* (Ricker et al. 2015) has led to a renewed interest in the modeling of stellar light curves, and, in particular, in understanding what these light curves can tell us about the surfaces of stars across the HR diagram. One of the dominant sources of stellar light curve variability is the modulation caused by starspots rotating in and out of view. Dark spots arise due to the suppression of convection in regions of intense magnetic field, resulting in a locally cooler (and hence darker) photosphere. Bright spots can similarly arise in the photosphere as faculae or in the chromosphere as plages, and are also magnetically driven (e.g., Berdyugina 2005). Constraining the sizes, contrasts, locations, and number of spots on stars can therefore reveal information about stellar magnetic activity, stellar interior structure, and how these quantities vary across spectral type and over time. A detailed understanding of the stellar surface is also crucial to mitigating systematics in the radial velocity search for exoplanets (e.g., Lanza et al. 2011) and in the spectroscopic characterization of their atmospheres (e.g., Rackham et al. 2018).

To date, most studies aimed at inferring stellar surface properties from light curves follow one of two broad approaches. The first is to model the stellar surface as a collection of one or more discrete, circular, uniform contrast dark or bright spots on a uniform intensity photosphere. The advantage of this approach is that the light curve can be computed efficiently and in some cases even analytically (e.g., Davenport et al. 2015; Morris et al. 2017; Morris 2020a). The second approach is to discretize the surface at some resolution and compute the emergent flux as a weighted sum of the visible pixel intensities. This approach is more flexible, since it is not limited to surfaces composed of distinct circular spots (e.g., Harmon & Crews 2000; Roettenbacher et al. 2017). Both approaches rely on an explicit *forward model*, a prescription for how to generate data given a set of parameters. In most cases, however, we are interested in the inverse problem: constraining the parameters given data. Unfortunately, the inverse problem is not only difficult—as it requires a large number of forward model evaluations to find the parameter values that are most consistent with the data—but also formally *ill-posed*. While the mapping from a stellar surface to a light curve (the forward problem) is unique, the mapping from a light curve to a surface (the inverse problem) is not: given any light curve, there exist an infinite number of surfaces that could have generated it. These degeneracies are illustrated in Figure 1, where six synthetic stellar surfaces are shown (rows) at twelve different phases (columns), all at an inclination  $I = 60^\circ$ . Each surface consists of a different number of dark spots on a brighter, heterogeneous background.

While the stellar surfaces are all distinct, containing between one (top) and six (bottom) large dark spots, **their rotational light curves are identical** (lower panel). This is true even at infinite signal to noise: the mapping from a stellar surface



**Figure 1.** The fundamental limitations of the mapping problem. Each row corresponds to a stellar surface with a different number of dark spots seen at various phases at an inclination  $I = 60^\circ$ ; all images are shown on the same color scale. The bottom panel shows the light curves of each of these stars. All six light curves are indistinguishable from each other, even at infinite signal to noise. See text for details.

to its rotational light curve is so degenerate that there exist an infinite number of solutions to the inverse problem. This fact has been pointed out recently in different contexts (e.g., Cowan et al. 2013; Luger et al. 2019; Basri & Shah 2020), but it dates back at least to Russell (1906), who demonstrated it by expanding the surface intensity of a celestial body in terms of spherical harmonics (see §2 and Figure 2). Russell (1906) showed that many of the modes comprising the intensity profile of a spherical object are in the *null space*, the set of surface features that have identically zero effect on the light curve. In fact, the *vast majority* of the modes are in the null space for rotational light curves of stars (§4.1). This is what allows us to construct pathological scenarios like that shown in Figure 1, where the light curve could be explained by any number of spots atop a heterogeneous bright background.

Stellar mapping studies tackle these degeneracies in different ways, but it usually comes down to a choice of prior: when the data is not sufficiently informative, prior assumptions are needed to discriminate between competing solutions. In discrete spot models like the ones discussed above, the degeneracy-breaking prior is (typically) the assumption that the spots must be circular, have uniform contrast, and sit atop an otherwise uniform photosphere. In gridded stellar surface models, it is common to assume a regularization prior such as the maximum entropy penalty (e.g., Vogt et al.

1987), which typically favors solutions with the fewest number of dark pixels (usually referred to as the “simplest” solution).

While these assumptions may be approximately valid in some cases, it is important to bear in mind that because of the light curve degeneracies discussed above, **most of the information about the stellar surface usually comes from the prior**, so it is very important to get the prior right. In general, sunspots are not circular and do not have uniform contrast throughout; nor do spots always arrange themselves in the highest entropy configuration. The amount of bias introduced by these assumptions will in general vary, but in principle it could be quite significant.

Fortunately, although individual light curves are not very constraining, we have *a lot of them* thanks to missions like *Kepler* and *TESS*. In this paper, we will thoroughly explore the power of ensemble analyses in breaking the degeneracies inherent to the light curve mapping problem. By jointly analyzing the light curves of many stars, we will show that it is often possible to uniquely constrain statistical properties about their surfaces. This idea was recently explored to some extent in Morris (2020b), who used ensemble analyses to derive constraints on spot coverage areas as a function of stellar age (see §4.7). We will go a step further and show that it is possible to constrain the distribution of spot sizes, active latitudes, and other statistical properties of stellar surfaces using a novel interpretable Gaussian process approach, which we motivate in the following section.

## 1.2. Gaussian processes

Over the past decade, Gaussian processes (GPs) have gained traction as a leading tool for modeling correlated signals in astrophysical timeseries (e.g., Rasmussen & Williams 2005; Brewer & Stello 2009; Luger et al. 2016; Foreman-Mackey et al. 2017; Angus et al. 2018). 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  $\mathbf{t}$  is “distributed as a GP” means that we may write

$$\mathbf{f} \sim \mathcal{N}(\mu, \Sigma), \quad (1)$$

where the elements of the mean and covariance are given by  $\mu_i = m(t_i)$  and  $\Sigma_{i,j} = k(t_i, t_j)$ , respectively.<sup>1</sup> Because of this relationship to multivariate Gaussians, GPs are easy to sample from.<sup>2</sup> But the real showstopper is the application of

<sup>1</sup> In this paper, we will use blackboard font (i.e.,  $\mathbf{f}$ ) to denote random variables and serif font (i.e.,  $\mathbf{f}$ ) to denote particular realizations of those variables. See Appendix A for a detailed explanation of our notation.

<sup>2</sup> Given a 1-d array `mean` and a 2-d array `cov` in Python, sampling from the corresponding GP (if it exists) can be done in a single line of code by calling `numpy.random.multivariate_normal(mean, cov)`.

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 18 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 dataset (i.e.,  $K \lesssim 10^4$  datapoints).

Another big advantage of GPs is their flexibility. GPs are often dubbed a class of “non-parametric” 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

$$\begin{aligned}\Sigma_{i,j} &= k(t_i, t_j) \\ &= k(|t_i - t_j|) \\ &\equiv k(\Delta t).\end{aligned}\tag{2}$$

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. Among a vast number of kernels are a few that are commonly used in modeling stellar variability: the exponential squared kernel,

$$k_{E^2}(\Delta t) = \sigma^2 \exp\left(-\frac{\Delta t^2}{2\tau}\right),\tag{3}$$

the Matérn-3/2 kernel,

$$k_{M^{\frac{3}{2}}}(\Delta t) = \sigma^2 \left(1 + \frac{\sqrt{3}\Delta t}{\tau}\right) \exp\left(-\frac{\sqrt{3}\Delta t}{\tau}\right),\tag{4}$$

the exponential sine squared kernel,

$$k_{S^2}(\Delta t) = \sigma^2 \exp\left[-\Gamma \sin^2\left(\frac{\pi\Delta t}{P}\right)\right],\tag{5}$$

the cosine kernel,

$$k_C(\Delta t) = \sigma^2 \cos\left(\frac{2\pi\Delta t}{P}\right),\tag{6}$$

or any linear combination and/or product of these. In the expressions above,  $\sigma$ ,  $\tau$ ,  $\Gamma$ , and  $P$  are the *hyperparameters* of the kernels. Specifically,  $\sigma^2$  is the variance (the value of the kernel when  $t_i = t_j$ ),  $\tau$  is a typical timescale over which data are correlated,  $\Gamma$  is a dimensionless scale parameter, and  $P$  is a period. Once a kernel has been chosen, these scalar quantities fully specify the covariance of the Gaussian process. GP-based inference entails optimizing or sampling over these few quantities, which in many cases can be done easily and efficiently.

Which brings us, finally, to an important drawback of GPs when applied to modeling stellar spot-induced variability. While it is straightforward to derive posterior constraints on the hyperparameters of a GP, it is often not obvious what those constraints actually tell us about the star or its surface. Specifically, it is difficult to constrain the relationship between quantities like the amplitude  $\sigma^2$  and the timescale  $\tau$  and physical properties of the surface. For instance, it may be tempting to interpret  $\sigma^2$  as some measure of the spot contrast or the total number of spots, or  $\tau$  as the timescale on which the spots evolve, 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 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](#)).<sup>3</sup> 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 to search for a planetary signal in a radial velocity dataset, a GP might be used to remove (or, better yet, to 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 is generally sufficient.

When the goal is to understand the stellar surface itself, however, GPs have been less useful to date. Instead, the common approach is that discussed in the previous section: to forward model the stellar surface and its effect on the light curve, then seek to constrain the surface properties by solving the inverse problem. As we discussed above, however, a unique solution to the surface map of the star from a rotational light curve 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.

<sup>3</sup> An exception to this is in the presence of strong differential rotation, in which case many periods may be present in the data, or when spots evolve coherently, which can also introduce weak periodicities in the light curve.

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 wouldn’t really know what to do with that information! Instead, we often treat (whether explicitly or not) the properties of a star spot 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 we were able 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 the spots as a whole. We could do this in the usual way described above, by modeling the properties of each of the spots in a posterior sampling scheme, but also including a few *hyperparameters* controlling the distribution of those properties across all spots (acting, therefore, as a prior term): 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 multi-modality 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 Gaussian processes to perform this marginalization for us!

### 1.3. Overview

**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.** Our Gaussian process analytically marginalizes over the (degenerate, 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.

The paper is organized as follows: we present an overview of the derivation of the GP in §2.1 and a suite of tests on synthetic data to show the model is calibrated in §3. We discuss our GP from an information theory perspective in §4, focusing on the various degeneracies our model can and cannot break when applied to ensembles of light curves. In §6 we present our implementation of the algorithm in the open-source `starry_process` package alongside speed and stability tests, and in §7 we summarize our results. Most of the math behind the algorithm is presented in the Appendix, followed by a series of supplementary figures (discussed in §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.

Finally, we note that all of the figures in this paper were auto-generated using the Azure Pipelines continuous integration (CI) service, which ensures 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. In this paper we also introduce the concept of equation “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  () 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.<sup>4</sup>

## 2. A GAUSSIAN PROCESS FOR STARSPOTS

In this section, we present an overview of the computation of our Gaussian process, which boils down to computing the mean and covariance of the stellar flux conditioned on certain physical properties of the star and its star spot distribution. Most of the math is folded into the Appendix for readability; readers may want to refer to Appendix A in particular for a discussion of the notation and conventions we adopt.

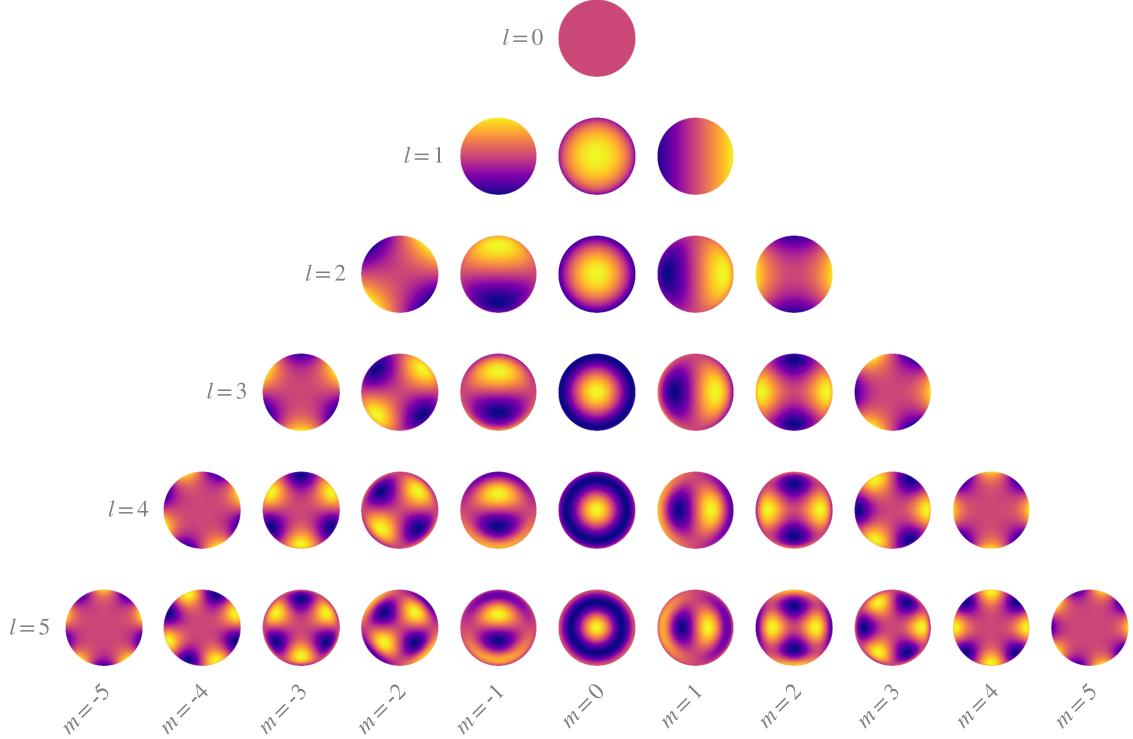
Before we dive in, it is useful to introduce the spherical harmonics, a set of orthogonal functions on the surface of the sphere which we will use to describe the intensity field on the surface of a star (Figure 2). As we will see below, the spherical harmonics are a particularly convenient basis in which to describe star spot distributions<sup>5</sup>, 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  $\mathbf{f}$  (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  $\mathbf{y}$  (up to a certain degree  $l_{\max}$ ), the flux is given by

$$\mathbf{f} = \mathbf{1} + \mathcal{A}(I, P, \mathbf{u}) \mathbf{y}, \quad (7)$$

where  $\mathbf{1}$  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  $\mathbf{u}$  (see Appendix B for details).

<sup>4</sup> These unit tests are certainly not foolproof: in particular, there is no guarantee against a mismatch in the L<sup>A</sup>T<sub>E</sub>X version of an equation and its Python implementation (e.g., due to an uncaught typo). However, they *do* ensure that the linked Python implementation is correct to within the accuracy of the numerical solution, providing readers with a valid implementation of the equation for purposes of reproducibility.

<sup>5</sup> There are, of course, drawbacks to using this basis: in particular, the spherical harmonics are smooth, continuous functions that struggle (at finite degree  $l$ ) to capture high resolution features such as small starspots. We discuss this point at length in §4.4, where we show that our model is useful even when applied to stars with spots smaller than the effective resolution of the GP.



**Figure 2.** The real spherical harmonics in the polar frame up to  $l = 5$ , where dark colors correspond to negative intensity and bright colors to positive intensity. Rows correspond to the degree  $l$  and columns to the order  $m$ . The set of all spherical harmonics forms a complete, orthogonal basis on the sphere.

### 2.1. Computing the GP

Let  $\mathbf{f} = (\mathbf{f}_0 \ \mathbf{f}_1 \ \cdots \ \mathbf{f}_{K-1})^\top$  denote a random vector of  $K$  flux measurements at times  $\mathbf{t} = (t_0 \ t_1 \ \cdots \ t_{K-1})^\top$ , 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  $\mathbf{u}$ , and on certain properties of the starspots  $\boldsymbol{\theta}_\bullet$  (including their number, sizes, positions, and contrasts), we wish to compute the mean  $\boldsymbol{\mu}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet)$  and covariance  $\boldsymbol{\Sigma}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet)$  of  $\mathbf{f}$ . Together, these specify a multidimensional Gaussian distribution, which we assume fully describes<sup>6</sup> how our flux measurements are distributed:

$$\mathbf{f}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \sim \mathcal{N}\left(\boldsymbol{\mu}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet), \boldsymbol{\Sigma}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet)\right) \quad (8)$$

<sup>6</sup> The *true* distribution of stellar light curves conditioned on  $I$ ,  $P$ ,  $\mathbf{u}$ , and  $\boldsymbol{\theta}_\bullet$  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.

As with any random variable, the mean and covariance may be computed from the expectation values of  $\mathbf{f}$  and  $\mathbf{f}\mathbf{f}^\top$ , respectively:

$$\boldsymbol{\mu}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = \mathbb{E}[\mathbf{f} \mid I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet] \quad (9)$$

$$\boldsymbol{\Sigma}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = \mathbb{E}[\mathbf{f}\mathbf{f}^\top \mid I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet] - \boldsymbol{\mu}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet)\boldsymbol{\mu}^\top(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet). \quad (10)$$

Given the linear relationship between flux and spherical harmonic coefficients (Equation 7), we may write the mean and covariance of our GP as

$$\boldsymbol{\mu}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = \mathbf{1} + \mathcal{A}(I, P, \mathbf{u}) \boldsymbol{\mu}_y(\boldsymbol{\theta}_\bullet) \quad (11)$$

$$\boldsymbol{\Sigma}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = \mathcal{A}(I, P, \mathbf{u}) \boldsymbol{\Sigma}_y(\boldsymbol{\theta}_\bullet) \mathcal{A}^\top(I, P, \mathbf{u}), \quad (12)$$

where

$$\boldsymbol{\mu}_y(\boldsymbol{\theta}_\bullet) = \mathbb{E}[\mathbf{y} \mid \boldsymbol{\theta}_\bullet] \quad (13)$$

$$\boldsymbol{\Sigma}_y(\boldsymbol{\theta}_\bullet) = \mathbb{E}[\mathbf{y}\mathbf{y}^\top \mid \boldsymbol{\theta}_\bullet] - \boldsymbol{\mu}_y(\boldsymbol{\theta}_\bullet)\boldsymbol{\mu}_y^\top(\boldsymbol{\theta}_\bullet) \quad (14)$$

are the mean and covariance of the distribution over spherical harmonic coefficient vectors  $\mathbf{y}$ . 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

$$\mathbb{E}[\mathbf{y} \mid \boldsymbol{\theta}_\bullet] = \int \mathbf{y}(\mathbf{x}) p(\mathbf{x} \mid \boldsymbol{\theta}_\bullet) d\mathbf{x} \quad (15)$$

$$\mathbb{E}[\mathbf{y}\mathbf{y}^\top \mid \boldsymbol{\theta}_\bullet] = \int \mathbf{y}(\mathbf{x})\mathbf{y}^\top(\mathbf{x}) p(\mathbf{x} \mid \boldsymbol{\theta}_\bullet) d\mathbf{x}, \quad (16)$$

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

$$\boldsymbol{\theta}_\bullet = (n \ c \ \mu_\phi \ \sigma_\phi \ r)^\top, \quad (17)$$

where  $n$  is the number of starspots,  $c$  is their contrast,  $\mu_\phi$  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, the spot contrast, and the number of spots are chosen to be delta functions centered at  $r$ ,  $c$ , and  $n$ , respectively (Appendices C.1 and C.4), while 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). Finally, the spot longitude is assumed to be uniformly distributed (Appendix C.3).

In this paper, we assume that the parameters  $\boldsymbol{\theta}_\bullet$  described above are the *physically interesting* ones. That is, given a set of  $M$  light curves  $(\mathbf{f}_0 \ \mathbf{f}_1 \ \cdots \ \mathbf{f}_{M-1})^\top$ , we wish to infer the statistical properties of the starspots, encoded in the entries of the vector  $\boldsymbol{\theta}_\bullet$ . This is typically a tall order, since it requires marginalizing over all the nuisance parameters, which include the nitty-gritty details of the size, contrast, and location of *every spot on every star* in the sample. Fortunately, however, the Gaussian process 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  $m^{\text{th}}$  dataset conditioned on a specific value of  $\boldsymbol{\theta}_\bullet$  (as well as  $I$ ,  $P$ , and  $\mathbf{u}$ ):

$$\begin{aligned} \ln \mathcal{L}_m(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = & -\frac{1}{2} \mathbf{r}_m^\top(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) [\Sigma(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) + \mathbf{C}_m]^{-1} \mathbf{r}_m(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \\ & - \frac{1}{2} \ln |\Sigma(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) + \mathbf{C}_m| - \frac{K_m}{2} \ln(2\pi), \end{aligned} \quad (18)$$

where

$$\mathbf{r}_m(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \equiv \mathbf{f}_m - \boldsymbol{\mu}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \quad (19)$$

is the residual vector,  $\mathbf{C}_m$  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),  $|\cdots|$  denotes the determinant, and  $K_m$  is the number of data points. In an ensemble analysis, the joint marginal likelihood of all datasets is simply the product of the individual likelihoods, so in log space we have

$$\ln \mathcal{L}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = \sum_m \ln \mathcal{L}_m(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet). \quad (20)$$

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 statistical properties of starspots from light curves with minimal bias.

## 2.2. Marginalizing over inclination

As we mentioned in the previous section, the equations for the mean and covariance of our GP (Equations 11 and 12, 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

$$N = 4M + 5, \quad (21)$$

since each of the stars will have their own set of 4 stellar properties (an inclination, a period, and usually two limb darkening coefficients) but will all share the same 5 spot properties  $\boldsymbol{\theta}_\bullet$  (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 distribution for the spot mapping problem can be. In addition to strong nonlinear degeneracies between some of the parameters (such as the contrast  $c$  and the number of spots  $n$ ), the posterior is often multimodal, especially in the stellar inclinations. While modern sampling schemes such as Hamiltonian Monte Carlo and Nested Sampling may in principle be able to deal with these issues, in practice it can be very difficult to obtain convergence in a reasonable amount of time.

One workaround is to fix the values of the stellar parameters. This could be done, for instance, to the rotational period  $P$ , which can often be estimated with fairly good accuracy from a periodogram. The limb darkening coefficients could be fixed at theoretical values, or perhaps their values could be shared among all stars (and sampled over), given the similarity assumption above.

The inclination, however, is a different matter. Absent prior information for a particular star (such as a measurement of its projected rotational velocity  $v \sin I$  or the knowledge that it hosts a transiting planet), it is simply not possible to reliably estimate the inclination in a pre-processing step. Any light curve statistic one might argue should scale with inclination—such as the amplitude of the variability—is invariably degenerate with the spot parameters  $\boldsymbol{\theta}_\bullet$ . If one knew  $\boldsymbol{\theta}_\bullet$ , then perhaps a decent point estimate of  $I$  could be obtained, but in that case the analysis wouldn't be needed in the first place!

Fortunately, there is a better way to reduce the number of parameters in the problem: we can explicitly marginalize over the stellar inclination. That is, we may write the mean and covariance of our GP as

$$\begin{aligned} \boldsymbol{\mu}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet) &= \mathbb{E}\left[\mathbf{f} \mid P, \mathbf{u}, \boldsymbol{\theta}_\bullet\right] \\ &= \mathbf{1} + \mathbf{e}_I \end{aligned} \quad (22)$$

$$\begin{aligned} \boldsymbol{\Sigma}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet) &= \mathbb{E}\left[\mathbf{f} \mathbf{f}^\top \mid P, \mathbf{u}, \boldsymbol{\theta}_\bullet\right] - \boldsymbol{\mu}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet)\boldsymbol{\mu}^\top(P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \\ &= \mathbf{E}_I - \mathbf{e}_I \mathbf{e}_I^\top \end{aligned} \quad (23)$$

where we define the inclination first moment integral

$$\mathbf{e}_I \equiv \int \mathcal{A}(\mathbb{I}, P, \mathbf{u}) \mathbb{E}\left[\mathbf{y} \mid \boldsymbol{\theta}_\bullet\right] p(\mathbb{I}) d\mathbb{I} \quad (24)$$

and the inclination second moment integral

$$\mathbf{E}_I \equiv \int \mathcal{A}(\mathbb{I}, P, \mathbf{u}) E\left[\mathbf{y}\mathbf{y}^\top \mid \boldsymbol{\theta}_\bullet\right] \mathcal{A}^\top(\mathbb{I}, P, \mathbf{u}) p(\mathbb{I}) d\mathbb{I}. \quad (25)$$

The expectations inside the integrals in the expressions for  $\mathbf{e}_I$  and  $\mathbf{E}_I$  are given by Equations (15) and (16), respectively, and are 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 Appendix 2.2, 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:

$$\begin{aligned} \boldsymbol{\mu}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet) &= (1 + e_I)\mathbf{1} \\ &\equiv \mu \mathbf{1}, \end{aligned} \quad (26)$$

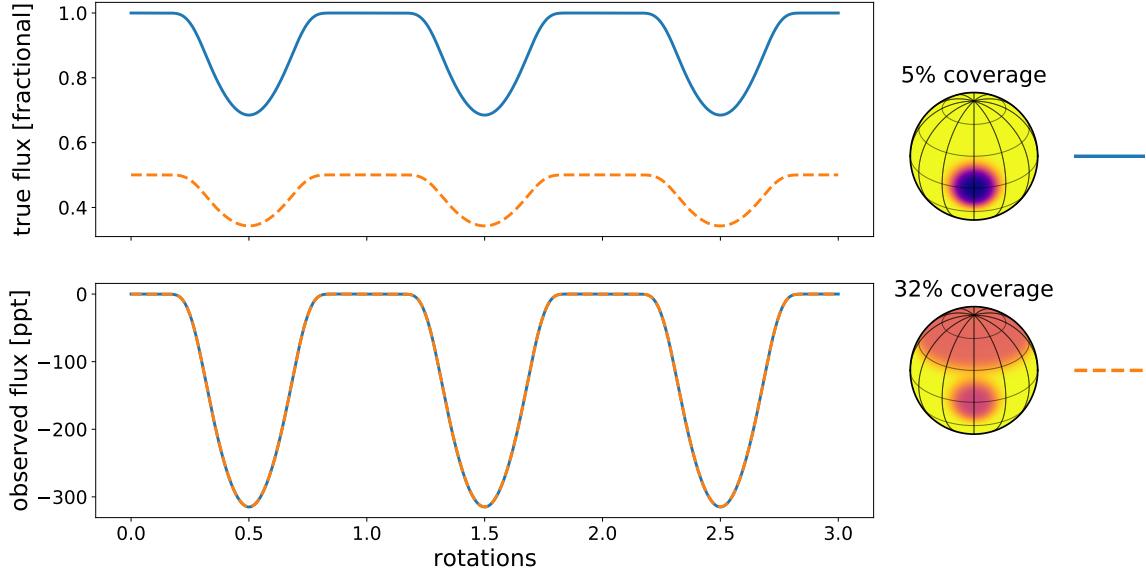
since by construction our GP is longitudinally isotropic (see Appendix D.2).

### 2.3. The baseline problem

There is one final subtle—but extremely important—point that remains to be discussed in our derivation of a Gaussian process for stellar rotation. Thus far we have cast the light curve problem in a typical GP inference setting: we model our observations  $\mathbf{f}$  as a multidimensional random variable distributed as a multivariate Gaussian with mean vector  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . However, in the context of relative photometry, the mean  $\boldsymbol{\mu}$  is typically *unobservable*. To understand why, consider what  $\boldsymbol{\mu}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet)$  actually represents: it is the expected value of the flux, *in units of the flux one would measure if the star had no spots*. We do not observe in these units: rather, we observe in units of counts on the detector, which depend on the luminosity of the star, the distance to the star, and various properties of the telescope. Even if we knew all these things, we would still need to know the intrinsic photospheric brightness in order to properly normalize our dataset for direct application of the Gaussian process.

Instead, what astronomers typically do is to self-normalize their data: i.e., divide the flux by the mean, median, maximum, or some similar statistic of the light curve. This operation folds the unknowability of the flux baseline under the rug and transforms the light curve into a *relative* measurement of the star’s temporal variability. While relative measurements are typically what we are interested in anyways, this normalization procedure has two significant drawbacks: (1) it introduces a new set of perfect degeneracies in the modeling and (2) it can substantially change the covariance structure of the dataset.

The first point is illustrated in the toy example in Figure 3. As it has been pointed out before (e.g., Basri 2018), we will not spend too much time discussing it, other than to say that the flux normalization introduces a degeneracy between the total



**Figure 3.** An example of the baseline problem. *Top:* Consider a star with a single equatorial spot of contrast  $c$  viewed at a certain inclination. The total flux (in some units) as a function of time is shown as the blue curve. Now, consider a second star, identical in all respects to the first, except that (1) the equatorial spot has half the contrast (i.e.,  $c/2$ ); and (2) there is a second, large spot centered on the pole. The corresponding light curve is shown as the dashed orange curve. The orange light curve is different from the blue one in two ways: (1) since the equatorial spot has half the contrast, the amplitude of the associated dips in the light curve is half that of the first star; and (2) since the polar spot is azimuthally symmetric, its only contribution is a net darkening at all phases. *Bottom:* The true baseline level of a stellar light curve, which corresponds to the flux one would measure in the absence of any spots, is almost always unknown. Photometric measurements are therefore meaningful only in a relative sense, i.e., as deviations from the mean, median, or maximum level of the light curve. The bottom panel shows the same two light curves, this time plotted as deviations in parts per thousand (ppt) from their respective maxima. To the observer, the two light curves are *indistinguishable*. In the absence of baseline information, there exists a perfect degeneracy between the total spot coverage and the contrast of any individual feature on the surface. As a consequence, the total spot coverage of a star cannot be inferred from relative photometry.

spot coverage of a star and the contrast of any individual feature on the surface. The two stars in the figure have different numbers of spots, different spot contrasts, and different total spot coverage. In “absolute” units, in which the flux is measured as a fraction of the flux of an unspotted star, the two light curves are distinct (top panel). However, we can only measure these light curves up to an unknown multiplicative factor. Normalizing these light curves to the “continuum” level yields the curves in the bottom panel, which are *perfectly* indistinguishable from each other. Normalization to the mean, median, or any other statistic of the light curve does not help break this degeneracy, which is inherent to relative photometry. While one could instead *fit* for the normalization as a latent variable, there is no information in the data that

can help break it, so the result will be entirely determined by the prior.<sup>7</sup> We discuss these issues further in §4.6.

The second point is of more immediate concern, since we must correct our expression for the covariance matrix  $\Sigma(P, \mathbf{u}, \boldsymbol{\theta}_\bullet)$  if we are to use our GP to model normalized light curves. For definiteness, let us consider only mean-normalized light curves.<sup>8</sup> Since we are dividing the flux by the mean, one might imagine that we could simply divide the covariance matrix by the square of the mean. However, this is incorrect, since we must be careful to distinguish between the *sample* mean and the *process* mean. When normalizing a light curve by the mean, the operation we perform is

$$\tilde{\mathbf{f}} = \frac{\mathbf{f}}{\langle \mathbf{f} \rangle}, \quad (27)$$

where  $\tilde{\mathbf{f}}$  is the normalized, unit-mean light curve,  $\mathbf{f}$  is the measured light curve (in detector counts), and  $\langle \mathbf{f} \rangle$  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,  $\mu(P, \mathbf{u}, \boldsymbol{\theta}_\bullet)$ , since the mean of a draw from the GP is itself normally distributed with a variance that scales with the GP variance.<sup>9</sup>

Computing the covariance of the normalized process is tricky, especially because the normalized process is *not* strictly Gaussian: the distribution has heavy tails due to the fact that  $\tilde{\mathbf{f}}$  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's because it is! A much safer approach is to resist the temptation to normalize the light curve and instead model the unknown multiplicative amplitude as a 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 Luger (2020) for the covariance of a normalized Gaussian process:

$$\tilde{\Sigma} \approx \frac{A}{\mu^2} \Sigma + z \left( (A + B) (\mathbf{1} - \mathbf{q}) (\mathbf{1} - \mathbf{q})^\top - A \mathbf{q} \mathbf{q}^\top \right), \quad \checkmark \quad (28)$$

<sup>7</sup> In principle, the maximum level of a light curve could set a lower limit on the location of the baseline. However, this applies only if the surface is known to be made up of *exclusively* dark spots. The presence of bright spots (faculae), which are common on the Sun, make it effectively impossible for one to infer the baseline from single-band relative photometry alone.

<sup>8</sup> In practice, the expressions derived below 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.

<sup>9</sup> Importantly, the sample mean and process mean 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 in general be different from the mean flux *across all stars* with similar surface properties (the process mean).

where

$$z \equiv \frac{\langle \Sigma \rangle}{\mu^2} \quad (29)$$

is the ratio of the average element in  $\Sigma$  to the square of the mean of the Gaussian process,  $\mathbf{q}$  is the ratio of the average of each row in  $\Sigma$  to the average element in  $\Sigma$ , and  $A, B$  are order unity and zero scalars, respectively, given by the optimally-truncated diverging series

$$A \equiv \sum_{i=0}^{i_{\max}} \frac{(2i+1)!}{2^i i!} z^i \quad (30)$$

$$B \equiv \sum_{i=0}^{i_{\max}} \frac{2i(2i+1)!}{2^i i!} z^i, \quad (31)$$

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

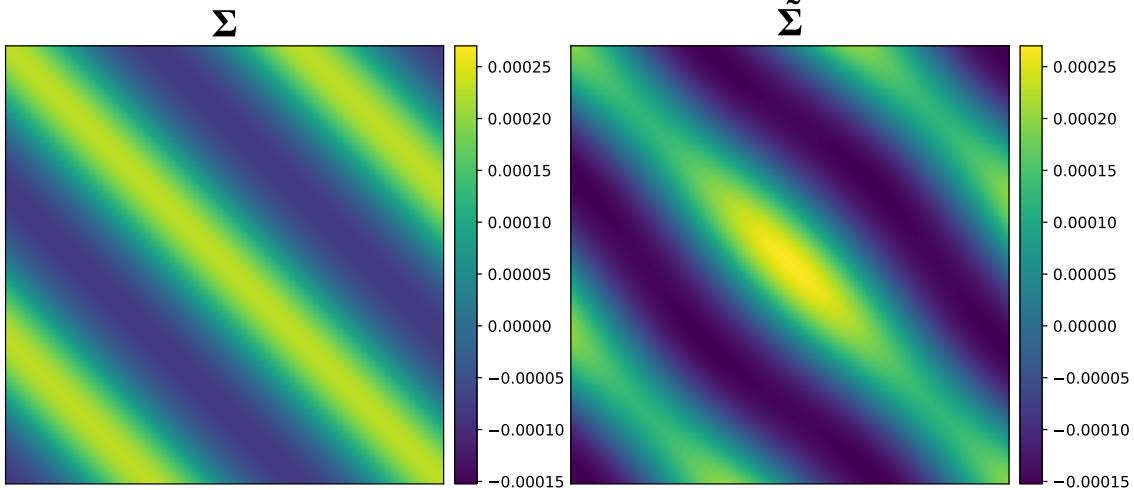
What Equation (28) allows us to do is effectively marginalize over the unknown baseline by modeling the *normalized* flux as a draw from a Gaussian process:

$$\tilde{\mathbf{f}}(P, \mathbf{u}, \boldsymbol{\theta}_*) \sim \mathcal{N}\left(\mathbf{1}, \tilde{\Sigma}(P, \mathbf{u}, \boldsymbol{\theta}_*)\right). \quad (32)$$

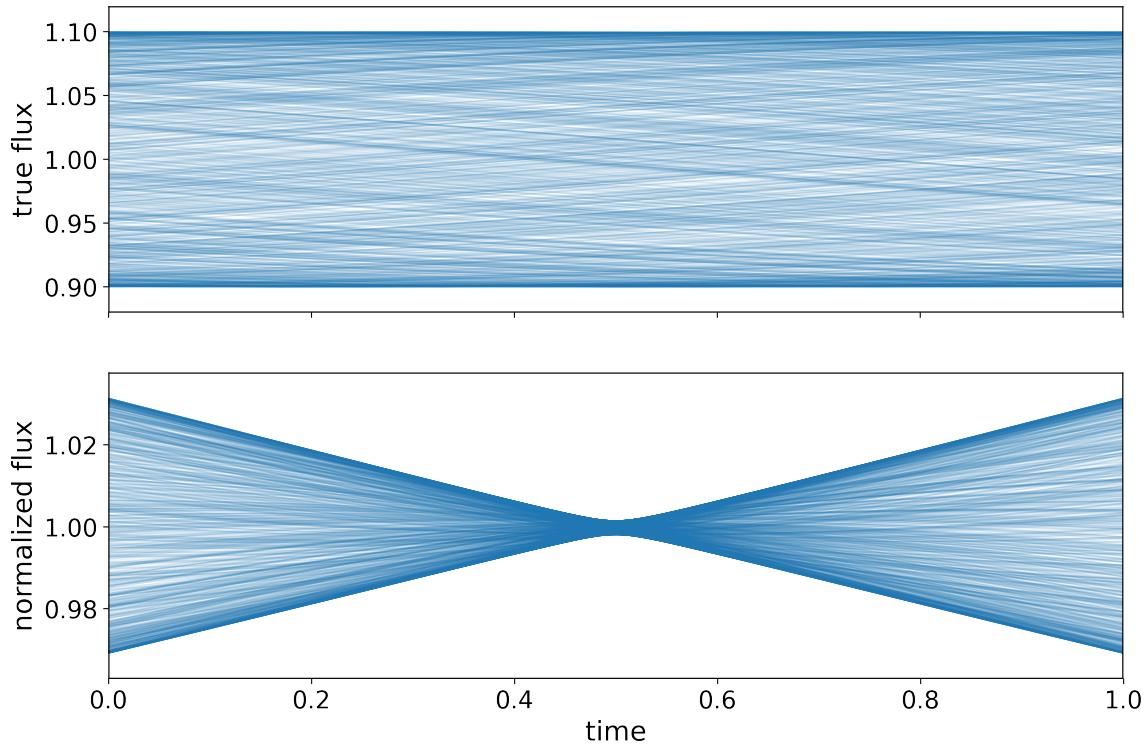
This is appropriate as long as  $z \ll 1$ , for which the true distribution of  $\tilde{\mathbf{f}}$  is approximately Gaussian. In practice, we recommend employing this trick only for  $z \lesssim 0.02$ , for which the error in the approximation to the covariance is less than  $10^{-6}$ . In cases where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

Figure 4 shows an example of a covariance matrix normalized according to the procedure outlined above. The principal difference between the normalized covariance and the original covariance is an overall scaling and a small offset. However, the normalization also results in the process becoming non-stationary: the covariance between two points in a light curve is now slightly dependent on their phases.

To understand why normalized light curves are non-stationary, it is useful to consider the toy example in Figure 5, in which we draw 1,000 samples from a sinusoid with period ten times longer than the observational baseline. In this limit, each light curve is approximately linear, which causes its mean value to roughly coincide with the midpoint of the observation window. Division by the mean value results in points near the midpoint being driven to unity and points near the edges (whose values differ the most from the mean) to be driven to both large and small values. The variance



**Figure 4.** An example of a flux covariance matrix  $\Sigma$  for a starry process (left) and the corresponding covariance of the normalized process (right), computed from Equation (28). In addition to an offset and an overall scaling relative to the original covariance matrix, the covariance of the normalized process is discernibly non-stationary (see Figure 5). [Download](#)



**Figure 5.** An example of why normalized light curves are non-stationary. The top panel shows 1,000 samples from a unit-mean sinusoid with an amplitude of 10% and a period of 10 days, much longer than the 1 day observation baseline. The bottom panel shows the same light curves, each normalized to its own mean. Because the mean tends to be near the center of the observation window, points near  $t = 0.5$  are driven to values very close to unity, while points near the edges have much larger scatter. [Download](#)

across all samples is now distinctly phase-dependent. In general, the non-stationarity will be most extreme in cases where the observation window is shorter than or comparable to one period, and asymptotically vanishes in the limit that a large number of cycles are observed.

#### 2.4. Summary

As the computation of the GP relies on many interdependent equations scattered throughout the previous sections and the Appendix, it is useful to summarize the procedure for the case where we marginalize over the inclination (§2.2) and the light curves are normalized to their means (§2.3), which is likely to be the primary use case for our algorithm.

We model the mean-normalized flux  $\tilde{\mathbf{f}}$  (Equation 27) as a Gaussian process:

$$\tilde{\mathbf{f}}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \sim \mathcal{N}\left(\mathbf{1}, \tilde{\Sigma}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet)\right). \quad (33)$$

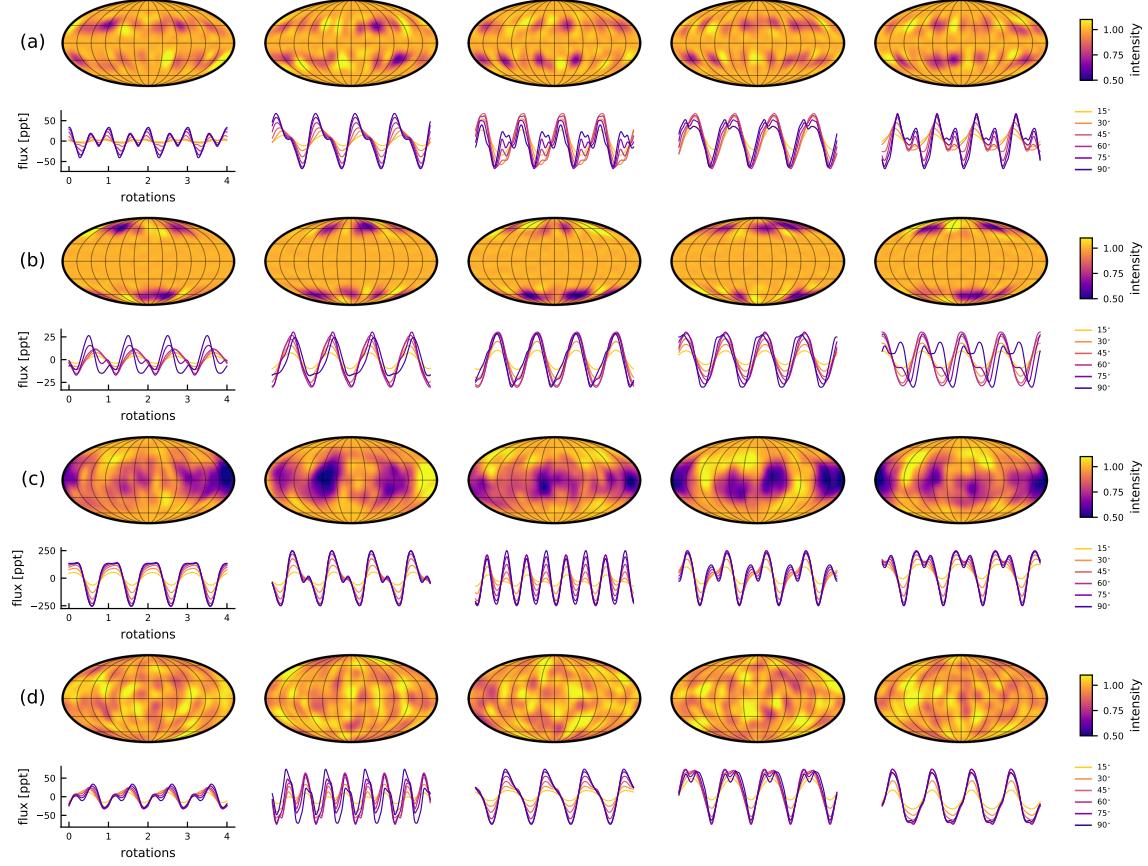
The hyperparameters of the GP are the stellar rotation period  $P$ , the vector of limb darkening coefficients  $\mathbf{u}$ , and the vector of parameters describing the spot distribution

$$\boldsymbol{\theta}_\bullet = (n \ c \ \mu_\phi \ \sigma_\phi \ r)^\top, \quad (34)$$

consisting of the number of spots  $n$ , their contrast  $c$ , the mode  $\mu_\phi$  and standard deviation  $\sigma_\phi$  of the latitude distribution, and the radius of the spots  $r$ . The quantity  $\tilde{\Sigma}$  is the covariance of the (normalized) process (Equation 28), which is a straightforward correction to the true covariance of the process, which accounts 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 (22) and (23), respectively. Those expressions in turn depend on the inclination expectation integrals  $e_I$  (Appendix D.2) and  $\mathbf{E}_I$  (Appendix D.3). Those, in turn, depend on the first and second moments of the distribution of spherical harmonic coefficient vectors,  $E[\mathbf{y}|\boldsymbol{\theta}_\bullet]$  and  $E[\mathbf{y}\mathbf{y}^\top|\boldsymbol{\theta}_\bullet]$ , given by Equations (C18) and (C19), respectively. To compute those, we must evaluate four nested integrals (Equations C20–C23 for the first moment and C24–C27 for the second moment), corresponding to integrals over the radius, latitude, longitude, and contrast distributions, respectively. 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.<sup>10</sup> Moreover, most of the terms in the expectation vectors and matrices may be computed recursively, and many may be pre-computed, as they do not depend on user inputs. It is therefore possible to evaluate  $\tilde{\Sigma}$  in an efficient manner. In §6 we discuss our implementation of the algorithm in a user-friendly Python package.

<sup>10</sup> The exception to this is the normalization correction (§2.3), which depends on a fast-to-evaluate series and thus adds negligible overhead to the computation.



**Figure 6.** Five random samples from the GP prior (columns) conditioned on four different values of  $\boldsymbol{\theta}_\bullet$  (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 mid-latitude spots; (b) small circumpolar spots; (c) large equatorial spots; and (d) small isotropic spots. See text for details. [Download](#)

## 2.5. An example

A concrete example of the GP derived above is presented in Figure 6, where we show random samples from the process evaluated up to spherical harmonic degree  $l_{\max} = 15$  and conditioned on different values of the hyperparameter vector  $\boldsymbol{\theta}_\bullet$ . Each column corresponds to a different random draw from the GP, while each row corresponds to a different value of  $\boldsymbol{\theta}_\bullet$ . 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 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  $15^\circ$  (yellow) to  $90^\circ$  (dark blue), and assuming no limb darkening (i.e.,  $\mathbf{u} = \mathbf{0}$ ). From top to bottom, the hyperparameter vectors  $\boldsymbol{\theta}_\bullet$  for

each row are

$$(n \ c \ \mu_\phi \ \sigma_\phi \ r)^\top = (10.0 \quad 0.10 \quad 30.0 \quad 5.00 \quad 10.0)^\top \quad (35a)$$

$$= (10.0 \quad 0.10 \quad 60.0 \quad 5.00 \quad 10.0)^\top \quad (35b)$$

$$= (10.0 \quad 0.05 \quad 0.00 \quad 5.00 \quad 30.0)^\top \quad (35c)$$

$$= (20.0 \quad 0.10 \quad 0.00 \quad 40.0 \quad 10.0)^\top. \quad (35d)$$

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  $30^\circ$  centered at  $0^\circ \pm 5^\circ$  latitude with a contrast of 5%; and (d) 20 spots of radius  $10^\circ$  centered at  $0^\circ \pm 40^\circ$  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 aren't 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 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 truncation error in the spherical harmonic expansion, to an intrinsic limitation of our

Gaussian approximation, and to 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.<sup>11</sup> 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 is sufficiently informative, it does not matter that the prior has finite support for unphysical configurations, as those will be confidently rejected.

In §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 in general unbiased and correctly estimates the posterior variance when used to infer the spot properties  $\boldsymbol{\theta}_\bullet$ .

But before we dive in, it is worth pausing for a moment to take another look at Figure 6. 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 Gaussian process, after all, is a non-parameteric *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; e.g., Wandelt 2012), the process is assumed to be isotropic, with no preferred angular direction. The fact that we are able to construct a GP that captures discrete features restricted to certain regions on the surface (albeit with some issues) is remarkable.

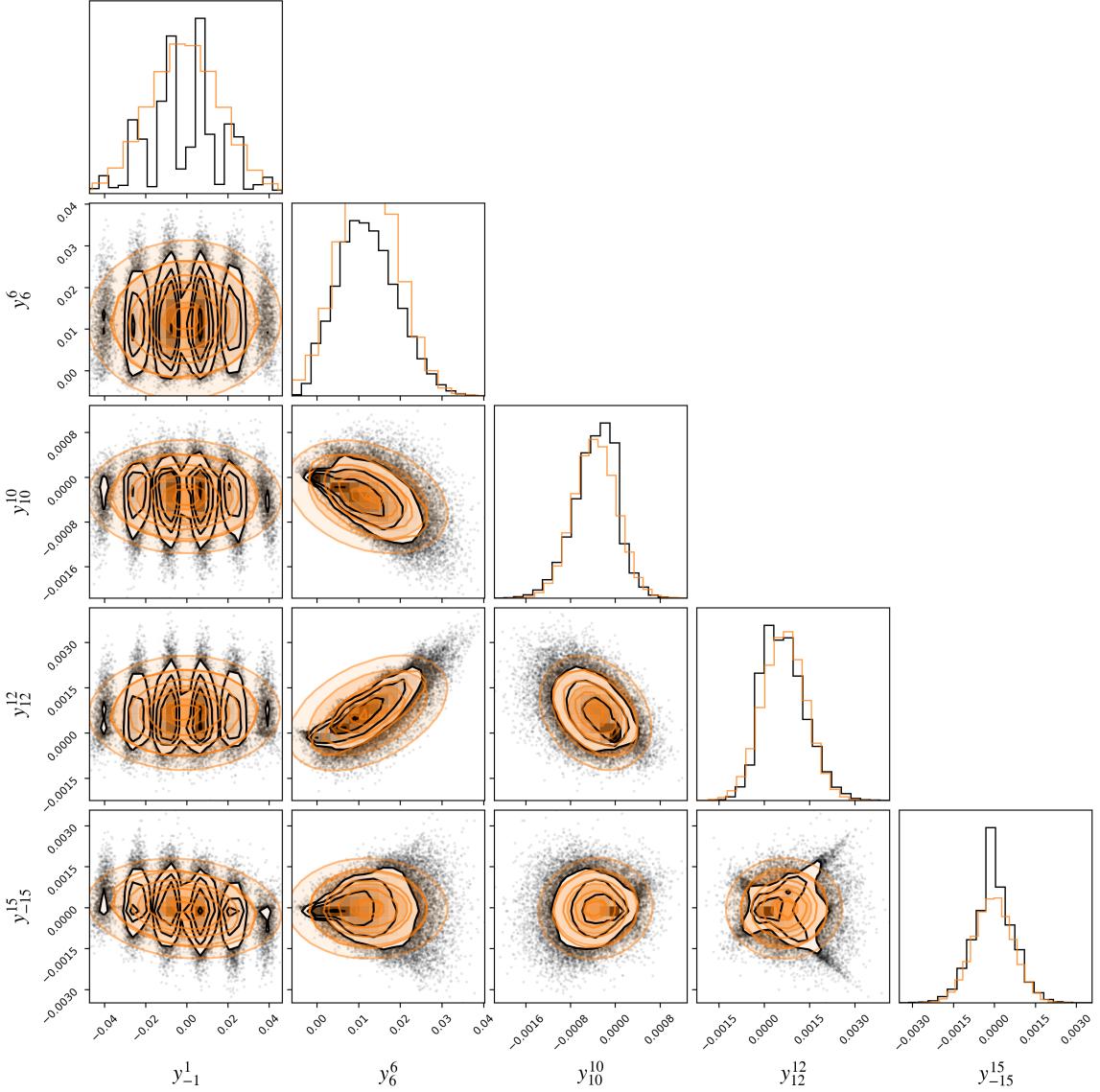
### 3. CALIBRATION

#### 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  $\boldsymbol{\theta}_\bullet$  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  $p(\tilde{\mathbf{f}} \mid P, \mathbf{u}, \boldsymbol{\theta}_\bullet)$  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  $\mu_\phi \pm \sigma_\phi = 30^\circ \pm 5^\circ$ . Let us then expand each surface in spherical harmonics and visualize the distribution of coefficients  $\mathbf{y}$ . Figure 7

<sup>11</sup> 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.



**Figure 7.** Corner plot showing the joint distributions of select spherical harmonic coefficients corresponding to the  $l = 15$  expansion of  $5 \times 10^4$  stellar surface maps drawn from a certain distribution of discrete circular spots (black points and contours). At  $l = 15$  there are 256 coefficients in total; we chose five of the coefficients with the most odd-looking distributions to illustrate the non-Gaussianity of the process. In addition to non-linear correlations, skewness, and the existence of points of very high curvature, some of the distributions are also multi-modal. The orange contours show slices of the Gaussian approximation to the joint distribution; this is the approximation adopted in this paper.

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	$\sim \mathcal{N}(30^\circ, 5^\circ)^2$
$\lambda$	spot longitude	$\sim \mathcal{U}(0^\circ, 360^\circ)$
$r$	spot radius	$\sim \mathcal{N}(15^\circ, 0^\circ)^2$
$i$	stellar inclination	$\sim \sin$
$P$	rotational period	1 day
$\mathbf{u}$	limb darkening coefficients	$(0 \ 0)^\top$
$\sigma_f$	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

**Table 1.** Default parameters used to generate synthetic light curves in the calibration tests.

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, non-linearly 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 in general an issue when doing inference with our GP, as the resulting posteriors are unbiased.

### 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  $M$  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 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 much higher degree ( $l_{\max}^{(0)} = 30$ ) than the degree we will use in the inference step ( $l_{\max} = 15$ ) to minimize potential ringing effects or other artefacts in the synthetic data. For reference, the chosen degree  $l_{\max} = 30$  is large enough to resolve features on the order of  $180^\circ/30 = 6^\circ$  across, but small enough that the algorithm for computing the light

curve is numerically stable.<sup>12</sup> We compute the light curve at  $K$  points equally spaced over a baseline  $\Delta t$  using the `starry` algorithm (Appendix B), assuming an inclination  $\mathbb{I}$ , a rotational period  $P$ , and limb darkening coefficients  $\mathbf{u}$ . Finally, we divide the light curve by the mean and add Gaussian noise with standard deviation  $\sigma_f$  to emulate photon noise.<sup>13</sup> The default values/distributions of each of the 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, the rotational period, and the 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.

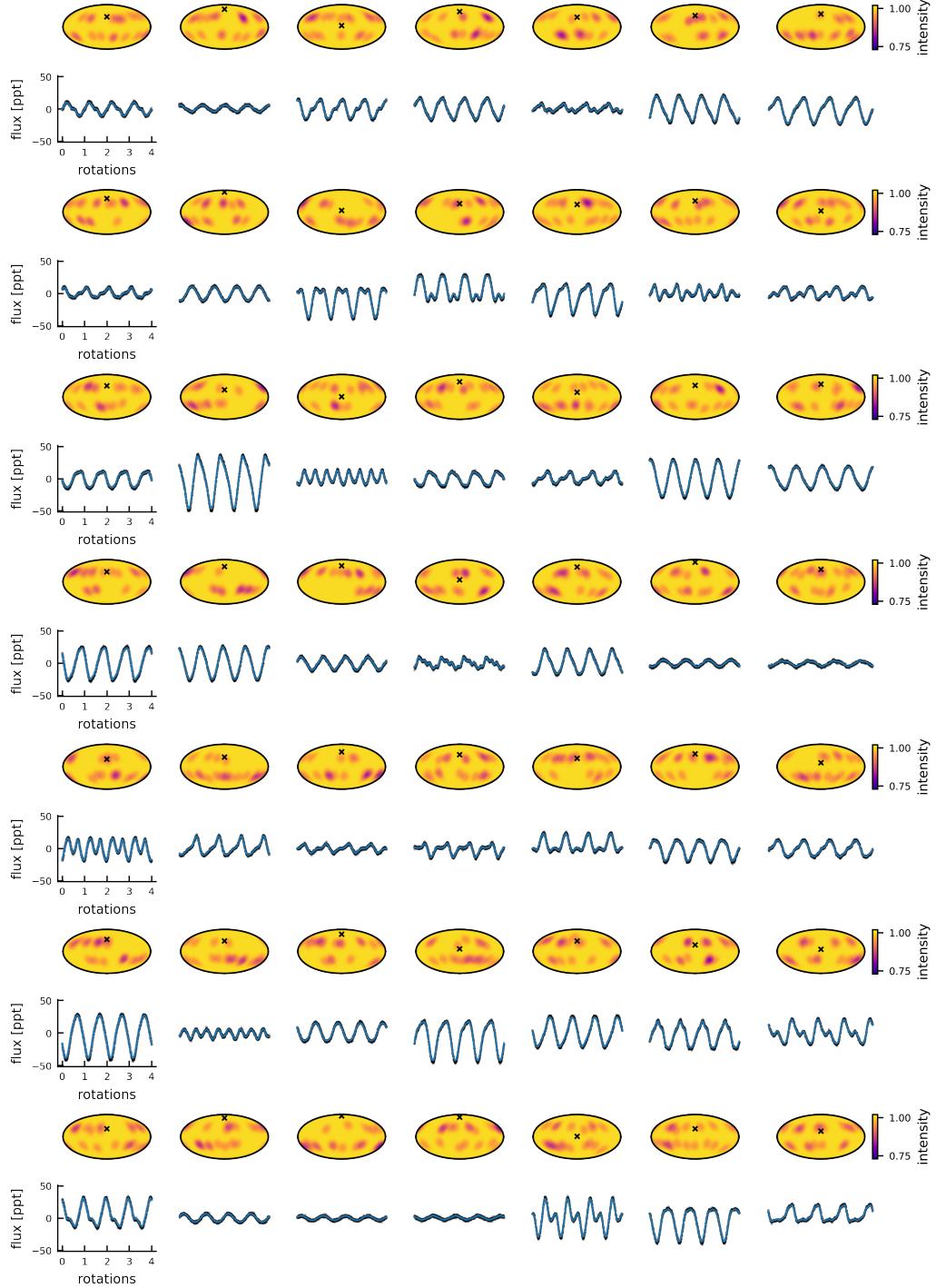
Figure 8 shows a synthetic dataset generated from the default values listed in Table 1. While the light curves correspond to surfaces with the same statistical spot properties, they all look qualitatively different: the mapping from star spot properties to flux is nontrivial. In the inference step below, we assume 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 (§6) to perform inference on the synthetic datasets. 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 §5.3. 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  $\boldsymbol{\theta}_\bullet$  (Equation 17). We experimented with three different methods for doing posterior inference on our synthetic datasets: No-U Turn Sampling, a variant of Hamiltonian Monte Carlo (NUTS; Duane et al. 1987; Hoffman & Gelman 2011), automatic differentiation variational inference (ADVI; Kucukelbir et al. 2016; Blei et al. 2016), and nested sampling (Skilling 2004, 2006). We obtained the best performance using the nested sampling algorithm implemented in the `dynesty` package (Speagle 2020), so that is what we will use below.

<sup>12</sup> 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 would still be subject to artefacts due to the discretization scheme. We have gone 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.

<sup>13</sup> 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 18). This would significantly increase the computational cost of our tests. Fortunately, in practice, the difference between these two approaches has 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 won't have this choice!



**Figure 8.** A synthetic dataset generated by adding discrete spots to a stellar surface with parameters given by the default values listed in Table 1. 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  $\times$  on each map indicates the sub-observer 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.

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  $\mu_\phi$  and standard deviation  $\sigma_\phi$ , provided we account for the Jacobian of the transformation (Equation C71) in our log probability function, which maps a uniform prior on  $a$  and  $b$  to a uniform prior on  $\mu_\phi$  and  $\sigma_\phi$ .

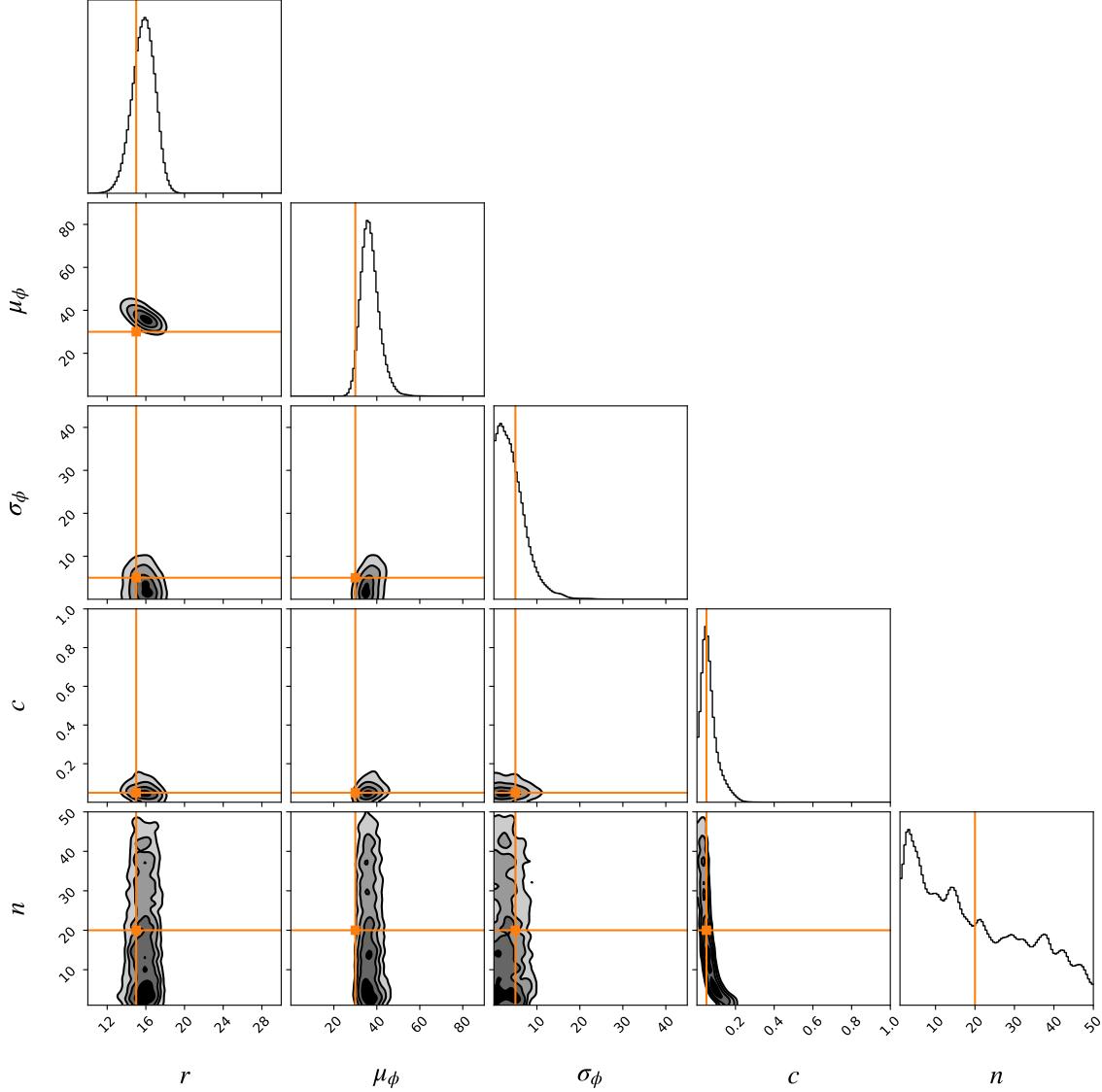
We place uniform priors on all five quantities, with support in  $1 \leq n \leq 50$ ,  $0 < c \leq 1$ ,  $10^\circ \leq r \leq 30^\circ$ ,  $0 \leq a \leq 1$ , and  $0 \leq b \leq 1$ . Note that while  $n$  formally represents an integer, its effect on the GP is a scaling of the covariance (see Equation C19); 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 §4.8.

We use Equation (18) as our log likelihood term, adding the log of the absolute value of Equation (C71) to enforce a uniform prior on  $\mu_\phi$  and  $\sigma_\phi$ . As we mentioned above, the fact that  $P$ ,  $\mathbf{u}$ , and  $\sigma_f$  are shared among all  $M$  light curves means that  $\Sigma + \mathbf{C}_m$  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 (28). Since we only consider light curves with variability limited to a few percent or less, the approximation for  $\tilde{\Sigma}$  is always valid. Finally, we restrict our spherical harmonic expansion to  $l_{\max} = 15$  as a compromise between resolution, computational speed, and numerical stability (see §6).

We use the standard implementation of the nested sampler, `dynesty.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  $\Delta \ln \mathcal{Z}$  drops below 0.5—is usually attained after 5,000 to 10,000 samples and within a couple hours on a typical machine for most of the trials we perform.

### 3.4. Default run

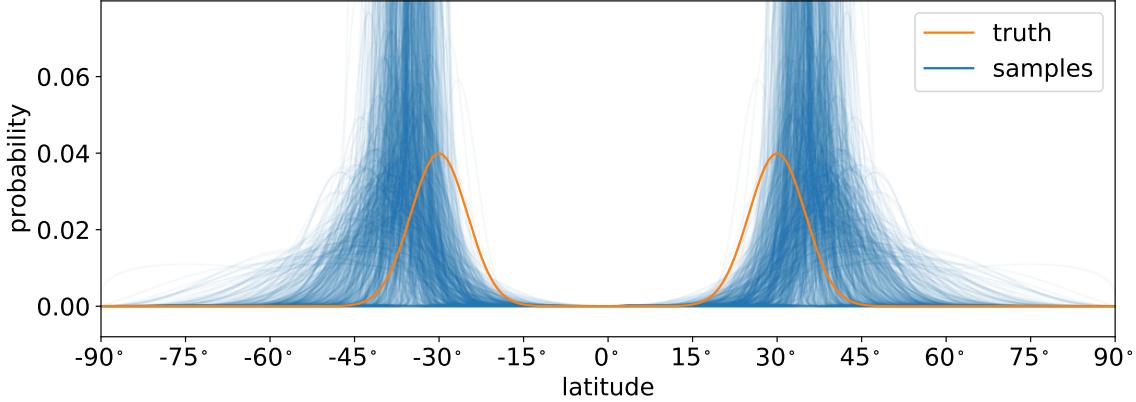
The input parameters for the default run are shown in Table 1, and the corresponding light curves in Figure 8. We run the nested sampler as described in the previous section and transform the posteriors in  $a$  and  $b$  to posteriors in  $\mu_\phi$  and  $\sigma_\phi$  via Equations (C63), (C66), and (C68). The results are shown in Figure 9, where we correctly infer all five parameters within 2 – 3 standard deviations. Posterior distributions for the spot radius  $r$ , the central spot latitude  $\mu_\phi$ , and the spot contrast  $c$  are fairly tight, while the distribution for the latitudinal scatter  $\sigma_\phi$  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 §4.6.



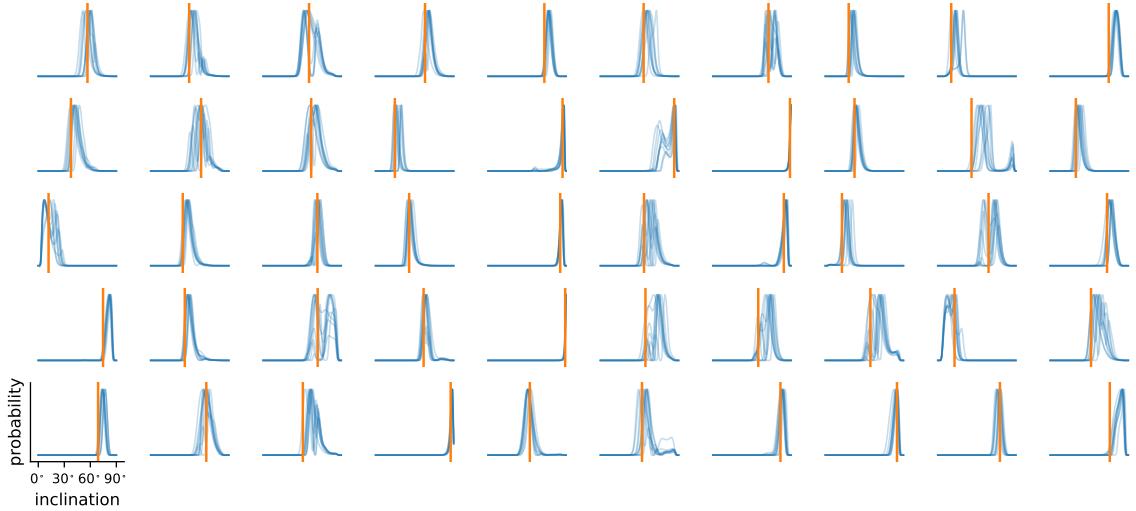
**Figure 9.** Posterior distributions for the spot parameters  $\theta_\bullet$  (radius  $r$  in degrees, latitude mode  $\mu_\phi$  and standard deviation  $\sigma_\phi$  in degrees, fractional contrast  $c$ , and number of spots  $n$ ) for the default run (Table 1 and Figure 8). The axes span the entire prior volume, and the orange lines and markers indicate the true (input) values.

Figure 10 shows samples from the spot latitude posterior (hyper)distribution. Since the parameters  $\mu_\phi$  and  $\sigma_\phi$  characterize a distribution over spot latitudes, uncertainty in their values translates to uncertainty in the actual shape of the spot latitude distribution. Thus, the ensemble of blue curves in Figure 10 quantifies our knowledge of how spots are distributed on the surfaces of the stars in the dataset. These distributions are again consistent with the true distribution used to generate the spots (orange curve) within less than 2 standard deviations.

Even though we explicitly marginalized over inclination, we can still derive posterior constraints on the inclinations of the individual stars in our sample by computing the log-likelihood as a function of  $I$  conditioned on the value of  $\theta_\bullet$  from a particular



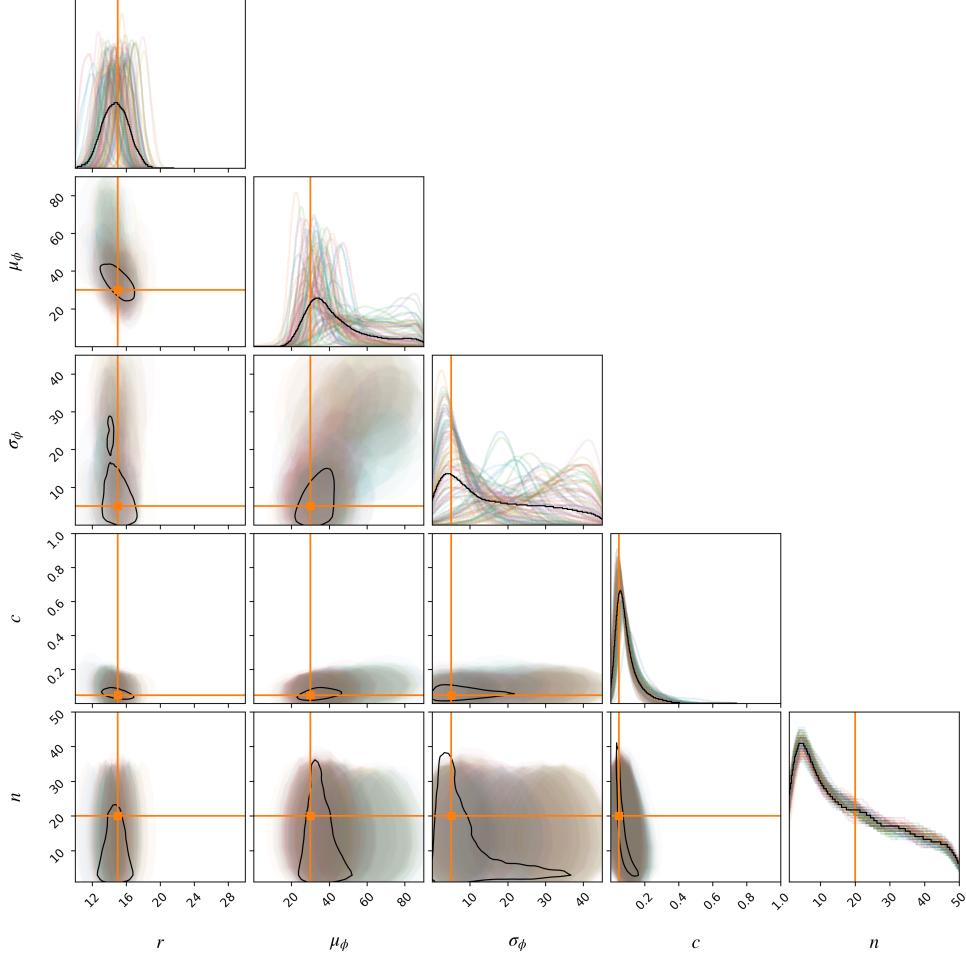
**Figure 10.** Posterior distributions for the spot latitudes for the default run. Each blue curve corresponds to the Beta distribution PDF (Equation C61) for the spot latitude with parameters drawn from the posterior in  $\mu_\phi$  and  $\sigma_\phi$  (Figure 9); the (hyper)distribution of blue curves quantifies our beliefs about how spots are distributed on any given star. The orange curve is the true distribution used to generate the spots (see Table 1).



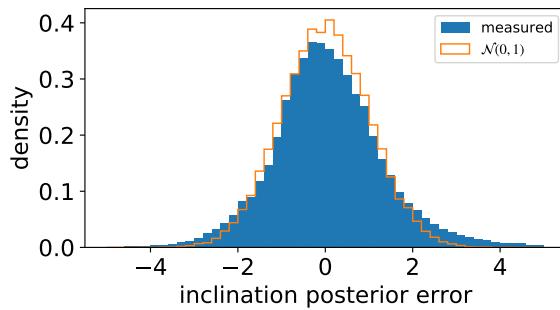
**Figure 11.** Posterior distributions for the inclinations of individual stars; individual panels correspond directly to those in Figure 8. As in Figure 10, each blue curve is a sample from the hyperdistribution of inclination PDFs, conditioned on a specific value of  $\theta_\bullet$  drawn from the posterior. The orange lines indicate the true inclination of each star.

draw from the posterior. We do this in Figure 11, where blue curves again correspond to samples from the posterior hyperdistribution, and orange lines indicate the true inclination; the panels are arranged in the same order as those in Figure 8. In almost all cases, we are able to constrain the inclinations of individual stars to within about  $10^\circ$ , consistent with the truth at less than  $2-3\sigma$ .

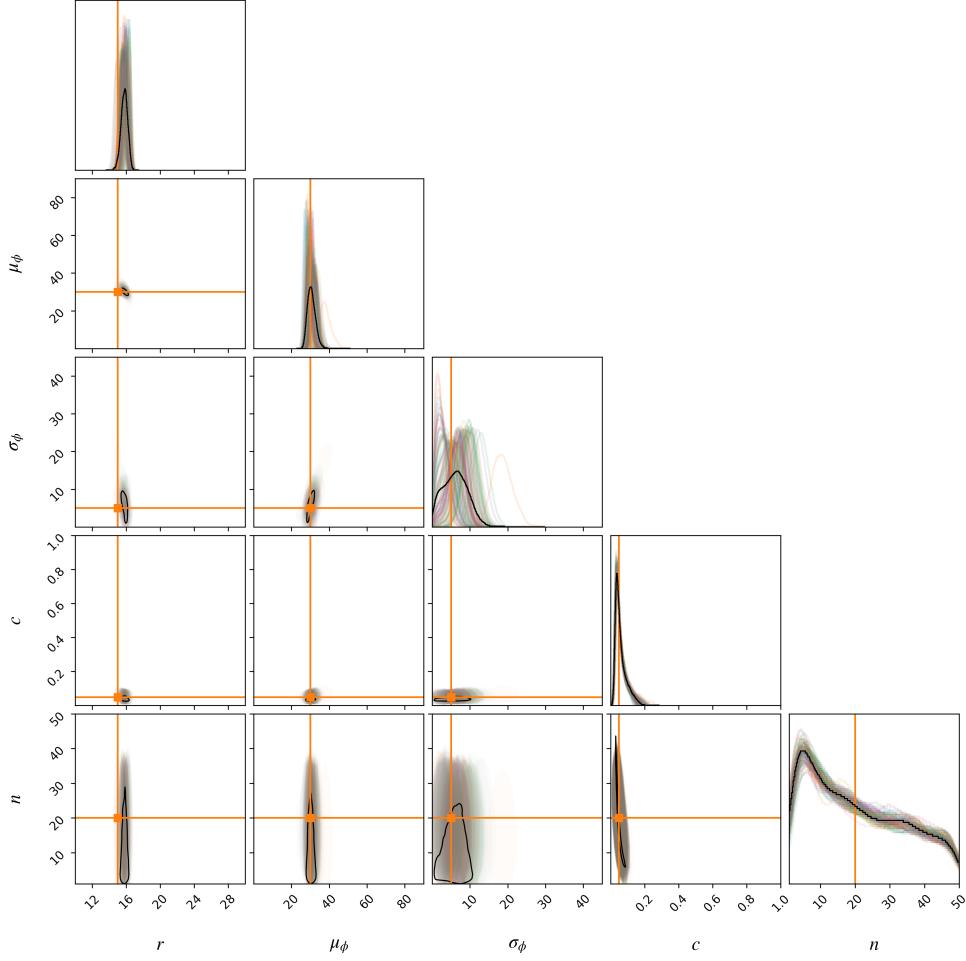
The results in Figures 9—11 are based on a single realization of the light curve ensemble conditioned on the properties of Table 1. To properly gauge potential biases in our model, it is useful to re-run the experiment under many different realizations of the dataset. We therefore generate 100 ensembles of light curves in exactly the same



**Figure 12.** Similar to Figure 9, but showing the posterior distribution for 100 different synthetic datasets, all generated from the same default input parameters (Table 1). Each colored histogram corresponds to a single run, with the corresponding  $1\sigma$  contours shown as shaded ellipses for each pair of parameters. The black histograms correspond to the distributions of all samples from each of the 100 runs, and the black curves again indicate  $1\sigma$  contours in the joint posterior.



**Figure 13.** Distribution of stellar inclination residuals normalized to the posterior standard deviation for all 5,000 stars across the 100 trials in Figure 12 (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.



**Figure 14.** Same as Figure 12, but showing the posterior distributions of 100 realizations of ensembles with  $M = 1,000$  (instead of 50) light curves each. The constraints on mosts of the parameters are much tighter, and the polar spot degeneracy is gone.

way as above and run inference on each of them. Figure 12 shows the marginal and joint posterior distributions for  $\theta_\bullet$  for all 100 trials. Posteriors for individual trials are shown as the translucent colored curves (in the marginal plots) and as ellipses bounding the  $1\sigma$  posterior level (in the joint posterior plots). The black curves show the marginal distributions of all samples across all trials, and the black contours show the corresponding  $1\sigma$  levels in the joint posterior plots.

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 baseline degeneracy discussed in §2.3 and illustrated in Figure 3: the total spottiness of a stellar surface cannot be probed in relative photometry. In particular, this means 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 mid-latitudes (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 (relative) light curve is negligible, and some spots at mid-latitudes, 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 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 a shrinkage prior on  $\sigma_\phi$  (§4.8), observe in a different mode that is sensitive to baseline information (§4.6), or simply collect more data. As we will see below, the degeneracy described above is not perfect: for very large ( $M \sim 1,000$ ) 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 12 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 baseline degeneracy: the total number of spots is not an observable in relative photometry (see §4.6).

There is one final distribution that is instructive to consider: the distribution of errors on the inferred stellar inclination. Figure 13 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  $\mathcal{N}(0, 1^2)$  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 14 shows the same posterior distributions as in Figure 12, but for 100 trials each with  $M = 1,000$  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 mid-latitude 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 (§3.2) is different from how we actually model these spots. While we generate the spots as compact circular disks expanded at high spherical harmonic degree, we model them as sigmoids (§C.1) expanded at 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 dataset, and the corresponding figures are presented at the end of the Appendix.

Figures S25—S28 show the results for different latitudinal distributions, keeping all other values in Table 1 the same. Specifically, Figure S25 corresponds to a trial with mid-latitude ( $\mu_\phi = 45^\circ$  and  $\sigma_\phi = 5^\circ$ ) spots, Figure S26 to a trial with high-latitude ( $\mu_\phi = 60^\circ$  and  $\sigma_\phi = 5^\circ$ ) spots, Figure S27 to a trial with equatorial ( $\mu_\phi = 0^\circ$  and  $\sigma_\phi = 5^\circ$ ) spots, and Figure S28 to a trial 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 spot contrast within  $2 - 3\sigma$ ; the number of spots is equally unconstrained in all runs. In Figure S26 and to a lesser extent in Figure S25, 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 S27 and S28 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 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 parametrization of the latitude distribution can approximate a zero-mean Gaussian. While the Beta distribution in  $\cos\phi$  *can* be unimodal in  $\phi$  (see Equation C61 and the first column of Figure 24), 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 S28. In both cases, the posterior approximates the true distribution as best it can given the constraints of the adopted PDF.

Figure S29 tests the performance of the model on 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  $r \gtrsim 10^\circ$  (see Figure 23), so we place zero

prior mass below this value. The figure shows the results of inference on a dataset generated from spots with  $r = 3^\circ$  (and an increased contrast  $c = 1$  to enforce a comparable signal-to-noise 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 the model is biased against small spots, this does not affect inference 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 S30 and S31 show results for the default run but with extreme values of the number of light curves in the ensemble:  $M = 1$  and  $M = 1,000$ , respectively. These two figures underscore the power of ensemble analyses: a single light curve (Figure S30) 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.

Figure S32—S34 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 S32, 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 odd harmonics at the expense of introducing strong degeneracies with the even harmonics (§4.3). In practice, this leads to higher uncertainty in the spot radii and latitudes relative to the same dataset without limb darkening (Figure 9). Fortunately, this uncertainty can be dramatically reduced with more data, as evident in Figure S33, which shows the results of the same run but with  $M = 1,000$  light curves in the ensemble. The constraints on  $r$ ,  $\mu_\phi$ ,  $\sigma_\phi$ , and  $c$  are now much tighter and in good agreement with the truth. Finally, Figure S34 shows the results of inference on limb-darkened light curves under the (wrong) assumption that limb darkening is not present ( $\mathbf{u} = \mathbf{0}$ ). Neglecting the effect of limb darkening leads to significant biases in the spot radius and latitude parameters. While the model still favors mid-latitude spots (at  $\sim 45^\circ$  instead of  $30^\circ$ ), the constraints are deceptively tight and discrepant by many standard deviations. We discuss these points in more detail in §4.3.

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 S35 and S36 show what happens when the model is applied to stars with  $n = 2$  and  $n = 1$  spots, respectively. Despite large portions of the light

curves being flat (and therefore extremely non-stationary) in these scenarios, the GP does surprisingly well, recovering the radii and latitude parameters within  $2 - 3\sigma$  in both cases. Note that in order to preserve the same signal-to-noise ratio relative to the other runs, the spots in Figure S35 have much higher contrast ( $c = 0.5$ ). Even though the contrast is degenerate with the number of spots (which is unobservable), the  $c$  posterior has a much heavier tail than in the other runs, showing that the model is in fact sensitive to the spot contrast to some degree.<sup>14</sup> In Figure S36, 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 (unmodelled) scatter in their sizes.

In Figure S37 we add variance to all 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 latitude variance 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 trial shown in Figure S37 while explicitly sampling over the distribution in  $\Delta r$  shifts the posterior mass to lower radii, mitigating the bias described above.

Our final trial is shown in Figure S38, in which we assume we know the true baseline 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 §2.3). In practice, this would require knowledge of the brightness (or temperature) of the unspotted photosphere, which is not an observable in single-band relative photometry. This value can be probed, however, in multi-band photometry (see §4.6), for which this trial is extremely relevant. We again recover the radii and latitude parameters to within  $2 - 3\sigma$ , but most importantly, we also infer the correct spot contrast *and* the correct number of spots with fairly high precision. In particular, knowledge of the baseline breaks the  $c - N$  degeneracy. Photometric measurements in multiple bands (even just two!) are therefore extremely useful when inferring spot properties. We discuss this point at length in §4.6.

## 4. DISCUSSION

### 4.1. *The null space*

In general, inferring the properties of a stellar surface from its light curve is not only difficult, but *formally impossible*. To understand why, consider an expansion of

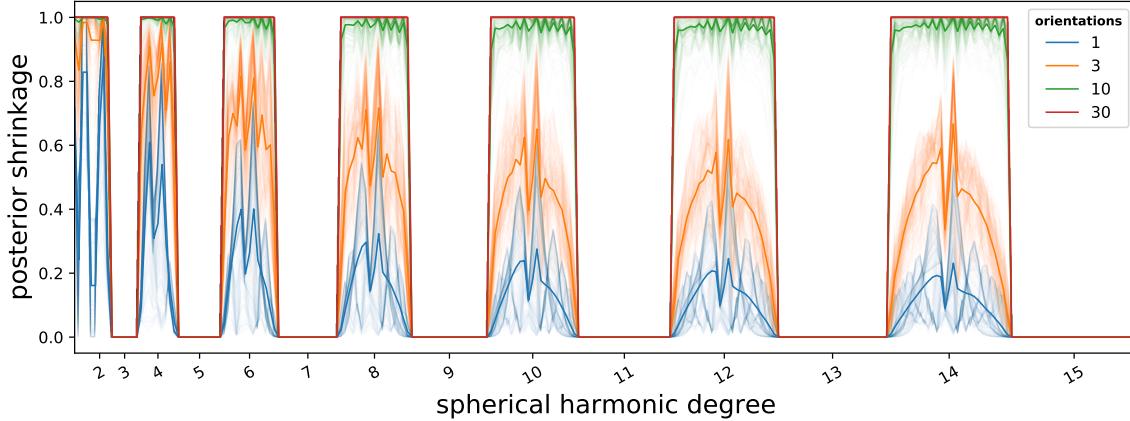
<sup>14</sup> The sensitivity to  $c$  is for the most part due to our hard lower prior bound  $n \geq 1$ .

the stellar surface intensity in the spherical harmonic basis (see Figure 2). Assuming (for the moment) that the star rotates about an axis that points up along the page, the observed light curve may be expressed as a weighted sum of the disk-integrated intensity of each of the spherical harmonics as they rotate about that same axis. However, not all spherical harmonics will contribute to the full light curve, as many (in fact, most) of the spherical harmonics are perfectly antisymmetric about the equator. This is the case for the  $l = 1, m = -1$  harmonic, which integrates to zero regardless of the phase at which it is viewed. The same is true, in fact, for all other harmonics of order  $m = -1$  and (perhaps less obviously) for all harmonics with odd  $l = 3, 5, 7, \dots$ . Furthermore, there exist many linear combinations of spherical harmonics that similarly integrate to zero at all rotational phases. Together, these modes constitute the *null space* of the problem: the set of modes on the surface that do not contribute to the observed light curve and therefore cannot be probed from photometry. As we will see below, the vast majority of the surface modes lie at least in part in the null space.

But what if the star rotates about a different axis? A useful property of the spherical harmonics is that rotation about any axis can only change the order  $m$  of the harmonic, while keeping the degree  $l$  fixed: in other words, rotation simply mixes the spherical harmonics along rows in Figure 2. Therefore, if we rotate the spherical harmonics in the figure about a different axis, we are still subject to just as many degeneracies (i.e., the size of the null space is unchanged), but the particular linear combination of modes in the null space will be different.

This all seems quite hopeless, and for analyses of individual stars, it pretty much is. Unless extremely strong (and likely incorrect) priors are employed (such as assuming the stellar surface consists of exactly two perfectly circular, uniform contrast spots), we simply cannot learn much about the surface of an individual star from photometry.

However, we can use the dependence of the null space on inclination to our advantage. If we could observe a star from many different vantage points, we would be able to break many of the degeneracies at play, since we would get different constraints on the amplitude of each mode when viewed at different inclinations. This, of course, is not possible (at least not yet!). But what we *can* do is observe many similar stars, each viewed at a different (random) inclination, and attempt to learn something about the properties of the ensemble of stars as a whole. This is exactly what we did in §3, where we showed that this kind of ensemble analyses is not only possible but can also be extremely constraining of the spot properties at the population level. In the following section, we explore the role of ensemble analyses in breaking the degeneracies of the mapping problem in more detail.



**Figure 15.** Posterior shrinkage as a function of spherical harmonic degree for a hypothetical scenario in which an observer can measure the light curves of 1 (blue), 3 (orange), 10 (green), and 30 (red) *identical* stars viewed at random orientations. The thicker curves correspond to the mean shrinkage for each experiment. The information content in the light curve of a star observed from a single vantage point approaches zero as  $l$  increases. However, observing many identical stars from different vantage points allows one to recover nearly all of the information in the even spherical harmonic modes. This is why ensemble analyses of many similar stars at different inclinations allows us to infer their surface properties.

#### 4.2. Why ensemble analyses work

To quantify the power of ensemble analyses, let us define the *posterior shrinkage*  $R$  for a given spherical harmonic coefficient as

$$R \equiv 1 - \lim_{\sigma_0^2 \rightarrow \infty} \frac{\sigma^2}{\sigma_0^2}, \quad (36)$$

where  $\sigma_0^2$  is the prior variance and  $\sigma^2$  is the posterior variance. The posterior shrinkage is a measure of how informative a measurement is about a given mode on the surface in the limit of infinite signal-to-noise ratio (SNR) and is independent of what the stellar surface actually looks like. If, at infinite SNR and with a completely uninformative prior, a particular mode can be learned exactly from a dataset, the posterior shrinkage is defined to be unity. Conversely, if the data is completely unconstraining of that mode (i.e., it is entirely in the null space),  $R$  will tend to zero.

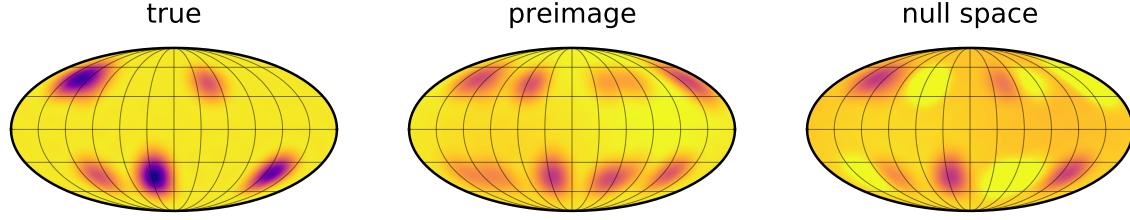
Figure 15 shows the posterior shrinkage as a function of spherical harmonic degree for a few different hypothetical scenarios in which we observe one or more stellar light curves with no limb darkening. Let us first consider the blue curves, in which a light curve is collected from a single star viewed at a single random inclination (which, of course, is all we can do from our fixed vantage point on Earth). Each thin blue curve corresponds to a particular random draw from the (isotropic) inclination distribution; the thick blue curve is the average over 300 trials. For some of the low-degree modes, the shrinkage is relatively high: it is fairly easy to constrain the dipole moment from a light curve, as this is usually the dominant sinusoidal signal. However, as the degree  $l$  increases, the shrinkage decreases dramatically: at  $l = 14$ , corresponding to features on scales of roughly  $15^\circ$ , the light curve can only tell us about  $\sim 10\%$

of the total information about what the surface looks like. As  $l$  increases further,  $R$  tends to zero. Another important feature of the shrinkage, which we hinted at above, is that it is exactly zero for all odd-degree modes above  $l = 1$ . This is a well-known fact: all odd spherical harmonics other than the dipole are in the null space *regardless of inclination* (e.g., Luger et al. 2019). In other words, these spherical harmonics are perfectly antisymmetric in projection over the unit disk when viewed from any orientation. Absent structure to break these symmetries (see below), we simply cannot learn anything about these modes from stellar light curves. If we average over the shrinkage for all modes up to  $l = 15$ , we find that a single light curve measurement can only tell us  $\sim 9\%$  of the information about the surface on those scales.

In an ensemble analysis, we assume we observe the lightcurves of many stars that are “similar” in some statistical sense. As a thought experiment, let us consider an extreme version of ensemble analysis in which all the stars in our dataset happen to have *identical* surfaces. We will still assume they are oriented at random inclinations, as we would expect for field stars. The orange curves in the top panel of the figure correspond to the case in which we analyze ensembles consisting of three identical stars viewed at three different random inclinations. Interestingly, the shrinkage increased at all even spherical harmonic degrees (the odd degrees, as we mentioned above, are always invisible). Note that since we are in the limit of infinite SNR, the fact that we have three times the data as we did in the previous scenario is irrelevant: the increase in the shrinkage is instead due to the fact that our observations from different vantage points broke some degeneracies in the problem. This is a consequence of the fact we mentioned in the previous section: the null space (for the even modes) is a strong function of the inclination. If we average over all modes, we obtain a total shrinkage of  $\sim 24\%$  for  $l \leq 15$  in this case.

Finally, the green and red curves correspond respectively to ensembles of ten and thirty identical stars observed at random inclinations. In these scenarios, the posterior shrinkage is nearly 100% for all even modes (it asymptotically approaches 100% as the number of light curves increases further). The total shrinkage is  $\sim 47\%$  and approaches 50% as both  $l_{\max}$  and the number of light curves tend to infinity. Thus, if we were able to measure light curves of identical stars from many different inclinations, the null space would consist *only* of the odd modes, and our data would tell us exactly half of all the information about the surfaces of these objects.

Figure 16 shows a spotted stellar surface decomposed into its *preimage* and its *null space* in the limiting case in which its light curve can be observed at a large number of inclinations. The former is the set of all modes that map onto the light curve (i.e., the complement of the null space). Because spots are compact features, they are necessarily made up of a continuum of spherical harmonic modes spanning many different values of  $l$ : they can therefore be seen in both the even modes (the preimage) and the odd modes (the null space). Interestingly, the symmetries at play



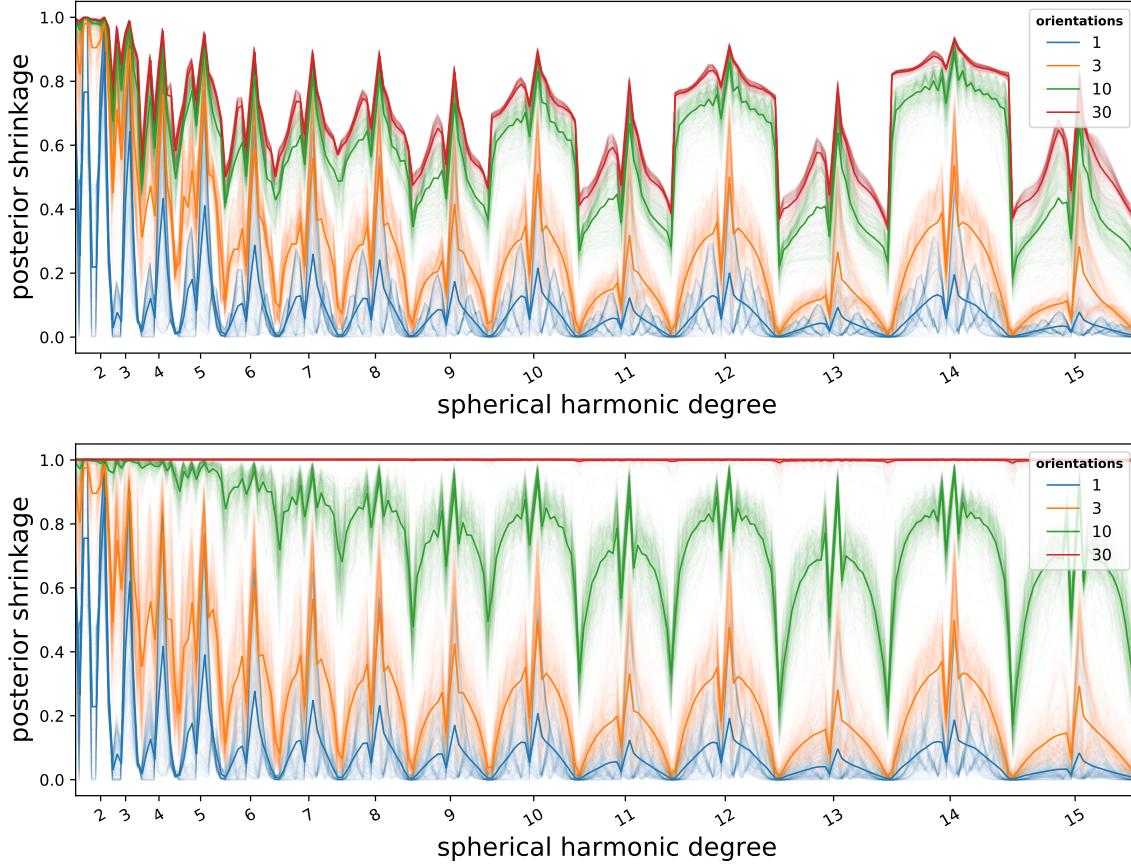
**Figure 16.** A spotted stellar surface seen in Mollweide projection (left), the same surface with all odd  $l > 2$  harmonics set to zero (center), and the same surface with  $l = 1$  and all even harmonics set to zero (right). The plot on the right represents the portion of the surface that is in the null space of the light curve problem at all inclinations; the plot in the center is the portion of the surface that can be inferred by combining observations made from all possible inclinations. [Download](#)

require spots to be paired with antipodal dark mirror images in the even modes, and with *bright* ones in the odd modes (which sum to perfectly cancel out in the true map). Since these reflections are at the same latitude (in the opposite hemisphere) as the true spots, they do not contribute any bias to our inference, which already assumes the latitude distribution is symmetric about the equator. Thus, even though we are only sensitive to the even modes in ensemble analyses, they are sufficient to learn about the properties we are interested in.

In practice, the fact that stars are not identical complicates this picture. We can still learn about the *distribution* of spot properties, but in general our ensemble must consist of far more light curves in order to place strong constraints on these properties. As we showed in §3, under ideal conditions an ensemble of  $M = 50$  similar stars may be sufficient, but in practice we may need on the order of hundreds or even a thousand light curves in order to mitigate the effects of variance in the sample.

#### 4.3. Limb darkening

As we saw in Figure S32, the presence of limb darkening makes it significantly more difficult to infer spot properties at fixed SNR. To understand why, let us repeat the experiment from the previous section, this time with moderate quadratic limb darkening ( $u_1 = 0.5$  and  $u_2 = 0.25$ ). The top panel of Figure 17 shows the posterior shrinkage plot (same as Figure 15, but with limb darkening). Interestingly, there is no longer a clean division of the null space between even and odd modes in the limit of large  $M$ . This is because limb darkening effectively lifts odd modes out of the null space, *at the expense of the even modes*. While no coefficient lies entirely in the null space ( $R = 0$ ) when limb darkening is present, no coefficient can be uniquely inferred ( $R = 1$ ), either. This can be understood by noting that a polynomial limb darkening law can be written exactly as a linear combination of the  $m = 0$  spherical harmonics up to a degree equal to the order of the limb darkening (in this case,  $l = 2$ ; Luger et al. 2019; Agol et al. 2020, and Appendix B.2). Since the limb darkening operation is a (multiplicative) downweighting of the surface intensity, the map seen



**Figure 17.** *Top:* Same as Figure 15, but for limb-darkened stars with quadratic coefficients  $u_1 = 0.5$  and  $u_2 = 0.25$ . Odd modes can now be probed, at the expense of the even modes. *Bottom:* Same as the top panel, but allowing for a 10% variation in  $u_1$  and  $u_2$  across different observations. For  $\gtrsim 30$  light curves, there is virtually no null space up to at least  $l_{\max} = 15$ . 

by the observer is just the product of the spherical harmonic representation of the surface ( $\mathbf{y}$ ) and the spherical harmonic representation of the limb darkening profile. And since spherical harmonics are just polynomials on the surface of the sphere, the product of spherical harmonics of degree  $l_1$  and  $l_2$  is a spherical harmonic of degree  $l_1 + l_2$ . This means that the linear limb darkening component ( $l = 1$ ) effectively raises the degree of all spherical harmonic coefficients of the surface map by one. This has the effect of reversing the null space: under *only* linear limb darkening, it is the *even* modes that would like in the null space. However, the quadratic limb darkening term ( $l = 2$ ) raises the degree of all spherical harmonics by two, so its presence ensures that the even modes can still be probed to some extent. In reality, the true limb darkening profile of a stellar surface is more complicated than a two-parameter quadratic model can capture; but one may still expand it as an arbitrary order polynomial, in which case the argument still applies—limb darkening mixes the null space and the preimage in a nontrivial way. The fact that no coefficient can be determined uniquely—i.e., there are perfect degeneracies involving *all* modes on the

surface—is what makes it more difficult in practice to perform ensemble analyses on limb-darkened stars. Nevertheless, it is still possible to uniquely and precisely infer the spot properties under limb darkening, but as a rule of thumb, the stronger the limb darkening, the larger the ensemble must be (see Figure S33).

In reality, it is unlikely that all stars in a given ensemble have exactly the same limb darkening coefficients, however “similar” the stars may be. The bottom panel of Figure 17 shows the same posterior shrinkage plot, but for the case where each star has coefficients  $u_1 = 0.5 \pm 0.05$  and  $u_2 = 0.25 \pm 0.025$ ; i.e., we add a scatter of 10% in the value of these coefficients. The plot shows the posterior shrinkage in the hypothetical case where we know the coefficients for each star exactly. Now, as the size of the ensemble increases, the posterior shrinkage approaches unity *for all spherical harmonic modes*. In the same way that the null space is a strong function of the inclination, allowing us to chip away at it with observations at different inclinations, the null space is also a strong function of the limb darkening law. Even a small amount of variance in the coefficients is sufficient to constrain all surface modes exactly (in the limit of infinite SNR and large  $M$ ).

In practice, of course, we will never know the limb darkening coefficients exactly, so we must either model them or marginalize over them (see §5.3). This will dilute our ability to infer surface properties exactly.

#### 4.4. Small spots

One of the biggest downsides of adopting a spherical harmonic representation of the stellar surface as the foundation 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  $l_{\max} = 15$ , which can model features only as small as  $\sim 180^\circ/15 = 12^\circ$  on the surface. Even on scales slightly larger than this, the presence of ringing can be seen (c.f. panel (b) in Figure 6, where ringing can just barely be seen 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 wave-like 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 (§C) that make the GP covariance evaluation tractable.

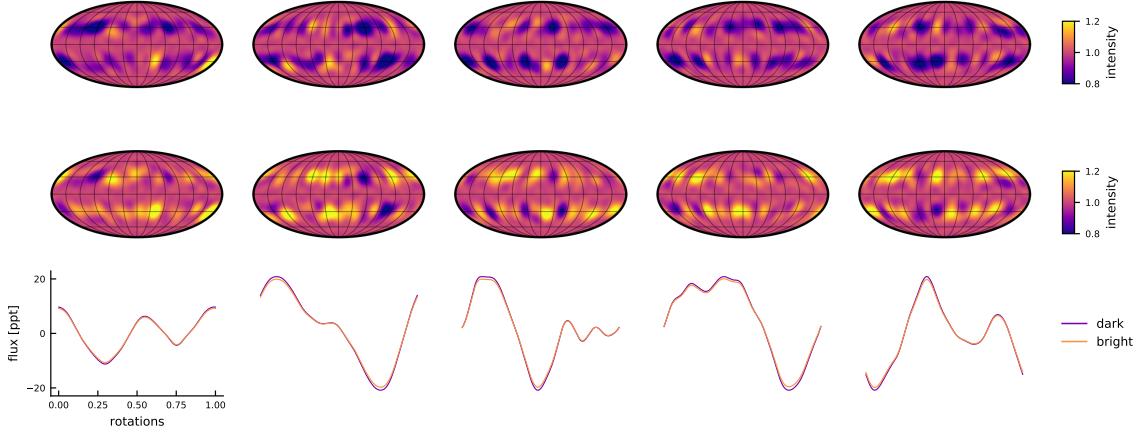
One option is to bypass the computation of the covariance on the surface of the star and to write down an expression for the flux directly in terms of the properties of a star spot. Circular spots of uniform intensity can be modeled as spherical caps, which are simply segments of ellipses when projected onto the sky. It is possible to

express in closed form the projected area they cover—and hence their contribution to the flux  $\mathbf{f}$ —even in the case where part of the spot is on the hemisphere facing away from the observer. This was done in Luger et al. (2017a), who solved this problem in closed form (see §3.3.2 and Appendix A.3 in that paper). Computing the GP is then a matter of integrating  $\mathbf{f}$  and  $\mathbf{f}\mathbf{f}^\top$ , weighted by the hyperparameter PDFs, as we did in §2.1. However, the expression for  $\mathbf{f}$  involves square roots and arctangents of functions of the spot latitude and longitude (c.f. Equation 40 in Luger et al. 2017a), so computing these integrals in closed form is likely to be very difficult. Even if a closed form solution can be found, incorporating limb darkening (which we argued can be extremely important) poses an even greater challenge. It is likely that several simplifications must be made in order to make this approach tractable. This was done to some extent in Morris (2020b), who derived a closed form expression for 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 §6, however, the algorithm presented here becomes unstable for  $l \gtrsim 15$ , so doing so would require a reparametrization of the equations in the Appendix. Importantly, however, as we showed in §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 S29, 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  $10^\circ$  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, mid-latitude spots. With the above caveats in mind, our GP can therefore still be used to model stars with small spots.

#### 4.5. Bright spots

All 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 18 show five random samples from the GP with dark spots ( $c = 0.1$ ) and bright spots ( $c = -0.1$ ), respectively; 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



**Figure 18.** Comparison of GP samples with dark spots (top) and bright spots (center) alongside their corresponding light curves viewed at  $I = 75^\circ$  (bottom). Without knowledge of the baseline (§2.3), it is very difficult to differentiate between the two from stellar light curves.

(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 relative photometry: we lack any information about the true baseline level of the light curve. Consider the dependence of the (unnormalized) GP covariance on the contrast: it enters in a single place, via Equation (C107), as  $c^2$ , meaning dark and bright spots 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 C106). However, as we argue in §2.3, the mean is not a direct observable. Instead, in relative photometry, we are only sensitive to the *ratio* of the covariance to the square of the mean (see Equation 28). 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.<sup>15</sup>

We therefore conclude that **relative single-band photometry is largely insensitive to the difference between bright spots and dark spots**. However, it is important to bear in mind that the degeneracy described above exists only for

<sup>15</sup> There is also a small additive term in Equation (28), 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.

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 spots 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 baseline level. 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. We discuss the ability of multi-band photometry to break these degeneracies in the following section.

#### 4.6. The baseline degeneracy

We showed in §3.3 that in relative photometry there exists a strong degeneracy between the contrast of individual spots  $c$  and the total number of spots  $n$  on a star. Inspection of the dependence of the GP on  $c$  and  $n$  (c.f. Appendix C) reveals that the mean and variance of the GP scale as

$$\begin{aligned}\mu &= 1 - c n g(I, \mathbf{u}, r, \mu_\phi, \sigma_\phi) \\ \sigma^2 &= c^2 n h(I, \mathbf{u}, r, \mu_\phi, \sigma_\phi),\end{aligned}\tag{37}$$

for some function  $g$  and some function  $h$  of the inclination, the limb darkening coefficients, the spot radius, and the spot latitude parameters. If our observations were sensitive to both the true mean and variance of stellar light curves, we could uniquely infer  $c$  and  $n$ . However, as we discussed in §2.3, to good approximation, relative photometry can only tell us about the *ratio*

$$\frac{\sigma^2}{\mu^2} \propto \frac{c^2 n}{(1 - c n g)^2},\tag{38}$$

where  $g$  is a complicated function of the (unknown) stellar and spot parameters; relative photometry cannot tell us about the values of  $c$  and  $n$  independently.

The reason for this is that the relationship between the amplitude of the light curve (measured in, e.g., counts per second on the detector) and the fractional intensity deficit due to spots on the surface of the star is not known. It is easy enough to normalize a light curve to its mean, median, or maximum to obtain what we typically refer to as “fractional” or “relative” units, but this is *not the same thing!* Consider the two stellar surfaces in Figure 3, whose true light curves—measured in units where an unspotted star would have unit flux—have different baselines and different amplitudes (top panel). This is not what we observe when we do photometry—instead, we observe these light curves multiplied by an unknown amplitude, which depends on

(potentially knowable) factors such as the luminosity of the star, the distance to it, and the efficiency of the detector but also on the (typically unknowable) unspotted photospheric brightness, which sets the true baseline level. Division of these light curves by their (observed) baseline levels (which is the operation typically done on stellar light curves) yields the “relative” light curves in the bottom panel, which are *indistinguishable* from each other. In this example, we are unable to distinguish between models of the surface with different spot contrasts and number of spots (as well as total spot “coverage”). In general, these quantities cannot be learned from single-band photometric measurements.

A direct consequence of the baseline degeneracy is that it is not in general possible to constrain the total spot coverage of a star from relative, single-band photometry. Assuming no spot overlap, the total spot coverage  $f_S$  is simply the (average) area of a spot times the total number of spots divided by the total area of the sphere, which may be expressed as

$$f_S = \frac{1}{2} (1 - \langle \cos r \rangle) n, \quad \checkmark \quad (39)$$

recalling that  $r$  is the *angular* radius of the spot. While  $r$  may be constrained from the covariance structure of the data (§3.3),  $n$  cannot.

The arguments above are based on the *Gaussian approximation* to the distribution of stellar light curves. As we discussed in §3.1, the true distribution has higher order moments that we do not consider here. It is possible in principle that these moments could encode information to break the  $c-N$  degeneracy, but these are in general much more difficult to constrain from the data. And even if this were the case, the fact remains that degeneracies like the polar spot effect are fundamental: recall Figure 3, in which the two stars have identical relative light curves, but very different spot coverage fractions.

Fortunately, it may be possible to break these degeneracies (at least to some extent) by collecting photometric data in multiple wavelength bands, an idea that has been explored recently (e.g., Gully-Santiago et al. 2017; Guo et al. 2018). Assuming the locations and sizes of a star’s spots are constant in wavelength, the amplitude of the light curve in different bands (and in particular its slope as a function of wavelength) can be used to directly constrain the temperature, and hence the contrast, of the spots. This effectively breaks the  $c-N$  degeneracy. In practice, the effective size of spots may be different at different wavelengths, which could complicate this picture somewhat. An extension of the formalism in this paper to multi-band photometry, and a detailed investigation of the information content of such observations, is left to future work.

## 4.7. Comparison to other work

### 4.7.1. GPs, GRFs, 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  $\boldsymbol{\theta}$ , one computes the forward model  $f(\boldsymbol{\theta})$  many times under different realizations of the noise, and adopts the sample mean and sample covariance as the mean and covariance of the Gaussian likelihood function  $p(f|\boldsymbol{\theta})$ . [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 a synthetic likelihood 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  $f$  conditioned on  $\boldsymbol{\theta}$  by sampling, we are able to actually compute the mean and covariance of  $f$  *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 cosmic microwave background (CMB). In particular, it is a type of Gaussian random field (GRF) on the sphere, which are 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.

Our method is also related to various families of approximate inference, such as variational inference (VI), in which a multivariate Gaussian is used to approximate the *posterior* distribution (e.g, [Blei et al. 2016](#)), or to approximate Bayesian computation (ABC), in which an (often intractable) likelihood function is replaced with an approximation computed from simulations from the prior (e.g., [Beaumont 2019](#)).

Finally, the methodology developed here is closely related to that in [Perger et al. \(2020\)](#), who studied the effect of different star spot configurations on the autocor-

relation and covariance of stellar radial velocity (RV) measurements. The authors of that study compared the performance of various commonly used quasi-periodic kernels when applied to synthetic RV datasets, arguing that a new four-parameter quasi-periodic 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 star spot inference (not just marginalization) problems. While the methodology presented here applies to photometry, it is possible to extend it to model RV datasets as well; we discuss this in §5.

#### 4.7.2. Spots and stellar variability

Recently, Morris (2020b) used *Kepler*, *K2*, and *TESS* light curves to derive a relationship between stellar age and spot coverage. Because of the intractability of the marginal likelihood function, that study used an approximate Bayesian computation (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, de-trended) 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. Morris (2020b) found that the spot coverage  $f_S$  as a function of age is well modeled by a simple power law, decreasing from  $\sim 10\%$  for the youngest ( $\sim 10$  Myr) stars to less than 1% for the oldest ( $\sim 5$  Gyr) stars. While this broadly agrees with the expectation that stellar magnetic activity decreases over time, our work strongly suggests that these results are predominantly prior-dominated. As we discussed at length in §4.6 (and found in our inference runs in §3.3, as well as in the toy model presented in Figure 3), the fractional spot coverage  $f_S$  is not an observable in single-band photometry. This is because the expression for  $f_S$  (Equation 39) depends on two quantities: the average spot radius  $r$ , which we showed can be constrained, and the total number of spots  $n$ , which we showed *cannot* be constrained from single-band light curves. In fact, Morris (2020b) assumed  $n = 3$  for simplicity when doing posterior inference. We therefore urge care in interpreting those results, as it is very likely this assumption does not hold across the large range of spectral types and stellar ages considered in that study. Nevertheless, the power

law scaling of the smoothed amplitude with age suggests it is still a good metric for stellar ages.

Moreover, the core idea in [Morris \(2020b\)](#) (and in related studies such as [Jackson & Jeffries 2013](#)) is very similar to that advocated here: the use of ensemble analyses to constrain population-level parameters when individual datasets are not sufficiently constraining. 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 by harnessing the covariance structure of the data when doing inference (as opposed to relying solely on the amplitude of the data).

Another recent paper relevant to our work is that of [Basri & Shah \(2020\)](#), who investigated the information content of stellar light curves, exploring what can and cannot be learned about star spot configurations from individual light curves. That paper strongly urges against the common practice of interpreting light curves with one or two dips as originating from one or two spots, respectively, a point we strongly agree with (c.f. Figure 1). It also reinforces our point about the additional degeneracies introduced by the unknown baseline inherent to single-band photometry. Finally, [Basri & Shah \(2020\)](#) present a large suite of forward models of light curves of spotted stars, which they use to discuss the (complicated) dependence of various light curve metrics on the physical spot parameters used to generate the data. They conclude that it is not possible to uniquely relate these metrics to the underlying star spot configuration. While we agree this is the case for individual stars, our work stresses that it is possible to break many of these degeneracies with ensemble analyses. [Basri & Shah \(2020\)](#) also conclude it is not in general 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 relies 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.

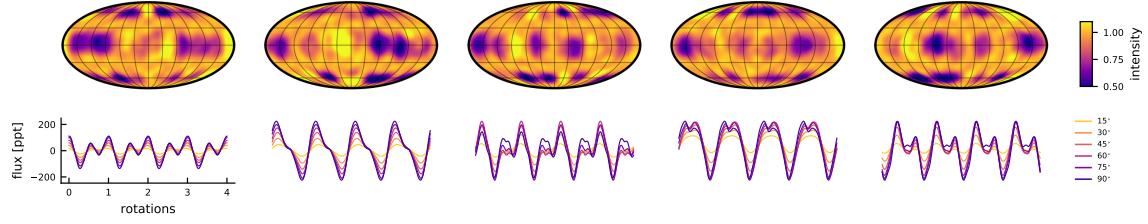
#### 4.8. *Caveats*

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

1. *The assumed latitude distribution is not Gaussian.* Because we require the first and second moment integrals (Equations C72 and C79) to have closed form solutions, there are restrictions on the probability density function 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 bi-modal Gaussian with mean  $\pm\mu_\phi$  and standard deviation  $\sigma_\phi$  (see Figure 24). 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 in general have exactly  $n$  spots (see, e.g., Figure 6). 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 §2.3. Modeling a light curve that has been normalized to its mean (or median) as a Gaussian process 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. 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 S26, 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 (§4.3). It is therefore extremely important to model it correctly; otherwise, there may be substantial bias in the inferred spot parameters.
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.



**Figure 19.** Prior samples from a sum of two `starry_process` models. The first model consists of small ( $r = 10^\circ$ ) circumpolar ( $\phi = 60^\circ \pm 3^\circ$ ) spots and the second model consists of larger ( $r = 20^\circ$ ) equatorial ( $\phi = 0^\circ \pm 3^\circ$ ) spots. The sum of two `starry_process` models is also a `starry_process`, making it easy to model more complex distributions of spots.

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 the sample selection step does not introduce bias. If there is reason to believe 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 (§C.2), which has exactly two modes at  $\pm\mu_\sigma$ . 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 Gaussian processes) 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  $\theta_{\bullet}$ . 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 19, 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 the Appendix we present a formulation of our GP that admits a radius distribution half-

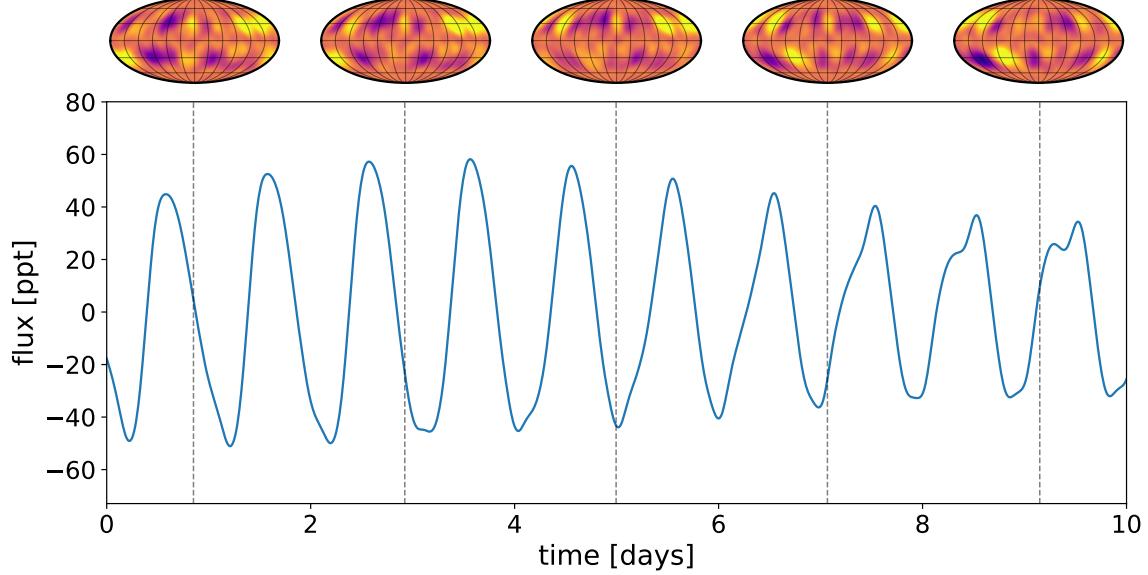
width parameter  $\Delta r$ ; this generalizes our delta function distribution to a uniform distribution between  $r - \Delta r$  and  $r + \Delta r$ . One may then compute the weighted sum of several GPs with half-widths  $\Delta \neq 0$  and central radii  $r = r_0$ ,  $r = r_0 + 2\Delta r$ ,  $r = r_0 + 4\Delta r$ , etc., 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 (§6), 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  $\theta_\bullet$  describe the *spatial* configuration of starspots, but they say nothing about their evolution in time. We know from observations of the Sun and of *Kepler* stars that temporal variability is extremely common: spots appear, disappear, and even migrate in latitude over time. While it may be possible to parametrize 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.
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.



**Figure 20.** Prior sample from a time-variable `starry_process`, assuming a rotation period  $P = 1$  day and a temporal evolution timescale  $\tau = 20$  days modeled with an exponential squared kernel. Evolving map samples are shown at the top, with the corresponding light curve viewed at an inclination  $I = 60^\circ$  at the bottom. The GP hyperparameters are set to their default values (Table 1).

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

$$\boldsymbol{\Sigma}_y^{(t)} = \mathbf{K} \otimes \boldsymbol{\Sigma}_y , \quad (40)$$

where  $\mathbf{K}$  is a  $(K \times K)$  matrix describing the covariance among the  $K$  points in time,  $\boldsymbol{\Sigma}_y$  is the  $(N \times N)$  matrix describing the covariance among the  $N \equiv (l_{\max} + 1)^2$  spherical harmonic coefficients (Equation 14), and  $\otimes$  denotes the Kronecker product. The quantity  $\boldsymbol{\Sigma}_y^{(t)}$  is the  $(NK \times NK)$  temporal-spatial covariance, whose coefficient at index  $(Nk + n, Nk' + n')$  is the covariance between the spherical harmonic coefficient  $y_n(t_k)$  and the spherical harmonic coefficient  $y_{n'}(t_{k'})$ , 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 (40) 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

$$\boldsymbol{\Sigma}^{(t)} = \mathcal{A}^\dagger \boldsymbol{\Sigma}_y^{(t)} \mathcal{A}^{\dagger \top} \quad (41)$$

where

$$\mathcal{A}^\ddagger \equiv \begin{pmatrix} \mathbf{a}_0^\top & & & \\ & \mathbf{a}_1^\top & & \\ & & \ddots & \\ & & & \mathbf{a}_{K-1}^\top \end{pmatrix} \quad (42)$$

is a design matrix constructed by staggering the rows of the standard design matrix  $\mathcal{A}$  (see Appendix B.1). The diagonal structure of  $\mathcal{A}^\ddagger$  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 (41) replaces the standard covariance  $\Sigma$ ; it can similarly be modified (§2.3) 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

$$\begin{aligned} \Sigma^{(t)} &= \mathcal{A}^\ddagger \Sigma_y^{(t)} \mathcal{A}^{\ddagger\top} \\ &= \mathcal{A}^\ddagger (\mathbf{K} \otimes \Sigma_y) \mathcal{A}^{\ddagger\top} \\ &= \Sigma \odot \mathbf{K}, \end{aligned} \quad \checkmark \quad (43)$$

that is, the flux covariance is just the elementwise  $\odot$  product of the GP covariance  $\Sigma$  and the temporal covariance  $\mathbf{K}$ . 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  $\mathbf{K}$ . 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 (Equation 3) or the Matérn-3/2 kernel (Equation 4) with variance parameter  $\sigma^2$  set to unity (since the variance is already specified within  $\Sigma$ ). In this case, the temporal covariance  $\mathbf{K}$  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.

Figure 20 shows a single prior sample from the temporal GP described above, assuming a squared exponential covariance with  $\tau = 20$  days and a rotation period  $P = 1$  day. We compute both the map samples at five different times (top) and the light curve sample over ten rotations (bottom). The flux variability is qualitatively similar to that seen in temporally-variable *Kepler* light curves.

### 5.3. Marginalizing over period and limb darkening

In §2.2 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 in general 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 pre-processing 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  $\boldsymbol{\theta}_\bullet$  (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 (B7) in the Appendix, 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 demonstrated how to perform analytically in Appendix D, entails integrating over the term  $\mathbf{R}_{\hat{x}}(-I)$ , 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  $\mathbf{R}_{\hat{z}}(\frac{2\pi}{P}t_k)$ , another Wigner matrix that rotates the star to the correct rotational phase at time  $t = t_k$ . Given the 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 (B7) 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  $u_2$  (c.f. Equation B9), 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 parametrizing 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, and given the fact that the mapping problem is particularly sensitive to the limb darkening coefficients (see §4.3), marginalizing over  $P$  and  $\mathbf{u}$  will be the subject of an upcoming paper.

#### 5.4. Modeling transits and radial velocity datasets

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  $\mathbf{d} = \mathbf{Ay}$ , then the mean and covariance of the GP model for  $\mathbf{d}$  are given by

$$\boldsymbol{\mu}_\mathbf{d} = \mathbf{A} \boldsymbol{\mu}_\mathbf{y} \quad (44)$$

$$\boldsymbol{\Sigma}_\mathbf{d} = \mathbf{A} \boldsymbol{\Sigma}_\mathbf{y} \mathbf{A}^\top, \quad (45)$$

respectively, where  $\boldsymbol{\mu}_\mathbf{y}$  and  $\boldsymbol{\Sigma}_\mathbf{y}$  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 radial velocity (RV) datasets, although this requires a bit more work. The instantaneous radial velocity shift  $v$  induced by a rotating spotted star may be approximated as

$$v = \frac{\iint_S I V \mathrm{d}S}{\iint_S I \mathrm{d}S} \quad (46)$$

where  $I$  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.<sup>16</sup> 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  $\mathbf{A}$  and  $\mathbf{B}$ , we may therefore write

$$\mathbf{v} = \frac{\mathbf{B} \mathbf{y}}{\mathbf{A} \mathbf{y}} \quad (47)$$

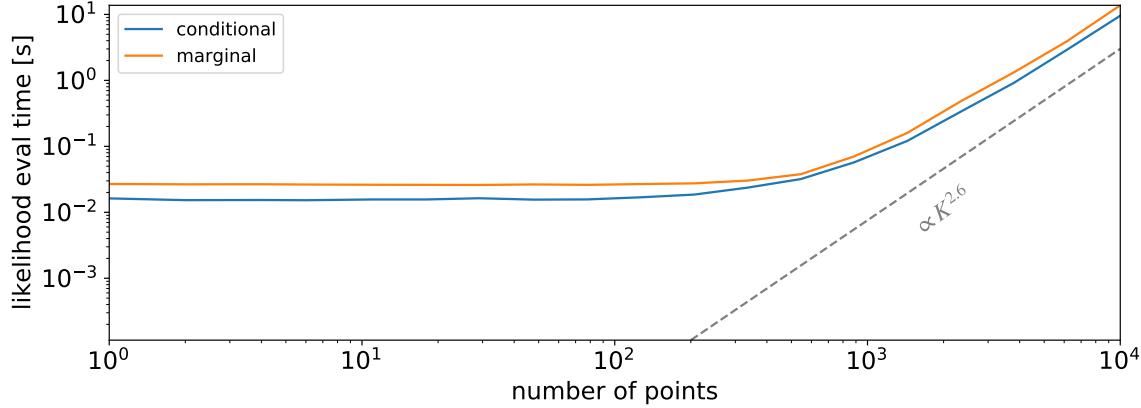
where  $\mathbf{v}$  is the random variable representing the observed radial velocity timeseries and  $\mathbf{y}$  is the Gaussian random variable representing the spherical harmonic coefficients describing the stellar surface. Since the  $\mathbf{v}$  is the ratio of two Gaussian random variables, its distribution is not Gaussian. However, we can still compute the first two moments of the distribution of  $\mathbf{v}$  to derive a Gaussian approximation to it (similar to what we did in §2.3), which will yield the mean and covariance of the Gaussian process representation of the radial velocity timeseries. Given the complexity of the operations described above, we defer this calculation (and the calibration of the resulting GP model) to future work.

## 6. IMPLEMENTATION

We implement our interpretable GP in the open-source, user-friendly **Python** package **starry\_process**, which can be installed via **pip** or from source on [GitHub](#). The code is thoroughly unit-tested and well documented, with examples on how to use the GP in custom inference problems. Users can choose, among other options, whether or not to marginalize over inclination (§2.2) and whether or not to model a normalized process (§2.3). Users can also choose the spherical harmonic degree of the expansion, although it is recommended to use  $l_{\max} = 15$  (see below). Users may compute the mean vector and covariance matrix in either the spherical harmonic basis or the flux basis, or they may sample from it or use it to compute marginal likelihoods. Arbitrary order limb darkening is implemented following [Agol et al. \(2020\)](#).

The code was designed to maximize the speed and numerical stability of the computation. Although most of the computations of the expectation integrals (see Appendix C) involve many layers of nested sums over all spherical harmonic coefficients, these may be expressed as high-dimensional tensor products, which can be evaluated efficiently on modern hardware. Many of the expressions can also be either pre-computed or computed recursively. To maximize the speed of the algorithm, the code is implemented in hybrid C++/Python using the just-in-time compilation capability of the **theano** package. Since all equations derived here have closed form expressions, these can be differentiated in a straightforward and numerically stable manner, enabling the computation of backpropagated gradients within **theano**. As

<sup>16</sup> Higher order effects such as differential rotation and convective blueshift can be easily modeled with a higher degree expansion of  $u$ .

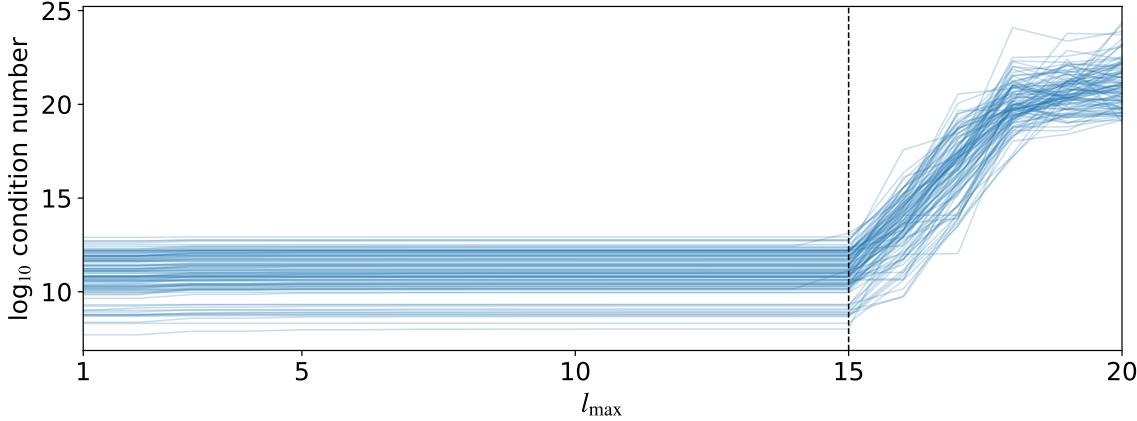


**Figure 21.** Evaluation time in seconds for a single log-likelihood evaluation as a function of the number of points  $K$  in each light curve when conditioning on a value of the inclination (blue) and when marginalizing over the inclination (orange). At  $l_{\max} = 15$ , computation of the covariance matrix of the GP takes about 20ms on a 2018 MacBook Pro. The dashed line shows the asymptotic scaling of the algorithm, which is due to the Cholesky factorization and solve operations.

such, `starry_process` is designed to work out-of-the box with `theano`-based inference tools such as `pymc3` for NUTS/HMC or ADVI sampling.

Figure 21 shows the computational scaling of the Python implementation of the algorithm for the case where we condition the GP on a specific value of the inclination (blue) and the case where we marginalize over inclination (orange). Both curves show the time in seconds to compute the likelihood (averaged over many trials) as a function of the number of points  $K$  in a single light curve. For  $K \lesssim 100$ , the computation time is constant at  $10 - 20$  ms for both algorithms. This is the approximate time (on a typical modern laptop) taken to compute the GP covariance matrix given a set of hyperparameters  $\theta_\bullet$ . For larger values of  $K$ , the cost approaches a scaling of  $K^{2.6}$ , which is dominated by the factorization of the covariance matrix and the solve operation to compute the likelihood. The likelihood marginalized over inclination is only slightly slower to compute, thanks to the tricks discussed in Appendix D.4.

Many modern GP packages (e.g., Ambikasaran et al. 2015; Foreman-Mackey et al. 2017) have significantly better asymptotic scalings, but these are usually due to specific structure imposed on the kernel functions, such as the assumption of stationarity. Our kernel structure is determined by the physics (or perhaps more accurately, the geometry) of stellar surfaces, and its nonstationarity is a consequence of the normalization step in relative photometry. Moreover, and unlike the kernels listed in §1, our kernel is a nontrivial function of the hyperparameters  $\theta_\bullet$ , so its computation is necessarily more expensive. Nevertheless, the fact that our GP may be used for likelihood evaluation in a small fraction of a second for typical datasets ( $K \sim 1,000$ ) makes it extremely useful for inference. Recall that we are implicitly marginalizing over all of the properties of *every spot* on the surface of the star.



**Figure 22.** Log of the condition number of the covariance in the spherical harmonic basis,  $\Sigma_y$ , as a function of the spherical harmonic degree of the expansion,  $l_{\max}$ . Different lines correspond to different values of  $\theta_\bullet$  drawn from a uniform prior (see text for details). In the majority of the cases, the matrix becomes ill-conditioned above  $l_{\max} = 15$ . [Cloud](#)

In ensemble analyses, we must compute the likelihood of each of the  $M$  light curves conditioned on  $\theta_\bullet$ . In practice, each star will have a different rotation period, different limb darkening coefficients, and different photometric uncertainty, meaning we must factorize  $M$  different covariance matrices. Fortunately, the spherical harmonic covariance,  $\Sigma_y$  (§2.1) need only be computed once; this can then be linearly transformed into the flux basis for each light curve. As the evaluation of  $\Sigma_y$  is the computationally-intensive step, the likelihood evaluation typically scales sub-linearly with  $M$ . Furthermore, as we discussed in §5.3, it is possible to marginalize over the period and limb darkening coefficients, which would remove the scaling with  $M$  entirely if the photometric precision were the same for all light curves. Even if it is not, the algorithm can still be greatly sped up, although an implementation of this is deferred to the next paper.

Our algorithm is also numerically stable over nearly all of the prior volume up to  $l_{\max} = 15$ . Figure 22 shows the log of the condition number of the covariance matrix in the spherical harmonic basis,  $\Sigma_y$ , as a function of the spherical harmonic degree of the expansion for 100 draws from a uniform prior over  $r \in [10^\circ, 45^\circ]$ ,  $\mu_\phi \in [0^\circ, 85^\circ]$ ,  $\sigma_\phi \in [5^\circ, 40^\circ]$ , and  $n \in [1, 50]$ . The condition number is nearly constant up to  $l_{\max} = 15$  in almost all cases; above this value, the algorithm suddenly becomes unstable and the covariance is ill-conditioned. The instability occurs within the computation of the latitude and longitude moment integrals (Appendix C) and is likely due to the large number of operations involving linear combinations of hypergeometric and gamma functions. While it may be possible to achieve stability at higher values of  $l_{\max}$  via careful reparametrization of some of those equations, we find that  $l_{\max} = 15$  is high enough for most practical purposes. We plan to revisit this point in future work.

Finally, instabilities can also occur if  $\sigma_\phi$  is too small and/or  $n$  is too large. Values of  $\sigma_\phi \lesssim 1^\circ$  lead to instabilities in the computation of the hypergeometric functions (Ap-

pendix C.2), while values of  $n \gtrsim 50$  can sometimes cause the Cholesky factorization of the covariance to fail (although this can be mitigated by adding a small quantity to the diagonal to ensure positive-semidefiniteness). In cases where the algorithm goes (very) unstable, the log-likelihood evaluation returns  $-\infty$ : in other words, they are silently rejected by an implicit prior. Fortunately, these cases are likely unphysical: in practice, there should always be some finite amount of variance in the latitudes of spots, and stars with more than 50 spots are likely too spotted for individual spots to be discernible in the first place. Instead, we are likely sensitive to *groups* of spots, which our GP is flexible enough to model.

## 7. CONCLUSIONS

We presented a new Gaussian process (GP) model for stellar variability whose hyperparameters explicitly correspond to 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 dataset consisting of  $K \sim 1,000$  points takes about 20ms on a modern laptop. Our algorithm is implemented in the open-source, user-friendly `Python` package `starry_process`, which is `pip`-installable and available on [GitHub](#). 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 backpropagated derivatives with respect to all input parameters for out-of-the-box usage with gradient-based inference and optimization tools such as Hamiltonian Monte Carlo (HMC), autodifferentiation variational inference (ADVI), and gradient-based nested sampling.

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

- 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 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. Our GP is 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  $5^\circ$ .
- 3. 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 post-processing 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.
- 4. Our GP can be used to model small, Sun-like spots.** The algorithm presented here is limited to a surface resolution of about  $10^\circ$ , corresponding to spots about an order of magnitude larger than typical sunspots. However, we have showed that when we apply our model to 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.
- 5. 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 Matérn-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.
- 6. Our GP can be used in exoplanet transit modeling and radial velocity datasets.** 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, Doppler tomography, or even radial velocity searches for exoplanets. The latter will be the subject of a future paper in this series.

In this paper we also demonstrated and discussed four points that apply more broadly to stellar light curve variability modeling in general:

1. **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  $K = 1,000$  cadence each and per-cadence precision of one part per thousand. Lower quality, shorter baseline observations, or datasets contaminated by outliers will in general 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 a thousand light curves is recommended.
2. **Limb darkening breaks degeneracies.** The presence of limb darkening changes the structure of the null space. Even small variations in the limb darkening coefficients among stars in an ensemble can effectively increase the information content of their light curves when analyzed jointly. However, this is both a blessing and a curse: it is very important to accurately model limb darkening to infer starspot properties without bias.
3. **Relative single-band photometry cannot uniquely constrain the spot contrast, the total number of spots, or the total spot coverage.** We showed that these two quantities are perfectly degenerate with each other in relative single-band photometry. Knowledge of the unspotted photospheric brightness (or temperature) can break this degeneracy, so analyses of multi-band light curves can in principle infer the total spottiness of a stellar surface.
4. **Care should be taken when analyzing normalized light curves.** We showed that the widely-adopted practice of normalizing light curves to their mean, median, or maximum changes the covariance structure of the data in potentially complex ways. In addition to changing the total variance of the data, this normalization step introduces non-stationarity in the covariance, which can lead to biases particularly when modeling light curves of stars with very long rotation periods.

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 (perhaps)

the rotation period. As such, it can be used to marginalize over stellar rotational variability signals in (say) transiting exoplanet searches, asteroseismic characterization of stars, radial velocity searches, etc. It can also be used to learn about stellar surfaces directly: to infer spot properties of main sequence stars as a function of spectral type and age, to differentiate between spot-dominated and plage-dominated stellar surfaces, to better understand chemically peculiar massive stars, and to better understand 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 (in no particular order)

- *A more rigorous treatment of time variability.* This 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.
- *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 inference for large ensembles of stellar light curves.
- *Extension of this formalism to radial velocity datasets.* As we discussed in §5.4, it is possible to extend the methodology presented here to model the contribution of stellar surface variability to radial velocity measurements, which can be used (for instance) to mitigate systematics in extreme precision radial velocity (EPRV) searches for exoplanets.
- *Extension of this formalism to Doppler imaging.* As we 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, and the Astronomical Data Group at the Center for Computational Astrophysics for many thought-provoking discussions that helped shape this paper.

## REFERENCES

- Agol, E., et al. 2020, AJ, 159, 123
- Ambikasaran, S., et al. 2015, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 252
- Angus, R., et al. 2018, MNRAS, 474, 2094
- Basri, G. 2018, ApJ, 865, 142
- Basri, G., & Shah, R. 2020, ApJ, 901, 14
- Beaumont, M. A. 2019, Annual Review of Statistics and Its Application, 6, 379. <https://doi.org/10.1146/annurev-statistics-030718-105212>
- Berdyugina, S. V. 2005, Living Reviews in Solar Physics, 2, 8
- Blei, D. M., et al. 2016, arXiv e-prints, arXiv:1601.00670
- Borucki, W. J., et al. 2010, Science, 327, 977
- Brewer, B. J., & Stello, D. 2009, MNRAS, 395, 2226
- Cantiello, M., & Braithwaite, J. 2019, ApJ, 883, 106
- Collado, J. R. A., et al. 1989, Computer Physics Communications, 52, 323
- Cowan, N. B., et al. 2013, MNRAS, 434, 2465
- Davenport, J. R. A., et al. 2015, ApJ, 806, 212
- Duane, S., et al. 1987, Physics Letters B, 195, 216
- Feroz, F., et al. 2009, MNRAS, 398, 1601
- Foreman-Mackey, D., et al. 2017, AJ, 154, 220
- Gully-Santiago, M. A., et al. 2017, ApJ, 836, 200
- Guo, Z., et al. 2018, ApJ, 868, 143
- Harmon, R. O., & Crews, L. J. 2000, AJ, 120, 3274
- Haywood, R. D., et al. 2014, MNRAS, 443, 2517
- Hoffman, M. D., & Gelman, A. 2011, arXiv e-prints, arXiv:1111.4246
- Jackson, R. J., & Jeffries, R. D. 2013, MNRAS, 431, 1883
- Kervella, P., et al. 2017, A&A, 597, A137
- Kipping, D. M. 2013, MNRAS, 435, 2152
- Kucukelbir, A., et al. 2016, arXiv e-prints, arXiv:1603.00788
- Lanza, A. F., et al. 2011, A&A, 533, A44
- Luger, R. 2020, in preparation
- Luger, R., et al. 2019, AJ, 157, 64
- . 2016, AJ, 152, 100
- . 2017a, ApJ, 851, 94
- . 2017b, Nature Astronomy, 1, 0129
- Morris, B. 2020a, The Journal of Open Source Software, 5, 2103
- Morris, B. M. 2020b, ApJ, 893, 67
- Morris, B. M., et al. 2018, ApJ, 857, 39
- . 2017, ApJ, 846, 99
- Perger, M., et al. 2020, arXiv e-prints, arXiv:2012.01862
- Price, L. F., et al. 2018, Journal of Computational and Graphical Statistics, 27, 1. <https://doi.org/10.1080/10618600.2017.1302882>
- Rackham, B. V., et al. 2018, ApJ, 853, 122
- Rajpaul, V., et al. 2015, MNRAS, 452, 2269
- Rasmussen, C. E., & Williams, C. K. I. 2005, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning) (The MIT Press)
- Ricker, G. R., et al. 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003
- Roettenbacher, R. M., et al. 2017, ApJ, 849, 120
- Russell, H. N. 1906, ApJ, 24, 1
- Schuessler, M., et al. 1996, A&A, 314, 503

- Skilling, J. 2004, in American Institute of Physics Conference Series, Vol. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. V. Toussaint, 395–405
- Skilling, J. 2006, Bayesian Anal., 1, 833.  
<https://doi.org/10.1214/06-BA127>
- Solanki, S. K., et al. 2006, Reports on Progress in Physics, 69, 563
- Speagle, J. S. 2020, MNRAS, 493, 3132
- Vogt, S. S., et al. 1987, ApJ, 321, 496
- Wandelt, B. 2012, Gaussian Random Fields in Cosmostatistics
- Wood, S. N. 2010, Nature, 466, 1102.  
<https://doi.org/10.1038/nature09319>

## APPENDIX

## A. NOTATION

Unless otherwise noted, we adopt the following conventions throughout this paper: integers are represented by italic uppercase letters (i.e.,  $K$ ), scalars are represented by italic lowercase letters (i.e.,  $x$ ), column vectors are represented by boldface lowercase letters ( $\mathbf{x}$ ), and matrices are represented by boldface capital letters ( $\mathbf{X}$ ). In general, the elements of a vector  $\mathbf{x}$  are denoted  $x_i$  and the elements of a matrix  $\mathbf{X}$  are denoted  $X_{i,j}$ . Importantly, we make a distinction between quantities like  $X_{i,j}$  and  $\mathbf{X}_{i,j}$ : 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, 4-dimensional) linear operator. Thus, lowercase bold symbols *always* represent vectors, and uppercase 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  $\mathbf{x}$  denotes a scalar random variable, while  $x$  denotes a particular realization of that variable. The same applies to vector-valued ( $\mathbf{x}$  is a realization of  $\mathbf{x}$ ) and matrix-valued random variables ( $\mathbf{X}$  is a realization of  $\mathbf{X}$ ).

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,  $x_m^l$ , and *four* indices to represent a scalar matrix element,  $X_{m,m'}^{l,l'}$ . The upper indices corresponds to the spherical harmonic degree,  $l \in [0, l_{\max}]$  and  $l' \in [0, l_{\max}]$ , 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  $\mathbf{x}$  representing a quantity in the spherical harmonic basis up to degree  $l_{\max}$  has components given by

$$\mathbf{x} = (x_0^0 \quad x_{-1}^1 \quad x_0^1 \quad x_1^1 \quad \cdots \quad x_{-l_{\max}}^{l_{\max}} \cdots \quad x_{l_{\max}}^{l_{\max}})^{\top}, \quad (\text{A1})$$

while a matrix  $\mathbf{X}$  in the same basis has components given by

$$\mathbf{X} = \begin{pmatrix} X_{0,0}^{0,0} & X_{0,-1}^{0,1} & X_{0,0}^{0,1} & X_{0,1}^{0,1} \\ X_{-1,0}^{1,0} & X_{-1,-1}^{1,1} & X_{-1,0}^{1,1} & X_{-1,1}^{1,1} \\ X_{0,0}^{1,0} & X_{0,-1}^{1,1} & X_{0,0}^{1,1} & X_{0,1}^{1,1} \\ X_{1,0}^{1,0} & X_{1,-1}^{1,1} & X_{1,0}^{1,1} & X_{1,1}^{1,1} \\ & & & \ddots \end{pmatrix}. \quad (\text{A2})$$

**Table 2.** List of common variables and symbols used throughout this paper.

Symbol	Description	Reference
$\mathbf{1}$	Vector of ones	—
$\sim$	Denotes a normalized vector-valued random variable	§2.3
$\odot$	Elementwise product	§5.2
$\otimes$	Kronecker product	§5.2
$a$	GP hyperparameter: spot latitude shape parameter	Appendix C.2
$\mathbf{a}^\top$	Row of the <i>starry</i> design matrix	Equation (B6)
$\mathbf{A}_1$	<i>starry</i> change of basis matrix	Appendix B
$\mathcal{A}$	<i>starry</i> design matrix	Equation (B6)
$\alpha$	GP hyperparameter: spot latitude shape parameter	Appendix C.2
$b$	GP hyperparameter: spot latitude shape parameter	Appendix C.2
$\beta$	GP hyperparameter: spot latitude shape parameter	Appendix C.2
$c$	GP hyperparameter: spot contrast	Appendix C.4
$\mathbf{c}$	Spot contrast (random variable)	Appendix C.4
$\Gamma(\dots)$	Gamma function	—
$\mathbf{D}_{\mathbf{u}}$	Complex Wigner rotation matrix about an axis $\mathbf{u}$	Appendix C.2.1
$\delta(\dots)$	Delta function	—
$\delta_{ij}$	Kronecker delta	—
$\Delta r$	GP hyperparameter: spot radius spread	Appendix C.1
$E[\dots]$	Expected value	Equation (9)
$\mathbf{e}_I$	First moment integral of the inclination	Equation (24)
$\mathbf{e}_r$	First moment integral of the radius	Equation (C20)
$\mathbf{e}_\phi$	First moment integral of the latitude	Equation (C21)
$\mathbf{e}_\lambda$	First moment integral of the longitude	Equation (C22)
$\mathbf{e}_c$	First moment integral of the contrast	Equation (C23)
$\mathbf{E}_I$	Second moment integral of the inclination	Equation (25)
$\mathbf{E}_r$	Second moment integral of the radius	Equation (C24)
$\mathbf{E}_\phi$	Second moment integral of the latitude	Equation (C25)
$\mathbf{E}_\lambda$	Second moment integral of the longitude	Equation (C26)
$\mathbf{E}_c$	Second moment integral of the contrast	Equation (C27)
$\mathbf{f}$	Flux vector	Equation (7)
$\mathbf{f}$	Flux vector (random variable)	Equation (1)
${}_2F_1(\dots)$	Gauss hypergeometric function	—
$\boldsymbol{\theta}_\bullet$	Vector of GP hyperparameters	Equation (17)
$I$	Stellar inclination	—
$\mathbb{I}$	Stellar inclination (random variable)	§2.2
$J$	Jacobian of the spot latitude transform	Equation (C71)
$k(\Delta t)$	GP kernel function	Equation (2)
$K$	Number of points in light curve	—
$\mathbf{K}$	Temporal covariance matrix	§5.2
$l$	Spherical harmonic degree	Appendix A
$\mathbf{L}$	Limb darkening operator	§B.2
$\mathcal{L}$	Likelihood function	Equation (18)

**Table 2.** (continued from previous page)

Symbol	Definition	Reference
$\lambda$	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 (26)
$\mu$	Flux GP mean vector	Equation (11)
$\mu_y$	Spherical harmonic GP mean vector	Equation (13)
$n$	GP hyperparameter: number of spots	Appendix C
$n$	Number of spots contrast (random variable)	Appendix C
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution: mean $\mu$ , variance $\sigma^2$	—
$p(\dots)$	Probability, probability density	—
$P$	Stellar rotation period	—
$r$	GP hyperparameter: spot radius	Appendix C.1
$\mathbf{r}^\top$	Integral over unit disk of polynomial basis	Appendix B
$r$	Spot radius (random variable)	Appendix C.1
$R$	Posterior shrinkage	Equation (36)
$\mathbf{R}_u$	Real wigner rotation matrix about an axis $\mathbf{u}$	Appendix C.2.1
$\mathbf{s}$	Spherical harmonic expansion of spot	Equation (C33)
$\sigma_f$	Photometric uncertainty	§3
$\sigma_\phi$	GP hyperparameter: spot latitude standard deviation	Appendix C.2
$\Sigma$	Flux GP covariance matrix	Equation (12)
$\Sigma^{(t)}$	Flux GP covariance matrix w/ temporal evolution	Equation (12)
$\tilde{\Sigma}$	Flux GP covariance matrix (normalized processs)	Equation (28)
$\Sigma_y$	Spherical harmonic GP covariance matrix	Equation (14)
$\Sigma_y^{(t)}$	Spherical harmonic GP covariance w/ temporal evolution	Equation (14)
$t$	time	—
$\tau$	GP hyperparameter: timescale	§5.2
$\mathbf{u}$	Limb darkening coefficient vector	Appendix B.2
$u_1, u_2$	Linear and quadratic limb darkening coefficients	Appendix B.2
$\mathbf{U}$	Complex-to-real basis change operator (Wigner matrices)	Appendix C.2.1
$\mathcal{U}(a, b)$	Uniform distribution between $a$ and $b$	—
$\mathbf{x}$	Random vector of spot properties	Equation (C13)
$\mathbf{y}$	Spherical harmonic coefficient vector	Equation (7)
$\mathbf{y}$	Spherical harmonic coefficient vector (random variable)	Equation (C14)
$\mu_\phi$	GP hyperparameter: spot latitude mode	Appendix C.2
$\phi$	Spot latitude (random variable)	Appendix C.2
$z$	GP normalization number	Equation (29)

For completeness, the element of a spherical harmonic vector  $\mathbf{x}$  with degree  $l$  and order  $m$  is at (flattened) index

$$n = l^2 + l + m. \quad (\text{A3})$$

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

$$\begin{aligned} l &= \lfloor \sqrt{n} \rfloor \\ m &= n - l^2 - l, \end{aligned} \quad (\text{A4})$$

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 should also 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 section in which they are presented.

## B. COMPUTING THE FLUX

### B.1. Basic expression

As we mentioned in §2.1, the flux  $\mathbf{f}$  is a purely linear function of the spherical harmonic coefficient vector  $\mathbf{y}$ :

$$\mathbf{f} = \mathbf{1} + \mathbf{A} \mathbf{y}. \quad (\text{B5})$$

Even though this is derived in detail in Luger et al. (2019), it is useful to expand on the computation of the design matrix  $\mathbf{A}$  in more detail here. Let  $\mathbf{a}_k^\top$  denote the  $k^{\text{th}}$  row of  $\mathbf{A}$ , such that

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_0^\top \\ \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_{K-1}^\top \end{pmatrix}. \quad (\text{B6})$$

The row vector  $\mathbf{a}_k^\top$  encodes how the spherical harmonic coefficient vector projects onto the  $k^{\text{th}}$  cadence in the flux timeseries, and may be computed from

$$\mathbf{a}_k^\top = \mathbf{r}^\top \mathbf{A}_1 \mathbf{R}_{\hat{\mathbf{x}}}(-I) \mathbf{R}_{\hat{\mathbf{z}}} \left( \frac{2\pi}{P} t_k \right) \mathbf{R}_{\hat{\mathbf{x}}} \left( \frac{\pi}{2} \right). \quad (\text{B7})$$

To understand the expression above, let us proceed from right to left, starting with the spherical harmonic vector  $\mathbf{y}$ , which we assume 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. The quantity  $\mathbf{R}_{\hat{\mathbf{x}}}$  is a Wigner rotation matrix (described in detail in §C.2.1), which in this case rotates the spherical harmonic representation of the star by an angle  $\pi/2$  counter-clockwise 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,  $\mathbf{R}_{\hat{\mathbf{z}}}$ , to rotate the star about  $\hat{\mathbf{z}}$  counter-clockwise (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  $I$  about  $\hat{\mathbf{x}}$ , where  $I$  is the stellar inclination ( $I = 0$  corresponding to a pole-on view and  $I = \pi/2$  corresponding to an edge-on view). With this last rotation, we are now in the observer's frame.<sup>17</sup>

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  $\mathbf{A}_1$  (c.f. 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  $(1 \; x \; z \; y \; x^2 \; xz \; xy \; yz \; y^2 \cdots)$  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), which we dot into our expression (and add one) to obtain the flux at the  $k^{\text{th}}$  cadence.

## B.2. *With limb darkening*

We must adjust our expression for the flux in the presence of limb darkening. For any polynomial limb darkening law of the form

$$\frac{I(\mu)}{I(\mu=1)} = 1 - \sum_{n=1}^{n_{\max}} u_n (1 - \mu)^n, \quad (\text{B8})$$

where  $I$  is the intensity on the stellar surface,  $\mu = z = \sqrt{1 - x^2 - y^2}$  is the radial coordinate on the projected disk, and  $u_n$  is a limb darkening coefficient, the effect of limb darkening on the stellar map can be expressed exactly as a linear operation on the spherical harmonic coefficient vector (Luger et al. 2019). This includes the popular linear and quadratic limb darkening laws and generalizes to *any* limb darkening law in the limit  $n_{\max} \rightarrow \infty$ . The linearity of the problem can be understood by noting that all terms in Equation (B8) are strictly polynomials in  $x$ ,  $y$ , and  $z$ , all of which can be expressed exactly as sums of spherical harmonics (Luger et al. 2019). When weighting the surface intensity by the limb darkening profile, the resulting intensity is simply a product of spherical harmonics, which is itself a linear combination of

<sup>17</sup> 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.

spherical harmonics. Thus, given a limb darkening law of degree  $n_{\max}$  with coefficients  $\mathbf{u}$ , we can construct a matrix  $\mathbf{L}(\mathbf{u})$  that transforms a spherical harmonic vector  $\mathbf{y}$  of degree  $l_{\max}$  to a limb-darkened spherical harmonic vector  $\mathbf{y}'$  of degree  $l_{\max} + n_{\max}$ . As an example, consider a map of degree  $l_{\max} = 1$  and the linear limb darkening law ( $n_{\max} = 1$ ) with coefficient vector  $\mathbf{u} = (u_1)$ . The transformation matrix from  $\mathbf{y}$  to  $\mathbf{y}'$  is

$$\mathbf{L}(\mathbf{u}) = \frac{1}{1 - \frac{u_1}{3}} \begin{pmatrix} 1 - u_1 & 0 & \frac{u_1}{\sqrt{3}} & 0 \\ 0 & 1 - u_1 & 0 & 0 \\ \frac{u_1}{\sqrt{3}} & 0 & 1 - u_1 & 0 \\ 0 & 0 & 0 & 1 - u_1 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{u_1}{\sqrt{5}} & 0 & 0 \\ 0 & 0 & \frac{2u_1}{\sqrt{15}} & 0 \\ 0 & 0 & 0 & \frac{u_1}{\sqrt{5}} \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad \checkmark \text{ (B9)}$$

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  $k^{\text{th}}$  row of the flux design matrix  $\mathcal{A}$  (Equation B7) with

$$\mathbf{a}_k^\top = \mathbf{r}^\top \mathbf{A}_1 \mathbf{L}(\mathbf{u}) \mathbf{R}_{\hat{x}}(-I) \mathbf{R}_{\hat{z}}\left(\frac{2\pi}{P}t_k\right) \mathbf{R}_{\hat{x}}\left(\frac{\pi}{2}\right) \quad (\text{B10})$$

for a given limb darkening coefficient vector  $\mathbf{u}$ .

### 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  $\mathbf{y}$ ,

$$\mathbb{E}[\mathbf{y} | \boldsymbol{\theta}_\bullet] = \int \mathbf{y}(\mathbf{x}) p(\mathbf{x} | \boldsymbol{\theta}_\bullet) d\mathbf{x} \quad (\text{C11})$$

$$\mathbb{E}[\mathbf{y}\mathbf{y}^\top | \boldsymbol{\theta}_\bullet] = \int \mathbf{y}(\mathbf{x})\mathbf{y}^\top(\mathbf{x}) p(\mathbf{x} | \boldsymbol{\theta}_\bullet) d\mathbf{x}, \quad (\text{C12})$$

which are linearly related to the mean and the covariance of our GP (§2.1). Recall that  $\mathbf{x}$  is a vector of parameters describing the exact configuration of features on the surface of a star, and  $p(\mathbf{x} | \boldsymbol{\theta}_\bullet)$  is its probability density function conditioned on hyperparameters  $\boldsymbol{\theta}_\bullet$ , 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

$$\mathbf{x} = (\mathbf{n} \ c_0 \ \cdots \ c_{n-1} \ \lambda_0 \ \cdots \ \lambda_{n-1} \ \phi_0 \ \cdots \ \phi_{n-1} \ r_0 \ \cdots \ r_{n-1})^\top \quad (\text{C13})$$

and

$$\mathbf{y}(\mathbf{x}) = - \sum_{i=0}^{n-1} c_i \mathbf{R}_{\hat{\mathbf{y}}}(\lambda_i) \mathbf{R}_{\hat{\mathbf{x}}}(\phi_i) \mathbf{s}(r_i), \quad (\text{C14})$$

where  $\mathbf{n}$  is the total number of spots,  $c_i$  is the contrast of the  $i^{\text{th}}$  spot,  $\lambda_i$  is its longitude,  $\phi_i$  is its latitude, and  $r_i$  is its radius. The vector function  $\mathbf{s}(r_i)$  returns the spherical harmonic expansion of a negative unit brightness circular spot of radius  $r_i$  at  $\lambda = \phi = 0$ ,  $\mathbf{R}_{\hat{\mathbf{x}}}(\phi_i)$  is the Wigner matrix that rotates the expansion about  $\hat{\mathbf{x}}$  such that the spot is centered at a latitude  $\phi_i$ , and  $\mathbf{R}_{\hat{\mathbf{y}}}(\lambda_i)$  is the Wigner matrix that then rotates the expansion about  $\hat{\mathbf{y}}$  such that the spot is centered at a longitude  $\lambda_i$ ; these three functions are detailed in the sections below. Equation (C14) 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  $\mathbf{y}$  relative to a baseline of zero: i.e., a star with no spots on it will have  $\mathbf{y} = \mathbf{0}$ . Second, and more importantly, we are not interested in any specific value of  $\mathbf{y}$ ; rather, we would like to know its expectation value under the probability distribution governing the different spot properties  $\mathbf{x}$ , i.e.,  $p(\mathbf{x} | \boldsymbol{\theta}_\bullet)$ .

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

$$p(\mathbf{n} | \boldsymbol{\theta}_\bullet) = \delta(\mathbf{n} - n), \quad (\text{C15})$$

where  $\delta$  is the delta function.<sup>18</sup> We further assume that  $p(\mathbf{x} | \boldsymbol{\theta}_\bullet)$  is separable in each of the four other spot properties, and that all of the spots are drawn from the same distribution:

$$p(\mathbf{x} | \boldsymbol{\theta}_\bullet) = \prod_{i=0}^{n-1} p(c_i | \boldsymbol{\theta}_c) p(\lambda_i | \boldsymbol{\theta}_\lambda) p(\phi_i | \boldsymbol{\theta}_\phi) p(r_i | \boldsymbol{\theta}_r), \quad (\text{C16})$$

where

$$\boldsymbol{\theta}_\bullet = (n \ \boldsymbol{\theta}_c \ \boldsymbol{\theta}_\lambda \ \boldsymbol{\theta}_\phi \ \boldsymbol{\theta}_r)^\top \quad (\text{C17})$$

is the vector of hyperparameters describing the process and  $\boldsymbol{\theta}_c$ ,  $\boldsymbol{\theta}_\lambda$ ,  $\boldsymbol{\theta}_\phi$ , and  $\boldsymbol{\theta}_r$  are yet to be specified. This allows us to rewrite the expectation integrals (C11) and (C12)

<sup>18</sup> 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  $\mathbf{n}$ . However, as we argue in the text,  $\mathbf{n}$  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 in any of the other hyperparameters.

as

$$\mathbb{E}[\mathbf{y} \mid \boldsymbol{\theta}_\bullet] = n \mathbf{e}_c \quad (\text{C18})$$

$$\mathbb{E}[\mathbf{y} \mathbf{y}^\top \mid \boldsymbol{\theta}_\bullet] = n \mathbf{E}_c \quad (\text{C19})$$

where we define the first moment integrals

$$\mathbf{e}_r \equiv \int \mathbf{s}(r) p(r \mid \boldsymbol{\theta}_r) dr \quad (\text{C20})$$

$$\mathbf{e}_\phi \equiv \int \mathbf{R}_{\hat{x}}(\phi) \mathbf{e}_r p(\phi \mid \boldsymbol{\theta}_\phi) d\phi \quad (\text{C21})$$

$$\mathbf{e}_\lambda \equiv \int \mathbf{R}_{\hat{y}}(\lambda) \mathbf{e}_\phi p(\lambda \mid \boldsymbol{\theta}_\lambda) d\lambda \quad (\text{C22})$$

$$\mathbf{e}_c \equiv - \int c \mathbf{e}_\lambda p(c \mid \boldsymbol{\theta}_c) dc \quad (\text{C23})$$

and the second moment integrals

$$\mathbf{E}_r \equiv \int \mathbf{s}(r) \mathbf{s}^\top(r) p(r \mid \boldsymbol{\theta}_r) dr \quad (\text{C24})$$

$$\mathbf{E}_\phi \equiv \int \mathbf{R}_{\hat{x}}(\phi) \mathbf{E}_r \mathbf{R}_{\hat{x}}^\top(\phi) p(\phi \mid \boldsymbol{\theta}_\phi) d\phi \quad (\text{C25})$$

$$\mathbf{E}_\lambda \equiv \int \mathbf{R}_{\hat{y}}(\lambda) \mathbf{E}_\phi \mathbf{R}_{\hat{y}}^\top(\lambda) p(\lambda \mid \boldsymbol{\theta}_\lambda) d\phi \quad (\text{C26})$$

$$\mathbf{E}_c \equiv \int c^2 \mathbf{E}_\lambda p(c \mid \boldsymbol{\theta}_c) dc. \quad (\text{C27})$$

In Equations (C18) and (C19), 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 ( $\mathbf{e}_r$ ,  $\mathbf{E}_r$ ) under a suitable spherical harmonic expansion  $\mathbf{s}(r)$  of the spot profile and a suitable probability distribution function for the spot radius,  $p(r \mid \boldsymbol{\theta}_r)$ .

### C.1.1. Spot profile

We model the brightness  $b$  an angle  $\vartheta$  away from the center of a spot of negative unit intensity and radius  $r$  as

$$b(r; \vartheta) = \frac{1}{1 + \exp\left(\frac{r - \vartheta}{s}\right)} - 1 \quad (\text{C28})$$

for some (constant) shape parameter  $s$ . In the limit  $s \rightarrow 0$ ,  $b$  approaches an inverted top-hat function with half-width equal to  $r$ , corresponding to a circular spot of uniform intensity. For  $s > 0$ , each half of  $b$  is a sigmoid with half-width at half-minimum equal to  $r$ . In our implementation of the algorithm we choose  $s = 0.2^\circ$ , which is small compared to features of interest but not too 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  $\hat{\mathbf{z}}$  (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  $\vartheta$  of  $K$  equally-spaced points between 0 and  $\pi$ , with coefficients given by

$$\vartheta_k = \frac{k\pi}{K-1} . \quad (\text{C29})$$

We wish to model the brightness evaluated at each  $\vartheta_k$  as a weighted combination of Legendre polynomials,

$$\mathbf{B} \mathbf{s}_0(r) = \mathbf{b}(r) \quad (\text{C30})$$

where  $\mathbf{b}(r)$  is computed by evaluating Equation (C28) at each of the  $\vartheta_k$ ,  $\mathbf{B}$  is a design matrix whose columns are the weighted Legendre polynomials,

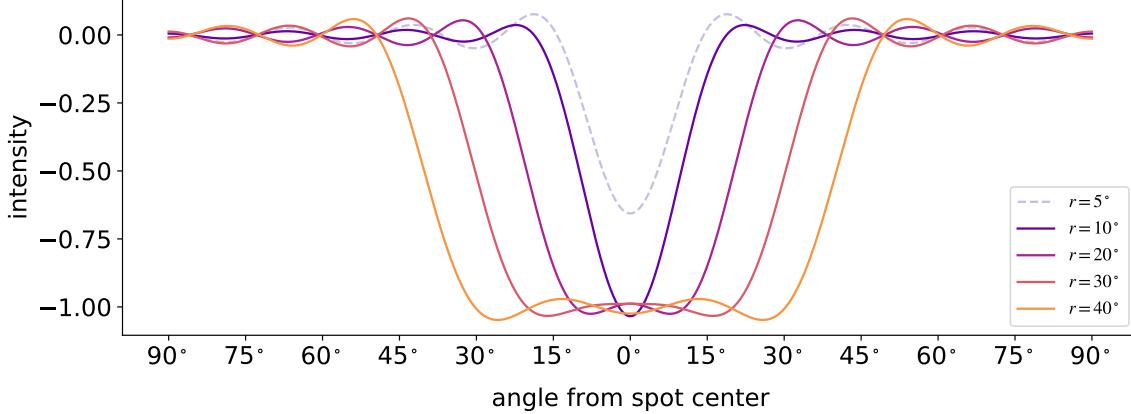
$$B_{k,l} = \sqrt{2l+1} P_l(\cos \vartheta_k) , \quad (\text{C31})$$

and  $\mathbf{s}_0(r)$  are the coefficients of the expansion. These are related to the full vector of spherical harmonic coefficients describing the spot,  $\mathbf{s}(r)$ , by

$$s_m^l(r) = s_0^l \delta_{m,0} , \quad (\text{C32})$$

or, in vector form,

$$\mathbf{s}(r) = \mathbf{I} \mathbf{s}_0(r) \quad (\text{C33})$$



**Figure 23.** Intensity profiles for spots with different radii  $r$  computed at spherical harmonic degree  $l_{\max} = 15$ . For  $r \gtrsim 10^\circ$ , the spherical harmonic expansion captures the spot shape and intensity reasonably well, albeit with some ringing due to the truncated expansion. [Download](#)

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

$$\mathcal{I}_{n,l} = \delta_{n,l^2+l}, \quad (\text{C34})$$

and  $\delta$  is the Kronecker delta function. To find the coefficients  $\mathbf{s}_0(r)$  (and hence  $\mathbf{s}(r)$ ), we solve the (linear) inverse problem,

$$\mathbf{s}_0(r) = \mathbf{B}^+ \mathbf{b}(r) \quad (\text{C35})$$

where

$$\mathbf{B}^+ = \mathbf{S} \left( \mathbf{B}^\top \mathbf{B} + \epsilon \mathbf{I} \right)^{-1} \mathbf{B}^\top \quad (\text{C36})$$

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

$$S_{k,l} = \exp \left[ -\frac{l(l+1)}{2\xi^2} \right] \delta_{k,l} \quad (\text{C37})$$

for smoothing strength  $\xi$ . For  $\epsilon \rightarrow 0$  and  $\xi \rightarrow \infty$ ,  $\mathbf{B}^+$  is the exact pseudo-inverse of  $\mathbf{B}$ . 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 23 shows the intensity profile for spots of different radii expanded to spherical harmonic degree  $l_{\max} = 15$ . The average intensity within the spots is close to  $-1$  and the half-widths at half-minimum are equal to the spot radii, 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  $r \lesssim 10^\circ$ , an expansion to  $l_{\max} = 15$  is insufficient

to correctly model the spot, as can be seen from the  $r = 5^\circ$  profile (dashed curve). Expansions to higher spherical harmonic degree allow one to model spots with radii smaller than  $10^\circ$ , although at increased computational cost and potential numerical stability issues; we discuss this point at length in §4.4.

### 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 a half-width  $\Delta r$ :

$$p(r | \boldsymbol{\theta}_r) = \begin{cases} \frac{1}{2\Delta r} & r - \Delta r \leq r \leq r + \Delta r \\ 0 & \text{otherwise,} \end{cases} \quad (\text{C38})$$

where the hyperparameters of the distribution are

$$\boldsymbol{\theta}_r = (r \ \Delta r)^\top. \quad (\text{C39})$$

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  $\Delta r \rightarrow 0$ , in which case the PDF becomes

$$p(r | \boldsymbol{\theta}_r, \Delta r = 0) = \delta(r - r), \quad (\text{C40})$$

where  $\delta$  is the delta function.

### C.1.3. First moment

The first moment of the radius distribution is (Equation C20)

$$\begin{aligned} \mathbf{e}_r &\equiv \int \mathbf{s}(r) p(r | \boldsymbol{\theta}_r) dr \\ &= \frac{1}{2\Delta r} \int_{r_0 - \Delta r}^{r_0 + \Delta r} \mathbf{s}(r) dr. \end{aligned} \quad (\text{C41})$$

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

$$\begin{aligned} (e_r)_m^l &= \frac{\delta_{m,0}}{2\Delta r} \int_{r-\Delta r}^{r+\Delta r} \tilde{s}_l(r) dr \\ &= \frac{\delta_{m,0}}{2\Delta r} \sum_{k=0}^{K-1} B_{l,k}^+ \int_{r-\Delta r}^{r+\Delta r} b_k(r) dr \\ &= \delta_{m,0} \sum_{k=0}^{K-1} B_{l,k}^+ c_k(r, \Delta r) \end{aligned} \quad \checkmark \quad (\text{C42})$$

where

$$c_k(r, \Delta r) = \frac{s}{2\Delta r} \ln \left( \frac{1 + \chi_k^-(r, \Delta r)}{1 + \chi_k^+(r, \Delta r)} \right) \quad \checkmark \quad (\text{C43})$$

and

$$\begin{aligned}\chi_k^-(r, \Delta r) &\equiv \exp\left(\frac{r - \Delta r - \vartheta_k}{s}\right) \\ \chi_k^+(r, \Delta r) &\equiv \exp\left(\frac{r + \Delta r - \vartheta_k}{s}\right).\end{aligned}\quad \checkmark \text{ (C44)}$$

In vector form, we may simply write

$$\mathbf{e}_r = \mathcal{I} \mathbf{B}^+ \mathbf{c}(r, \Delta r). \quad \checkmark \text{ (C45)}$$

Note, finally, that in the limit  $\Delta r \rightarrow 0$ ,

$$\lim_{\Delta r \rightarrow 0} \mathbf{e}_r = \mathcal{I} \mathbf{B}^+ \mathbf{b}(r). \quad \checkmark \text{ (C46)}$$

#### C.1.4. Second moment

The second moment of the radius distribution is (Equation C24)

$$\begin{aligned}\mathbf{E}_r &\equiv \int \mathbf{s}(\mathbf{r}) \mathbf{s}^\top(\mathbf{r}) p(\mathbf{r} \mid \boldsymbol{\theta}_r) d\mathbf{r} \\ &= \frac{1}{2\Delta r} \int_{r-\Delta r}^{r+\Delta r} \mathbf{s}(\mathbf{r}) \mathbf{s}^\top(\mathbf{r}) d\mathbf{r}.\end{aligned}\quad \text{(C47)}$$

As before, its components may be written

$$\begin{aligned}(E_r)_{m,m'}^{l,l'} &= \frac{\delta_{m,0}\delta_{m',0}}{2\Delta r} \int_{r-\Delta r}^{r+\Delta r} \tilde{s}_m^l(\mathbf{r}) \tilde{s}_{m'}^{l'}(\mathbf{r}) d\mathbf{r} \\ &= \frac{\delta_{m,0}\delta_{m',0}}{2\Delta r} \sum_{k=0}^{K-1} B_{l,k}^+ \sum_{k'=0}^{K-1} B_{l',k'}^+ \int_{r-\Delta r}^{r+\Delta r} b_k(\mathbf{r}) b_{k'}(\mathbf{r}) d\mathbf{r} \\ &= \delta_{m,0}\delta_{m',0} \sum_{k=0}^{K-1} B_{l,k}^+ \sum_{k'=0}^{K-1} B_{l',k'}^+ C_{k,k'}(r, \Delta r),\end{aligned}\quad \checkmark \text{ (C48)}$$

where

$$C_{k,k'}(r, \Delta r) \equiv \frac{s}{2\Delta r} \begin{cases} \frac{\exp\left(\frac{\vartheta_k - \vartheta_{k'}}{s}\right) \ln\left(\frac{1+\chi_k^-}{1+\chi_k^+}\right) - \ln\left(\frac{1+\chi_{k'}^-}{1+\chi_{k'}^+}\right)}{1 - \exp\left(\frac{\vartheta_k - \vartheta_{k'}}{s}\right)} & k \neq k' \\ \frac{1}{1+\chi_k^+} + \frac{\chi_k^-}{1+\chi_k^-} - \ln\left(\frac{1+\chi_k^-}{1+\chi_k^+}\right) - 1 & k = k'. \end{cases}\quad \checkmark \text{ (C49)}$$

In vector form, this may be written

$$\mathbf{E}_r = \mathcal{I} \mathbf{B}^+ \mathbf{C}(r, \Delta r) \mathbf{B}^{+\top} \mathcal{I}^\top. \quad \checkmark \text{ (C50)}$$

Finally, in the limit  $\Delta r \rightarrow 0$ ,

$$\lim_{\Delta r \rightarrow 0} \mathbf{E}_r = \mathcal{I} \mathbf{B}^+ \mathbf{b}(r) \mathbf{b}^\top(r) \mathbf{B}^{+\top} \mathcal{I}^\top. \quad \checkmark \text{ (C51)}$$

### C.2. The Latitude Integrals

Our goal in this section is to compute the first and second moments of the latitude distribution ( $\mathbf{e}_\phi$  and  $\mathbf{E}_\phi$ , given by Equations C21 and C25, respectively). These involve integrals over the terms in the Wigner rotation matrix for spherical harmonics, which we discuss below.

#### C.2.1. Rotation matrices

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

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}^0 & & & \\ & \mathbf{R}^1 & & \\ & & \mathbf{R}^2 & \\ & & & \ddots \\ & & & & \mathbf{R}^{l_{\max}} \end{pmatrix}, \quad (\text{C52})$$

where

$$\mathbf{R}^l = \mathbf{U}^{l-1} \mathbf{D}^l \mathbf{U}^l \quad (\text{C53})$$

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

$$\mathbf{U}^l = \frac{1}{\sqrt{2}} \begin{pmatrix} \cdots & & & & & & \cdots \\ & i & & & & & 1 \\ & & i & & & & 1 \\ & & & i & & 1 & \\ & & & & \sqrt{2} & & \\ & & & i & & -1 & \\ & & -i & & & 1 & \\ & i & & & & & -1 \\ \cdots & & & & & & \cdots \end{pmatrix} \quad (\text{C54})$$

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

$$D_{m,m'}^l(\alpha, \beta, \gamma) = \exp(-i m' \alpha) d_{m,m'}^l(\beta) \exp(-i m \gamma) \quad (\text{C55})$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  are the Euler angles describing the rotation in the  $\hat{\mathbf{z}}-\hat{\mathbf{y}}-\hat{\mathbf{z}}$  convention and  $i$  is the imaginary unit (Collado et al. 1989). The terms of the  $d$ -matrix depend on powers of  $\sin(\beta/2)$  and  $\cos(\beta/2)$  (c.f. Equation C15 in Luger et al. 2019), but it is convenient to use the half-angle formula to express these terms instead as

$$d_{m,m'}^l(\beta) = \sum_{i=0}^{2l} c_{m,m',i}^l \operatorname{sgn}(\sin \beta)^i (1 - \cos \beta)^{\frac{2l-i}{2}} (1 + \cos \beta)^{\frac{i}{2}} \quad \checkmark \quad (\text{C56})$$

where

$$c_{m,m',i}^l = \begin{cases} \frac{(-1)^{\frac{2l-m+m'-i}{2}} \sqrt{(l-m)!(l+m)!(l-m')!(l+m')!}}{2^l \left(\frac{i-m-m'}{2}\right)! \left(\frac{i+m+m'}{2}\right)! \left(\frac{2l-i-m+m'}{2}\right)! \left(\frac{2l-i+m-m'}{2}\right)!} & m - m' - i \text{ even} \\ 0 & m - m' - i \text{ odd} \end{cases} \quad \checkmark \quad (\text{C57})$$

### C.2.2. Probability density function

The latitude integrals (Equations C21 and C25) involve rotations by an angle  $\phi$  about  $\hat{\mathbf{x}}$ , which may be accomplished by choosing Euler angles  $\alpha = \pi/2$ ,  $\beta = \phi$ , and  $\gamma = -\pi/2$ , such that

$$\mathbf{R}_{\hat{\mathbf{x}}}^l(\phi) = \mathbf{U}^{l-1} \mathbf{D}_{\hat{\mathbf{x}}}^l(\phi) \mathbf{U}^l \quad (\text{C58})$$

with

$$\mathbf{D}_{\hat{\mathbf{x}}}^l(\phi) = \mathbf{D}^l \left( \frac{\pi}{2}, \phi, -\frac{\pi}{2} \right). \quad (\text{C59})$$

From the expressions above, it is clear that all terms in  $\mathbf{R}_{\hat{\mathbf{x}}}(\phi)$  are equal to (weighted) sums of powers of  $(1 \pm \cos \phi)$ . Since our goal is to compute integrals of these terms multiplied by a probability density function, 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 in  $\cos \phi$  has hyperparameters  $\alpha$  and  $\beta$  (not to be confused with the Euler angles  $\alpha$  and  $\beta$ ) and PDF given by

$$p(\cos \phi | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (\cos \phi)^{\alpha-1} (1 - \cos \phi)^{\beta-1}, \quad (\text{C60})$$

where  $\Gamma$  is the Gamma function. The implied distribution for  $\phi$  may be computed by a straightforward change of variable:

$$p(\phi | \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{2\Gamma(\alpha)\Gamma(\beta)} |\sin \phi| (\cos \phi)^{\alpha-1} (1 - \cos \phi)^{\beta-1}, \quad \checkmark \quad (\text{C61})$$

for  $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ . 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

$$\begin{aligned} a &\equiv \frac{\ln \alpha - K_{00}}{K_{10} - K_{00}} \\ b &\equiv \frac{\ln \beta - K_{10}}{K_{11} - K_{10}} \end{aligned} \quad (\text{C62})$$

with inverse transform

$$\begin{aligned} \alpha &= \exp(K_{00} + (K_{10} - K_{00})a) \\ \beta &= \exp(K_{10} + (K_{11} - K_{10})b), \end{aligned} \quad (\text{C63})$$

where the matrix

$$\mathbf{K} = \begin{pmatrix} 0 & 10 \\ \ln \frac{1}{2} & 10 \end{pmatrix} \quad (\text{C64})$$

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

$$\boldsymbol{\theta}_\phi = (a \ b)^\top. \quad (\text{C65})$$

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 parametrize 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 choose instead

$$\boldsymbol{\theta}_\phi = (\mu_\phi \ \sigma_\phi)^\top. \quad (\text{C65}^*)$$

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 (C61) with respect to  $\phi$ , set the expression equal to zero, and solve for  $\phi$  to obtain

$$\mu_\phi = 2 \tan^{-1} \left( \sqrt{2\alpha + \beta - 2 - \sqrt{4\alpha^2 - 8\alpha - 6\beta + 4\alpha\beta + \beta^2 + 5}} \right). \quad \checkmark \quad (\text{C66})$$

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:

$$\sigma^2 = - \left( \frac{d^2 \ln \varphi(\phi; \mu, \sigma^2)}{d\phi^2} \right)^{-1} \quad (\text{C67})$$

We therefore twice differentiate the log of Equation (C61), 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:

$$\sigma_\phi = \frac{\sin \mu_\phi}{\sqrt{1 - \alpha + \beta + (\beta - 1) \cos \mu_\phi + \frac{\alpha - 1}{\cos \mu_\phi^2}}} . \quad \checkmark \quad (\text{C68})$$

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

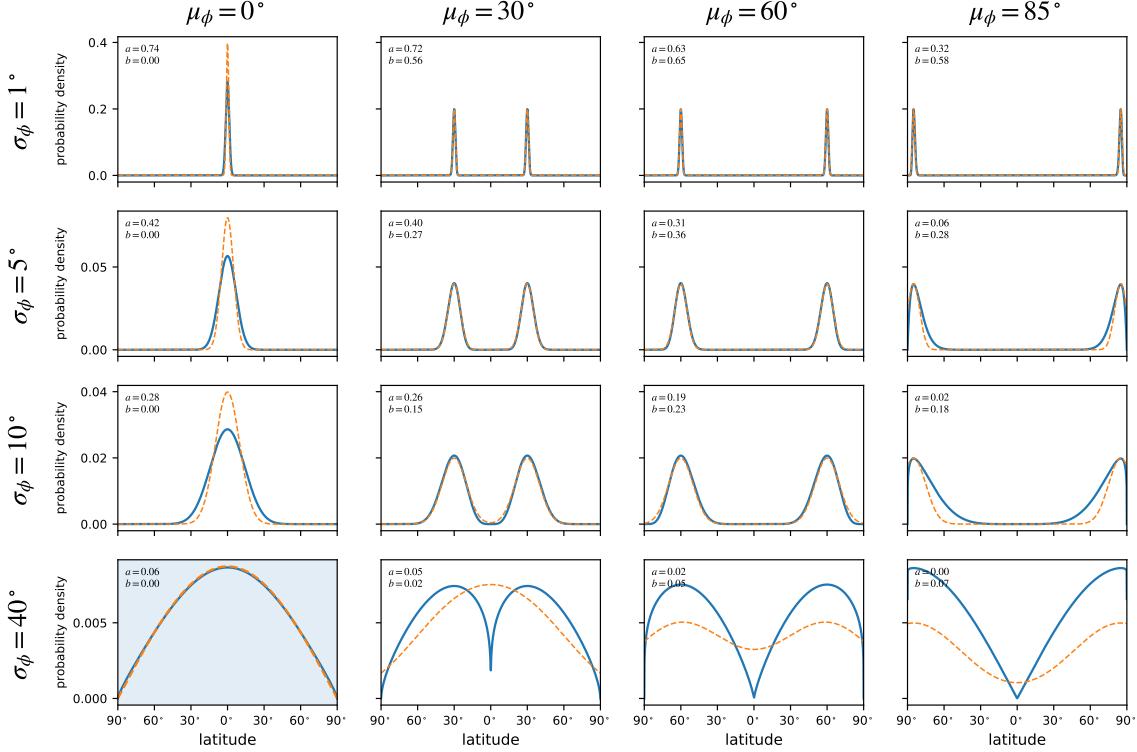
$$\alpha = \frac{2 + 4\sigma_\phi^2 + (3 + 8\sigma_\phi^2) \cos \mu_\phi + 2 \cos(2\mu_\phi) + \cos(3\mu_\phi)}{16\sigma_\phi^2 \cos\left(\frac{\mu_\phi}{2}\right)^4} \quad \checkmark \quad (\text{C69})$$

$$\beta = \frac{\cos \mu_\phi + 2\sigma_\phi^2(3 + \cos(2\mu_\phi)) - \cos(3\mu_\phi)}{16\sigma_\phi^2 \cos\left(\frac{\mu_\phi}{2}\right)^4} . \quad \checkmark \quad (\text{C70})$$

For reference, Figure 24 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 in the top left of 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 to the poles, the curvature of the distribution changes significantly as a function of  $\phi$ , so the variance is somewhat underestimated by the approximation; and for  $\sigma_\phi$  large, 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 parametrization 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



**Figure 24.** Probability density function for the spot latitude (blue curves) for different values of the mode  $\mu_\phi$  (columns) and local standard deviation  $\sigma_\phi$  (rows). The corresponding values of  $a$  and  $b$  are indicated within each panel. The bimodal normal distribution with mean  $\mu_\phi$  and standard deviation  $\sigma_\phi$  is shown as the dashed orange curves; for mid-latitude modes and low standard deviations, the Gaussian approximation is quite good. The shaded panel in the lower right ( $\mu_\phi = 0^\circ$ ,  $\sigma_\phi = 0^\circ$ ) corresponds to an approximately isotropic distribution of spots over the surface of the star; in this panel only, the dashed orange curve corresponds to a cosine distribution in  $\phi$  (i.e., the exact isotropic distribution).

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

$$\begin{aligned} J &= \frac{\partial \mu_\phi}{\partial a} \frac{\partial \sigma_\phi}{\partial b} - \frac{\partial \mu_\phi}{\partial b} \frac{\partial \sigma_\phi}{\partial a} \\ &= \frac{(K_{10} - K_{00})(K_{11} - K_{10})\alpha\beta(1 + \cos \mu_\phi)^3 \sin(2\mu_\phi)^3}{\sigma_\phi (2\alpha + \beta - 3 + (2\alpha + \beta - 1)\cos \mu_\phi) (2(\alpha + \beta - 1) + 3(\beta - 1)\cos \mu_\phi - 2(\alpha - \beta - 1)\cos(2\mu_\phi) + (\beta - 1)\cos(3\mu_\phi))^2} \end{aligned} \quad \checkmark \text{ (C71)}$$

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 parametrization, enforcing instead a uniform prior on the quantities  $\mu_\phi$  and  $\sigma_\phi$ .

### 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

$$\begin{aligned}\mathbf{e}_\phi &\equiv \int \mathbf{R}_{\hat{\mathbf{x}}}(\phi) \mathbf{e}_r p(\phi | \boldsymbol{\theta}_\phi) d\phi \\ &= (\mathbf{e}_\phi^0 \ \mathbf{e}_\phi^1 \ \mathbf{e}_\phi^2 \ \cdots \ \mathbf{e}_\phi^{l_{\max}})^\top,\end{aligned}\quad \checkmark \text{ (C72)}$$

where

$$\begin{aligned}\mathbf{e}_\phi^l &= \int \mathbf{R}_{\hat{\mathbf{x}}}^l(\phi) \mathbf{e}_r^l p(\phi | \boldsymbol{\theta}_\phi) d\phi \\ &= \mathbf{U}^{l-1} \mathbf{p}_\phi^l,\end{aligned}\quad \checkmark \text{ (C73)}$$

and we define

$$\mathbf{p}_\phi^l \equiv \int \mathbf{D}_{\hat{\mathbf{x}}}^l(\phi) \bar{\mathbf{e}}_r^l p(\phi | \boldsymbol{\theta}_\phi) d\phi \quad \text{(C74)}$$

$$\bar{\mathbf{e}}_r^l \equiv \mathbf{U}^l \mathbf{e}_r^l. \quad \text{(C75)}$$

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

$$\begin{aligned}(p_\phi^l)_m &= \int \sum_{\mu=-l}^l (D_{\hat{\mathbf{x}}}^l)_{m,\mu}(\phi) (\bar{e}_r^l)_\mu p(\phi | \boldsymbol{\theta}_\phi) d\phi \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \sum_{\mu=-l}^l (\bar{e}_r^l)_\mu \exp\left[\frac{i\pi}{2}(m - \mu)\right] \sum_{i=0}^{2l} c_{m,\mu,i}^l (q_\phi^l)_i(\boldsymbol{\theta}_\phi),\end{aligned}\quad \text{(C76)}$$

where

$$\begin{aligned}(q_\phi^l)_i(\boldsymbol{\theta}_\phi) &\equiv \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \operatorname{sgn}(\sin \phi)^i |\sin \phi| (\cos \phi)^{\alpha-1} (1 - \cos \phi)^{l+\beta-\frac{i}{2}-1} (1 + \cos \phi)^{\frac{i}{2}} d\phi \\ &= \begin{cases} \int_0^1 x^{\alpha-1} (1-x)^{l+\beta-\frac{i}{2}-1} (1+x)^{\frac{i}{2}} dx & i \text{ even} \\ 0 & i \text{ odd}, \end{cases} \quad \checkmark \text{ (C77)}\end{aligned}$$

and in the last line we made use of the transformation  $x = \cos \phi$ . The integral in the expression above has a closed-form solution in terms of the hypergeometric function  ${}_2F_1$ :

$$(q_\phi^l)_i(\boldsymbol{\theta}_\phi) = \begin{cases} \frac{\Gamma(\alpha)\Gamma(l+\beta-\frac{i}{2})}{\Gamma(l+\alpha+\beta-\frac{i}{2})} {}_2F_1\left(-\frac{i}{2}, \alpha; l+\alpha+\beta-\frac{i}{2}; -1\right) & i \text{ even} \\ 0 & i \text{ odd}. \end{cases} \quad \checkmark \text{ (C78)}$$

In order to compute the integral  $\mathbf{e}_\phi$  (Equation C21) we must evaluate Equation (C78) 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 function and the hypergeometric function.

### C.2.4. Second moment

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

$$\mathbf{E}_\phi \equiv \int \mathbf{R}_{\hat{\mathbf{x}}}(\phi) \mathbf{E}_r \mathbf{R}_{\hat{\mathbf{x}}}^\top(\phi) p(\phi | \boldsymbol{\theta}_\phi) d\phi$$

$$= \begin{pmatrix} \mathbf{E}_\phi^{0,0} & \mathbf{E}_\phi^{0,1} & \mathbf{E}_\phi^{0,2} & \dots & \mathbf{E}_\phi^{0,l_{\max}} \\ \mathbf{E}_\phi^{1,0} & \mathbf{E}_\phi^{1,1} & \mathbf{E}_\phi^{1,2} & \dots & \mathbf{E}_\phi^{1,l_{\max}} \\ \mathbf{E}_\phi^{2,0} & \mathbf{E}_\phi^{2,1} & \mathbf{E}_\phi^{2,2} & \dots & \mathbf{E}_\phi^{2,l_{\max}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}_\phi^{l_{\max},0} & \mathbf{E}_\phi^{l_{\max},1} & \mathbf{E}_\phi^{l_{\max},2} & \dots & \mathbf{E}_\phi^{l_{\max},l_{\max}} \end{pmatrix}, \quad \checkmark \text{ (C79)}$$

where

$$\mathbf{E}_\phi^{l,l'} = \int \mathbf{R}_{\hat{\mathbf{x}}}^l(\phi) \mathbf{E}_r^{l,l'} \mathbf{R}_{\hat{\mathbf{x}}}^{l'}(\phi) p(\phi | \boldsymbol{\theta}_\phi) d\phi$$

$$= \mathbf{U}^{l-1} \mathbf{P}_\phi^{l,l'} \mathbf{U}^{l'-1}{}^\top \quad \checkmark \text{ (C80)}$$

and we define

$$\mathbf{P}_\phi^{l,l'} \equiv \int \mathbf{D}_{\hat{\mathbf{x}}}^l(\phi) \bar{\mathbf{E}}_r^{l,l'} \mathbf{D}_{\hat{\mathbf{x}}}^{l'}(\phi) p(\phi | \boldsymbol{\theta}_\phi) d\phi \quad \text{(C81)}$$

$$\bar{\mathbf{E}}_r^{l,l'} \equiv \mathbf{U}^l \mathbf{E}_r^{l,l'} \mathbf{U}^{l'}{}^\top. \quad \text{(C82)}$$

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

$$(P_\phi^{l,l'})_{m,m'} = \int \sum_{\mu=-l}^l \sum_{\mu'=-l'}^{l'} (D_{\hat{\mathbf{x}}}^l)_{m,\mu}(\phi) (\bar{E}_r^{l,l'})_{\mu,\mu'} (D_{\hat{\mathbf{x}}}^{l'})_{m',\mu'}(\phi) p(\phi | \boldsymbol{\theta}_\phi) d\phi \quad \text{(C83)}$$

$$= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \sum_{\mu=-l}^l \sum_{\mu'=-l'}^{l'} (\bar{E}_r^{l,l'})_{\mu,\mu'} \exp\left[\frac{i\pi}{2}(m - \mu + m' - \mu')\right] \sum_{i=0}^{2l} \sum_{i'=0}^{2l'} c_{m,\mu,i}^l c_{m',\mu',i'}^{l'} (Q_\phi^{l,l'})_{i,i'}(\boldsymbol{\theta}_\phi),$$

where, similarly to before,

$$(Q_\phi^{l,l'})_{i,i'}(\boldsymbol{\theta}_\phi) \equiv \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \operatorname{sgn}(\sin \phi)^{i+i'} |\sin \phi| (\cos \phi)^{\alpha-1} (1 - \cos \phi)^{l+l'+\beta - \frac{i+i'}{2} - 1} (1 + \cos \phi)^{\frac{i+i'}{2}} d\phi$$

$$= \begin{cases} \int_0^1 \mathbf{x}^{\alpha-1} (1 - \mathbf{x})^{l+l'+\beta - \frac{i+i'}{2} - 1} (1 + \mathbf{x})^{\frac{i+i'}{2}} d\mathbf{x} & i + i' \text{ even} \\ 0 & i + i' \text{ odd.} \end{cases} \quad \checkmark \text{ (C84)}$$

We may again express this integral in closed form:

$$(Q_\phi^{l,l'})_{i,i'}(\boldsymbol{\theta}_\phi) = \begin{cases} \frac{\Gamma(\alpha)\Gamma(l+l'+\beta - \frac{i+i'}{2})}{\Gamma(l+l'+\alpha+\beta - \frac{i+i'}{2})} {}_2F_1\left(-\frac{i+i'}{2}, \alpha; l+l'+\alpha+\beta - \frac{i+i'}{2}; -1\right) & i + i' \text{ even} \\ 0 & i + i' \text{ odd.} \end{cases} \quad \checkmark \text{ (C85)}$$

As before, this integral may be evaluated recursively to efficiently compute all of the terms in  $\mathbf{E}_\phi$ .

### C.3. The Longitude Integrals

In this section we will compute the first and second moments of the longitude distribution ( $\mathbf{e}_\lambda$  and  $\mathbf{E}_\lambda$ , given by Equations C22 and C26, respectively). The math here is very similar to that in the previous section, as we are again dealing with integrals of Wigner matrices (§C.2.1).

#### C.3.1. Probability density function

The longitude integrals (Equations C22 and C26) involve rotations by an angle  $\lambda$  about  $\hat{\mathbf{y}}$ , which may be accomplished by choosing Euler angles  $\alpha = 0$ ,  $\beta = \lambda$ , and  $\gamma = 0$ , such that

$$\mathbf{R}_{\hat{\mathbf{y}}}^l(\lambda) = \mathbf{U}^{l-1} \mathbf{D}_{\hat{\mathbf{y}}}^l(\lambda) \mathbf{U}^l \quad (\text{C86})$$

with

$$\mathbf{D}_{\hat{\mathbf{y}}}^l(\lambda) = \mathbf{D}^l(0, \lambda, 0) . \quad (\text{C87})$$

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  $\lambda \in [-\pi, \pi]$ :

$$p(\lambda | \boldsymbol{\theta}_\lambda) = \begin{cases} \frac{1}{2\pi} & -\pi \leq \lambda < \pi \\ 0 & \text{otherwise.} \end{cases} \quad (\text{C88})$$

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

$$\boldsymbol{\theta}_\lambda = () . \quad (\text{C89})$$

#### C.3.2. First moment

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

$$\begin{aligned} \mathbf{e}_\lambda &\equiv \int \mathbf{R}_{\hat{\mathbf{y}}}^l(\lambda) \mathbf{e}_\phi p(\lambda | \boldsymbol{\theta}_\lambda) d\lambda \\ &= (\mathbf{e}_\lambda^0 \ \mathbf{e}_\lambda^1 \ \mathbf{e}_\lambda^2 \ \cdots \ \mathbf{e}_\lambda^{l_{\max}})^\top , \end{aligned} \quad \checkmark \quad (\text{C90})$$

where

$$\begin{aligned} \mathbf{e}_\lambda^l &= \int \mathbf{R}_{\hat{\mathbf{y}}}^l(\lambda) \mathbf{e}_\phi^l p(\lambda | \boldsymbol{\theta}_\lambda) d\lambda \\ &= \mathbf{U}^{l-1} \mathbf{p}_\lambda^l , \end{aligned} \quad \checkmark \quad (\text{C91})$$

and we define

$$\mathbf{p}_\lambda^l \equiv \int \mathbf{D}_{\hat{\mathbf{x}}}^l(\lambda) \bar{\mathbf{e}}_\phi^l p(\lambda | \boldsymbol{\theta}_\lambda) d\lambda \quad (\text{C92})$$

$$\bar{\mathbf{e}}_\phi^l \equiv \mathbf{U}^l \mathbf{e}_\phi^l . \quad (\text{C93})$$

The integral  $\mathbf{p}_\lambda^l$  defined above has a closed-form solution. To show this, we write the terms of  $\mathbf{p}_\lambda^l$  as

$$\begin{aligned} (p_\lambda^l)_m &= \int \sum_{\mu=-l}^l (D_{\hat{\mathbf{y}}}^l)_{m,\mu}(\lambda) (\bar{e}_\phi^l)_\mu p(\lambda | \boldsymbol{\theta}_\lambda) d\lambda \\ &= \sum_{\mu=-l}^l (\bar{e}_\phi^l)_\mu \sum_{i=0}^{2l} c_{m,\mu,i}^l (q_\lambda^l)_i, \end{aligned} \quad (\text{C94})$$

where

$$\begin{aligned} (q_\lambda^l)_i &\equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{sgn}(\sin \lambda)^i (1 - \cos \lambda)^{l-\frac{i}{2}} (1 + \cos \lambda)^{\frac{i}{2}} d\lambda \\ &= \begin{cases} \frac{2^l \Gamma\left(\frac{1+i}{2}\right) \Gamma\left(l + \frac{1-i}{2}\right)}{\pi \Gamma(l+1)} & i \text{ even} \\ 0 & i \text{ odd}, \end{cases} \end{aligned} \quad \checkmark \quad (\text{C95})$$

whose terms may easily be computed by upward recursion. Since  $\mathbf{q}_I^l$  does not depend on any user inputs, it may be computed a single time as a pre-processing step for efficiency.

### C.3.3. Second moment

We write the second moment integral as

$$\begin{aligned} \mathbf{E}_\lambda &\equiv \int \mathbf{R}_{\hat{\mathbf{y}}}(\lambda) \mathbf{E}_\phi \mathbf{R}_{\hat{\mathbf{y}}}^\top(\lambda) p(\lambda | \boldsymbol{\theta}_\lambda) d\lambda \\ &= \begin{pmatrix} \mathbf{E}_\lambda^{0,0} & \mathbf{E}_\lambda^{0,1} & \mathbf{E}_\lambda^{0,2} & \dots & \mathbf{E}_\lambda^{0,l_{\max}} \\ \mathbf{E}_\lambda^{1,0} & \mathbf{E}_\lambda^{1,1} & \mathbf{E}_\lambda^{1,2} & \dots & \mathbf{E}_\lambda^{1,l_{\max}} \\ \mathbf{E}_\lambda^{2,0} & \mathbf{E}_\lambda^{2,1} & \mathbf{E}_\lambda^{2,2} & \dots & \mathbf{E}_\lambda^{2,l_{\max}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}_\lambda^{l_{\max},0} & \mathbf{E}_\lambda^{l_{\max},1} & \mathbf{E}_\lambda^{l_{\max},2} & \dots & \mathbf{E}_\lambda^{l_{\max},l_{\max}} \end{pmatrix}, \end{aligned} \quad \checkmark \quad (\text{C96})$$

where

$$\begin{aligned} \mathbf{E}_\lambda^{l,l'} &= \int \mathbf{R}_{\hat{\mathbf{y}}}^l(\lambda) \mathbf{E}_\phi^{l,l'} \mathbf{R}_{\hat{\mathbf{y}}}^{l'\top}(\lambda) p(\lambda | \boldsymbol{\theta}_\lambda) d\lambda \\ &= \mathbf{U}^{l-1} \mathbf{P}_\lambda^{l,l'} \mathbf{U}^{l'-1\top} \end{aligned} \quad \checkmark \quad (\text{C97})$$

and we define

$$\mathbf{P}_\lambda^{l,l'} \equiv \int \mathbf{D}_{\hat{\mathbf{x}}}^l(\lambda) \bar{\mathbf{E}}_\phi^{l,l'} \mathbf{D}_{\hat{\mathbf{x}}}^{l'\top}(\lambda) p(\lambda | \boldsymbol{\theta}_\lambda) d\lambda \quad (\text{C98})$$

$$\bar{\mathbf{E}}_\phi^{l,l'} \equiv \mathbf{U}^l \mathbf{E}_\phi^{l,l'} \mathbf{U}^{l'\top}. \quad (\text{C99})$$

We then express the terms of  $\mathbf{P}_\lambda^{l,l'}$  as

$$\begin{aligned} (P_\lambda^{l,l'})_{m,m'} &= \int \sum_{\mu=-l}^l \sum_{\mu'=-l'}^{l'} (D_{\hat{\mathbf{x}}}^l)_{m,\mu}(\lambda) (\bar{E}_\phi^{l,l'})_{\mu,\mu'} (D_{\hat{\mathbf{x}}}^{l'})_{m',\mu'}(\lambda) p(\lambda | \boldsymbol{\theta}_\lambda) d\lambda \\ &= \sum_{\mu=-l}^l \sum_{\mu'=-l'}^{l'} (\bar{E}_\phi^{l,l'})_{\mu,\mu'} \sum_{i=0}^{2l} \sum_{i'=0}^{2l'} c_{m,\mu,i}^l c_{m',\mu',i'}^{l'} (Q_\lambda^{l,l'})_{i,i'}, \end{aligned} \quad (\text{C100})$$

where

$$\begin{aligned} (Q_\lambda^{l,l'})_{i,i'} &\equiv \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{sgn}(\sin \lambda)^{i+i'} (1 - \cos \lambda)^{l+l'-\frac{i+i'}{2}} (1 + \cos \lambda)^{\frac{i+i'}{2}} d\lambda \\ &= \begin{cases} \frac{2^{l+l'} \Gamma\left(\frac{1+i+i'}{2}\right) \Gamma\left(l+l'+\frac{1-i-i'}{2}\right)}{\pi \Gamma(l+l'+1)} & i+i' \text{ even} \\ 0 & i+i' \text{ odd}, \end{cases} \end{aligned} \quad \checkmark \quad (\text{C101})$$

whose terms may again be computed by upward recursion in a single pre-processing step.

#### C.4. The Contrast Integrals

The final integrals we must take in our computation of  $E[\mathbf{y} | \boldsymbol{\theta}_\bullet]$  and  $E[\mathbf{y}\mathbf{y}^\top | \boldsymbol{\theta}_\bullet]$  are the integrals over the spot contrast distribution,  $\mathbf{e}_c$  and  $\mathbf{E}_c$ . 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  $\mathbf{e}_\lambda$  and  $\mathbf{E}_\lambda$  out of the integrals in Equations (C23) and (C27) to write

$$\mathbf{e}_c \equiv -\mathbf{e}_\lambda \int \mathbf{c} p(\mathbf{c} | \boldsymbol{\theta}_c) d\mathbf{c} \quad (\text{C102})$$

$$\mathbf{E}_c \equiv \mathbf{E}_\lambda \int \mathbf{c}^2 p(\mathbf{c} | \boldsymbol{\theta}_c) d\mathbf{c}. \quad (\text{C103})$$

These integrals may be computed analytically for any choice of probability density function  $p(\mathbf{c} | \boldsymbol{\theta}_c)$  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 (see §4.6). In our implementation of the algorithm, we therefore choose the simplest possible probability distribution, a delta function:

$$p(\mathbf{c} | \boldsymbol{\theta}_c) = \delta(\mathbf{c} - c) \quad (\text{C104})$$

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

$$\boldsymbol{\theta}_c = (c)^\top. \quad (\text{C105})$$

The moment integrals are then trivial to evaluate:

$$\mathbf{e}_c = -c \mathbf{e}_\lambda \quad (\text{C106})$$

$$\mathbf{E}_c = c^2 \mathbf{E}_\lambda. \quad (\text{C107})$$

## D. INCLINATION

In this section we will compute the first and second moment integrals of the inclination distribution (Equations 24 and 25), which allow us to compute the mean and covariance of the process that describes the flux marginalized over all values of the inclination (Equations 22 and 23).

### D.1. Probability density function

Similar to the latitude integrals, the process of inclining a star relative to the observer (see Equation B7) involves rotations by an angle  $-\mathbb{I}$  about  $\hat{\mathbf{x}}$ , which may be accomplished by choosing Euler angles  $\alpha = \pi/2$ ,  $\beta = -\mathbb{I}$ , and  $\gamma = -\pi/2$ , such that

$$\mathbf{R}_{\hat{\mathbf{x}}}^l(-\mathbb{I}) = \mathbf{U}^{l-1} \mathbf{D}_{\hat{\mathbf{x}}}^l(-\mathbb{I}) \mathbf{U}^l \quad (\text{D108})$$

with

$$\mathbf{D}_{\hat{\mathbf{x}}}^l(-\mathbb{I}) = \mathbf{D}^l \left( \frac{\pi}{2}, -\mathbb{I}, -\frac{\pi}{2} \right). \quad (\text{D109})$$

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

$$p(\mathbb{I}) = \sin \mathbb{I} \quad (\text{D110})$$

for  $\mathbb{I} \in [0, \frac{\pi}{2}]$ .

### D.2. First moment

The expression for the first moment is

$$\mathbf{e}_I \equiv \int \mathcal{A}(\mathbb{I}, P, \mathbf{u}) \mathbf{e}_y p(\mathbb{I}) d\mathbb{I} \quad (\text{D111})$$

where

$$\mathbf{e}_y \equiv E[\mathbf{y} | \boldsymbol{\theta}_\bullet] \quad (\text{D112})$$

is the first moment of the distribution over spherical harmonic coefficients (Equation 15). We can use Equations (B6) and (B7) to express the element at index  $k$  (corresponding to the mean of the GP at time  $t = t_k$ ) as

$$\begin{aligned} (e_I)_k &= \int \mathbf{a}_k^\top(\mathbb{I}) \mathbf{e}_y p(\mathbb{I}) d\mathbb{I} \\ &= \mathbf{r}^\top \mathbf{A}_1 \int \mathbf{R}_{\hat{\mathbf{x}}}(-\mathbb{I}) (\mathbf{e}_{y'})_k p(\mathbb{I}) d\mathbb{I}, \end{aligned} \quad (\text{D113})$$

where we define

$$(\mathbf{e}_{y'})_k \equiv \mathbf{R}_{\hat{\mathbf{z}}} \left( \frac{2\pi}{P} t_k \right) \mathbf{R}_{\hat{\mathbf{x}}} \left( \frac{\pi}{2} \right) \mathbf{e}_y \quad (\text{D114})$$

as the expectation of  $\mathbf{y}$  in the polar frame at time  $t = t_k$ .<sup>19</sup> 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  $\hat{\mathbf{z}}$  (i.e., the rotational axis of the star) therefore cannot change the expectation of  $\mathbf{y}$ , so

$$\begin{aligned} (\mathbf{e}_{y'})_k &= (\mathbf{e}_{y'})_0 \\ &= \mathbf{R}_{\hat{\mathbf{x}}} \left( \frac{\pi}{2} \right) \mathbf{e}_y \\ &\equiv \mathbf{e}_{y'} . \end{aligned} \quad (\text{D115})$$

We therefore have

$$\mathbf{e}_I \equiv e_I \mathbf{1} \quad (\text{D116})$$

where

$$e_I = \mathbf{r}^\top \mathbf{A}_1 \mathbf{e}_{y''} \quad (\text{D117})$$

and

$$\begin{aligned} \mathbf{e}_{y''} &= \int \mathbf{R}_{\hat{\mathbf{x}}}(-\mathbb{I}) \mathbf{e}_{y'} p(\mathbb{I}) d\mathbb{I} \\ &= (\mathbf{e}_{y''}^0 \ \mathbf{e}_{y''}^1 \ \mathbf{e}_{y''}^2 \ \cdots \ \mathbf{e}_{y''}^{l_{\max}})^\top \end{aligned} \quad (\text{D118})$$

is the expectation of  $\mathbf{y}$  in the observer's frame, and as before we explicitly separate it out by spherical harmonic degree. As in §C.2, we may write

$$\begin{aligned} \mathbf{e}_{y''}^l &= \int \mathbf{R}_{\hat{\mathbf{x}}}^l(-\mathbb{I}) \mathbf{e}_{y'}^l p(\mathbb{I}) d\mathbb{I} \\ &= \mathbf{U}^{l-1} \mathbf{p}_I^l , \end{aligned} \quad \checkmark \quad (\text{D119})$$

and we define

$$\mathbf{p}_I^l \equiv \int \mathbf{D}_{\hat{\mathbf{x}}}^l(-\mathbb{I}) \bar{\mathbf{e}}_{y'}^l p(\mathbb{I}) d\mathbb{I} \quad (\text{D120})$$

$$\bar{\mathbf{e}}_{y'}^l \equiv \mathbf{U}^l \mathbf{e}_{y'}^l . \quad (\text{D121})$$

<sup>19</sup> In the presence of limb darkening, we must include the limb darkening operator  $\mathbf{L}(\mathbf{u})$  in Equation (D113); see §B.2.

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

$$\begin{aligned} (p_I^l)_m &= \int \sum_{\mu=-l}^l (D_{\hat{\mathbf{x}}}^l)_{m,\mu}(-\mathbb{I}) (\bar{e}_{y'}^l)_\mu p(\mathbb{I}) d\mathbb{I} \\ &= \sum_{\mu=-l}^l (\bar{e}_{y'}^l)_\mu \exp \left[ \frac{i\pi}{2}(m-\mu) \right] \sum_{i=0}^{2l} c_{m,\mu,i}^l (q_I^l)_i, \end{aligned} \quad \checkmark \quad (\text{D122})$$

where

$$\begin{aligned} (q_I^l)_i &\equiv \int_0^{\frac{\pi}{2}} (-1)^i (1 - \cos \mathbb{I})^{\frac{2l-i}{2}} (1 + \cos \mathbb{I})^{\frac{i}{2}} \sin \mathbb{I} d\mathbb{I} \\ &= (-1)^i \int_0^1 (1 - \mathbf{x})^{\frac{2l-i}{2}} (1 + \mathbf{x})^{\frac{i}{2}} d\mathbf{x} \\ &= \frac{(-1)^i}{l - \frac{i}{2} + 1} {}_2F_1 \left( 1, -\frac{i}{2}; 2 + l - \frac{i}{2}; -1 \right), \end{aligned} \quad \checkmark \quad (\text{D123})$$

which may easily be computed recursively. As with the longitude integrals, the vector  $\mathbf{q}_I^l$  need only be computed a single time as a pre-processing step, as it does not depend on any user inputs.

### D.3. Second moment

The expression for the second moment is

$$\mathbf{E}_I \equiv \int \mathbf{A}(\mathbb{I}, P, \mathbf{u}) \mathbf{E}_y \mathbf{A}^\top(I, P, \mathbf{u}) p(\mathbb{I}) d\mathbb{I} \quad (\text{D124})$$

where

$$\mathbf{E}_y \equiv E[\mathbf{y} \mathbf{y}^\top | \boldsymbol{\theta}_\bullet] \quad (\text{D125})$$

is the second moment of the distribution over spherical harmonic coefficients (Equation 16). We can use Equations (B6) and (B7) 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

$$\begin{aligned} (E_I)_{k,k'} &= \int \mathbf{a}_k^\top(\mathbb{I}) \mathbf{E}_y \mathbf{a}_{k'}(\mathbb{I}) p(\mathbb{I}) d\mathbb{I} \\ &= \mathbf{r}^\top \mathbf{A}_1 (\mathbf{E}_{y''})_{k,k'} \mathbf{A}_1^\top \mathbf{r}, \end{aligned} \quad (\text{D126})$$

where

$$(\mathbf{E}_{y''})_{k,k'} = \int \mathbf{R}_{\hat{\mathbf{x}}}(-\mathbb{I}) (\mathbf{E}_{y'})_{k,k'} \mathbf{R}_{\hat{\mathbf{x}}}^\top(-\mathbb{I}) p(\mathbb{I}) d\mathbb{I} \quad (\text{D127})$$

is the expectation of  $\mathbf{y}\mathbf{y}^\top$  in the observer's frame at times  $t = t_k$  and  $t' = t_{k'}$  and

$$(\mathbf{E}_{y'})_{k,k'} \equiv \mathbf{R}_{\hat{\mathbf{z}}} \left( \frac{2\pi}{P} t_k \right) \mathbf{R}_{\hat{\mathbf{x}}} \left( \frac{\pi}{2} \right) \mathbf{E}_y \mathbf{R}_{\hat{\mathbf{x}}}^\top \left( \frac{\pi}{2} \right) \mathbf{R}_{\hat{\mathbf{z}}}^\top \left( \frac{2\pi}{P} t_{k'} \right) \quad (\text{D128})$$

is the expectation of  $\mathbf{y}\mathbf{y}^\top$  in the polar frame at times  $t = t_k$  and  $t' = t_{k'}$ . The rest of the computation follows what we did in §C.2, except that the number of operations required to compute  $\mathbf{E}_I$  is a factor of  $K^2$  larger than in the computation of expectations like  $\mathbf{E}_\phi$  (Equation C79). That is because we must compute the integral of all terms of a matrix for *each* of the  $K^2$  elements of  $\mathbf{E}_I$ . We discuss in §D.4 below strategies that can drastically improve the computational scaling of marginalizing over the inclination.

Let us write Equation (D127) in terms of its spherical harmonic components:

$$(\mathbf{E}_{y''})_{k,k'} = \begin{pmatrix} (\mathbf{E}_{y''}^{0,0})_{k,k'} & (\mathbf{E}_{y''}^{0,1})_{k,k'} & (\mathbf{E}_{y''}^{0,2})_{k,k'} & \cdots & (\mathbf{E}_{y''}^{0,l_{\max}})_{k,k'} \\ (\mathbf{E}_{y''}^{1,0})_{k,k'} & (\mathbf{E}_{y''}^{1,1})_{k,k'} & (\mathbf{E}_{y''}^{1,2})_{k,k'} & \cdots & (\mathbf{E}_{y''}^{1,l_{\max}})_{k,k'} \\ (\mathbf{E}_{y''}^{2,0})_{k,k'} & (\mathbf{E}_{y''}^{2,1})_{k,k'} & (\mathbf{E}_{y''}^{2,2})_{k,k'} & \cdots & (\mathbf{E}_{y''}^{2,l_{\max}})_{k,k'} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (\mathbf{E}_{y''}^{l_{\max},0})_{k,k'} & (\mathbf{E}_{y''}^{l_{\max},1})_{k,k'} & (\mathbf{E}_{y''}^{l_{\max},2})_{k,k'} & \cdots & (\mathbf{E}_{y''}^{l_{\max},l_{\max}})_{k,k'} \end{pmatrix}, \quad \checkmark \quad (\text{D129})$$

where

$$\begin{aligned} (\mathbf{E}_{y''}^{l,l'})_{k,k'} &= \int \mathbf{R}_{\hat{\mathbf{x}}}^l(-\mathbb{I}) \left( \mathbf{E}_{y'}^{l,l'} \right)_{k,k'} \mathbf{R}_{\hat{\mathbf{x}}}^{l'\top}(-\mathbb{I}) p(\mathbb{I}) d\mathbb{I} \\ &= \mathbf{U}^{l-1} \left( \mathbf{P}_I^{l,l'} \right)_{k,k'} \mathbf{U}^{l'-1\top} \end{aligned} \quad \checkmark \quad (\text{D130})$$

and we define

$$\left( \mathbf{P}_I^{l,l'} \right)_{k,k'} \equiv \int \mathbf{D}_{\hat{\mathbf{x}}}^l(-\mathbb{I}) \left( \bar{\mathbf{E}}_{y'}^{l,l'} \right)_{k,k'} \mathbf{D}_{\hat{\mathbf{x}}}^{l'\top}(-\mathbb{I}) p(\mathbb{I}) d\mathbb{I} \quad (\text{D131})$$

$$\left( \bar{\mathbf{E}}_{y'}^{l,l'} \right)_{k,k'} \equiv \mathbf{U}^l \left( \mathbf{E}_{y'}^{l,l'} \right)_{k,k'} \mathbf{U}^{l'\top}. \quad (\text{D132})$$

As before, we may express the solution to the integral in Equation (D131) in closed form. Let us write its terms as

$$\begin{aligned} \left[ \left( P_I^{l,l'} \right)_{k,k'} \right]_{m,m'} &= \sum_{\mu=-l}^l \sum_{\mu'=-l}^l (D_{\hat{\mathbf{x}}}^l)_{m,\mu}(-\mathbb{I}) \left[ \left( \bar{E}_{y'}^{l,l'} \right)_{k,k'} \right]_{\mu,\mu'} (D_{\hat{\mathbf{x}}}^{l'})_{\mu',m'}(-\mathbb{I}) p(\mathbb{I}) d\mathbb{I} \quad (\text{D133}) \\ &= \sum_{\mu=-l}^l \sum_{\mu'=-l'}^{l'} \left[ \left( \bar{E}_{y'}^{l,l'} \right)_{k,k'} \right]_{\mu,\mu'} \exp \left[ \frac{i\pi}{2}(m - \mu + m' - \mu') \right] \sum_{i=0}^{2l} \sum_{i'=0}^{2l'} c_{m,\mu,i}^l c_{m',\mu',i'}^{l'} (Q_I^{l,l'})_{i,i'}, \end{aligned}$$

where, similarly to before,

$$\begin{aligned}
(Q_I^{l,l'})_{i,i'} &\equiv \int_0^{\frac{\pi}{2}} (-1)^{i+i'} (1 - \cos \alpha)^{l+l'-\frac{i+i'}{2}} (1 + \cos \alpha)^{\frac{i+i'}{2}} \sin \alpha d\alpha \\
&= (-1)^{i+i'} \int_0^1 (1 - x)^{l+l'-\frac{i+i'}{2}} (1 + x)^{\frac{i+i'}{2}} dx \\
&= \frac{(-1)^{i+i'}}{l + l' - \frac{i+i'}{2} + 1} {}_2F_1 \left( 1, -\frac{i+i'}{2}; 2 + l + l' - \frac{i+i'}{2}; -1 \right), \quad \checkmark \quad (\text{D134})
\end{aligned}$$

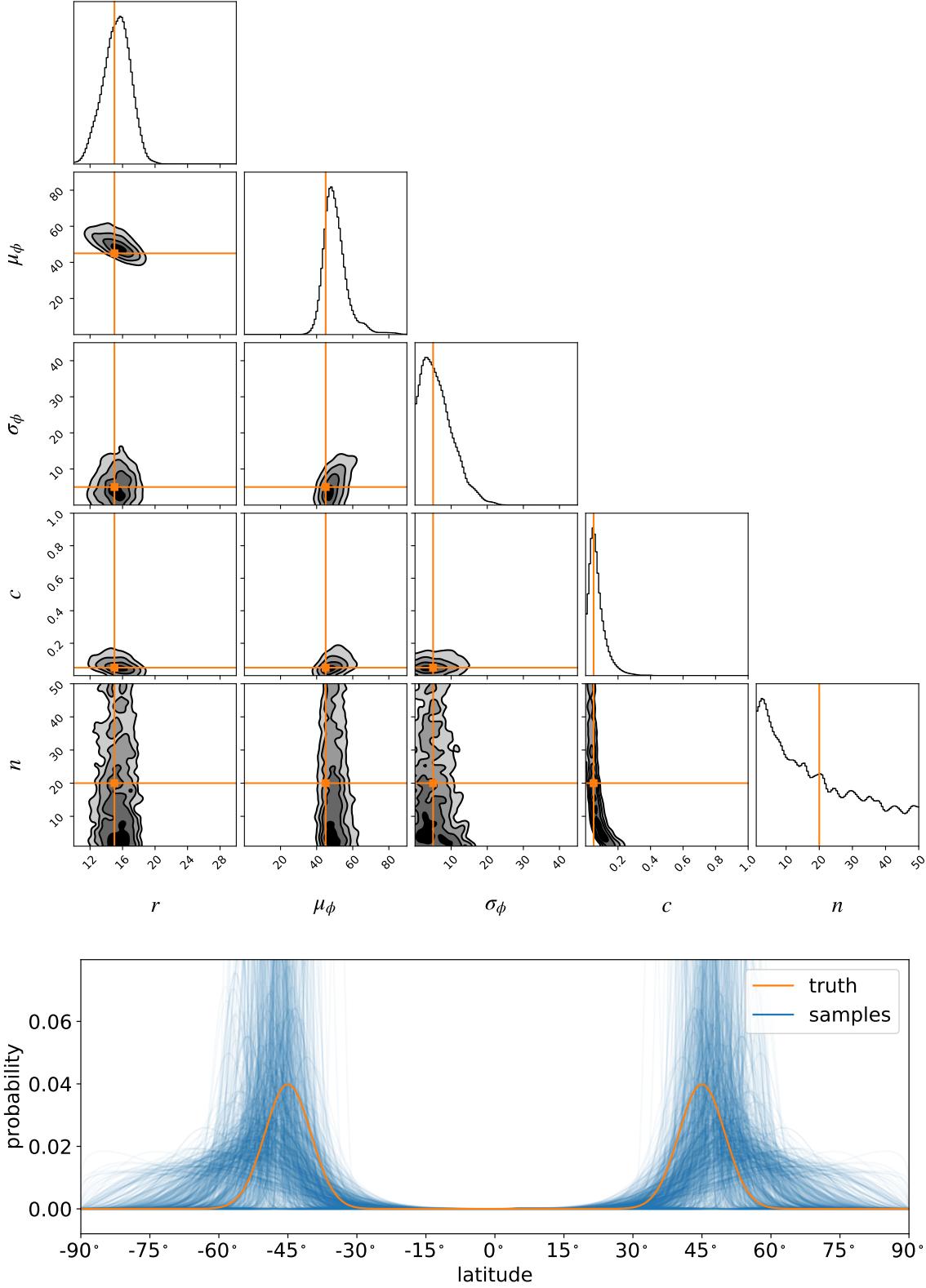
which may again be computed recursively in a pre-processing step.

#### D.4. Speeding up the computation

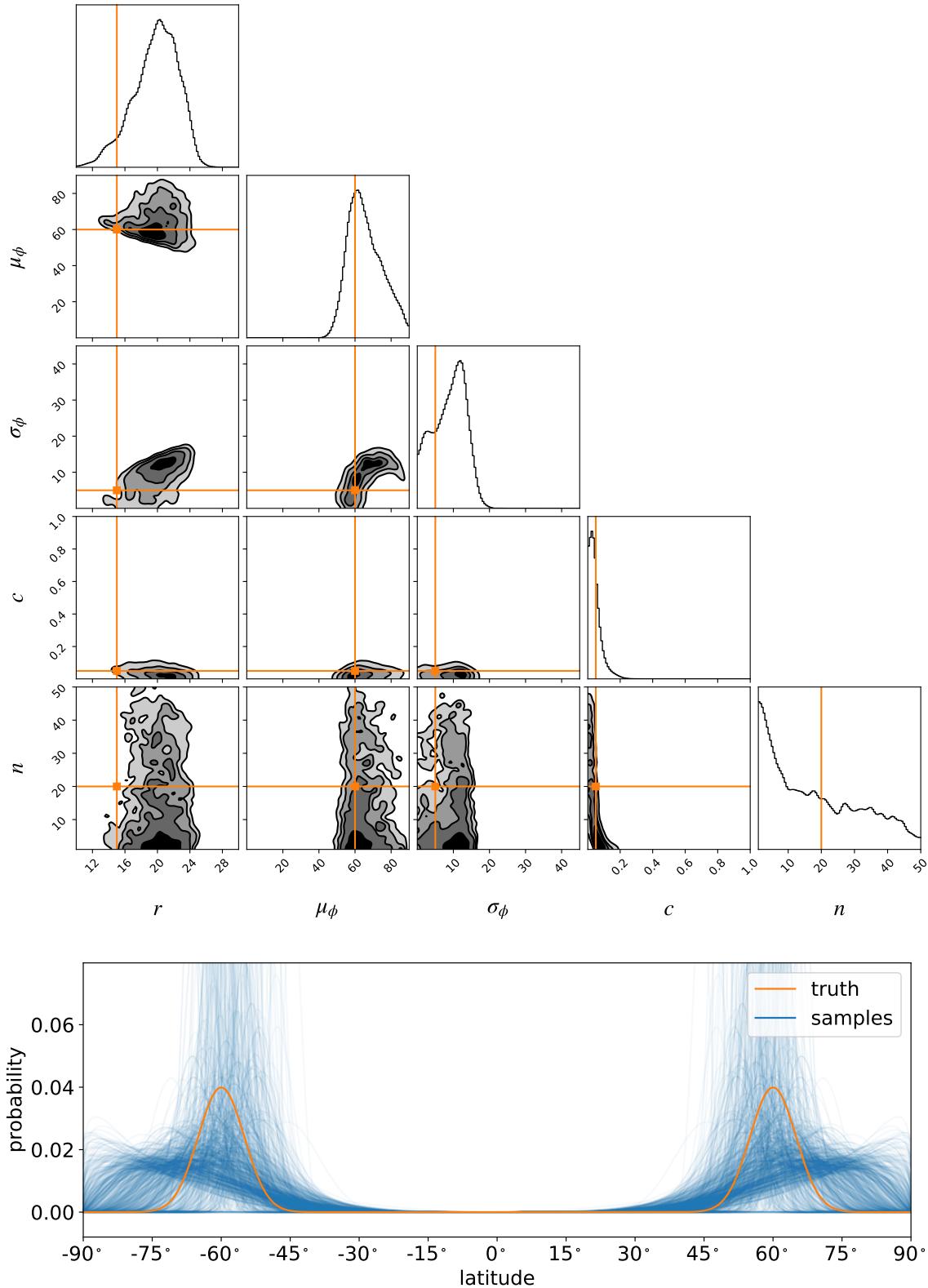
The expressions in the previous section are a bit of a nightmare, particularly because of the dimensionality of some of the linear operators involved. The complexity of the expressions is due to the fact that the second moment of the spherical harmonic vector projected onto the sky (Equation D127) is 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 (D133). In addition to being cumbersome to evaluate, the full second moment matrix  $\mathbf{E}_I$  (and hence the flux covariance matrix) is costly to compute. It is helpful that Equation (D134) does not depend on any user inputs and thus may be pre-computed, but even still we require evaluating the four nested sums in Equation (D127)  $\mathcal{O}(l_{\max}^3)$  times for each entry in the  $(K \times K)$  matrix  $\mathbf{E}_I$ .

Fortunately, the inner two sums in Equation (D127) do not depend on user inputs, so those may be pre-computed, and Equation (D127) 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 §D.2: any two entries  $(E_I)_{k,k'}$  and  $(E_I)_{j,j'}$  are the same if  $t_k - t_{k'} = t_j - t_{j'}$ . 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 which is fully specified by a single vector of length  $K$ .

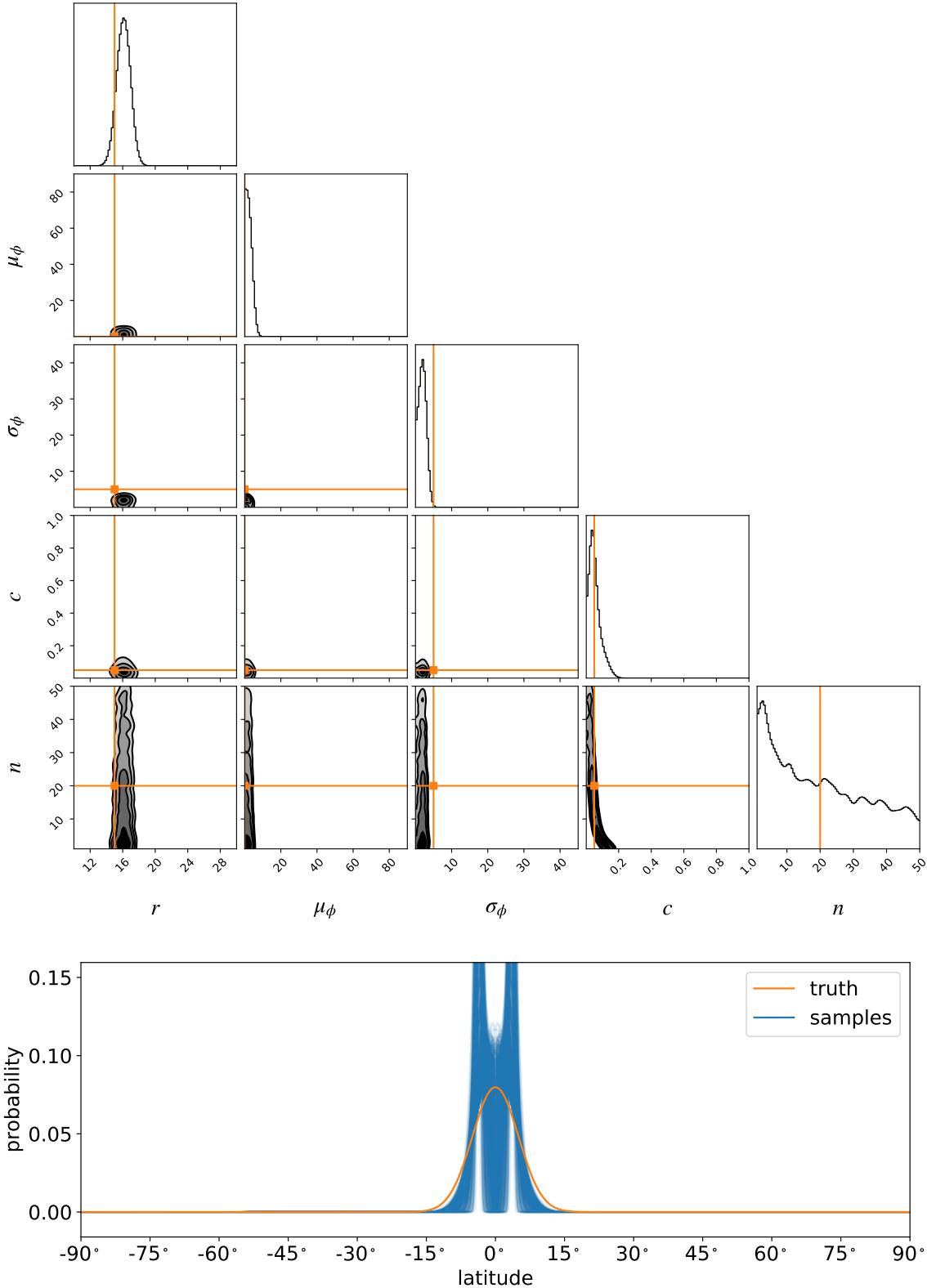
In the more general case where the data are not evenly sampled, we may still evaluate the covariance at a fixed number of points  $K' < K^2$  and approximate the full covariance matrix via interpolation. As long as the data are roughly evenly sampled, as is the case with *Kepler* or *TESS* light curves, 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 (c.f. §2.3), the non-stationary correction to the covariance is applied *after* the step where we marginalize over the inclination, so this approach is still valid.



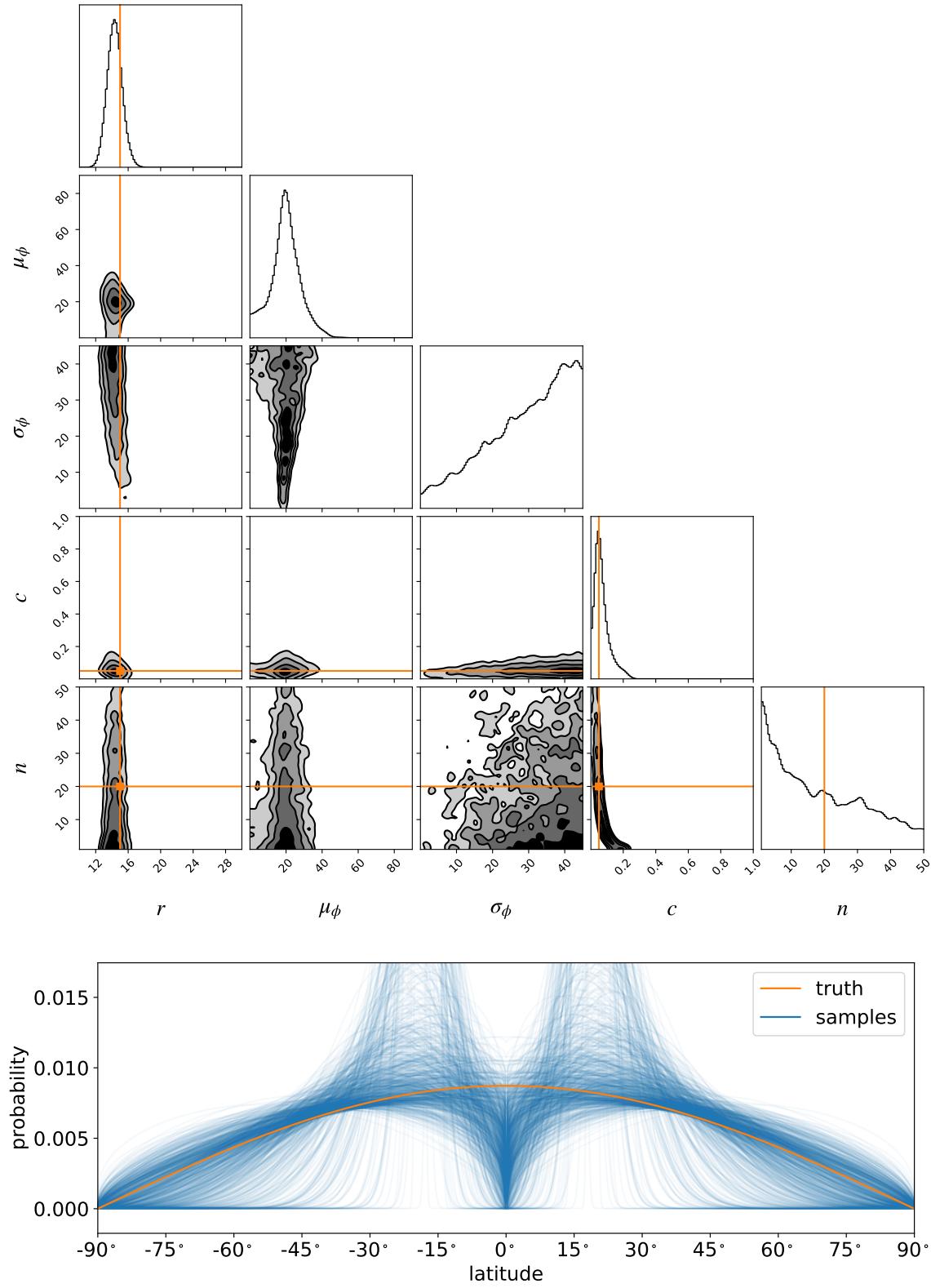
**Figure S25.** Same as Figures 9 and 10, but for mid-latitude spots with  $\mu_\phi = 45^\circ$ . The radius and latitude parameters are again inferred correctly.



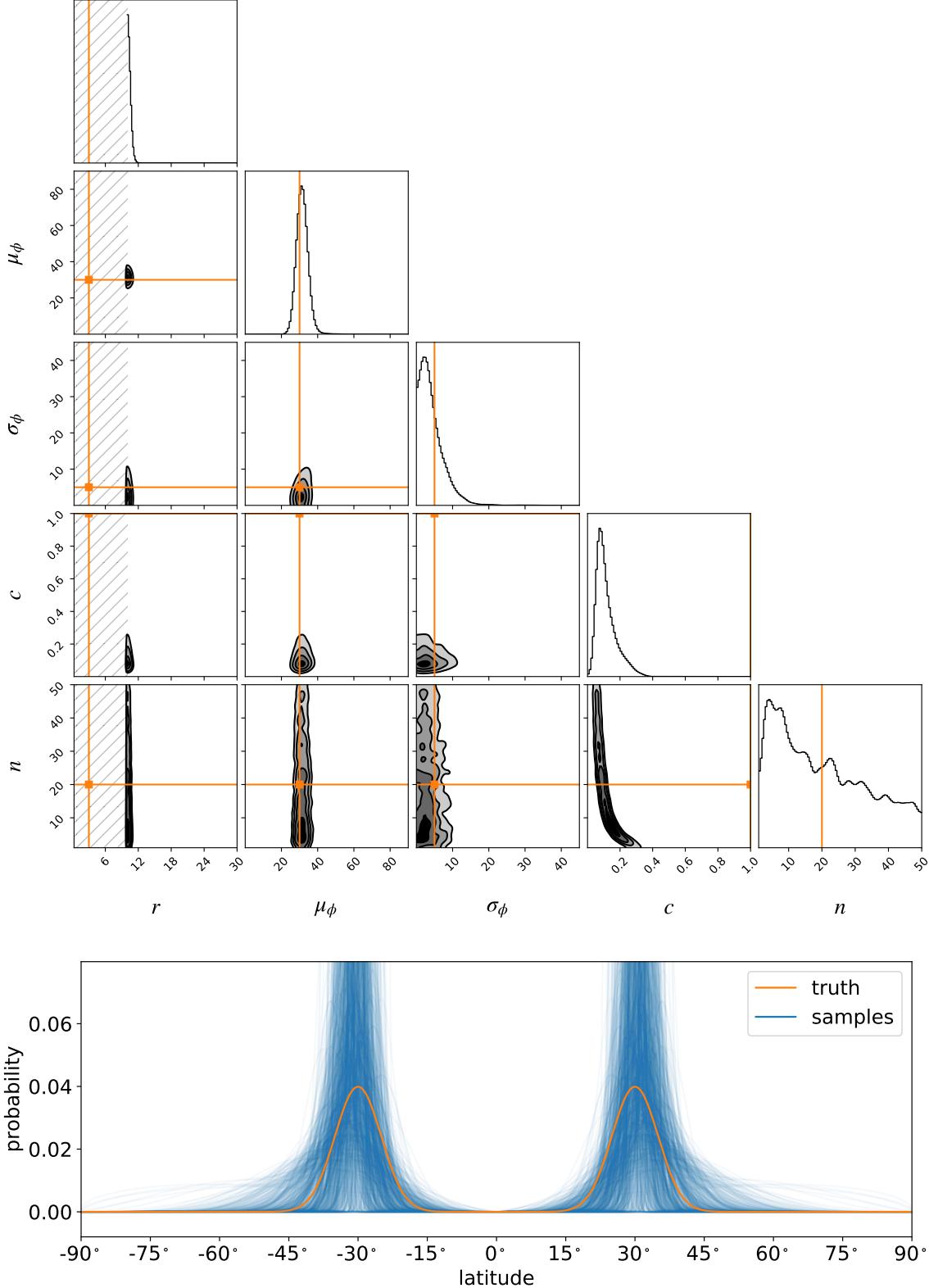
**Figure S26.** Same as Figure S25, but for high-latitude spots with  $\mu_\phi = 60^\circ$ . The radius and latitude parameters are again inferred correctly, although the model cannot rule out the presence of polar spots.



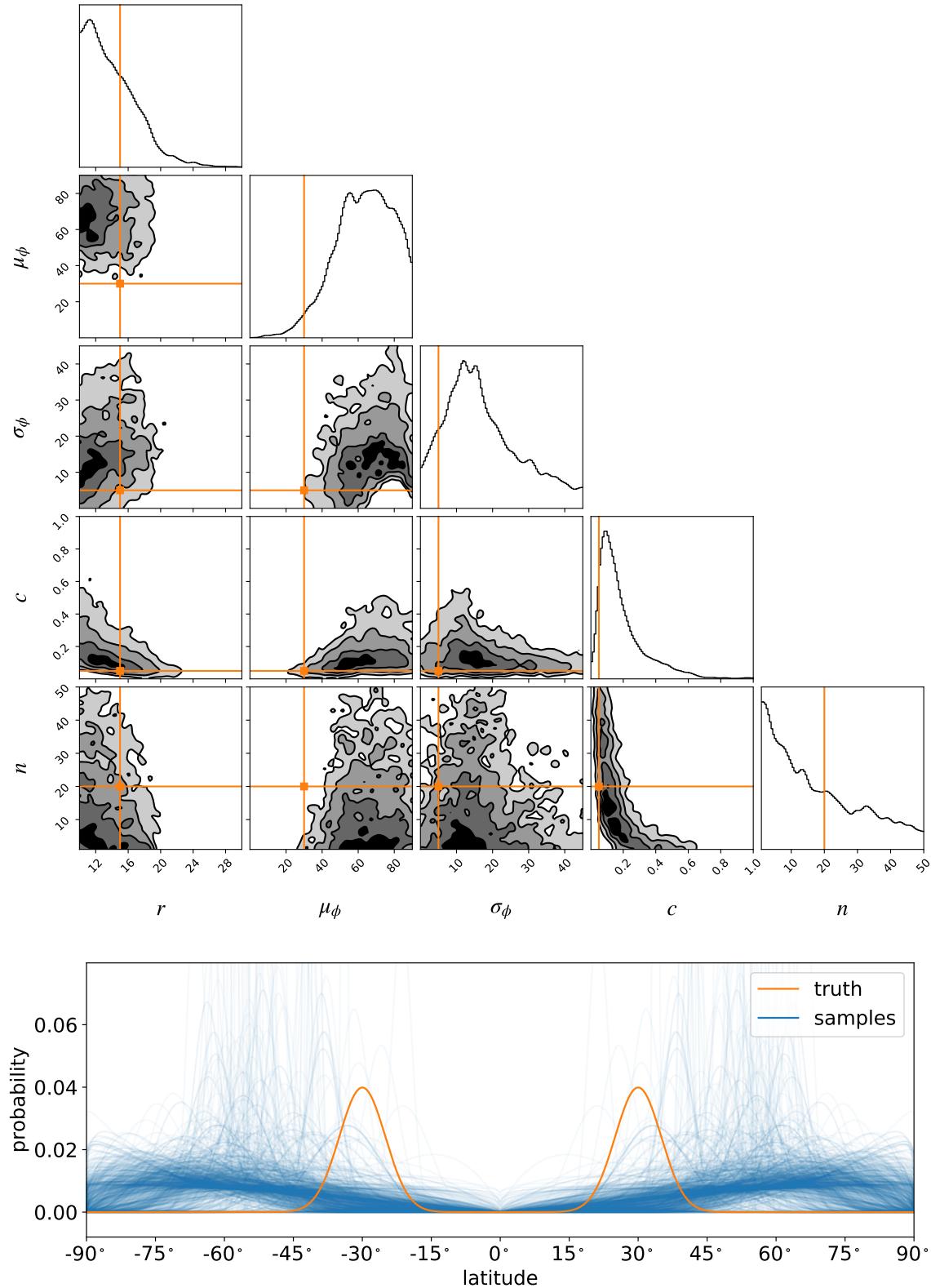
**Figure S27.** Same as Figure S25, but for equatorial spots with  $\mu_\phi = 0^\circ$ . Even though the model favors a bimodal distribution at low latitudes, the posterior strongly supports the presence of equatorial spots.



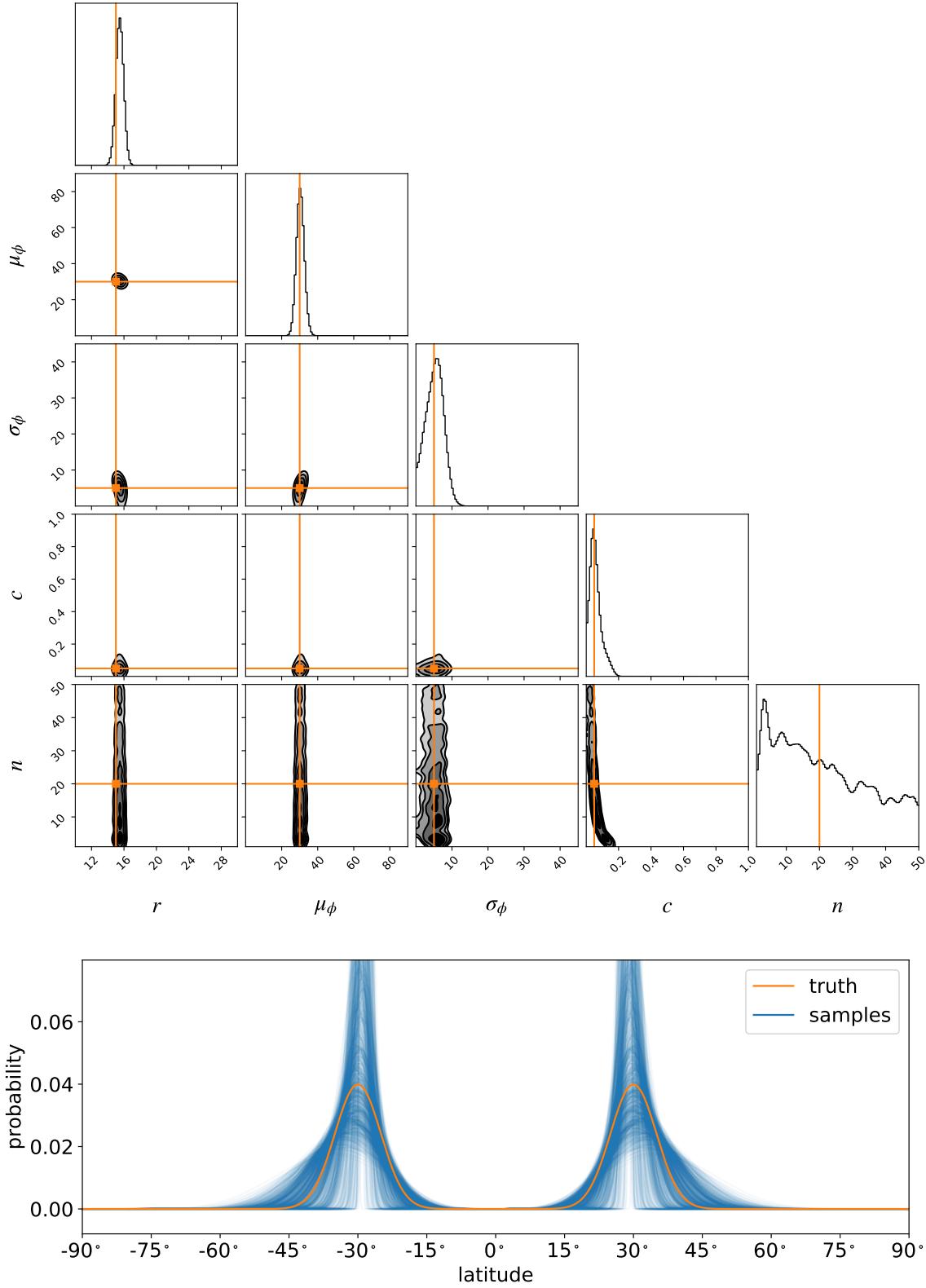
**Figure S28.** Same as Figure S25, but for isotropically-distributed spots. The posterior accurately captures the cosine-like distribution of spot latitudes.



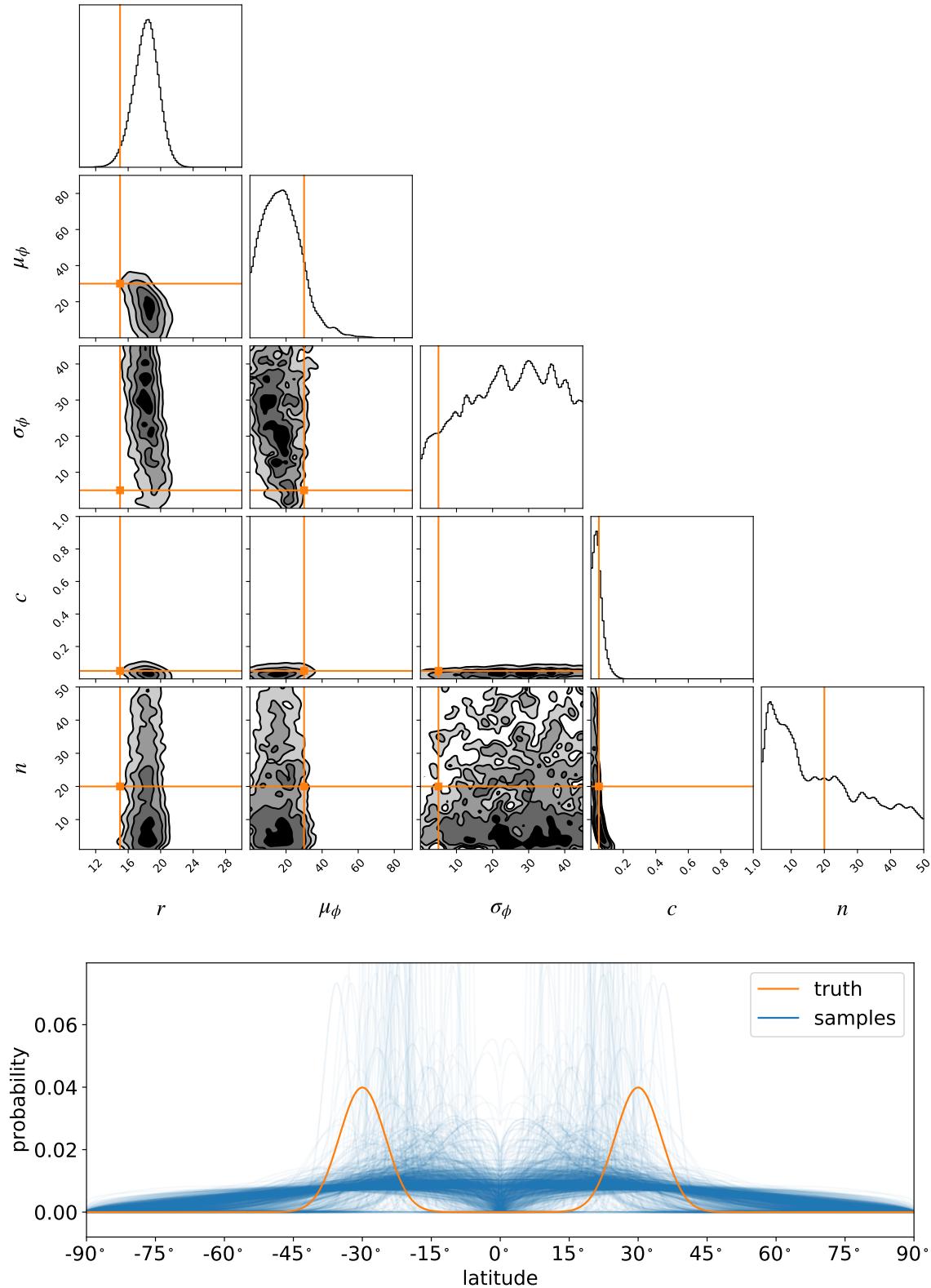
**Figure S29.** Same as Figures 9 and 10, but for high contrast ( $c = 1$ ) small ( $r = 3^\circ$ ) spots (significantly lower than the effective resolution of the model). The hatched regions in the posterior plots for the radius ( $r < 10^\circ$ ) are excluded by the prior, since the model cannot capture features that small. Despite this, the spot latitude distribution is still inferred correctly, although the spot contrast is off by more than  $10\sigma$ .



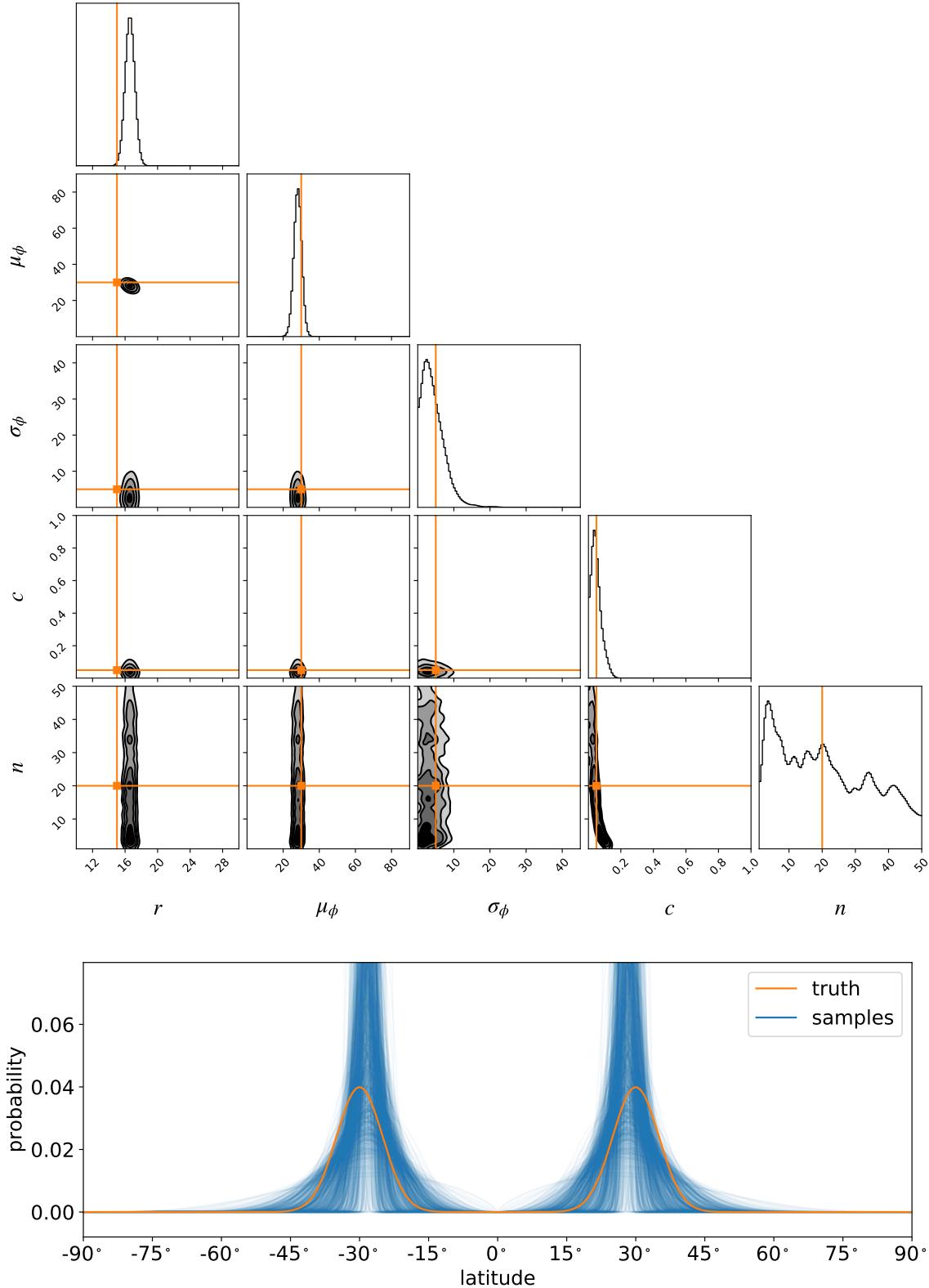
**Figure S30.** Same as Figures 9 and 10, but for inference based on a single light curve ( $M = 1$ ). The constraints on all of the parameters are dramatically weaker. 🤖



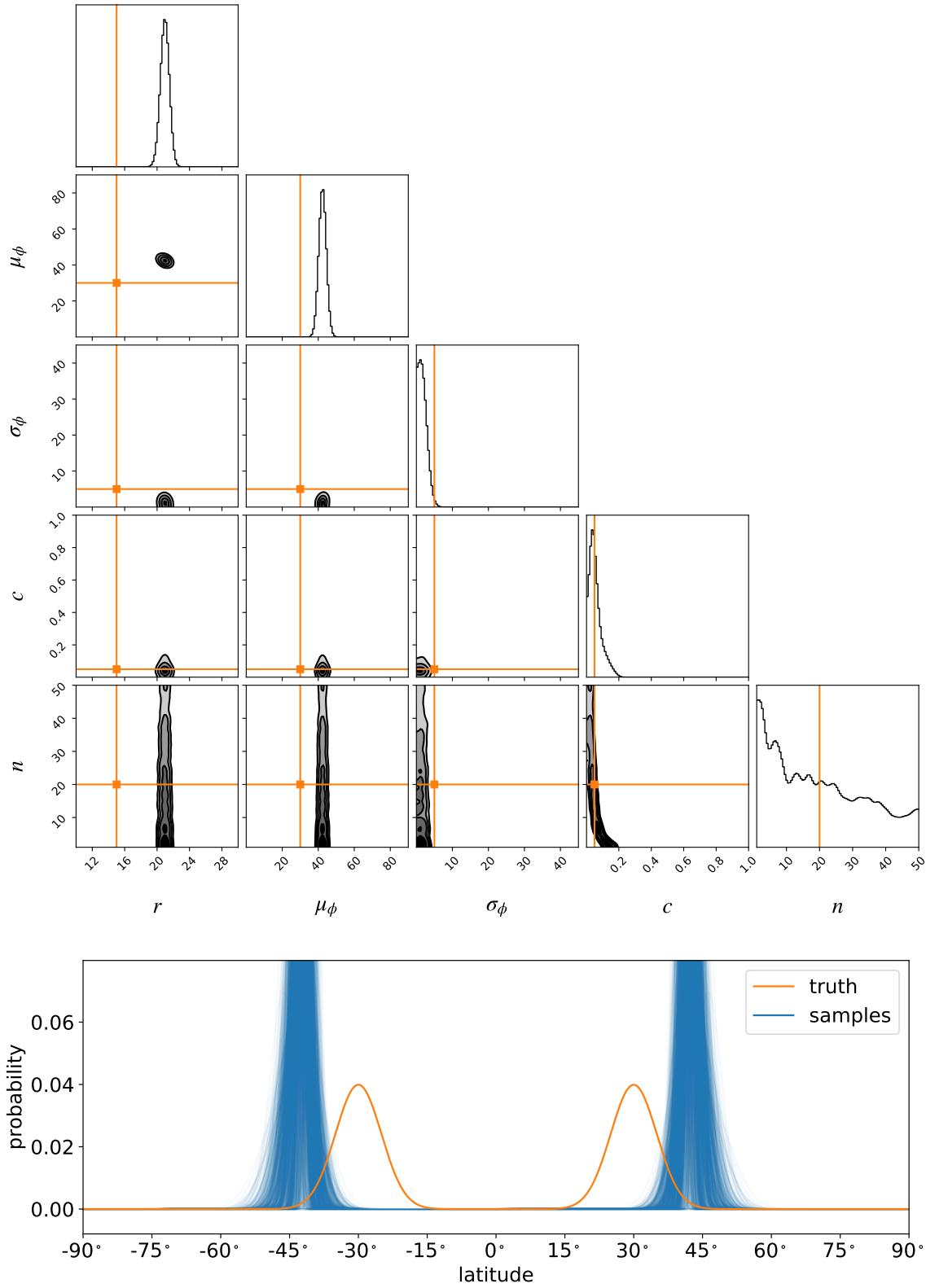
**Figure S31.** Same as Figures 9 and 10, but for inference based on one thousand light curves ( $M = 1,000$ ). The constraints on the radius and the latitude parameters are dramatically tighter. 



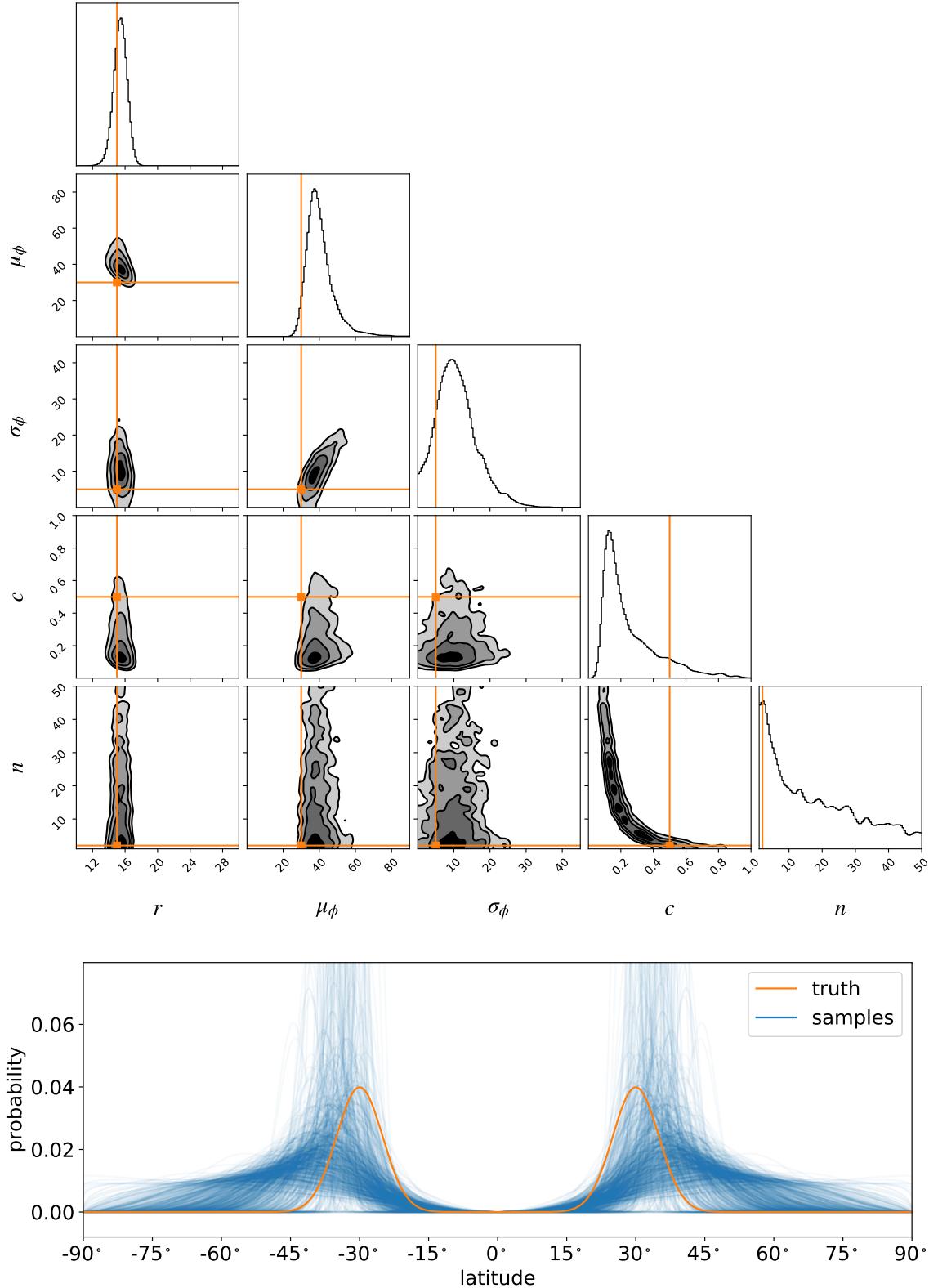
**Figure S32.** Same as Figures 9 and 10, but for stars with (assumed known) limb darkening coefficients  $u_1 = 0.50$  and  $u_2 = 0.25$ . Limb darkening makes it much harder to infer the variance of the distribution of star spot latitudes.



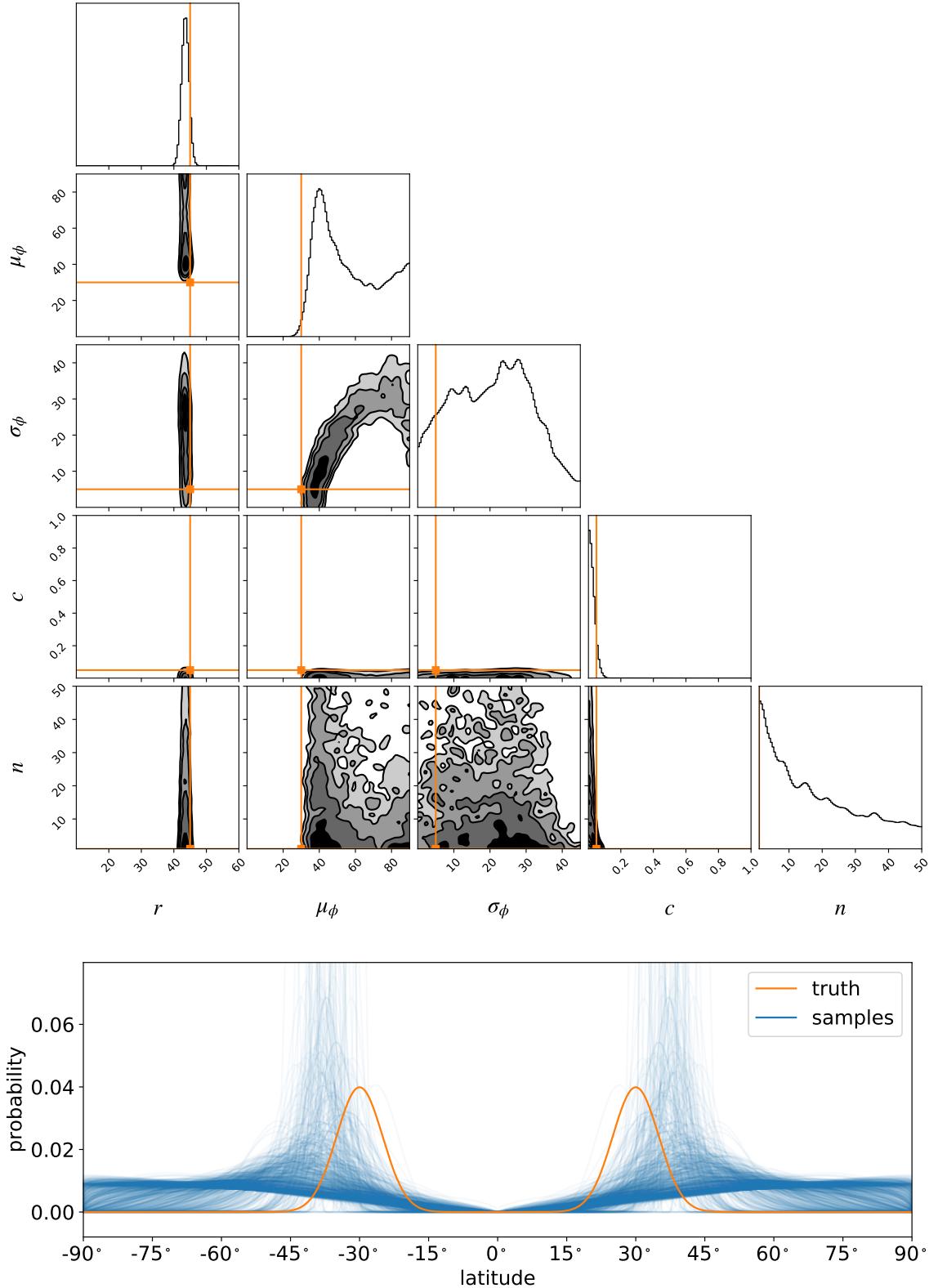
**Figure S33.** Same as Figure S32, but for an ensemble consisting of  $M = 1,000$  light curves. For a sufficiently large ensemble, it is possible to correctly infer the spot radii and latitudes in the presence of limb darkening.



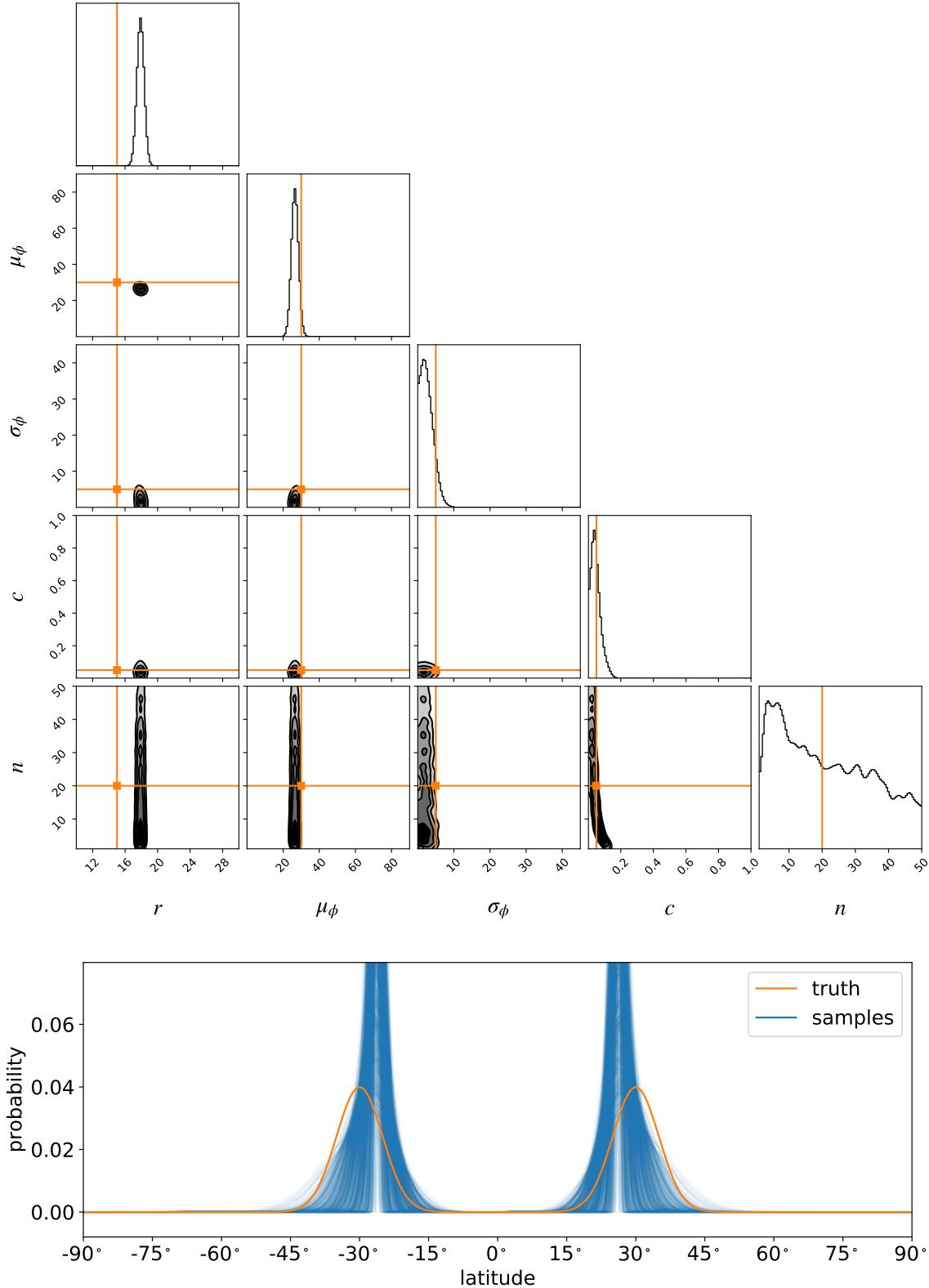
**Figure S34.** Same as Figure S32, but for  $M = 500$  light curves and assuming no limb darkening when doing inference. Neglecting limb darkening leads to significant bias in the inferred spot radii and to a lesser extent in the mean spot latitude.



**Figure S35.** Same as Figures 9 and 10, 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.

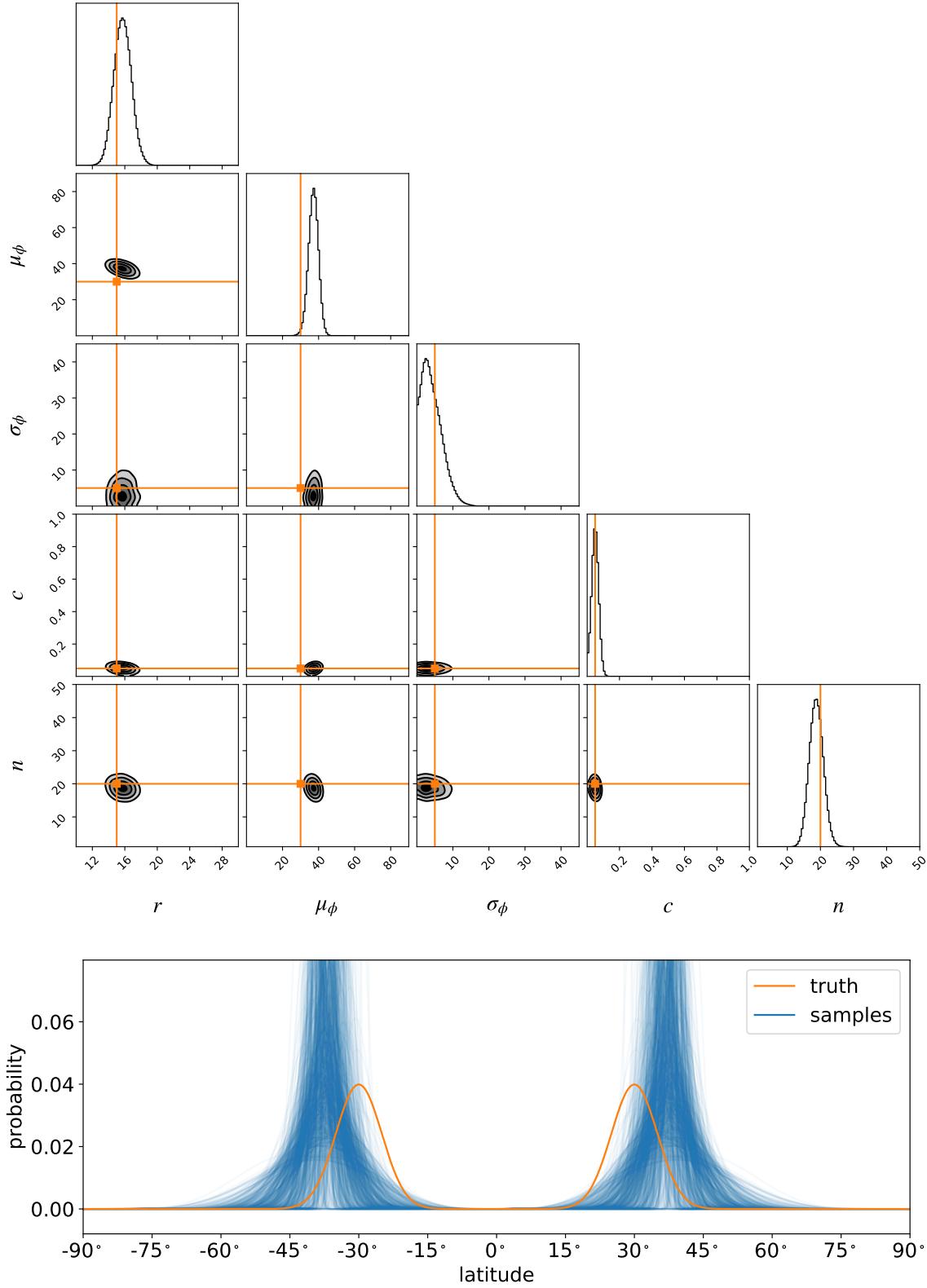


**Figure S36.** Same as Figures 9 and 10, but for a single ( $n = 1$ ) large ( $r \sim \mathcal{N}(45^\circ, 5^\circ)$ ) spot on each star. Our model correctly infers the larger radius, and infers the latitude within  $3\sigma$ , albeit with large uncertainty. 🤞



**Figure S37.** Same as Figures 9 and 10, but for stars with variance in their number of spots,  $n \sim \mathcal{N}(20, 3^2)$ , the spot radii,  $r \sim \mathcal{N}(15^\circ, 3^\circ)^2$ , and the spot contrasts,  $c \sim \mathcal{N}(0.05, 0.01^2)$ .





**Figure S38.** Same as Figures 9 and 10, but assuming the light curves are not normalized and the true baseline is known. Knowledge of the baseline breaks the  $c - N$  degeneracy and allows us to infer the total number of spots.