

<sup>1</sup> DRAFT VERSION MARCH 31, 2021  
Typeset using L<sup>A</sup>T<sub>E</sub>X **modern** style in AASTeX62

## Mapping stellar surfaces

### II: An interpretable Gaussian process model for light curves

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

<sup>3</sup> *Center for Computational Astrophysics, Flatiron Institute, New York, NY*

<sup>4</sup> *Virtual Planetary Laboratory, University of Washington, Seattle, WA*

<sup>5</sup> *Bay Area Environmental Research Institute, Moffett Field, CA*

<sup>6</sup> *NASA Ames Research Center, Moffett Field, CA*

## ABSTRACT

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

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

*Features:* open-source figures ; equation unit tests: 17 passed , 0 failed 

## 31 1. INTRODUCTION

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

44 A major source of stellar variability in both light curves and radial velocity datasets  
 45 is the modulation induced by magnetically-driven surface features like starspots ro-  
 46 tating in and out of view. While GPs excel at marginalizing over stellar rotational  
 47 variability, they have been less useful when the goal is to make inferences about the  
 48 actual source of this variability, such as the properties of starspots and the magnetic  
 49 processes that generate them. **While it is straightforward to derive posterior**  
 50 **constraints on the hyperparameters of an effective GP model for observa-**  
 51 **tions of a star, it is not clear what those constraints actually tell us about**  
 52 **the stellar surface.** Specifically, in all but a few restricted cases, there is no *first*  
 53 *principles* relationship between the descriptive parameters of a typical GP model (see  
 54 §2.1) and the physical properties of the stellar surface that is being observed. For  
 55 instance, it may be tempting to interpret the GP amplitude hyperparameter as some  
 56 measure of the spot contrast or the total number of spots, or the GP timescale hy-  
 57 perparameter as the spot lifetime, but there are no guarantees these interpretations  
 58 will hold in general. After all, the choice of kernel is quite often *ad hoc*, providing  
 59 an *effective*—as opposed to *interpretable*—description of the physics. There are two  
 60 important exceptions to this: asteroseismic studies, in which the the GP hyperparam-  
 61 eters can offer direct insight into the behavior of complex pulsation modes and thus  
 62 physical properties of the stellar interior (e.g., Brewer & Stello 2009; Foreman-Mackey  
 63 et al. 2017); and stellar rotation period studies, in which the period hyperparameter  
 64  $P$  can usually be associated with the rotation period of the star (e.g., Angus et al.  
 65 2018).<sup>1</sup> For spot-induced variability, on the other hand, GPs are usually used when  
 66 the variability itself is a nuisance parameter. For example, if the goal is to constrain  
 67 the properties of a transiting exoplanet or to search for a planetary signal in a radial  
 68 velocity dataset, a GP might be used to remove (or, better yet, to marginalize over)

<sup>1</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.

the stellar variability (e.g., Haywood et al. 2014; Rajpaul et al. 2015; Luger et al. 2017b). In this case, the physics behind the variability is irrelevant, so an effective model of this sort may be sufficient.

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

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

When the goal is to learn about the stellar surface, the common approach in the literature has not been to use GPs, but to explicitly *forward model* the surface. Such a model allows one to compute a stellar light curve or spectral timeseries conditioned on certain surface properties, a procedure that must then be inverted in order to constrain the surface given a dataset. We discussed this approach for rotational light curves of stars in Luger et al. (2021b) (hereafter Paper I), where we argued a unique solution

to the surface map of the star is not possible without the use of aggressive (and often *ad hoc*) priors. **The degeneracies at play make it effectively impossible for one to know the exact configuration of starspots and other features on the surface of a star from its rotational light curve alone.**

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

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

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

While our GP can be used to model light curves of individual stars, it is particularly useful for **ensemble analyses of light curves of many similar stars.** As we showed in Paper I, the joint information content of the light curves of many similar stars can be harnessed to constrain statistical properties of the surfaces of those stars, even in the presence of degeneracies that preclude knowledge about the surfaces of

individual stars. By “similar”, we do not mean stars that *look* similar, but whose spot properties are drawn from the same parent distribution. The parameters of this parent distribution are the ones we can constrain; these are also usually the physically interesting ones, such as the typical spot sizes or typical active latitudes and the variance in those quantities across the population. Ensembles may thus comprise light curves of stars in a narrow spectral type, metallicity, and rotation period bin, which we might reasonably expect to have *statistically* similar surfaces. We encourage readers to read Paper I to better understand this and other points regarding the information theory behind stellar rotational light curves.

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

Most of the math behind the algorithm is presented in the 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. The algorithm developed in this paper is fully implemented in the `starry_process` code, which is available on GitHub and is described in more detail in Luger et al. (2021a).

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. As in Paper I, the principal equations are accompanied by “unit tests”: `pytest`-compatible test scripts associated with the principal equations that pass (fail) if the equation is correct (wrong), in which case a clickable  () is shown next to the equation label. In most cases, the validity of an equation is gauged by comparison to a numerical solution. Like the figure scripts, the equation unit tests are run on Azure Pipelines upon every commit of the code.

## 2. A GAUSSIAN PROCESS FOR STARSPOTS

In this section, we provide a brief overview of Gaussian processes (§2.1) and spherical harmonics (§2.2), followed by an outline of the derivation of our interpretable GP (§2.3). This derivation boils down to computing the mean and covariance of the stellar flux conditioned on certain physical properties of the star and its starspot distribution. In §2.4 and §2.5 we derive useful extensions of the model. **For convenience, we summarize the results of this entire section in §2.6.** 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.

### 2.1. Brief overview of Gaussian processes

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

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

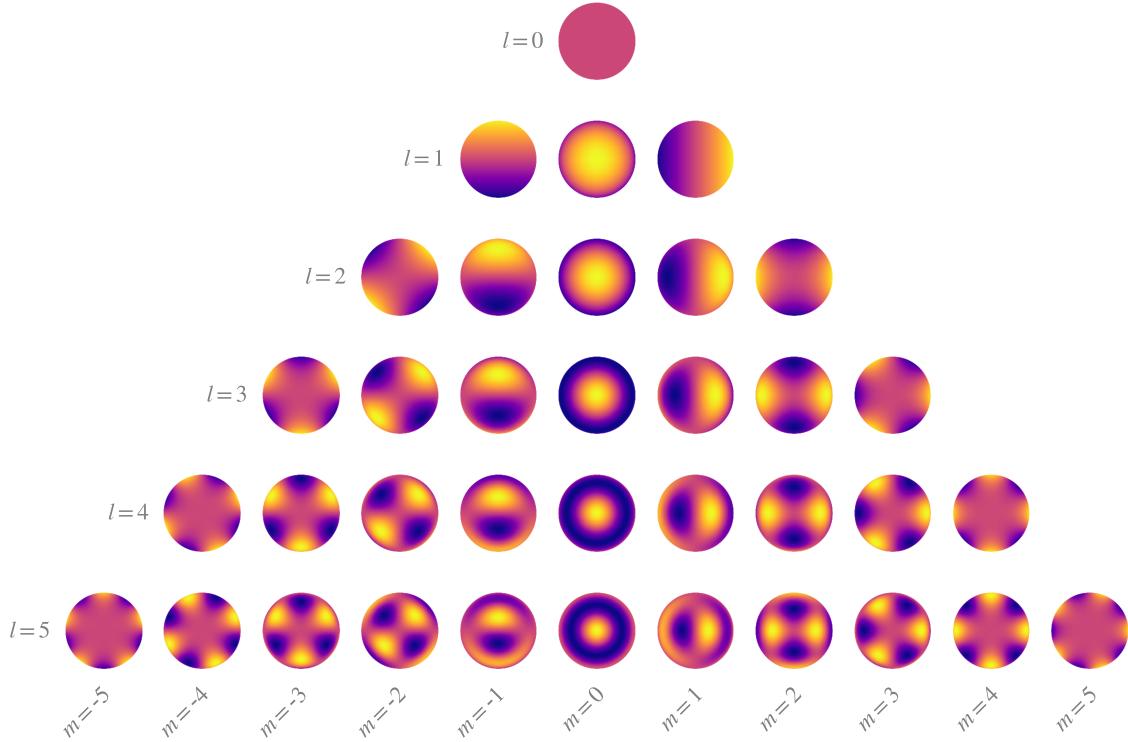
where the elements of the mean and covariance are given by  $\mu_i = m(t_i)$  and  $\Sigma_{i,j} = k(t_i, t_j)$ , respectively.<sup>2</sup> Because of this relationship to multivariate Gaussians, GPs are easy to sample from.<sup>3</sup> But, as we alluded to earlier, the real showstopper is the application of GPs to inference problems. Multivariate Gaussian distributions have a closed-form (marginal) likelihood function, so it is easy to compute the probability of one’s data conditioned on a given value of  $\mu$  and  $\Sigma$  (i.e., the “likelihood”; see Equation 14 below). This can in turn be maximized to infer the optimal values of the model parameters or used in a numerical sampling scheme to compute the probability of those parameters given the data (i.e., the “posterior”). Thanks to modern computer architectures, linear algebra packages, and GP algorithms, evaluating the GP likelihood may typically be done in a fraction of a second for a reasonably-sized 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  $\mathbf{f}$ . 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

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

<sup>2</sup> In this paper, we will use blackboard font (i.e.,  $\mathbf{f}$ ) to denote random variables and serif font (i.e.,  $f$ ) to denote particular realizations of those variables. See Appendix A for a detailed explanation of our notation.

<sup>3</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)`.



**Figure 1.** 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. [Cloud](#)

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

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

## 232 2.2. Spherical harmonics

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

to describe starspot distributions<sup>4</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

$$\mathbf{f} = \mathbf{1} + \mathcal{A}(I, P, \mathbf{u}) \mathbf{y}, \quad (3)$$

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  $\mathbf{u}$ , as well as the observation times (see Appendix B for details).

### 2.3. Computing the GP

Let  $\mathbf{f} = (\mathbf{f}_0 \ \mathbf{f}_1 \ \dots \ \mathbf{f}_{K-1})^\top$  denote a random vector of  $K$  flux measurements at times  $\mathbf{t} = (t_0 \ t_1 \ \dots \ t_{K-1})^\top$ , defined in units such that a star with no spots on it will have unit flux.<sup>5</sup> 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:

$$\mathbf{f}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \sim \mathcal{N}\left(\boldsymbol{\mu}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet), \boldsymbol{\Sigma}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet)\right) \quad (4)$$

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:

$$\boldsymbol{\mu}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = \mathbb{E}\left[\mathbf{f} \mid I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet\right] \quad (5)$$

$$\boldsymbol{\Sigma}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = \mathbb{E}\left[\mathbf{f} \mathbf{f}^\top \mid I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet\right] - \boldsymbol{\mu}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \boldsymbol{\mu}^\top(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet). \quad (6)$$

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

$$\boldsymbol{\mu}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = \mathbf{1} + \mathcal{A}(I, P, \mathbf{u}) \boldsymbol{\mu}_y(\boldsymbol{\theta}_\bullet) \quad (7)$$

$$\boldsymbol{\Sigma}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = \mathcal{A}(I, P, \mathbf{u}) \boldsymbol{\Sigma}_y(\boldsymbol{\theta}_\bullet) \mathcal{A}^\top(I, P, \mathbf{u}), \quad (8)$$

<sup>4</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 §4.1, where we show that our model is useful even when applied to stars with spots smaller than the effective resolution of the GP.

<sup>5</sup> Note, importantly, that these are not the units we observe in! See Paper I and §2.5 below.

<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.

261 where

$$\mu_y(\theta_\bullet) = E[y | \theta_\bullet] \quad (9)$$

$$\Sigma_y(\theta_\bullet) = E[yy^\top | \theta_\bullet] - \mu_y(\theta_\bullet)\mu_y^\top(\theta_\bullet) \quad (10)$$

262 are the mean and covariance of the distribution over spherical harmonic coefficient  
 263 vectors  $\mathbf{y}$ . The bulk of the math in this paper (Appendix C) is devoted to computing  
 264 the expectations in the expressions above, which are given by the integrals

$$E[y | \theta_\bullet] = \int y(\mathbf{x}) p(\mathbf{x} | \theta_\bullet) d\mathbf{x} \quad (11)$$

$$E[yy^\top | \theta_\bullet] = \int yy^\top(\mathbf{x}) p(\mathbf{x} | \theta_\bullet) d\mathbf{x}, \quad (12)$$

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

$$\theta_\bullet = (n \ c \ \mu_\phi \ \sigma_\phi \ r)^\top, \quad (13)$$

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

288 In this paper, we assume that the parameters  $\theta_\bullet$  described above are the *physically interesting*  
 289 ones. That is, given a light curve  $\mathbf{f}$  or an ensemble of  $M$  light curves of statistically similar stars  
 290  $(\mathbf{f}_0 \ \mathbf{f}_1 \ \dots \ \mathbf{f}_{M-1})^\top$ , 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* (and, if  $M > 1$ , on *every star* in the ensemble). 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  $\mathbf{u}$ ):

$$\begin{aligned} \ln \mathcal{L}_m(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = & -\frac{1}{2} \mathbf{r}_m^\top(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) [\Sigma(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) + \mathbf{C}_m]^{-1} \mathbf{r}_m(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \\ & -\frac{1}{2} \ln |\Sigma(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) + \mathbf{C}_m| - \frac{K}{2} \ln(2\pi), \end{aligned} \quad (14)$$

where

$$\mathbf{r}_m(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \equiv \mathbf{f}_m - \boldsymbol{\mu}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \quad (15)$$

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  $K$  is the number of data points in each light curve.<sup>7</sup> 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

$$\ln \mathcal{L}(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet) = \sum_m \ln \mathcal{L}_m(I, P, \mathbf{u}, \boldsymbol{\theta}_\bullet). \quad (16)$$

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.

#### 2.4. Marginalizing over inclination

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

<sup>7</sup> In Equation (14) we implicitly assume all stars in the ensemble are observed at the same set of times  $\mathbf{t}$ . If this is not the case, the mean and covariance of the GP for each star must be computed from Equations (5) and (6) with the flux design matrix  $\mathbf{A}(\mathbf{t}_m)$  evaluated at the particular observation times  $\mathbf{t}_m$ .

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

$$N = 4M + 5, \quad (17)$$

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  $\theta_\bullet$  (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  $\theta_\bullet$ . If one knew  $\theta_\bullet$ , 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

$$\begin{aligned} \mu(P, \mathbf{u}, \theta_\bullet) &= E[\mathbf{f} \mid P, \mathbf{u}, \theta_\bullet] \\ &= \mathbf{1} + \mathbf{e}_I \end{aligned} \quad (18)$$

$$\begin{aligned} \Sigma(P, \mathbf{u}, \theta_\bullet) &= E[\mathbf{f} \mathbf{f}^\top \mid P, \mathbf{u}, \theta_\bullet] - \mu(P, \mathbf{u}, \theta_\bullet) \mu^\top(P, \mathbf{u}, \theta_\bullet) \\ &= \mathbf{E}_I - \mathbf{e}_I \mathbf{e}_I^\top \end{aligned} \quad (19)$$

349 where we define the inclination first moment integral

$$\mathbf{e}_I \equiv \int \mathcal{A}(\mathbb{I}, P, \mathbf{u}) E\left[\mathbf{y} \mid \boldsymbol{\theta}_{\bullet}\right] p(\mathbb{I}) d\mathbb{I} \quad (20)$$

350 and the inclination second moment integral

$$\mathbf{E}_I \equiv \int \mathcal{A}(\mathbb{I}, P, \mathbf{u}) E\left[\mathbf{y} \mathbf{y}^T \mid \boldsymbol{\theta}_{\bullet}\right] \mathcal{A}^T(\mathbb{I}, P, \mathbf{u}) p(\mathbb{I}) d\mathbb{I}, \quad (21)$$

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

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

$$\begin{aligned} \boldsymbol{\mu}(P, \mathbf{u}, \boldsymbol{\theta}_{\bullet}) &= (1 + e_I) \mathbf{1} \\ &\equiv \mu \mathbf{1}, \end{aligned} \quad (22)$$

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

### 360 2.5. Normalization correction

361 In Paper I we discussed a subtle but important point about stellar light curves:  
 362 the common procedure of normalizing light curves to their mean or median level  
 363 changes the covariance structure of the data, since it correlates all the observations  
 364 in a nontrivial way. When normalizing a light curve by the mean,<sup>8</sup> the operation we  
 365 perform is

$$\tilde{\mathbf{f}} = \frac{\mathbf{f}}{\langle \mathbf{f} \rangle}, \quad (23)$$

366 where  $\tilde{\mathbf{f}}$  is the normalized, unit-mean light curve,  $\mathbf{f}$  is the measured light curve (in  
 367 detector counts), and  $\langle \mathbf{f} \rangle$  is the *sample* mean: i.e., the average value of a given star's  
 368 light curve (which we model as a sample from our GP). This may be close to but  
 369 is in general different from the *process* mean,  $\boldsymbol{\mu}(P, \mathbf{u}, \boldsymbol{\theta}_{\bullet})$ , since the mean of a draw  
 370 from the GP is itself normally distributed with a variance that scales with the GP  
 371 variance.<sup>9</sup>

372 When modeling normalized light curves, we must correct our expression for the  
 373 covariance matrix  $\Sigma$  of the GP. Computing the new covariance matrix is tricky, es-  
 374 specially because the normalized process is *not* strictly Gaussian: the distribution of

<sup>8</sup> In practice, the expressions derived here also work well for median-normalized light curves, since the distribution of the GP sample median is usually close to the distribution of the sample mean.

<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).

normalized light curves 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) amplitude of the data as a multiplicative latent variable. However, this would require an extra parameter *for every light curve*, so the computational savings we achieved by marginalizing out the inclination would be gone. Fortunately, in practice, the variance of a stellar light curve is usually small compared to its mean: stellar variability amplitudes are typically at the level of a few percent or lower. When this is the case, the probability of drawing a GP sample whose mean is close to zero is extremely small, and we can make use of the approximate expression derived in Luger (2021) for the covariance of a normalized Gaussian process:

$$\tilde{\Sigma} \approx \frac{A}{\mu^2} \Sigma + z \left( (A + B) (\mathbf{1} - \mathbf{q}) (\mathbf{1} - \mathbf{q})^\top - A \mathbf{q} \mathbf{q}^\top \right), \quad \checkmark \quad (24)$$

where

$$z \equiv \frac{\langle \Sigma \rangle}{\mu^2} \quad (25)$$

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

$$A \equiv \sum_{i=0}^{i_{\max}} \frac{(2i+1)!}{2^i i!} z^i \quad (26)$$

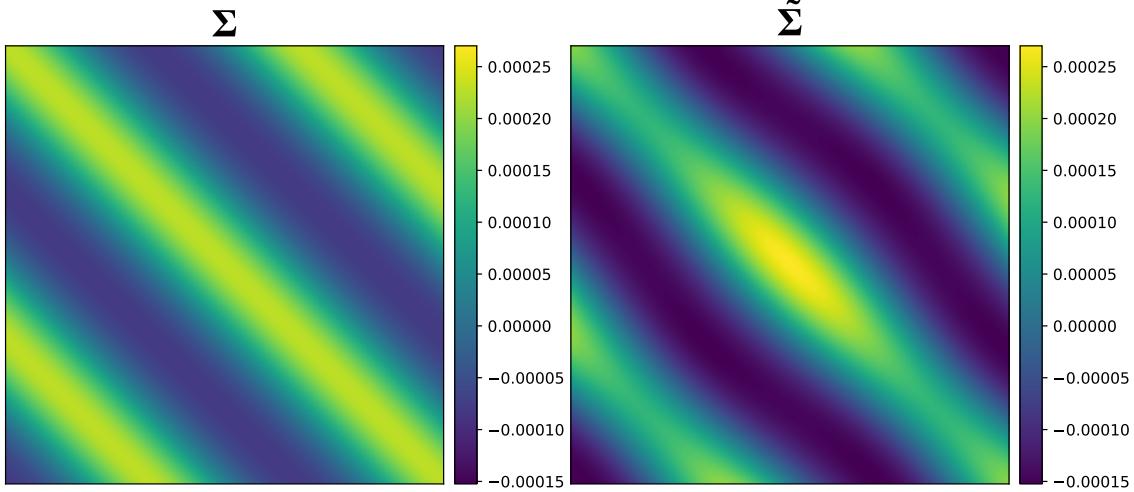
$$B \equiv \sum_{i=0}^{i_{\max}} \frac{2i(2i+1)!}{2^i i!} z^i, \quad (27)$$

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

$$\tilde{\mathbf{f}}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \sim \mathcal{N}\left(\mathbf{1}, \tilde{\Sigma}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet)\right). \quad (28)$$

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  $z \lesssim 0.02$ ,



**Figure 2.** 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 (24). 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.

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 2 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.

## 2.6. 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 (§2.4) and the light curves are normalized to their means (§2.5), which is likely to be the primary use case for our algorithm.

We model the mean-normalized flux  $\tilde{\mathbf{f}}$  (Equation 23) as a Gaussian process:

$$\tilde{\mathbf{f}}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet) \sim \mathcal{N}\left(\mathbf{1}, \tilde{\Sigma}(P, \mathbf{u}, \boldsymbol{\theta}_\bullet)\right). \quad (29)$$

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

$$\boldsymbol{\theta}_\bullet = (n \ c \ \mu_\phi \ \sigma_\phi \ r)^\top, \quad (30)$$

consisting of the number of spots  $n$ , their contrast  $c$  (the fractional intensity difference between the background and the spot), 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 24), which is a straightforward correction to the true covariance of the process, accounting for changes in scale and phase introduced by the common process of normalizing light curves to a mean of unity. It depends on the true (constant) mean  $\mu$  and true covariance  $\Sigma$ , given by Equations (18) and (19), respectively. Those expressions in turn depend on the inclination expectation integrals  $e_I$  (Appendix D.2) and  $\mathbf{E}_I$  (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 the companion paper (Luger et al. 2021a), we discuss our implementation of the algorithm in a user-friendly Python package.

## 2.7. An example

A concrete example of the GP derived above is presented in Figure 3, where we show random samples from the process evaluated up to spherical harmonic degree  $l_{\max} = 15$  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

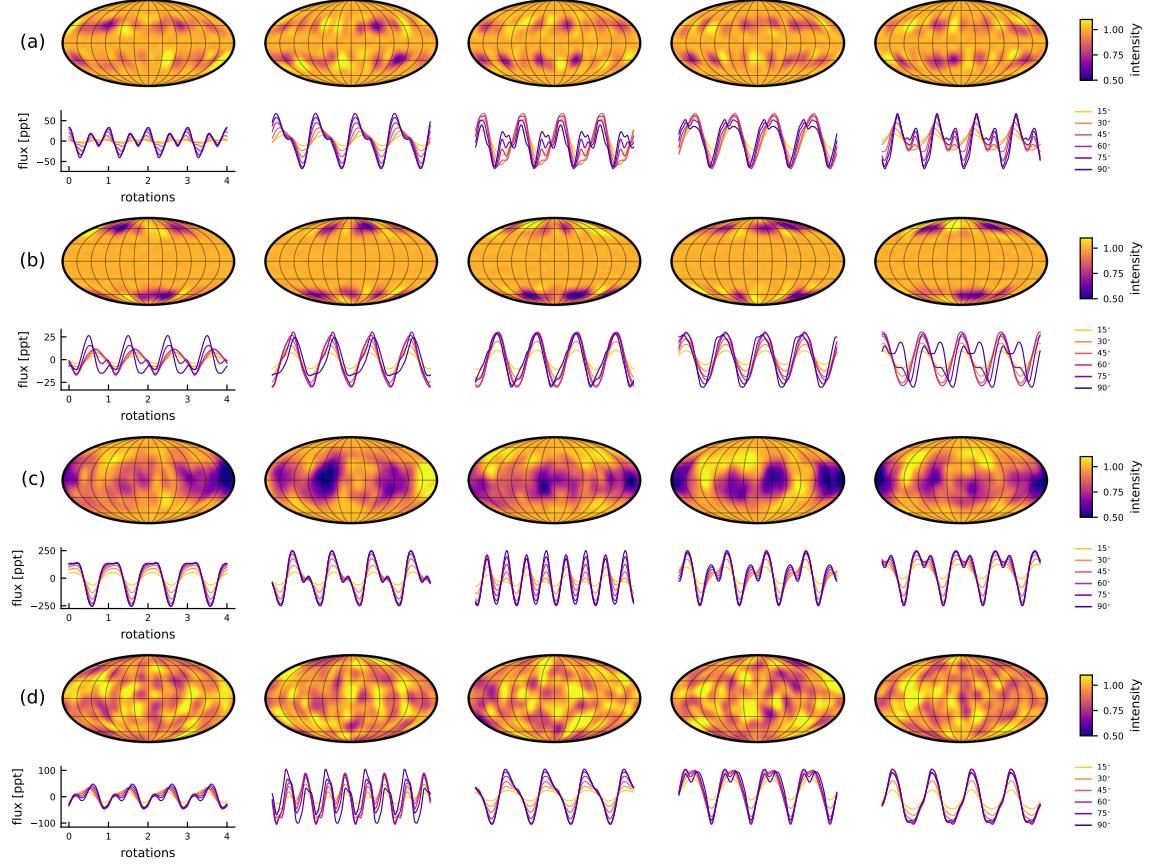
$$(n \ c \ \mu_\phi \ \sigma_\phi \ r)^\top = (10.0 \quad 0.10 \quad 30.0 \quad 5.00 \quad 10.0)^\top \quad (31a)$$

$$= (10.0 \quad 0.10 \quad 60.0 \quad 5.00 \quad 10.0)^\top \quad (31b)$$

$$= (10.0 \quad 0.05 \quad 0.00 \quad 5.00 \quad 30.0)^\top \quad (31c)$$

$$= (20.0 \quad 0.10 \quad 0.00 \quad 40.0 \quad 10.0)^\top. \quad (31d)$$

<sup>10</sup> The exception to this is the normalization correction (§2.5), which depends on a fast-to-evaluate series and thus adds negligible overhead to the computation.



**Figure 3.** Five random samples from the GP prior (columns) conditioned on four different values of  $\theta_\bullet$  (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](#)

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.

- 464    2. *There is significant variance in contrast from one spot to another*, even though  
 465    our model implicitly treats  $c$  as constant. Within spots, the contrast is also not  
 466    constant, even though (again) our model implicitly treats it as such.
- 467    3. *There is ringing in the background*. This is apparent to some extent in row (b),  
 468    where there are small fluctuations in the brightness at low latitudes where no  
 469    spots are present.
- 470    4. *There aren't exactly 10 (or 20) spots in those maps*. This is most obvious in  
 471    row (c), where only a few large distinct spots, plus maybe a few smaller ones,  
 472    are visible.
- 473    5. *There are bright spots in addition to dark spots*. This may be the most glaring  
 474    issue. We explicitly model spots as being dark, and yet there are (almost)  
 475    just as many bright spots, particularly in row (d). While bright spots (such as  
 476    plages) certainly exist in reality, we did not explicitly ask for them here!

477    While these may appear to be critical shortcomings of our model, it is important to  
 478    keep in mind that a model consisting of discrete, circular, constant contrast spots  
 479    is likely just as far (or perhaps even farther) from the truth. In fact, points (1)  
 480    and (2) above suggest our model is more flexible than the discrete spot model and  
 481    thus (potentially) better suited to modeling real stellar surfaces. Points (3), (4),  
 482    and (5), on the other hand, are more concerning, since they are due, respectively, to  
 483    truncation error in the spherical harmonic expansion, to an intrinsic limitation of our  
 484    Gaussian approximation, and to the fact that Gaussian distributions are symmetric  
 485    about the mean: a positive deviation is just as likely as a negative deviation of  
 486    the same magnitude.<sup>11</sup> However, as we have argued before, the true power of the  
 487    GP is in its applicability to inference problems. In other words, while our GP has  
 488    some undesirable features when used as a prior sampling distribution, the real test  
 489    of the GP is when it is faced with data in an inference setting. As long as the data  
 490    is sufficiently informative, it does not matter that the prior has finite support for  
 491    unphysical configurations, as those will be confidently rejected.

492    In §3 below, we exhaustively test the performance of our GP as an inference tool  
 493    when used to model synthetic light curves. We will show that, despite the issues  
 494    raised above, the GP is in general unbiased and correctly estimates the posterior  
 495    variance when used to infer the spot properties  $\theta_*$ .

496    But before we dive into calibration tests, it is worth pausing for a moment to take  
 497    another look at Figure 3. While we focused on the shortcomings of the GP as a prior in  
 498    the discussion above, it is important to appreciate that it even works in the first place!  
 499    A Gaussian process, after all, is a non-parameteric *process* describing a smooth and  
 500    continuous function only via its covariance structure. The GP knows nothing about

<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.

501 the existence of discrete spots—only how any two points on the surface are correlated.  
 502 Because spherical harmonics are smooth functions with support over the entire sphere,  
 503 the GP also does not know about features restricted to certain latitudes; in fact, in  
 504 most applications of GPs to mapping problems in astronomy (such as in models of  
 505 the cosmic microwave background; e.g., Wandelt 2012), the process is assumed to be  
 506 isotropic, with no preferred angular direction. However, by prescribing the correct  
 507 structure to the covariance matrix, we are able to approximately model compact  
 508 spot-like features with given sizes and restricted to particular latitudes.

### 509 3. CALIBRATION

#### 510 3.1. *Why we need calibration tests*

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

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

#### 529 3.2. *Setup*

530 We seek to demonstrate that our model is correctly calibrated by testing it on  
 531 synthetic data, which we generate as follows. For each of  $M$  synthetic light curves  
 532 in a given ensemble, we create a rectangular ( $150 \times 300$ ) latitude-longitude grid of  
 533 surface intensity values, all initialized to zero. We then add  $n$  spots to this grid,  
 534 each of fractional contrast  $c$  and radius  $r$  centered at latitude  $\phi$  and longitude  $\lambda$ , by  
 535 decreasing the intensity at all points within an angular distance  $r$  (measured along  
 536 the surface of the sphere) by an amount  $c$ . In order to compute the corresponding  
 537 light curve, we expand the surface in spherical harmonics, although at much higher  
 538 degree ( $l_{\max}^{(0)} = 30$ ) than the degree we will use in the inference step ( $l_{\max} = 15$ )

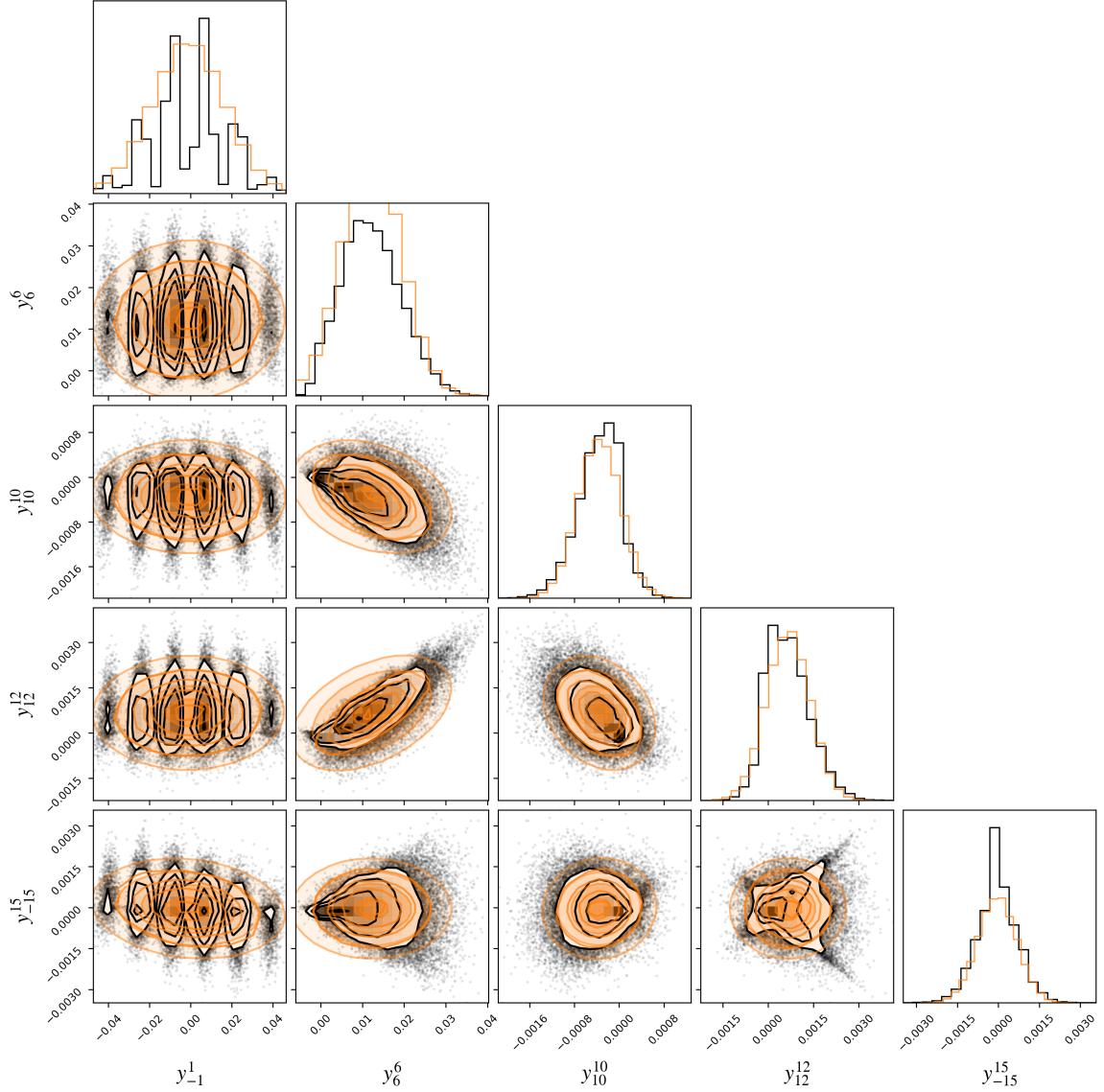
Symbol	Description	Value
$n$	number of spots	$\sim \mathcal{N}(20, 0^2)$
$c$	spot contrast	$\sim \mathcal{N}(0.05, 0^2)$
$\phi$	spot latitude	$\sim \mathcal{N}(30^\circ, 5^\circ)^2$
$\lambda$	spot longitude	$\sim \mathcal{U}(0^\circ, 360^\circ)$
$r$	spot radius	$\sim \mathcal{N}(15^\circ, 0^\circ)^2$
$I$	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.

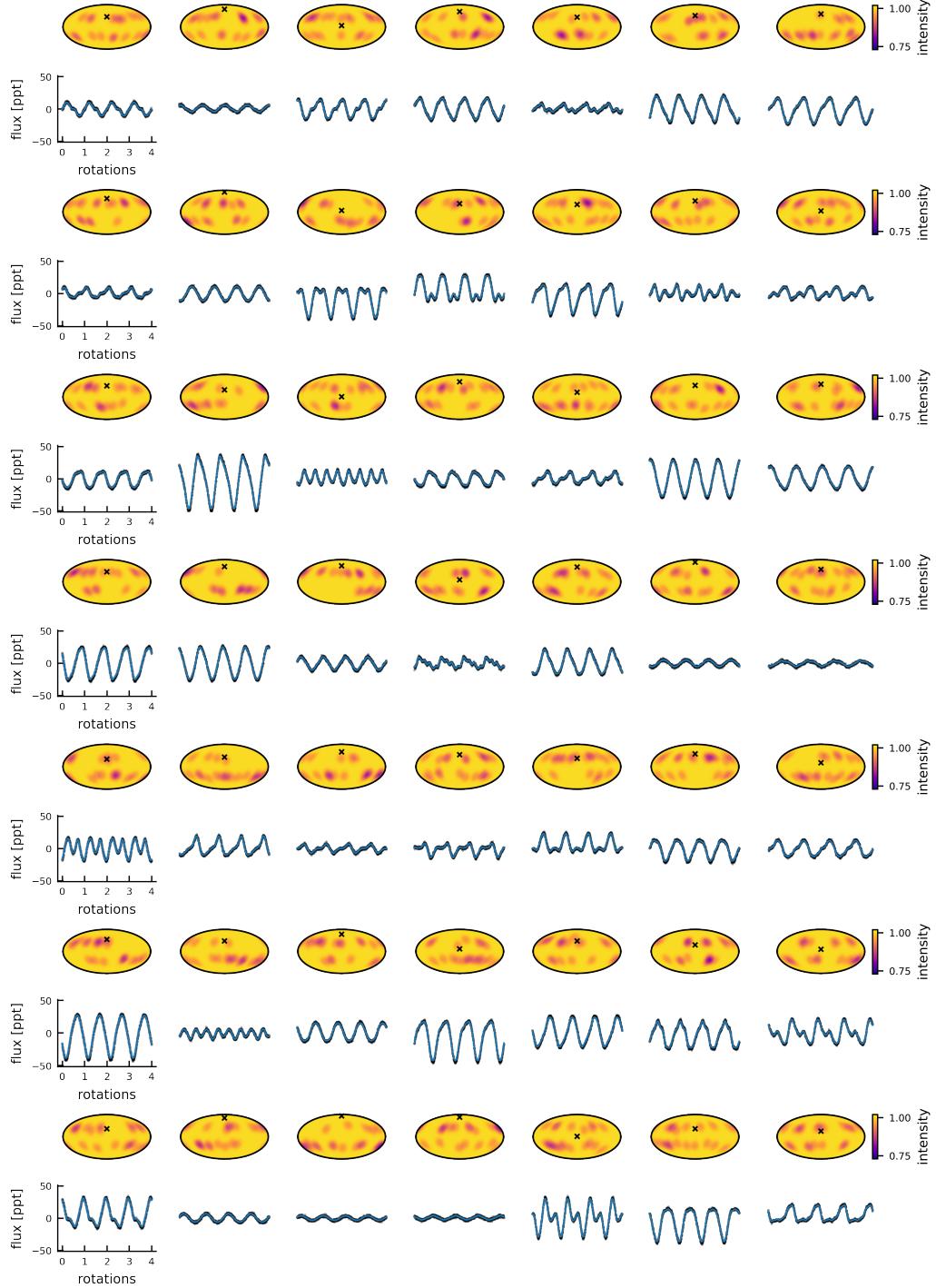
539 to minimize potential ringing effects or other artefacts in the synthetic data. For  
 540 reference, the chosen degree  $l_{\max} = 30$  is large enough to resolve features on the order  
 541 of  $180^\circ/30 = 6^\circ$  across, but small enough that the algorithm for computing the light  
 542 curve is numerically stable.<sup>12</sup> We compute the light curve at  $K$  points equally spaced  
 543 over a baseline  $\Delta t$  using the `starry` algorithm (Appendix B), assuming an inclination  
 544  $I$ , a rotational period  $P$ , and limb darkening coefficients  $\mathbf{u}$ . Finally, we divide the light  
 545 curve by the mean and add Gaussian noise with standard deviation  $\sigma_f$  to emulate  
 546 photon noise.<sup>13</sup> The default values/distributions of each of the parameters mentioned  
 547 above are given in Table 1. Some, like the number of spots, their contrast, etc., are  
 548 drawn from fiducial distributions, while others, like the photometric uncertainty, the  
 549 rotational period, and the limb darkening coefficients, are fixed across all  $M$  light  
 550 curves in an ensemble. These fixed values are not realistic, but they greatly speed  
 551 up the inference step, since they allow us to invert a single covariance matrix for all  
 552 light curves when computing the log likelihood.

<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 14). 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 4.** 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. 



**Figure 5.** 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.

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

### 559                 3.3. *Inference*

560     We use our Python-based implementation of the GP (Luger et al. 2021a) to per-  
 561     form inference on the synthetic datasets. For simplicity, we assume we know the value  
 562     of the period  $P$ , which is fixed at unity for all stars, as well as the value of the limb  
 563     darkening coefficients (fixed at zero for the default run). In practice these will not be  
 564     known exactly; we discuss this further in §5.3. Since we explicitly marginalize over  
 565     the inclinations of individual stars, the only quantities we must constrain are the five  
 566     parameters in the spot parameter vector  $\theta_\bullet$ . (Equation 13). We experimented with  
 567     three different methods for doing posterior inference on our synthetic datasets: No-  
 568     U Turn Sampling, a variant of Hamiltonian Monte Carlo (NUTS; Duane et al. 1987;  
 569     Hoffman & Gelman 2011), automatic differentiation variational inference (ADVI; Ku-  
 570     cukelbir et al. 2016; Blei et al. 2016), and nested sampling (Skilling 2004, 2006). We  
 571     obtained the best performance using the nested sampling algorithm implemented in  
 572     the `dynesty` package (Speagle 2020), so that is what we will use below.

573     Our sampling parameters are the number of spots  $n$ , their contrast  $c$ , their radius  $r$ ,  
 574     and the Beta distribution parameters  $a$  and  $b$  describing their distribution in latitude.  
 575     As we discuss in Appendix C.2, the parameters  $a$  and  $b$  are easier to sample in than  
 576     the mode  $\mu_\phi$  and standard deviation  $\sigma_\phi$ , provided we account for the Jacobian of the  
 577     transformation (Equation C71) in our log probability function, which maps a uniform  
 578     prior on  $a$  and  $b$  to a uniform prior on  $\mu_\phi$  and  $\sigma_\phi$ .

579     We place uniform priors on all five quantities, with support in  $1 \leq n \leq 50$ ,  
 580      $0 < c \leq 1$ ,  $10^\circ \leq r \leq 30^\circ$ ,  $0 \leq a \leq 1$ , and  $0 \leq b \leq 1$ . Note that while  $n$   
 581     formally represents an integer, its effect on the GP is a scaling of the covariance (see  
 582     Equation C19); as such, it has support over all real numbers within the bounds listed  
 583     above. We *could* restrict it to integer values, but this would make sampling quite  
 584     tricky. Moreover, in practice it is useful to allow for noninteger values to add some  
 585     flexibility to the model; we discuss this in more detail in §4.4.

586     We use Equation (14) as our log likelihood term, adding the log of the absolute  
 587     value of Equation (C71) to enforce a uniform prior on  $\mu_\phi$  and  $\sigma_\phi$ . As we mentioned  
 588     above, the fact that  $P$ ,  $\mathbf{u}$ , and  $\sigma_f$  are shared among all  $M$  light curves means that  
 589      $\Sigma + \mathbf{C}_m$  is the same for all of them, greatly speeding up the likelihood evaluation since  
 590     we need only invert (or factorize) it a single time per sample.<sup>14</sup> Our covariance is the

<sup>14</sup> Re-factorization would incur a computational cost on the order of 10ms per target per likelihood evaluation for the datasets we consider here.

591 covariance of the *normalized* process, given by Equation (24). Since we only consider  
 592 light curves with variability limited to a few percent or less, the approximation for  $\tilde{\Sigma}$   
 593 is always valid. Finally, we restrict our spherical harmonic expansion to  $l_{\max} = 15$  as  
 594 a compromise between resolution, computational speed, and numerical stability (see  
 595 Luger et al. 2021a).

596 We use the standard implementation of the nested sampler, `dynesty.NestedSampler`,  
 597 with all arguments set to their default values (multi-ellipsoidal decomposition for  
 598 bounds determination (Feroz et al. 2009), uniform sampling within the bounds, 500  
 599 live points, and no gradients), to perform our inference step. Convergence—defined  
 600 as when the estimate of the remaining evidence  $\Delta \ln \mathcal{Z}$  drops below 0.5—is usually  
 601 attained after 5,000 to 10,000 samples and within a couple hours on a typical machine  
 602 for most of the trials we perform.

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

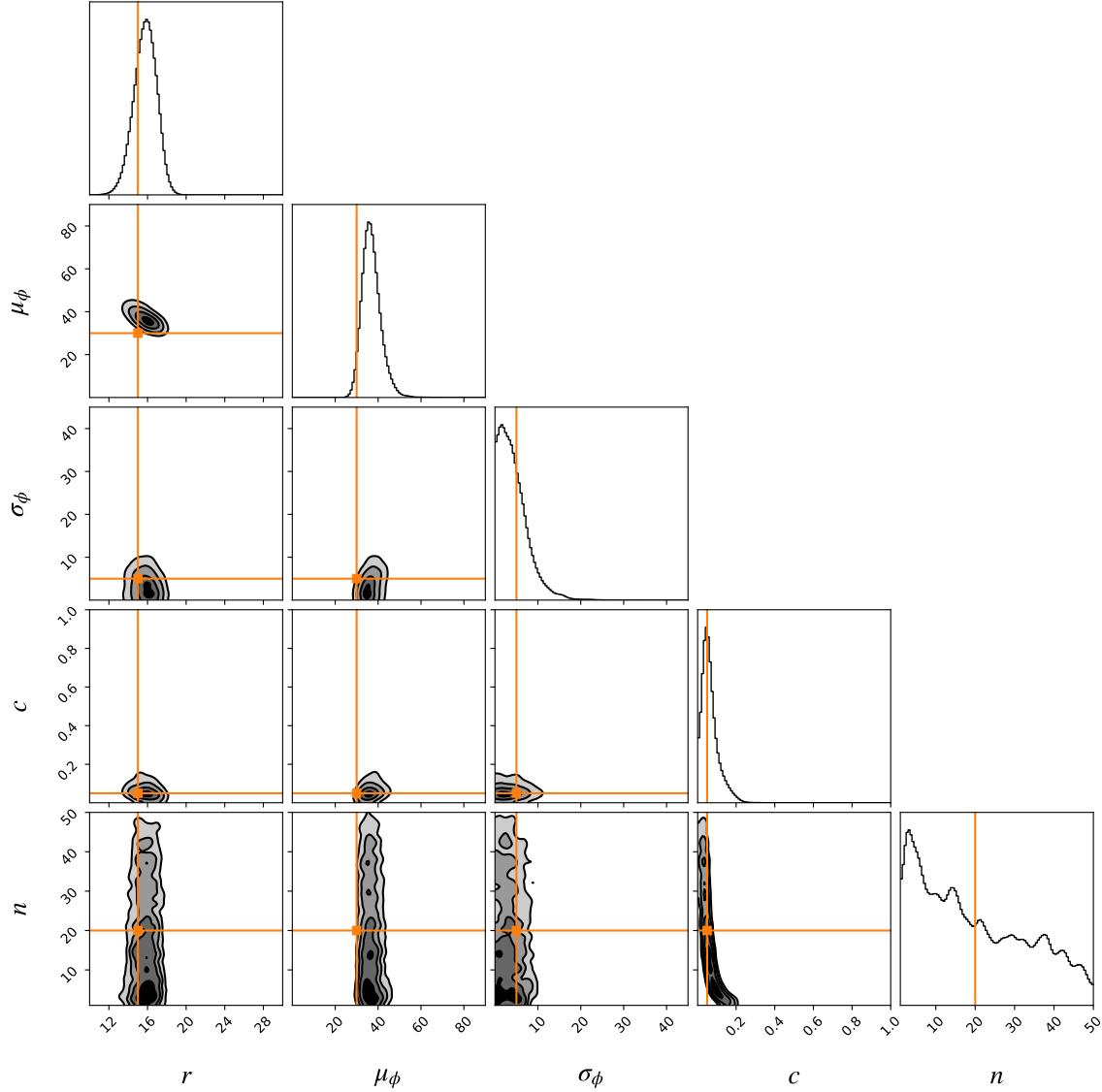
#### 606 3.4. Default run

607 The input parameters for the default run are shown in Table 1, and the corre-  
 608 sponding light curves in Figure 5. We run the nested sampler as described in the  
 609 previous section and transform the posteriors in  $a$  and  $b$  to posteriors in  $\mu_\phi$  and  $\sigma_\phi$   
 610 via Equations (C63), (C66), and (C68). The results are shown in Figure 6, where we  
 611 correctly infer all five parameters within 2 – 3 standard deviations. Posterior distri-  
 612 butions for the spot radius  $r$ , the central spot latitude  $\mu_\phi$ , and the spot contrast  $c$  are  
 613 fairly tight, while the distribution for the latitudinal scatter  $\sigma_\phi$  has wider tails and  
 614 the distribution for the number of spots  $n$  is very poorly constrained. The latter, in  
 615 particular, is degenerate with the spot contrast  $c$ ; we discuss this at length in Paper  
 616 I.

617 Figure 7 shows samples from the spot latitude posterior (hyper)distribution. Since  
 618 the parameters  $\mu_\phi$  and  $\sigma_\phi$  characterize a distribution over spot latitudes, uncertainty  
 619 in their values translates to uncertainty in the actual shape of the spot latitude dis-  
 620 tribution. Thus, the collection of blue curves in Figure 7 quantifies our knowledge  
 621 of how spots are distributed on the surfaces of the stars in the dataset. These dis-  
 622 tributions are again consistent with the true distribution used to generate the spots  
 623 (orange curve) within less than 2 standard deviations.<sup>15</sup>

624 Even though we explicitly marginalized over inclination, we can still derive pos-  
 625 terior constraints on the inclinations of the individual stars in our ensemble by com-  
 626 puting the log-likelihood as a function of  $I$  conditioned on the value of  $\theta_\bullet$  from a  
 627 particular draw from the posterior. We do this in Figure 8, where blue curves again

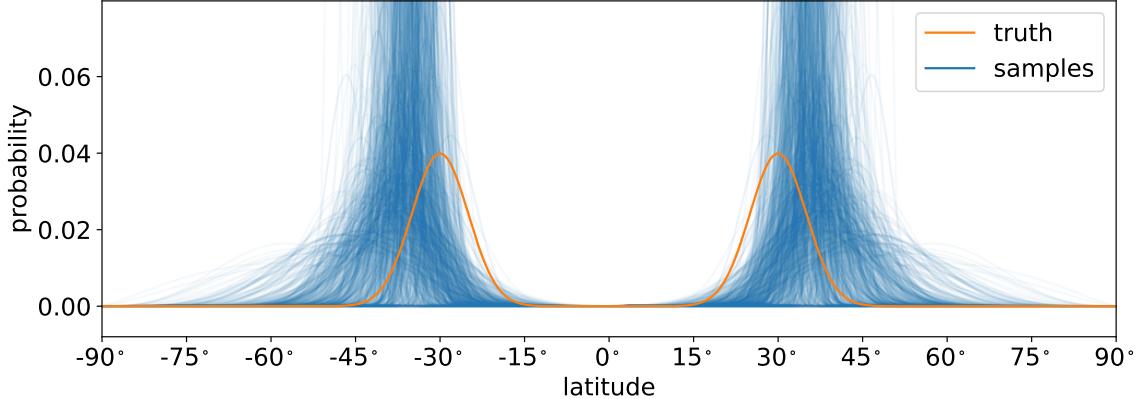
<sup>15</sup> If the results in Figure 7 seem biased, recall from Figure 6 that the mean of the latitude distribution is consistent with the truth at  $2 - 3\sigma$ . That is roughly the difference between the orange curve and the average of the blue curves. As we will see, inference with a larger ensemble (Figure S23) allows us to infer the mean latitude to within about two degrees.



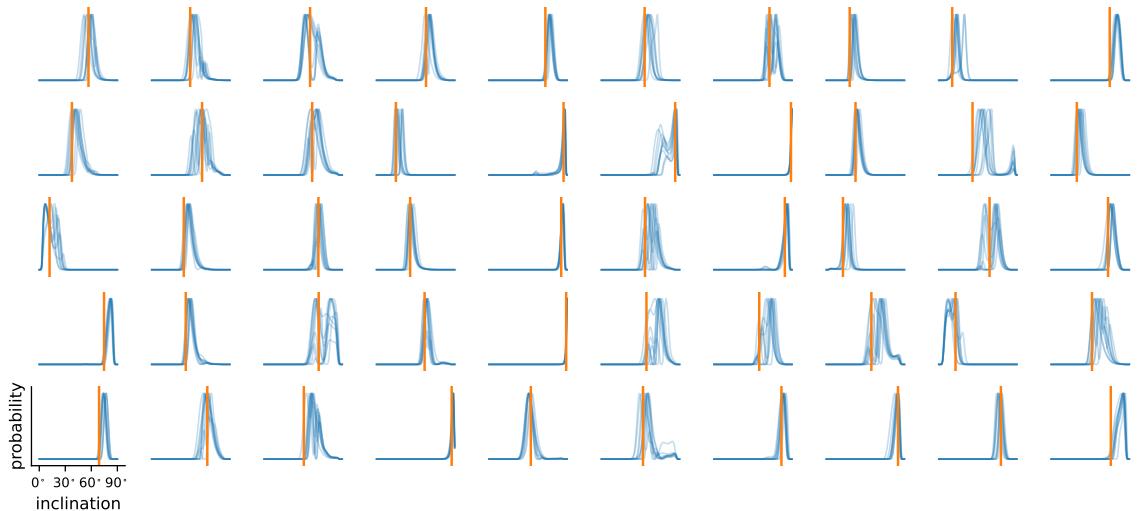
**Figure 6.** Posterior distributions for the spot parameters  $\theta_\bullet$  (radius  $r$  in degrees, latitude mode  $\mu_\phi$  and standard deviation  $\sigma_\phi$  in degrees, fractional contrast  $c$ , and number of spots  $n$ ) for the default run (Table 1 and Figure 5). The axes span the entire prior volume, and the orange lines and markers indicate the true (input) values.

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 5. 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 6—8 are based on a single run: i.e., a single realization of the light curve ensemble conditioned on the properties of Table 1. To properly gauge potential biases in our model, it is useful to perform the run under many different realizations of the dataset. We therefore generate 100 ensembles of light curves in exactly the same way as above and perform inference on each of them. Figure 9 shows the marginal and joint posterior distributions for  $\theta_\bullet$  for all 100 trials. Posteriors for



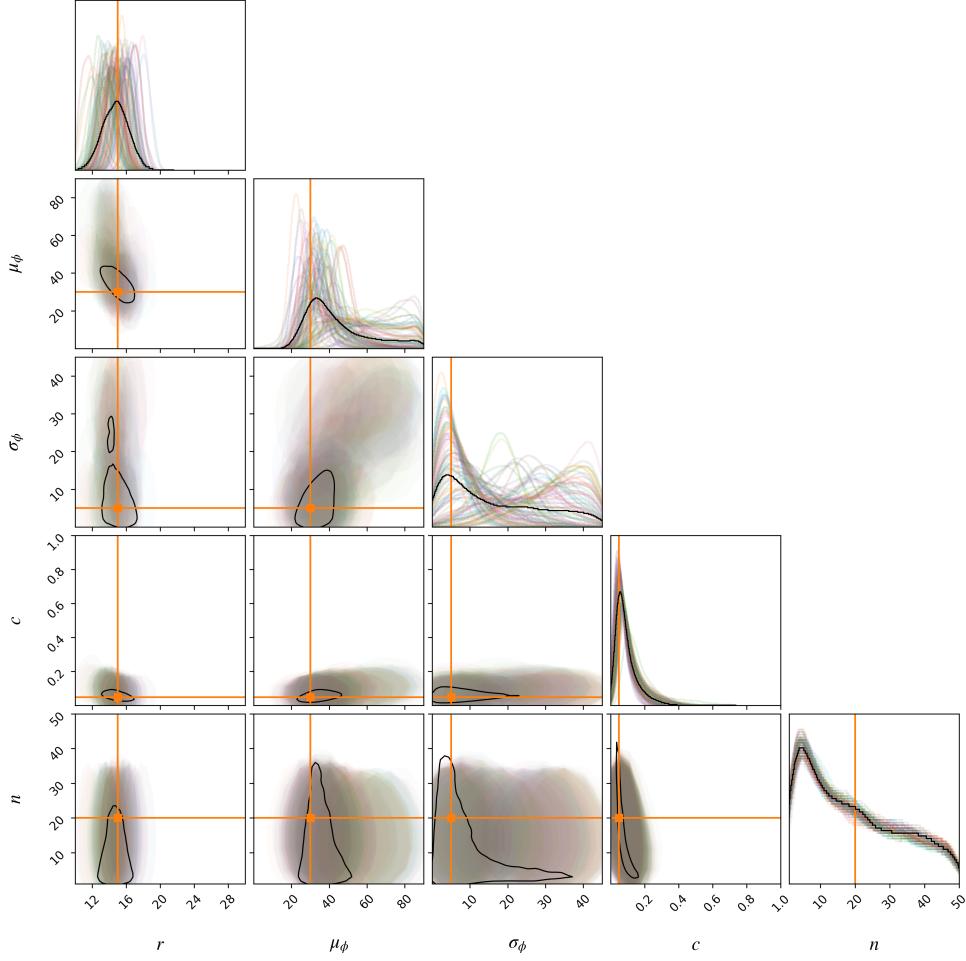
**Figure 7.** Posterior distributions for the spot latitudes for the default run. Each blue curve corresponds to the Beta distribution PDF (Equation C61) for the spot latitude with parameters drawn from the posterior in  $\mu_\phi$  and  $\sigma_\phi$  (Figure 6); the (hyper)distribution of blue curves 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).



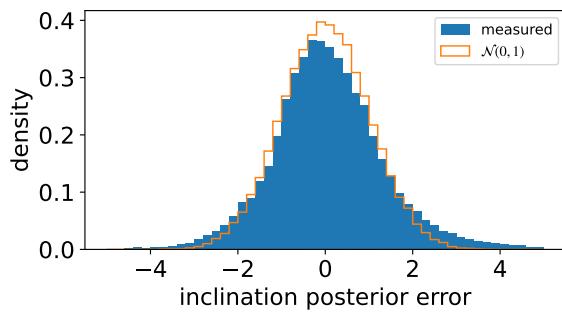
**Figure 8.** Posterior distributions for the inclinations of individual stars; individual panels correspond directly to those in Figure 5. As in Figure 7, each blue curve is a sample from the hyperdistribution of inclination PDFs, conditioned on a specific value of  $\theta_\bullet$  drawn from the posterior. The orange lines indicate the true inclination of each star.

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.

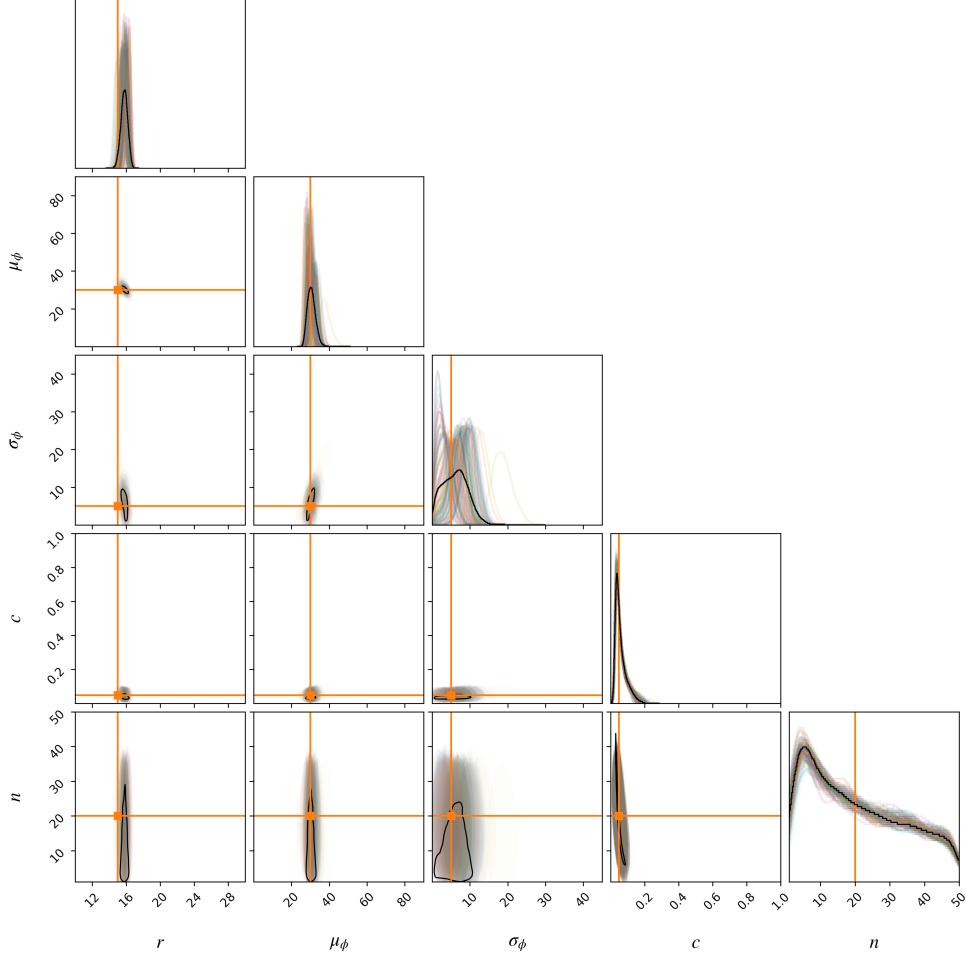
If our model is truly unbiased, in the limit of an infinite number of realizations of the data, the expectation value of the distribution of samples across all ensembles (the mean of the black curves in the marginal posterior plots) should coincide with the true values (orange lines). This is approximately the case with the spot size  $r$  posterior: on average, the posterior distributions are centered on the correct value.



**Figure 9.** Similar to Figure 6, 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 10.** Distribution of stellar inclination residuals normalized to the posterior standard deviation for all 5,000 stars across the 100 trials in Figure 9 (blue). The standard normal distribution is shown in orange for comparison. The inclination posteriors inferred with our GP are largely unbiased and have the expected variance.



**Figure 11.** Same as Figure 9, 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.

However, this is not the case for the spot latitude parameters  $\mu_\phi$  and  $\sigma_\phi$ , for which our posterior means are biased high. While the *modes* of their posteriors are very close to the true values, the distributions have long tails extending to high latitudes and high variance, respectively.

The reason for this bias has to do with the normalization degeneracy discussed in Paper I: the total spottiness of a stellar surface is not an observable in single-band 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 stronger priors on  $\mu_\phi$  and  $\sigma_\phi$  (§4.4), observe in multiple wavelength bands (Paper I), or simply collect more data. As we will see below, the particular degeneracy described above is not perfect: for very large ( $M \sim 1,000$ ) 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 9 reveals that it too is positively correlated with the mean spot latitude and therefore suffers from the same degeneracy as above. And while the mean of the spot number distribution is roughly correct, the posterior is nearly unchanged across all runs, and equally uninformative in all of them. This is yet another manifestation of the normalization degeneracy: the total number of spots is not an observable in single-band photometry (Paper I).

There is one final distribution that is instructive to consider: the distribution of errors on the inferred stellar inclination. Figure 10 shows a histogram of stellar inclination residuals (posterior mean minus true value) normalized to the posterior standard deviation for all stars across the 100 trials described above. For a correctly calibrated model, this distribution should equal the standard normal  $\mathcal{N}(0, 1^2)$  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 11 shows the same posterior distributions as in Figure 9, but for 100 runs 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 S17—S20 show the results for different latitudinal distributions, keeping all other values in Table 1 the same. Specifically, Figure S17 corresponds to a run with mid-latitude ( $\mu_\phi = 45^\circ$  and  $\sigma_\phi = 5^\circ$ ) spots, Figure S18 to a run with high-latitude ( $\mu_\phi = 60^\circ$  and  $\sigma_\phi = 5^\circ$ ) spots, Figure S19 to a run with equatorial ( $\mu_\phi = 0^\circ$  and  $\sigma_\phi = 5^\circ$ ) spots, and Figure S20 to a run with isotropically-distributed spots ( $\phi \sim \cos$ ). The results are largely consistent with those of the default run: in all cases we infer the correct spot radius, the mean and standard deviation of the spot latitude, and the spot contrast within  $2 - 3\sigma$ ; the number of spots is equally unconstrained in all runs. In Figure S18 and to a lesser extent in Figure S17, 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 S19 and S20 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 16), this happens *only* when  $\beta = 0$ , which occupies an infinitesimally thin hyperplane in parameter space. In practice, the majority of the posterior mass will be close to but not exactly at  $\beta = 0$ , leading to the bimodality in the figure. The same argument applies to Figure S20. In both cases, the posterior approximates the true distribution as best it can given the constraints of the adopted PDF.

Figure S21 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 15), 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 S22 and S23 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 S22) 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 S24—S26 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 S24, 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 (Paper I). In practice, this leads to higher uncertainty in the spot radii and latitudes relative to the same dataset without limb darkening (Figure 6). Fortunately, this uncertainty can be dramatically reduced with more data, as evident in Figure S25, which shows the results of the same run but with  $M = 1,000$  light curves in the ensemble. The constraints on  $r$ ,  $\mu_\phi$ , and  $\sigma_\phi$  are now much tighter and in good agreement with the truth. **While there is evidence for slight bias in the inferred radius, this is likely due to the difference between the model used to generate the spots and that used for inference, as discussed in the previous section.** Finally, Figure S26 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 can lead to 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 §4.4.

The runs so far correspond to stars with many ( $n = 20$ ) spots, for which the resulting light curves are smooth due to the fact that many spots are in view at any given time. Figures S27 and S28 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, we gave the spots in Figure S27 a much higher contrast ( $c = 0.5$ ). Even though the contrast is degenerate with the number of spots (which is very poorly constrained), the  $c$  posterior has a much heavier tail than in the other runs. Thus, in spite of the arguments in Luger et al. (2021b) about the difficulty in constraining  $c$  and  $n$  from single-band photometry, it is evident that the full covariance structure of

the data encodes *some* information about the contrast and—to a much lesser extent—the number of spots. In Figure S28, we compensate for the smaller number of spots by increasing the spot radius to  $r = 45^\circ \pm 5^\circ$  instead, showing that the model can accurately model large spots, even in the presence of some (unmodeled) scatter in their sizes.

In Figure S29 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 variance of the latitude distribution in our model. We correctly infer the latitude parameters and the contrast, but our radii appear to be biased high. This is likely due to the fact that larger spots have a bigger impact on the signal, so our inferred radius is a weighted average of all spot radii. In Appendix C.1 we derive an expression for the moment integrals of the spot size distribution assuming a uniform distribution between  $r - \Delta r$  and  $r + \Delta r$  (instead of a delta function at  $r$ ), which can be used to compute the GP if one wishes to explicitly account for scatter in the spot sizes. We find that repeating the run shown in Figure S29 while explicitly sampling over the distribution in  $\Delta r$  shifts the posterior mass to lower radii, mitigating the bias described above.

Our final run is shown in Figure S30, in which we assume we know the true normalization of each light curve. That is, we assume that we can measure all light curves in units of the flux we would measure if the stars had no spots on them, and we *do not* normalize them (see §2.5). In practice, this would require knowledge of the brightness (or temperature) of the unspotted photosphere, which is not an observable in single-band photometry. This value can in principle be probed, however, in multi-band photometry (e.g., Gully-Santiago et al. 2017; Guo et al. 2018), for which this run is extremely relevant. We again recover the radii and latitude parameters to within  $2 - 3\sigma$ , 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 correct normalization breaks the  $c - n$  degeneracy. Photometric measurements in multiple bands (even just two!) are therefore extremely useful when inferring spot properties. We discussed this point in Paper I.

## 4. DISCUSSION

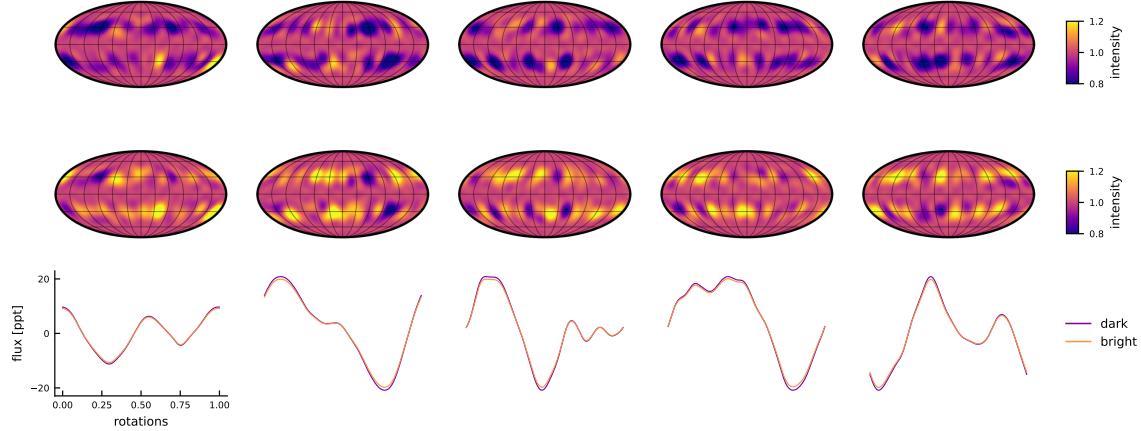
### 4.1. Small spots

One of the biggest downsides of adopting a spherical harmonic representation of the stellar surface as the foundation of 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  $\sim 180^\circ / 15 = 12^\circ$  across. Even on scales slightly larger than this, the presence of ringing can be seen (see panel (b) in Figure 3, where ringing is just barely noticeable in the

equatorial region of the maps). The spherical harmonic basis consists of global modes, all of which contribute to the intensity everywhere on the surface. Localized features require constructive interference of modes inside and destructive interference of modes outside, often leading to a 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 starspot. 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 §2.3. However, the expression for  $\mathbf{f}$  involves square roots and arctangents of functions of the spot latitude and longitude (see 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 (Luger et al. 2021a), 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 S21, 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



**Figure 12.** 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 correct normalization (§2.5), it is very difficult to differentiate between the two from stellar light curves.

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.

#### 4.2. 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 12 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 (bottom panel), which are almost identical. There is a slight difference in amplitude between the two cases: surfaces with dark spots have slightly higher light curve amplitudes than surfaces with bright spots of the same contrast magnitude. However, the magnitude of the bright spots can be increased slightly to get a near-perfect match to the dark spot light curves, meaning it may be difficult (if not impossible) to tell the difference between dark and bright spots via the GP approach.

The reason for this degeneracy is rooted (once again) in the fundamental issue with photometry: we lack any information about the correct normalization of the light curve. Consider the dependence of the (unnormalized) GP covariance on the contrast: it enters in a single place, via Equation (C107), as  $c^2$ , meaning dark and bright spot models have exactly the same covariance. These models differ only in the mean of the unnormalized process, since that is proportional to  $c$  (via Equation C106).

However, as we argued in Paper I, the mean is not a direct observable. Instead, in single-band photometry, we are only sensitive to the *ratio* of the covariance to the square of the mean (see Equation 24). From that equation, we can deduce that stars with dark spots (for which the light curve mean  $\mu < 1$ ) will therefore have larger variance than stars with bright spots ( $\mu > 1$ ), leading to the slight difference seen in the figure. However, since this is strictly a multiplicative factor affecting the covariance, it is degenerate with the two other properties that scale the covariance: the magnitude of the contrast and the total number of spots.<sup>16</sup>

We therefore conclude that **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 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 correct normalization. Recently, Morris et al. (2018) used approximately coeval *Kepler* and *Spitzer* light curves of TRAPPIST-1 to argue that a bright spot model for the star is more consistent with the data. A detailed exploration of the effect of multi-band photometry on the degeneracies of the mapping problem is deferred to a future paper in this series.

### 4.3. Comparison to other work

#### 4.3.1. Synthetic likelihoods, random fields, and approximate inference

The core idea behind the methodology presented in this paper—to compute the Gaussian approximation to an intractable multidimensional distribution in order to obtain a likelihood function for inference—is not new. Although the method likely goes by different names in different fields, it is a popular technique particularly in the field of ecological population dynamics, where it is referred to as the synthetic likelihood (SL; Wood 2010) or Bayesian synthetic likelihood (BSL; Price et al. 2018) method. In many ecological systems, population growth is a chaotic process; observations of the size of a population over time can be dominated by steep spikes and drops in the population that occur due to sudden, random environmental pressure. While population growth can be forward modeled with ease, it is very difficult to use forward models to constrain basic growth parameters in an inference setting, since

<sup>16</sup> There is also a small additive term in Equation (24), but this, too, depends only on the ratio of entries in the covariance matrix to the mean, so it is of little help in breaking the degeneracy.

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 (usually of a summary statistic of the data) 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 some function of  $f$  conditioned on  $\boldsymbol{\theta}$  by sampling, we are able to actually compute the mean and covariance of  $f$  directly *in closed form*. While traditional SL methods are inherently noisy, our method employs the *exact* Gaussian approximation to the likelihood function.

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

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).

#### 4.3.2. Starspots and stellar variability

The methodology developed in this paper is closely related to that in Perger et al. (2020), who studied the effect of different starspot configurations on the autocorrelation 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 starspot 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.

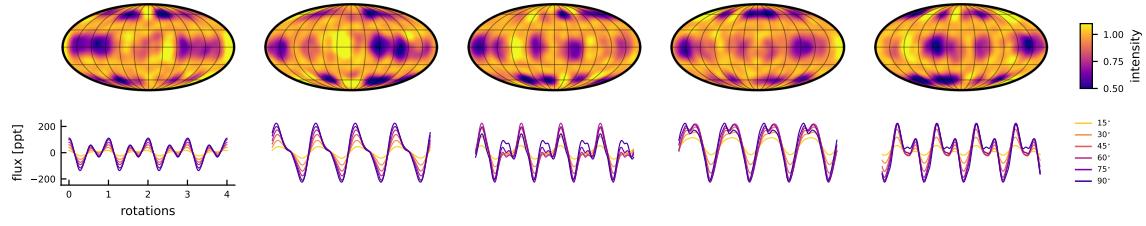
Recently, Morris (2020b) used *Kepler*, *K2*, and *TESS* light curves to derive a relationship between stellar age and spot coverage using an ensemble analysis similar to that proposed here. Because of the intractability of the marginal likelihood function, that study used an approximate Bayesian computation (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. While we believe the spot coverage results of that paper are predominantly driven by the prior (due to the strong degeneracy between the spot contrast and the number of spots; see §4.2 in Luger et al. 2021b), the ensemble analysis employed in that paper is nevertheless a powerful technique to infer spot properties. Our work builds on that of Morris (2020b) by deriving a closed form solution to the likelihood function (as opposed to a sample-based likelihood-free inference algorithm) and by harnessing the covariance structure of the data when doing inference (as opposed to relying solely on the amplitude of the data).

Finally, Basri & Shah (2020) recently presented a large suite of forward models of light curves of spotted stars, which they used to discuss the (complicated) dependence of various light curve metrics on the physical spot parameters used to generate the data. They concluded that it is not possible to uniquely relate these metrics to the underlying starspot configuration. While we agree this is the case for individual stars, our work stresses that it is possible to circumvent many of these degeneracies with ensemble analyses. Basri & Shah (2020) also concluded 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 relied on the effect these processes have on simple light curve metrics, which are almost certainly not sufficient statistics of the data. Inference that takes into account the full covariance structure of the data, while considering large ensembles of light curves, could in principle break this degeneracy. While we do not explicitly model differential rotation in this paper, it will be the subject of a future paper in this series.

1005                          4.4. *Caveats*

1006                          We conclude our discussion with a list of several notes and caveats that should be  
 1007                          kept in mind when using our algorithm and its Python implementation.

- 1008                          1. *The assumed latitude distribution is not Gaussian.* Because we require the  
   1009                          first and second moment integrals (Equations C72 and C79) to have closed  
   1010                          form solutions, there are restrictions on the probability density function we can  
   1011                          assume for the spot latitude. We find that a Beta distribution in the cosine  
   1012                          of the latitude is integrable in closed form and can be evaluated efficiently in  
   1013                          terms of recursion relations. In many cases, particularly when  $\mu_\phi \lesssim 75^\circ$  and  
   1014                           $\sigma_\phi \lesssim 10^\circ$ , the distribution in the spot latitude is close to a bi-modal Gaussian  
   1015                          with mean  $\pm\mu_\phi$  and standard deviation  $\sigma_\phi$  (see Figure 16). In general, however,  
   1016                           $\mu_\phi$  is formally equal to the *mode* (as opposed to the mean) of the distribution,  
   1017                          and  $\sigma_\phi$  is the Laplace approximation to the local standard deviation at the  
   1018                          mode.
- 1019                          2. *The number of spots  $n$  does not have to be an integer.* Samples from our GP  
   1020                          prior will not in general have exactly  $n$  spots (see, e.g., Figure 3). This is due  
   1021                          to the fact that our model is only an approximation to the true distribution of  
   1022                          stellar surfaces conditioned on the spot properties. A corollary of this point is  
   1023                          that  $n$  need not be an integer, which makes it easier in practice to sample over  
   1024                          using modern inference techniques such as MCMC, HMC, ADVI, and nested  
   1025                          sampling.
- 1026                          3. *Care should be taken when modeling large-amplitude light curves.* We discussed  
   1027                          this point at length in §2.5. Modeling a light curve that has been normalized  
   1028                          to its mean (or median) as a Gaussian process is conceptually a bad idea when  
   1029                          the amplitude of variability is large compared to the mean. As a rule of thumb,  
   1030                          if the amplitude of variability exceeds  $\sim 10\%$ , we recommend not normalizing  
   1031                          the light curve in this way, and instead modeling the normalization amplitude  
   1032                          as a latent variable.
- 1033                          4. *Keep in mind the polar spot degeneracy.* Even when modeling ensembles of  
   1034                          light curves, there are still strong degeneracies at play (Paper I). In particular,  
   1035                          spots centered on the poles are always in the null space, so it can be difficult in  
   1036                          practice to rule out their presence. This can be seen in Figure S18, in which the  
   1037                          model cannot distinguish between spots localized at  $60^\circ$  and polar spots with  
   1038                          high latitude variance. It may thus be advisable to adopt a prior that favors  
   1039                          small values of  $\sigma_\phi$ , such as the common inverse gamma prior for the variance.  
   1040                          Alternatively, one could place an isotropic prior on the latitude (with density  
   1041                          proportional to  $\cos \mu_\phi$ ) to downweight very high latitude spots.
- 1042                          5. *Limb darkening matters!* The null space is extremely sensitive to limb darkening  
   1043                          (Paper I). It is therefore extremely important to model it correctly; otherwise,



**Figure 13.** 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.

there may be substantial bias in the inferred spot parameters. For stars with transiting exoplanets, it may be possible to infer the limb darkening coefficients empirically, but in general we recommend modeling them as latent variables with priors informed by theoretical models.

- 6. *Careful with the data.* In general, covariances can be very hard to estimate from noisy data. This makes it especially important to ensure one is correctly modeling the noise. When applying our GP to model real data, we recommend the usual inference practices of clipping outliers, modeling a latent white noise (jitter) term, and modeling a small latent additive offset term to minimize the risk of bias in the posteriors of interest.
- 7. *Careful with the sample selection.* When performing an ensemble analysis of stellar variability, it is tempting to only analyze light curves that show variability in the first place. *This is extremely dangerous*, since the lack of variability could simply be due to low inclinations. It is extremely important to ensure 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 hy-

perparameter 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 13, where samples are drawn from a GP with small circumpolar spots and large equatorial spots (see caption for details). Since the composite GP inherits all of the properties of the standard GP, it can be used to do inference under more complex priors than those presented here.

It is also worth noting that this technique may be employed to model arbitrary distributions for parameters like the spot radius and the number of spots. In 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 (Luger et al. 2021a), in which case adding components to the model will result in negligible overhead.

## 5.2. Time evolution

Another big limitation of the base algorithm is the implicit assumption that stellar surfaces are static. Our GP hyperparameters  $\theta_\bullet$  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. Stars may still have active longitudes under this assumption, but there is no preferred longitude *across all stars*.
2. *The temporal and spatial covariances are independent.* This implies that the evolution of each spherical harmonic mode in time is independent of the evolution of any of the other modes in time.

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

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

$$\boldsymbol{\Sigma}_y^{(t)} = \mathbf{K} \otimes \boldsymbol{\Sigma}_y, \quad (32)$$

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 10), 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.

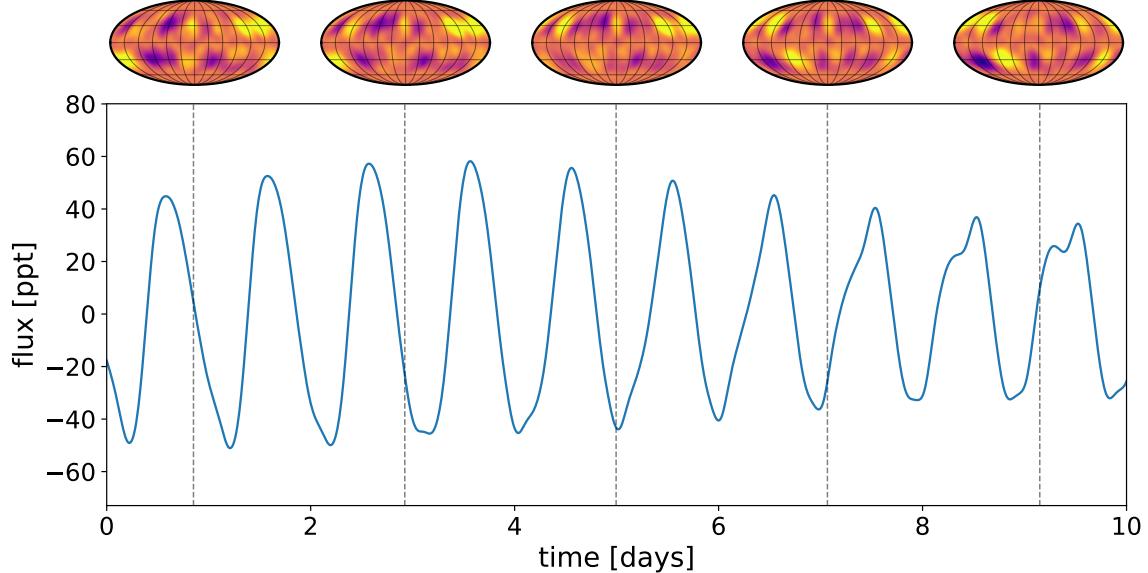
The covariance matrix in Equation (32) can be sampled from to yield time-variable surface maps, or it can be transformed into flux space for sampling light curves or computing likelihoods. The covariance in flux space is given by

$$\boldsymbol{\Sigma}^{(t)} = \mathcal{A}^\dagger \boldsymbol{\Sigma}_y^{(t)} \mathcal{A}^{\dagger \top} \quad (33)$$

where

$$\mathcal{A}^\dagger \equiv \begin{pmatrix} \mathbf{a}_0^\top & & & \\ & \mathbf{a}_1^\top & & \\ & & \ddots & \\ & & & \mathbf{a}_{K-1}^\top \end{pmatrix} \quad (34)$$

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



**Figure 14.** 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). 

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

$$\begin{aligned}\Sigma^{(t)} &= \mathcal{A}^\ddagger \Sigma_y^{(t)} \mathcal{A}^{\ddagger\top} \\ &= \mathcal{A}^\ddagger (\mathbf{K} \otimes \Sigma_y) \mathcal{A}^{\ddagger\top} \\ &= \Sigma \odot \mathbf{K},\end{aligned}\quad \checkmark \quad (35)$$

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

$$k_{E^2}(\Delta t) = \sigma^2 \exp\left(-\frac{\Delta t^2}{2\tau}\right) \quad (36)$$

or the Matérn-3/2 kernel

$$k_{M\frac{3}{2}}(\Delta t) = \sigma^2 \left(1 + \frac{\sqrt{3}\Delta t}{\tau}\right) \exp\left(-\frac{\sqrt{3}\Delta t}{\tau}\right) \quad (37)$$

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 14 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 §2.4 we discussed the value of marginalizing over the stellar inclination when computing the GP covariance (and thus the likelihood). When jointly analyzing the light curves of many stars (which we have repeatedly argued is the best way to infer their spot properties), it can be extremely useful to minimize the number of latent variables associated with individual stars by analytically marginalizing over them. This can dramatically reduce the number of free parameters, turning a difficult inference problem in (possibly) hundreds or thousands of dimensions into a much easier problem in a handful of dimensions. We showed how it is possible to marginalize away the dependence of our GP on the inclinations of individual stars, which is a huge step in this direction. However, our GP remains a function of two other quantities that will 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

1195 integration over the term  $\mathbf{R}_{\hat{\mathbf{z}}} \left( \frac{2\pi}{P} t_k \right)$ , another Wigner matrix that rotates the star to the  
 1196 correct rotational phase at time  $t = t_k$ . Given an appropriate prior on the rotational  
 1197 angular frequency  $\frac{2\pi}{P}$ , it should be possible to follow the same procedure outlined in  
 1198 Appendix D to analytically compute the first two moments of the distribution of light  
 1199 curves marginalized over the rotation period.

1200 Equation (B7) also makes clear the dependence of the GP on the limb darkening  
 1201 coefficients, which enter via the limb darkening operator  $\mathbf{L}(\mathbf{u})$ . The coefficients of this  
 1202 matrix are linear in the limb darkening coefficients  $u_1$  and  $u_2$  (see Equation B9), so  
 1203 it should be possible to derive closed form solutions to the relevant integrals over  
 1204  $\mathbf{L}(\mathbf{u})$  to yield the GP covariance marginalized over  $\mathbf{u}$ . It may be possible to compute  
 1205 the marginalized covariance even when parametrizing the quadratic limb darkening  
 1206 coefficients in terms of the uncorrelated  $q_1$  and  $q_2$  parameters from Kipping (2013),  
 1207 which would allow us to incorporate hard constraints on positivity and strict limb  
 1208 darkening (as opposed to brightening) directly into the prior.

1209 Given the complexity of the operations involved, performing these marginalizations  
 1210 is beyond the scope of the present paper. However, given the significant computational  
 1211 benefits of this marginalization, as well as the fact that the mapping problem is  
 1212 particularly sensitive to the limb darkening coefficients (Paper I), marginalizing over  
 1213  $P$  and  $\mathbf{u}$  will be the subject of an upcoming paper.

#### 1214 5.4. Modeling transits and radial velocity datasets

1215 Finally, we would like to note that the GP developed in this paper is not limited  
 1216 to modeling rotational light curves of stars. We derived the covariance structure  
 1217 of spotted stellar surfaces in the spherical harmonic basis, which we then linearly  
 1218 transformed into flux space to obtain the light curve GP presented above. However,  
 1219 our GP may be used to model *any* kind of observation that is linearly related to the  
 1220 spherical harmonic representation. If  $\mathbf{A}$  is the linear operator that transforms the  
 1221 spherical harmonic representation  $\mathbf{y}$  to the data vector  $\mathbf{d}$  via  $\mathbf{d} = \mathbf{A}\mathbf{y}$ , then the mean  
 1222 and covariance of the GP model for  $\mathbf{d}$  are given by

$$\boldsymbol{\mu}_d = \mathbf{A} \boldsymbol{\mu}_y \quad (38)$$

$$\Sigma_d = \mathbf{A} \Sigma_y \mathbf{A}^\top, \quad (39)$$

1223 respectively, where  $\boldsymbol{\mu}_y$  and  $\Sigma_y$  are the mean and covariance in the spherical harmonic  
 1224 basis derived in this paper. In Luger et al. (2019) we showed that occultation light  
 1225 curves are also linearly related to the spherical harmonic representation of the stellar  
 1226 surface, so it is straightforward to use our GP to model transits of planets across  
 1227 spotted stars, either to marginalize over the spot variability or to constrain the surface  
 1228 map of the star. In this context,  $\mathbf{A}$  may be computed via the `design_matrix()` method  
 1229 of a `Map` instance of the `starry` package.

1230 The formalism developed here may also be extended to model radial velocity (RV)  
 1231 datasets, although this requires a bit more work. The instantaneous radial velocity

shift  $v$  induced by a rotating spotted star may be approximated as

$$v = \frac{\iint_S I V dS}{\iint_S I dS} \quad (40)$$

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>17</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

$$\mathbf{v} = \frac{\mathbf{B} \mathbf{y}}{\mathbf{A} \mathbf{y}} \quad (41)$$

where  $\mathbf{v}$  is the random variable representing the observed radial velocity time series,  $\mathbf{y}$  is the Gaussian random variable representing the spherical harmonic coefficients describing the stellar surface, and the division is performed elementwise. Since  $\mathbf{v}$  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 §2.5), which will yield the mean and covariance of the Gaussian process representation of the radial velocity time series. Given the complexity of the operations described above, we defer this calculation (and the calibration of the resulting GP model) to future work.

## 6. CONCLUSIONS

This paper is the second in a series devoted to the development of statistically rigorous techniques to model stellar surfaces based on unresolved photometric and spectroscopic measurements. Here, we presented a new 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.

<sup>17</sup> Higher order effects such as differential rotation and convective blueshift can be easily modeled with a higher degree expansion of  $u$ .

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 **between 30ms and 60ms** on a modern laptop. Our algorithm is implemented in the open-source, user-friendly Python package `starry_process`, which is pip-installable, available on [GitHub](#), and described in Luger et al. (2021a). The algorithm is implemented in a combination of C++ and Python, linked using the `theano` package. Because our GP covariance has an exact representation, so too do its derivatives. We therefore implement 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 starspot properties. Below we list our main results:

1. **Our GP works best for ensemble analyses.** The light curve mapping problem is extremely degenerate, as light curves contain a vanishingly small fraction of the total information about a stellar surface. However, the degenerate surface modes are a strong function of the observer's viewing angle, so light curves of stars seen at different inclinations constrain different components of the surface. We have shown that if we jointly analyze the light curves of many stars, we can break many of the degeneracies at play and uniquely infer the statistical properties of the spots across the ensemble. This type of analysis works best if the stars in the ensemble are statistically similar: i.e., the properties of their spots are all drawn from the same parent distribution, whose parameters we can constrain.
2. **Typically, an ensemble of at least  $M \sim 50$  light curves is needed to place meaningful constraints on starspot properties.** This estimate is based on ensemble analyses of light curves with  $K = 1,000$  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.
3. **Our GP is in most cases an unbiased estimator for the spot radius and latitude distributions.** We showed that our GP can accurately infer the angular size of spots and the mode and standard deviation of their distribution in latitude from stellar light curves. For the fiducial ensemble of  $M = 50$  light curves described above, we are able to constrain the average spot radii to within a couple degrees and the average spot latitudes to within  $5^\circ$ .

- 1305     **4. Our GP can accurately infer stellar inclinations.** We presented two  
 1306     versions of our GP: one conditioned on a specific value of the stellar inclination,  
 1307     and one marginalized over inclination under an isotropic prior. In both cases,  
 1308     we find that we can accurately infer the inclinations of individual stars in an  
 1309     ensemble analysis (in the latter case, a simple post-processing step can yield the  
 1310     inclination posterior distribution). While the inclination is not an observable  
 1311     for an individual stellar light curve, the population-level constraints on the  
 1312     spot properties achieved by the GP can break the degeneracies involving the  
 1313     inclination, allowing us to usually infer it to within about  $10^\circ$  and without bias.
- 1314     **5. Our GP can be used to model small, Sun-like spots.** The algorithm  
 1315     presented here is limited to a surface resolution of about  $10^\circ$ , corresponding to  
 1316     spots about an order of magnitude larger than typical sunspots. However, we  
 1317     have showed that when we apply our model to light curves of stars with small  
 1318     ( $r \sim 3^\circ$ ) spots, we can still infer their latitudinal distribution without bias, as  
 1319     well as the presence of spots below our resolution limit.
- 1320     **6. Our GP can be extended to model time-variable surfaces.** The algo-  
 1321     rithm presented here was derived for static stellar surfaces, corresponding to  
 1322     perfectly periodic light curves. However, time variability can easily be modeled  
 1323     as the product of the `starry_process` kernel and a kernel describing the covari-  
 1324     ance of the process in time, such as a simple exponential squared or Matérn-3/2  
 1325     kernel. The hyperparameters of the temporal covariance are then strictly tied  
 1326     to the timescale on which the surface evolves. More complex temporal variabil-  
 1327     ity, such as that induced by differential rotation, will be the subject of a future  
 1328     paper in this series.
- 1329     **7. Our GP can be used in exoplanet transit modeling and extended to  
 1330       radial velocity datasets.** At its core, the `starry_process` GP defines a distribu-  
 1331       tion over spherical harmonic representations of stellar surfaces. It can therefore  
 1332       be used as a physically interpretable prior when modeling transits of exoplan-  
 1333       ets across spotted stars, either to marginalize over the stellar inhomogeneity or  
 1334       to explicitly infer spot properties. When combined with the `starry` package, it  
 1335       can also be used as a prior on the stellar surface in Doppler imaging, Doppler  
 1336       tomography, or even radial velocity searches for exoplanets. The latter will be  
 1337       the subject of a future paper in this series.

1338     The GP presented here has far-ranging applications for stellar light curve studies.  
 1339     It serves as a drop-in replacement for commonly used GP kernels for stellar vari-  
 1340     ability, which currently do not have physically interpretable parameters other than  
 1341     the rotation period and, in some cases, a spot evolution timescale. As such, it can  
 1342     be used to marginalize over stellar rotational variability signals in (say) transiting  
 1343     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, Adam Jermyn, Megan Bedell, Will Farr, Dylan Simon, and the Astronomical Data Group at the Center for Computational Astrophysics for their help and for many thought-provoking discussions that made this paper possible.

## REFERENCES

- 1381 Aigrain, S., et al. 2016, MNRAS, 459, 1430  
 1382 2408 1431
- 1383 Ambikasaran, S., et al. 2015, IEEE 1432  
 1384 Transactions on Pattern Analysis and 1433  
 1385 Machine Intelligence, 38, 252 1434
- 1386 Angus, R., et al. 2018, MNRAS, 474, 209435  
 1387 —. 2019, AJ, 158, 173 1436
- 1388 Barnes, S., et al. 2001, ApJ, 548, 1071 1437
- 1389 Basri, G., & Shah, R. 2020, ApJ, 901, 141438
- 1390 Beaumont, M. A. 2019, Annual Review of 1439  
 1391 Statistics and Its Application, 6, 379 1440
- 1392 Blei, D. M., et al. 2016, arXiv e-prints, 1441  
 1393 arXiv:1601.00670 1442
- 1394 Brewer, B. J., & Stello, D. 2009, MNRAS 1443  
 1395 395, 2226 1444
- 1396 Cantiello, M., & Braithwaite, J. 2019, 1445  
 1397 ApJ, 883, 106 1446
- 1398 Collado, J. R. A., et al. 1989, Computer 1447  
 1399 Physics Communications, 52, 323 1448
- 1400 Damasso, M., et al. 2019, MNRAS, 489, 1449  
 1401 2555 1450
- 1402 Duane, S., et al. 1987, Physics Letters B, 1451  
 1403 195, 216 1452
- 1404 Feroz, F., et al. 2009, MNRAS, 398, 1601453
- 1405 Foreman-Mackey, D., et al. 2017, AJ, 154454  
 1406 220 1455
- 1407 Fuller, J., et al. 2015, Science, 350, 423 1456
- 1408 Gilbertson, C., et al. 2020, The 1457  
 1409 Astrophysical Journal, 905, 155. <https://doi.org/10.3847/1538-4357/abc627> 1458
- 1410 Gough, D. O., & Tayler, R. J. 1966, 1460  
 1411 Monthly Notices of the Royal 1461  
 1412 Astronomical Society, 133, 85. <https://doi.org/10.1093/mnras/133.1.85> 1462
- 1413 Gully-Santiago, M. A., et al. 2017, ApJ, 1464  
 1414 836, 200 1465
- 1415 Guo, Z., et al. 2018, ApJ, 868, 143 1466
- 1416 Haywood, R. D., et al. 2014, MNRAS, 1467  
 1417 443, 2517 1468
- 1418 Hoffman, M. D., & Gelman, A. 2011, 1469  
 1419 arXiv e-prints, arXiv:1111.4246 1470
- 1420 Ireland, L. G., & Browning, M. K. 2018, 1471  
 1421 ApJ, 856, 132 1472
- 1422 Jones, D. E., et al. 2017, arXiv e-prints, 1473  
 1423 arXiv:1711.01318 1474
- 1424 Kervella, P., et al. 2017, A&A, 597, A137475
- 1425 Kipping, D. M. 2013, MNRAS, 435, 2152476
- 1426 Kucukelbir, A., et al. 2016, arXiv e-prints, 1477  
 1427 arXiv:1603.00788 1478
- 1428 Luger, R. 2021, in preparation
- 1429 Luger, R., et al. 2019, AJ, 157, 64
- . 2016, AJ, 152, 100
- . 2021a, arXiv e-prints, arXiv:2102.01774
- . 2021b, arXiv e-prints, arXiv:2102.00007
- . 2017a, ApJ, 851, 94
- . 2017b, Nature Astronomy, 1, 0129
- Miesch, M. S., & Toomre, J. 2009, Annual Review of Fluid Mechanics, 41, 317.  
<https://doi.org/10.1146/annurev.fluid.010908.165215>
- 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
- 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
- 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)
- Robertson, P., et al. 2020, ApJ, 897, 125
- Schuessler, M., et al. 1996, A&A, 314, 503
- Sikora, J., et al. 2018, Monthly Notices of the Royal Astronomical Society, 483, 3127.  
<https://doi.org/10.1093/mnras/sty2895>
- 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
- Solanki, S. K., et al. 2006, Reports on Progress in Physics, 69, 563
- Speagle, J. S. 2020, MNRAS, 493, 3132

- 1479 Turcotte, S. 2003, in Astronomical 1488  
1480 Society of the Pacific Conference Series 1489  
1481 Vol. 305, Magnetic Fields in O, B and 1490  
1482 A Stars: Origin and Connection to 1491  
1483 Pulsation, Rotation and Mass Loss, ed. 1492  
1484 L. A. Balona, H. F. Henrichs, & 1492  
1485 R. Medupe, 199 1493  
1486 Vanderburg, A., et al. 2016, MNRAS, 1493  
1487 459, 3565
- Wandelt, B. 2012, Gaussian Random  
Fields in Cosmostatistics  
Weber, M. A., & Browning, M. K. 2016,  
ApJ, 827, 95  
Wood, S. N. 2010, Nature, 466, 1102  
Yadav, R. K., et al. 2015, A&A, 573, A68

**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	§2.5
$\odot$	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
$\mathbf{A}$	<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 (5)
$\mathbf{e}_I$	First moment integral of the inclination	Equation (20)
$\mathbf{e}_r$	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 (21)
$\mathbf{E}_r$	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 (3)
$\mathbf{f}$	Flux vector (random variable)	Equation (1)
${}_2F_1(\dots)$	Gauss hypergeometric function	—
$\boldsymbol{\theta}_\bullet$	Vector of GP hyperparameters	Equation (13)
$I$	Stellar inclination	—
$\mathbb{I}$	Stellar inclination (random variable)	§2.4
$J$	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 (14)

**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 (22)
$\mu$	Flux GP mean vector	Equation (7)
$\mu_y$	Spherical harmonic GP mean vector	Equation (9)
$n$	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
$\mathbf{R}_u$	Real wigner rotation matrix about an axis $\mathbf{u}$	Appendix C.2.1
$s$	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 (8)
$\Sigma^{(t)}$	Flux GP covariance matrix w/ temporal evolution	Equation (8)
$\tilde{\Sigma}$	Flux GP covariance matrix (normalized processs)	Equation (24)
$\Sigma_y$	Spherical harmonic GP covariance matrix	Equation (10)
$\Sigma_y^{(t)}$	Spherical harmonic GP covariance w/ temporal evolution	Equation (10)
$t$	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 (3)
$\mathbf{y}$	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 (25)

## 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

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

1524 For completeness, the element of a spherical harmonic vector  $\mathbf{x}$  with degree  $l$  and  
 1525 order  $m$  is at (flattened) index

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

1526 Conversely, the element at (flattened) index  $n$  has degree and order

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

1527 respectively. Note, finally, that our use of upper and lower indices is purely a nota-  
 1528 tional convenience, and should not be confused with exponentiation or a distinction  
 1529 between covariant and contravariant tensors. It should also not be confused with the  
 1530 notation used for the complex spherical harmonics, which also uses upper and lower  
 1531 indexing.

1532 For reference, Table 2 lists the principal symbols, operators, and variables used  
 1533 throughout the paper, with links to the equations and/or section in which they are  
 1534 presented.

## 1535 B. COMPUTING THE FLUX

### 1536 B.1. Basic expression

1537 As we mentioned in §2.3, the flux  $\mathbf{f}$  is a purely linear function of the spherical  
 1538 harmonic coefficient vector  $\mathbf{y}$ :

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

1539 **In this expression, we assume  $\mathbf{y}$  describes the surface intensity of the star**  
 1540 **at time  $t = 0$  in a frame where  $\hat{\mathbf{x}}$  points to the right,  $\hat{\mathbf{y}}$  points up, and  $\hat{\mathbf{z}}$**   
 1541 **points out of the page.** Even though this is derived in detail in Luger et al. (2019),  
 1542 it is useful to expand on the computation of the design matrix  $\mathcal{A}$ , **which transforms**  
 1543 **from spherical harmonics to flux.** Let  $\mathbf{a}_k^\top$  denote the  $k^{\text{th}}$  row of  $\mathcal{A}$ , such that

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

1544 The row vector  $\mathbf{a}_k^\top$  encodes how the spherical harmonic coefficient vector projects onto  
 1545 the  $k^{\text{th}}$  cadence in the flux time series, 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 consider how each of its terms operate on the spherical harmonic representation  $\mathbf{y}$  from right to left. 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>18</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  $\mathbf{A}_1$  (see Appendix B in Luger et al. 2019) projects the stellar map into the *polynomial basis* (Equation 7 in Luger et al. 2019), comprised of the sequence of monomials in Cartesian coordinates  $(1 \; x \; z \; y \; x^2 \; xz \; xy \; yz \; y^2 \cdots)$  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

<sup>18</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.

is simply a product of spherical harmonics, which is itself a linear combination of 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  $\mathbf{y}$ ,

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

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

which are linearly related to the mean and the covariance of our GP (§2.3). Recall that  $\mathbf{x}$  is a vector of parameters describing the exact configuration of features on the surface of a star, and  $p(\mathbf{x} \mid \boldsymbol{\theta}_\bullet)$  is its probability density function conditioned on hyperparameters  $\boldsymbol{\theta}_\bullet$ , which describe the *distribution* of the features on the surface of

1600 one or many stars. As we are specifically interested in modeling the effect of starspots  
 1601 on stellar light curves, we let

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

1602 and

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

1603 where  $\mathfrak{n}$  is the total number of spots,  $c_i$  is the contrast of the  $i^{\text{th}}$  spot,  $\lambda_i$  is its  
 1604 longitude,  $\phi_i$  is its latitude, and  $r_i$  is its radius. The vector function  $\mathbf{s}(r_i)$  returns the  
 1605 spherical harmonic expansion of a negative unit brightness circular spot of radius  $r_i$   
 1606 at  $\lambda = \phi = 0$ ,  $\mathbf{R}_{\hat{\mathbf{x}}}(\phi_i)$  is the Wigner matrix that rotates the expansion about  $\hat{\mathbf{x}}$  such  
 1607 that the spot is centered at a latitude  $\phi_i$ , and  $\mathbf{R}_{\hat{\mathbf{y}}}(\lambda_i)$  is the Wigner matrix that then  
 1608 rotates the expansion about  $\hat{\mathbf{y}}$  such that the spot is centered at a longitude  $\lambda_i$ ; these  
 1609 three functions are detailed in the sections below. Equation (C14) thus provides a  
 1610 way of converting a random variable  $\mathbf{x}$  describing the size, brightness, and position of  
 1611 spots to the corresponding representation in terms of spherical harmonics. Regarding  
 1612 this equation, two things should be noted. First, we define  $\mathbf{y}$  relative to a baseline  
 1613 of zero: i.e., a star with no spots on it will have  $\mathbf{y} = \mathbf{0}$  (which is why we add unity  
 1614 in the expression for the flux in Equation 3). Second, and more importantly, we are  
 1615 not interested in any specific value of  $\mathbf{y}$ ; rather, we would like to know its expectation  
 1616 value under the probability distribution governing the different spot properties  $\mathbf{x}$ , i.e.,  
 1617  $p(\mathbf{x} | \boldsymbol{\theta}_\bullet)$ .

1618 For simplicity, we assume that the total number of spots is fixed to a value  $n$ , i.e.,

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

1619 where  $\delta$  is the delta function.<sup>19</sup> We further assume that  $p(\mathbf{x} | \boldsymbol{\theta}_\bullet)$  is separable in each  
 1620 of the four other spot properties, and that all of the spots are drawn from the same  
 1621 distribution:

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

1622 where

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

<sup>19</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  $\mathfrak{n}$ . However, as we argue in the text,  $\mathfrak{n}$  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.

1623 is the vector of hyperparameters describing the process and  $\boldsymbol{\theta}_c$ ,  $\boldsymbol{\theta}_\lambda$ ,  $\boldsymbol{\theta}_\phi$ , and  $\boldsymbol{\theta}_r$  are yet  
 1624 to be specified. This allows us to rewrite the expectation integrals (C11) and (C12)  
 1625 as

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

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

1626 where we define the first moment integrals

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

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

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

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

1627 and the second moment integrals

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

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

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

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

1628 In Equations (C18) and (C19), we used the fact that both the mean and the variance  
 1629 of the sum of  $n$  independent, identically-distributed random variables are equal to  $n$   
 1630 times the individual mean and the variance, respectively.

1631 We devote the remainder of this section to the computation of these eight integrals.

### 1632 C.1. The Radius Integrals

1633 Below we compute the first and second moments of the radius distribution ( $\mathbf{e}_r$ ,  $\mathbf{E}_r$ )  
 1634 under a suitable spherical harmonic expansion  $\mathbf{s}(r)$  of the spot profile and a suitable  
 1635 probability distribution function for the spot radius,  $p(r \mid \boldsymbol{\theta}_r)$ .

1636                   C.1.1. *Spot profile*

1637     We model the brightness  $b$  an angle  $\vartheta$  away from the center of a spot of negative  
 1638     unit intensity and radius  $r$  as

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

1639     for some (constant) shape parameter  $s$ . In the limit  $s \rightarrow 0$ ,  $b$  approaches an inverted  
 1640     top-hat function with half-width equal to  $r$ , corresponding to a circular spot of uni-  
 1641     form intensity. For  $s > 0$ , each half of  $b$  is a sigmoid with half-width at half-minimum  
 1642     equal to  $r$ . In our implementation of the algorithm we choose  $s = 0.2^\circ$ , which is small  
 1643     compared to features of interest but not too small as to create numerical issues when  
 1644     computing model gradients (which would be undefined at the spot boundary if the  
 1645     spot profile were truly an inverted top-hat).

1646     Our goal now is to expand the function above in spherical harmonics. To that end,  
 1647     we note that in a frame where the spot is centered on  $\hat{\mathbf{z}}$  (i.e., at polar angle  $\vartheta = 0$ ),  
 1648     the brightness profile is azimuthally symmetric, so the only nonzero coefficients in  
 1649     the spherical harmonic expansion are those with order  $m = 0$ . The corresponding  
 1650     spherical harmonics are simply proportional to the Legendre polynomials in  $\cos \vartheta$ , so  
 1651     our task is simplified to finding the Legendre polynomial expansion of  $b$ . Define a  
 1652     vector  $\vartheta$  of  $K$  equally-spaced points between 0 and  $\pi$ , with coefficients given by

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

1653     We wish to model the brightness evaluated at each  $\vartheta_k$  as a weighted combination of  
 1654     Legendre polynomials,

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

1655     where  $\mathbf{b}(r)$  is computed by evaluating Equation (C28) at each of the  $\vartheta_k$ ,  $\mathbf{B}$  is a design  
 1656     matrix whose columns are the weighted Legendre polynomials,

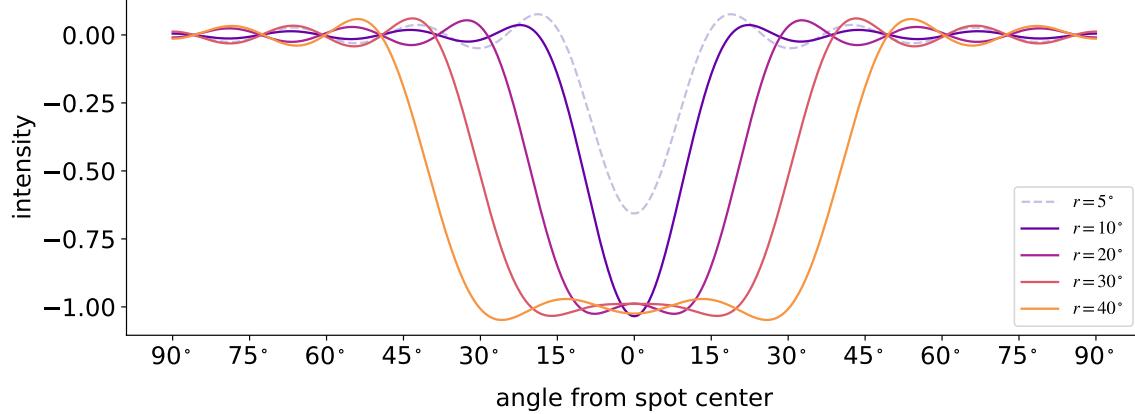
$$B_{k,l} = \sqrt{2l + 1} P_l(\cos \vartheta_k), \quad (\text{C31})$$

1657     and  $\mathbf{s}_0(r)$  are the coefficients of the expansion. These are related to the full vector of  
 1658     spherical harmonic coefficients describing the spot,  $\mathbf{s}(r)$ , by

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

1659     or, in vector form,

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



**Figure 15.** 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.

1660 where  $\mathbf{I}$  is a rectangular  $((l_{\max} + 1)^2 \times (l_{\max} + 1))$  identity-like matrix with components  
1661

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

1662 and  $\delta$  is the Kronecker delta function. To find the coefficients  $\mathbf{s}_0(r)$  (and hence  $\mathbf{s}(r)$ ),  
1663 we solve the (linear) inverse problem,

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

1664 where

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

1665 is the smoothed pseudo-inverse of  $\mathbf{B}$  with small regularization parameter  $\epsilon$ ,  $\mathbf{I}$  is the  
1666 identity matrix, and  $\mathbf{S}$  is a diagonal smoothing matrix with coefficients

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

1667 for smoothing strength  $\xi$ . For  $\epsilon \rightarrow 0$  and  $\xi \rightarrow \infty$ ,  $\mathbf{B}^+$  is the exact pseudo-inverse of  
1668  $\mathbf{B}$ . However,  $\epsilon > 0$  is chosen for improved numerical stability and  $\xi > 0$  is chosen to  
1669 mitigate the effect of ringing in the solution. In practice, we obtain good results with  
1670  $\epsilon \approx 10^{-9}$  and  $\xi \approx 15$ .

1671 Figure 15 shows the intensity profile for spots of different radii expanded to spherical  
1672 harmonic degree  $l_{\max} = 15$ . The average intensity within the spots is close to  
1673  $-1$  and the half-widths at half-minimum are equal to the spot radii, as expected.  
1674 The effect of ringing due to the truncated spherical harmonic expansion is evident,  
1675 although it is strongly suppressed compared to an expansion without the smoothing  
1676 term (i.e.,  $\xi = \infty$ ). However, for  $r \lesssim 10^\circ$ , an expansion to  $l_{\max} = 15$  is insufficient

1677 to correctly model the spot, as can be seen from the  $r = 5^\circ$  profile (dashed curve).  
 1678 Expansions to higher spherical harmonic degree allow one to model spots with radii  
 1679 smaller than  $10^\circ$ , although at increased computational cost and potential numerical  
 1680 stability issues; we discuss this point at length in §4.1.

### 1681 C.1.2. Probability density function

1682 For simplicity, we will adopt a uniform probability distribution for the spot radius,  
 1683 characterized by a mean radius  $r$  and a half-width  $\Delta r$ :

$$1684 p(\mathbf{r} \mid \boldsymbol{\theta}_r) = \begin{cases} \frac{1}{2\Delta r} & r - \Delta r \leq r \leq r + \Delta r \\ 0 & \text{otherwise,} \end{cases} \quad (\text{C38})$$

where the hyperparameters of the distribution are

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

1685 As we argue in the text, in practice it is often difficult to constrain the moments  
 1686 of the radius distribution above the first (the mean). It is therefore useful to also  
 1687 consider the limiting case of the radius distribution as  $\Delta r \rightarrow 0$ , in which case the  
 1688 PDF becomes

$$1689 p(\mathbf{r} \mid \boldsymbol{\theta}_r, \Delta r = 0) = \delta(\mathbf{r} - r), \quad (\text{C40})$$

where  $\delta$  is the delta function.

### 1690 C.1.3. First moment

1691 The first moment of the radius distribution is (Equation C20)

$$1692 \mathbf{e}_r \equiv \int \mathbf{s}(\mathbf{r}) p(\mathbf{r} \mid \boldsymbol{\theta}_r) d\mathbf{r} \\ = \frac{1}{2\Delta r} \int_{r_0 - \Delta r}^{r_0 + \Delta r} \mathbf{s}(\mathbf{r}) d\mathbf{r}. \quad (\text{C41})$$

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

$$1693 (e_r)_m^l = \frac{\delta_{m,0}}{2\Delta r} \int_{r-\Delta r}^{r+\Delta r} \tilde{s}_l(r) dr \\ = \frac{\delta_{m,0}}{2\Delta r} \sum_{k=0}^{K-1} B_{l,k}^+ \int_{r-\Delta r}^{r+\Delta r} b_k(r) dr \\ = \delta_{m,0} \sum_{k=0}^{K-1} B_{l,k}^+ c_k(r, \Delta r) \quad \checkmark \quad (\text{C42})$$

where

$$1694 c_k(r, \Delta r) = \frac{s}{2\Delta r} \ln \left( \frac{1 + \chi_k^-(r, \Delta r)}{1 + \chi_k^+(r, \Delta r)} \right) \quad \checkmark \quad (\text{C43})$$

1694 and

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

1695 In vector form, we may simply write

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

1696 Note, finally, that in the limit  $\Delta r \rightarrow 0$ ,

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

#### 1697 C.1.4. Second moment

1698 The second moment of the radius distribution is (Equation C24)

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

1699 As before, its components may be written

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

1700 where

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

1701 In vector form, this may be written

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

1702 Finally, in the limit  $\Delta r \rightarrow 0$ ,

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

1703                    C.2. *The Latitude Integrals*

1704     Our goal in this section is to compute the first and second moments of the latitude  
 1705     distribution ( $\mathbf{e}_\phi$  and  $\mathbf{E}_\phi$ , given by Equations C21 and C25, respectively). These involve  
 1706     integrals over the terms in the Wigner rotation matrix for spherical harmonics, which  
 1707     we discuss below.

1708                    C.2.1. *Rotation matrices*

1709     The Wigner rotation matrix for real spherical harmonics up to degree  $l_{\max}$  may be  
 1710     written as the block-diagonal matrix

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

1711     where

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

1712     is the Wigner rotation matrix for a single spherical harmonic degree,

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

1713     describes the transformation from complex to real spherical harmonics, and  $\mathbf{D}$  is  
 1714     the Wigner matrix for complex spherical harmonics, whose terms are given by the  
 1715     expression

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

1716 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  
 1717 depend on powers of  $\sin(\beta/2)$  and  $\cos(\beta/2)$  (see Equation C15 in Luger et al. 2019),  
 1718 but it is convenient to use the half-angle formula to express these terms instead as  
 1719

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

1720 where

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

### 1721 C.2.2. Probability density function

1722 The latitude integrals (Equations C21 and C25) involve rotations by an angle  $\phi$   
 1723 about  $\hat{\mathbf{x}}$ , which may be accomplished by choosing Euler angles  $\alpha = \pi/2$ ,  $\beta = \phi$ , and  
 1724  $\gamma = -\pi/2$ , such that

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

1725 with

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

1726 From the expressions above, it is clear that all terms in  $\mathbf{R}_{\hat{\mathbf{x}}}(\phi)$  are equal to (weighted)  
 1727 sums of powers of  $(1 \pm \cos \phi)$ . Since our goal is to compute integrals of these terms  
 1728 multiplied by a probability density function, it is convenient to model  $\cos \phi$  as a Beta-  
 1729 distributed variable. As we will see, this choice will allow us to analytically compute  
 1730 the first two moments of the distribution of  $\phi$  conditioned on  $\theta_\phi$ .

1731 The Beta distribution in  $\cos \phi$  has hyperparameters  $\alpha$  and  $\beta$  (not to be confused  
 1732 with the Euler angles  $\alpha$  and  $\beta$ ) and PDF given by

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

1733 where  $\Gamma$  is the Gamma function. The implied distribution for  $\phi$  may be computed by  
 1734 a straightforward change of variable:

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

1735 for  $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ . Both  $\alpha$  and  $\beta$  are restricted to  $(0, \infty)$ . However, in practice it  
 1736 is necessary to limit the values of these parameters to a finite range to ensure the  
 1737 numerical stability of the algorithm. It is also convenient to work with the log of

1738 these quantities because of their large dynamic range. We therefore introduce the  
1739 modified parameters

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

1740 with inverse transform

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

1741 where the matrix

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

1742 defines the minimum and maximum values of  $\ln \alpha$  (top row) and  $\ln \beta$  (bottom row)  
1743 we adopt in our implementation of the algorithm. The lower limits correspond to  
1744  $\alpha > 1$  and  $\beta > \frac{1}{2}$ , which excludes distributions with unphysically sharp peaks at  
1745  $\phi = 90^\circ$ . Both  $a$  and  $b$  are restricted to the domain  $(0, 1)$ , and together comprise the  
1746 hyperparameter vector

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

1747 Because of their trivial domain, these parameters are convenient to sample in when  
1748 doing inference (provided we account for their implied prior on the spot latitudes; see  
1749 below). However,  $a$  and  $b$  do not intuitively relate to physical quantities of interest.  
1750 In many cases it is more desirable to parametrize the latitude distribution in terms  
1751 of a parameter  $\mu_\phi$  controlling the central latitude and a parameter  $\sigma_\phi$  controlling the  
1752 dispersion in latitude among the spots. In this case, we may choose instead

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

1753 Over most of the parameter space in  $a$  and  $b$ , the spot latitude distribution defined  
1754 above is well approximated by a bimodal Gaussian. In particular, there exists a one-  
1755 to-one relationship between  $a$  and  $b$  and the mean  $\mu_\phi$  and standard deviation  $\sigma_\phi$  of a  
1756 normal approximation to the distribution. Moreover, we find that if we let  $\mu_\phi$  be the  
1757 *mode* of the PDF and  $\sigma_\phi^2$  be a *local* approximation to the variance of the PDF, the  
1758 relationship has a convenient closed form.

1759 To compute  $\mu_\phi$ , we differentiate Equation (C61) with respect to  $\phi$ , set the expres-  
1760 sion equal to zero, and solve for  $\phi$  to obtain

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

1761 To compute  $\sigma_\phi$ , we note that the variance of a Gaussian distribution  $\varphi(\phi; \mu, \sigma^2)$  is  
 1762 the negative reciprocal of its curvature in log space:

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

1763 We therefore twice differentiate the log of Equation (C61), negate it, take the re-  
 1764 ciprocal, and evaluate it at  $\phi = \mu_\phi$  to obtain a local approximation to the standard  
 1765 deviation of the distribution at the mode:

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

1766 For completeness, the inverse transform also has a closed form:

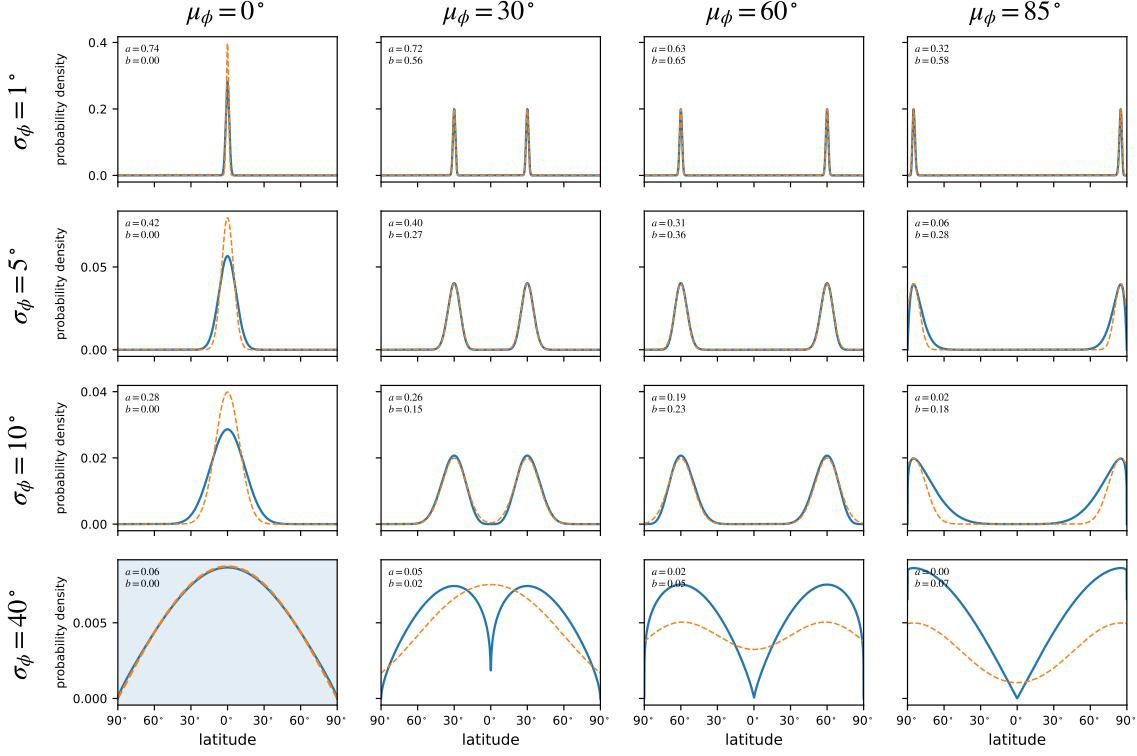
$$\alpha = \frac{2 + 4\sigma_\phi^2 + (3 + 8\sigma_\phi^2) \cos \mu_\phi + 2 \cos(2\mu_\phi) + \cos(3\mu_\phi)}{16\sigma_\phi^2 \cos\left(\frac{\mu_\phi}{2}\right)^4} \quad \checkmark \quad (\text{C69})$$

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

1767 For reference, Figure 16 shows the latitude PDF and the corresponding Gaussian  
 1768 approximation for different values of  $\mu_\phi$  and  $\sigma_\phi$ . The corresponding values of  $a$  and  
 1769  $b$  are indicated in the top left of each panel. For  $\mu_\phi$  at intermediate latitudes and  
 1770 moderate values of  $\sigma_\phi$ , the approximation is quite good. However, for  $\mu_\phi$  very close  
 1771 to the equator or to the poles, the curvature of the distribution changes significantly  
 1772 as a function of  $\phi$ , so the variance is somewhat underestimated by the approximation;  
 1773 and for  $\sigma_\phi$  large, the distribution becomes noticeably non-Gaussian.

1774 The shaded panel at the lower left is a special case of the distribution ( $\mu_\phi = 0^\circ$ ,  
 1775  $\sigma_\phi \approx 40^\circ$ ; or equivalently,  $a \approx 0.06$ ,  $b = 0$ ), which is approximately isotropic in  
 1776 latitude. In this panel, the orange curve instead corresponds to an isotropic (cosine)  
 1777 distribution in  $\phi$ ; note the excellent agreement. Thus, in addition to having closed-  
 1778 form moments, the Beta distribution is quite flexible and, via the transforms outlined  
 1779 above, intuitive in how it affects the distribution of spots on the surface of a star.

1780 In the following sections we derive expressions for the moments of the distribution  
 1781 in terms of  $\alpha$  and  $\beta$ , as this is somewhat more convenient; these can easily be  
 1782 transformed into expressions involving either  $\mu_\phi$  and  $\sigma_\phi$  or  $a$  and  $b$  via the equations  
 1783 above. The former parametrization is convenient when these properties are known  
 1784 and can be fixed; i.e., when using the GP as a restrictive prior for the light curve  
 1785 of a star whose spot distribution is already understood. However, for the purposes  
 1786 of posterior inference—that is, when trying to constrain the hyperparameters of the  
 1787 GP—we recommend sampling in the parameters  $a$  and  $b$ , since their domains are  
 1788 trivial, with uncorrelated boundaries. Posterior constraints on these quantities may



**Figure 16.** 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).

1789 easily be transformed into constraints on  $\mu_\phi$  and  $\sigma_\phi$  via the equations above. Note,  
1790 importantly, that this requires us to explicitly add a Jacobian term to the likelihood  
1791 to account for the prior implied by sampling uniformly in  $a$  and  $b$  in the range  $(0, 1)$ .  
1792 The Jacobian is given by

$$J = \frac{\partial \mu_\phi}{\partial a} \frac{\partial \sigma_\phi}{\partial b} - \frac{\partial \mu_\phi}{\partial b} \frac{\partial \sigma_\phi}{\partial a} \quad \checkmark \text{ (C71)}$$

$$= \frac{(K_{10} - K_{00})(K_{11} - K_{10})\alpha\beta(1 + \cos \mu_\phi)^3 \sin(2\mu_\phi)^3}{\sigma_\phi (2\alpha + \beta - 3 + (2\alpha + \beta - 1)\cos \mu_\phi) (2(\alpha + \beta - 1) + 3(\beta - 1)\cos \mu_\phi - 2(\alpha - \beta - 1)\cos(2\mu_\phi) + (\beta - 1)\cos(3\mu_\phi))^2}$$

1793 Adding the log of the absolute value of  $J$  to the log likelihood corrects for the *ad hoc*  
1794 prior on the latitude parameters introduced by our particular choice of parametrization,  
1795 enforcing instead a uniform prior on the quantities  $\mu_\phi$  and  $\sigma_\phi$ .

### 1796 C.2.3. First moment

1797 Since the Wigner matrices are block diagonal, we may evaluate the moments of  
1798 the distribution one spherical harmonic degree at a time. To that end, let us write

1799 the first moment integral as

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

1800 where

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

1801 and we define

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

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

1802 The integral  $\mathbf{p}_\phi^l$  defined above has a closed-form solution. To show this, we write the  
1803 terms of  $\mathbf{p}_\phi^l$  as

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

1804 where

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

1805 and in the last line we made use of the transformation  $x = \cos \phi$ . The integral in the  
1806 expression above has a closed-form solution in terms of the hypergeometric function  
1807  ${}_2F_1$ :

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

1808 In order to compute the integral  $\mathbf{e}_\phi$  (Equation C21) we must evaluate Equation (C78)  
1809 for all  $0 \leq l \leq l_{\max}$ ,  $0 \leq i \leq 2l$ , which can be done efficiently via upward recursion  
1810 relations for both the gamma function and the hypergeometric function.

1811 C.2.4. *Second moment*

1812 Similarly as before, let us write the second moment integral as

$$\begin{aligned} \mathbf{E}_\phi &\equiv \int \mathbf{R}_{\hat{\mathbf{x}}}(\phi) \mathbf{E}_r \mathbf{R}_{\hat{\mathbf{x}}}^\top(\phi) p(\phi | \boldsymbol{\theta}_\phi) d\phi \\ &= \begin{pmatrix} \mathbf{E}_\phi^{0,0} & \mathbf{E}_\phi^{0,1} & \mathbf{E}_\phi^{0,2} & \dots & \mathbf{E}_\phi^{0,l_{\max}} \\ \mathbf{E}_\phi^{1,0} & \mathbf{E}_\phi^{1,1} & \mathbf{E}_\phi^{1,2} & \dots & \mathbf{E}_\phi^{1,l_{\max}} \\ \mathbf{E}_\phi^{2,0} & \mathbf{E}_\phi^{2,1} & \mathbf{E}_\phi^{2,2} & \dots & \mathbf{E}_\phi^{2,l_{\max}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{E}_\phi^{l_{\max},0} & \mathbf{E}_\phi^{l_{\max},1} & \mathbf{E}_\phi^{l_{\max},2} & \dots & \mathbf{E}_\phi^{l_{\max},l_{\max}} \end{pmatrix}, \quad \checkmark \text{ (C79)} \end{aligned}$$

1813 where

$$\begin{aligned} \mathbf{E}_\phi^{l,l'} &= \int \mathbf{R}_{\hat{\mathbf{x}}}^l(\phi) \mathbf{E}_r^{l,l'} \mathbf{R}_{\hat{\mathbf{x}}}^{l'\top}(\phi) p(\phi | \boldsymbol{\theta}_\phi) d\phi \\ &= \mathbf{U}^{l-1} \mathbf{P}_\phi^{l,l'} \mathbf{U}^{l'-1\top} \quad \checkmark \text{ (C80)} \end{aligned}$$

1814 and we define

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

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

1815 As before, we may express the solution to the integral  $\mathbf{P}_\phi^{l,l'}$  in closed form. Let us  
1816 write the terms of  $\mathbf{P}_\phi^{l,l'}$  as

$$\begin{aligned} (P_\phi^{l,l'})_{m,m'} &= \int \sum_{\mu=-l}^l \sum_{\mu'=-l'}^{l'} (D_{\hat{\mathbf{x}}}^l)_{m,\mu}(\phi) (\bar{E}_r^{l,l'})_{\mu,\mu'} (D_{\hat{\mathbf{x}}}^{l'})_{m',\mu'}(\phi) p(\phi | \boldsymbol{\theta}_\phi) d\phi \quad \text{(C83)} \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \sum_{\mu=-l}^l \sum_{\mu'=-l'}^{l'} (\bar{E}_r^{l,l'})_{\mu,\mu'} \exp\left[\frac{i\pi}{2}(m - \mu + m' - \mu')\right] \sum_{i=0}^{2l} \sum_{i'=0}^{2l'} c_{m,\mu,i}^l c_{m',\mu',i'}^{l'} (Q_\phi^{l,l'})_{i,i'}(\boldsymbol{\theta}_\phi), \end{aligned}$$

1817 where, similarly to before,

$$\begin{aligned} (Q_\phi^{l,l'})_{i,i'}(\boldsymbol{\theta}_\phi) &\equiv \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \operatorname{sgn}(\sin \phi)^{i+i'} |\sin \phi| (\cos \phi)^{\alpha-1} (1 - \cos \phi)^{l+l'+\beta-\frac{i+i'}{2}-1} (1 + \cos \phi)^{\frac{i+i'}{2}} d\phi \\ &= \begin{cases} \int_0^1 \mathbf{x}^{\alpha-1} (1 - \mathbf{x})^{l+l'+\beta-\frac{i+i'}{2}-1} (1 + \mathbf{x})^{\frac{i+i'}{2}} d\mathbf{x} & i + i' \text{ even} \\ 0 & i + i' \text{ odd.} \end{cases} \quad \checkmark \text{ (C84)} \end{aligned}$$

1818 We may again express this integral in closed form:

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

1819 As before, this integral may be evaluated recursively to efficiently compute all of the  
1820 terms in  $\mathbf{E}_\phi$ .

1821                    C.3. *The Longitude Integrals*

1822     In this section we will compute the first and second moments of the longitude  
 1823     distribution ( $\mathbf{e}_\lambda$  and  $\mathbf{E}_\lambda$ , given by Equations C22 and C26, respectively). The math  
 1824     here is very similar to that in the previous section, as we are again dealing with  
 1825     integrals of Wigner matrices (§C.2.1).

1826                    C.3.1. *Probability density function*

1827     The longitude integrals (Equations C22 and C26) involve rotations by an angle  $\lambda$   
 1828     about  $\hat{\mathbf{y}}$ , which may be accomplished by choosing Euler angles  $\alpha = 0$ ,  $\beta = \lambda$ , and  
 1829      $\gamma = 0$ , such that

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

1830     with

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

1831     Since we expect the longitudinal distribution of features on the surfaces of stars to  
 1832     be (on average) isotropic, we will place a uniform prior on  $\lambda \in [-\pi, \pi]$ :

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

1833     We therefore have no hyperparameters controlling the longitudinal distribution, i.e.,

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

1834                    C.3.2. *First moment*

1835     As before, we will solve for the terms of the moment integrals one spherical har-  
 1836     monic degree at a time:

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

1837     where

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

1838     and we define

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

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

1839 The integral  $\mathbf{p}_\lambda^l$  defined above has a closed-form solution. To show this, we write the  
 1840 terms of  $\mathbf{p}_\lambda^l$  as

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

1841 where

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

1842 whose terms may easily be computed by upward recursion. Since  $\mathbf{q}_I^l$  does not depend  
 1843 on any user inputs, it may be computed a single time as a pre-processing step for  
 1844 efficiency.

### 1845 C.3.3. Second moment

1846 We write the second moment integral as

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

1847 where

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

1848 and we define

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

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

1849 We then express the terms of  $\mathbf{P}_\lambda^{l,l'}$  as

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

1850 where

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

1851 whose terms may again be computed by upward recursion in a single pre-processing  
1852 step.

#### 1853 C.4. The Contrast Integrals

1854 The final integrals we must take in our computation of  $E[\mathbf{y} | \boldsymbol{\theta}_\bullet]$  and  $E[\mathbf{y}\mathbf{y}^\top | \boldsymbol{\theta}_\bullet]$   
1855 are the integrals over the spot contrast distribution,  $\mathbf{e}_c$  and  $\mathbf{E}_c$ . These are by far  
1856 the easiest, since the spot contrast is a scalar multiplier of the spherical harmonic  
1857 coefficient vector, so we can pull the terms  $\mathbf{e}_\lambda$  and  $\mathbf{E}_\lambda$  out of the integrals in Equations  
1858 (C23) and (C27) to write

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

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

1859 These integrals may be computed analytically for any choice of probability density  
1860 function  $p(\mathbf{c} | \boldsymbol{\theta}_c)$  with closed-form moments. However, in practice, it is quite difficult  
1861 to constrain the spot contrast from light curves, let alone higher moments of its  
1862 distribution; this is due largely to the fact that the contrast is extremely degenerate  
1863 with the total number of spots (Paper I). In our implementation of the algorithm, we  
1864 therefore choose the simplest possible probability distribution, a delta function:

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

1865 characterized by a single parameter, the contrast of the spots:

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

1866 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 20 and 21), which allow us to compute the mean and covariance of the process that describes the flux marginalized over all values of the inclination (Equations 18 and 19).

### 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 11). 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

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

1887 where we define

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

1888 as the expectation of  $\mathbf{y}$  in the polar frame at time  $t = t_k$ .<sup>20</sup> At this point, it is  
 1889 convenient to invoke the fact that our GP is longitudinally isotropic: there is no  
 1890 preferred longitude on the surface of the star, or, equivalently, no preferred phase in  
 1891 the light curve. The rotation about  $\hat{\mathbf{z}}$  (i.e., the rotational axis of the star) therefore  
 1892 cannot change the expectation of  $\mathbf{y}$ , so

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

1893 We therefore have

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

1894 where

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

1895 and

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

1896 is the expectation of  $\mathbf{y}$  in the observer's frame, and as before we explicitly separate it  
 1897 out by spherical harmonic degree. As in §C.2, we may write

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

1898 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>20</sup> In the presence of limb darkening, we must include the limb darkening operator  $\mathbf{L}(\mathbf{u})$  in Equation (D113); see §B.2.

1899 The integral  $\mathbf{p}_I^l$  defined above has a closed-form solution. To show this, we write the  
 1900 terms of  $\mathbf{p}_I^l$  as

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

1901 where

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

1902 which may easily be computed recursively. As with the longitude integrals, the vector  
 1903  $\mathbf{q}_I^l$  need only be computed a single time as a pre-processing step, as it does not depend  
 1904 on any user inputs.

### 1905 D.3. Second moment

1906 The expression for the second moment is

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

1907 where

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

1908 is the second moment of the distribution over spherical harmonic coefficients (Equation  
 1909 12). We can use Equations (B6) and (B7) to express the element at index  $k, k'$   
 1910 (corresponding to the covariance of the GP between times  $t = t_k$  and  $t' = t_{k'}$ ) as

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

1911 where

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

1912 is the expectation of  $\mathbf{y}\mathbf{y}^\top$  in the observer's frame at times  $t = t_k$  and  $t' = t_{k'}$  and

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

1913 is the expectation of  $\mathbf{y}\mathbf{y}^\top$  in the polar frame at times  $t = t_k$  and  $t' = t_{k'}$ . The  
 1914 rest of the computation follows what we did in §C.2, except that the number of  
 1915 operations required to compute  $\mathbf{E}_I$  is a factor of  $K^2$  larger than in the computation of  
 1916 expectations like  $\mathbf{E}_\phi$  (Equation C79). That is because we must compute the integral  
 1917 of all terms of a matrix for *each* of the  $K^2$  elements of  $\mathbf{E}_I$ . We discuss in §D.4 below  
 1918 strategies that can drastically improve the computational scaling of marginalizing  
 1919 over the inclination.

1920 Let us write Equation (D127) in terms of its spherical harmonic components:

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

1921 where

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

1922 and we define

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

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

1923 As before, we may express the solution to the integral in Equation (D131) in closed  
 1924 form. Let us write its terms as

$$\left[ \left( P_I^{l,l'} \right)_{k,k'} \right]_{m,m'} = \int \sum_{\mu=-l}^l \sum_{\mu'=-l}^l (D_{\hat{\mathbf{x}}}^l)_{m,\mu}(-\mathbb{I}) \left[ \left( \bar{E}_{y'}^{l,l'} \right)_{k,k'} \right]_{\mu,\mu'} (D_{\hat{\mathbf{x}}}^{l'})_{\mu',m'}(-\mathbb{I}) p(\mathbb{I}) d\mathbb{I} \quad (\text{D133})$$

$$= \sum_{\mu=-l}^l \sum_{\mu'=-l'}^{l'} \left[ \left( \bar{E}_{y'}^{l,l'} \right)_{k,k'} \right]_{\mu,\mu'} \exp \left[ \frac{i\pi}{2} (m - \mu + m' - \mu') \right] \sum_{i=0}^{2l} \sum_{i'=0}^{2l'} c_{m,\mu,i}^l c_{m',\mu',i'}^{l'} (Q_I^{l,l'})_{i,i'},$$

1925 where, similarly to before,

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

1926 which may again be computed recursively in a pre-processing step.

#### 1927 D.4. Speeding up the computation

1928 The expressions in the previous section are a bit of a nightmare, particularly be-  
 1929 cause of the dimensionality of some of the linear operators involved. The complexity  
 1930 of the expressions is due to the fact that the second moment of the spherical har-  
 1931 monic vector projected onto the sky (Equation D127) is time-dependent: it changes  
 1932 as the star rotates. Computing the second moment of the *flux* requires computing  
 1933 the outer product of this tensor with itself, leading to multi-indexed quantities like  
 1934 those in Equation (D133). In addition to being cumbersome to evaluate, the full  
 1935 second moment matrix  $\mathbf{E}_I$  (and hence the flux covariance matrix) is costly to com-  
 1936 pute. It is helpful that Equation (D134) does not depend on any user inputs and thus  
 1937 may be pre-computed, but even still we require evaluating the four nested sums in  
 1938 Equation (D127)  $\mathcal{O}(l_{\max}^3)$  times for each entry in the  $(K \times K)$  matrix  $\mathbf{E}_I$ .

1939 Fortunately, the inner two sums in Equation (D127) do not depend on user inputs,  
 1940 so those may be pre-computed, and Equation (D127) may be cast as a straightforward  
 1941 matrix dot product. In practice we also find it helpful to take advantage of the phase  
 1942 independence (i.e., stationarity) of the covariance of our GP, as we did in §D.2: any  
 1943 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  
 1944 to be evenly sampled, such that the time difference between adjacent cadences is  
 1945 constant, then we need only compute the covariance at a total of  $K$  points (as opposed  
 1946 to  $K^2$ ), as the covariance is a circulant matrix which is fully specified by a single vector  
 1947 of length  $K$ .

1948 In the more general case where the data are not evenly sampled, we may still  
 1949 evaluate the covariance at a fixed number of points  $K' < K^2$  and approximate the  
 1950 full covariance matrix via interpolation. As long as the data are roughly evenly  
 1951 sampled, as is the case with *Kepler* or *TESS* light curves, this approximation leads  
 1952 to negligible error when  $K' \approx K$ , affording the same  $\mathcal{O}(K)$  computational savings.  
 1953 Note that even in the case where the flux is normalized (see §2.5), the non-stationary  
 1954 correction to the covariance is applied *after* the step where we marginalize over the  
 1955 inclination, so this approach is still valid.

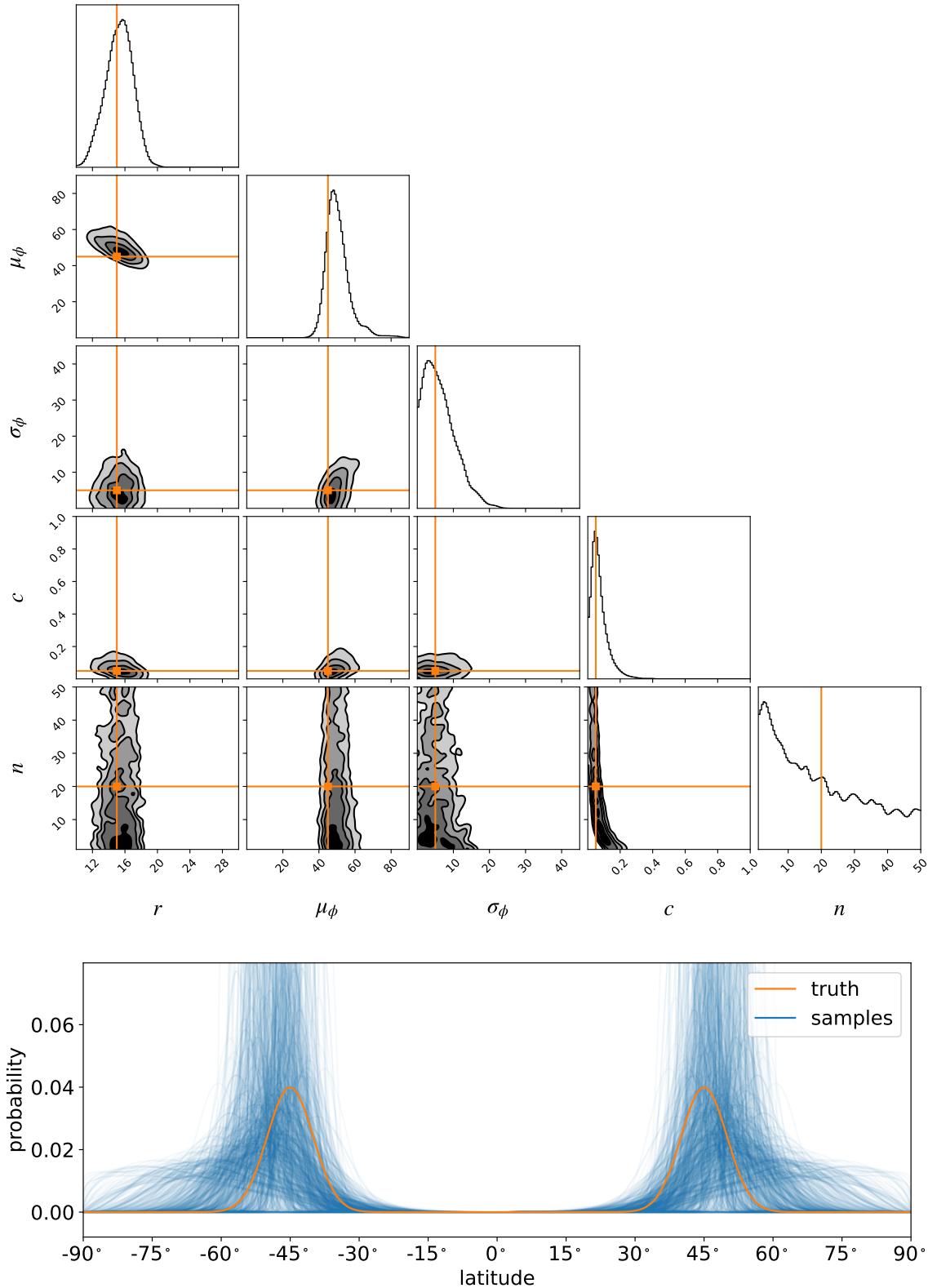
1956

## S. SUPPLEMENTARY FIGURES

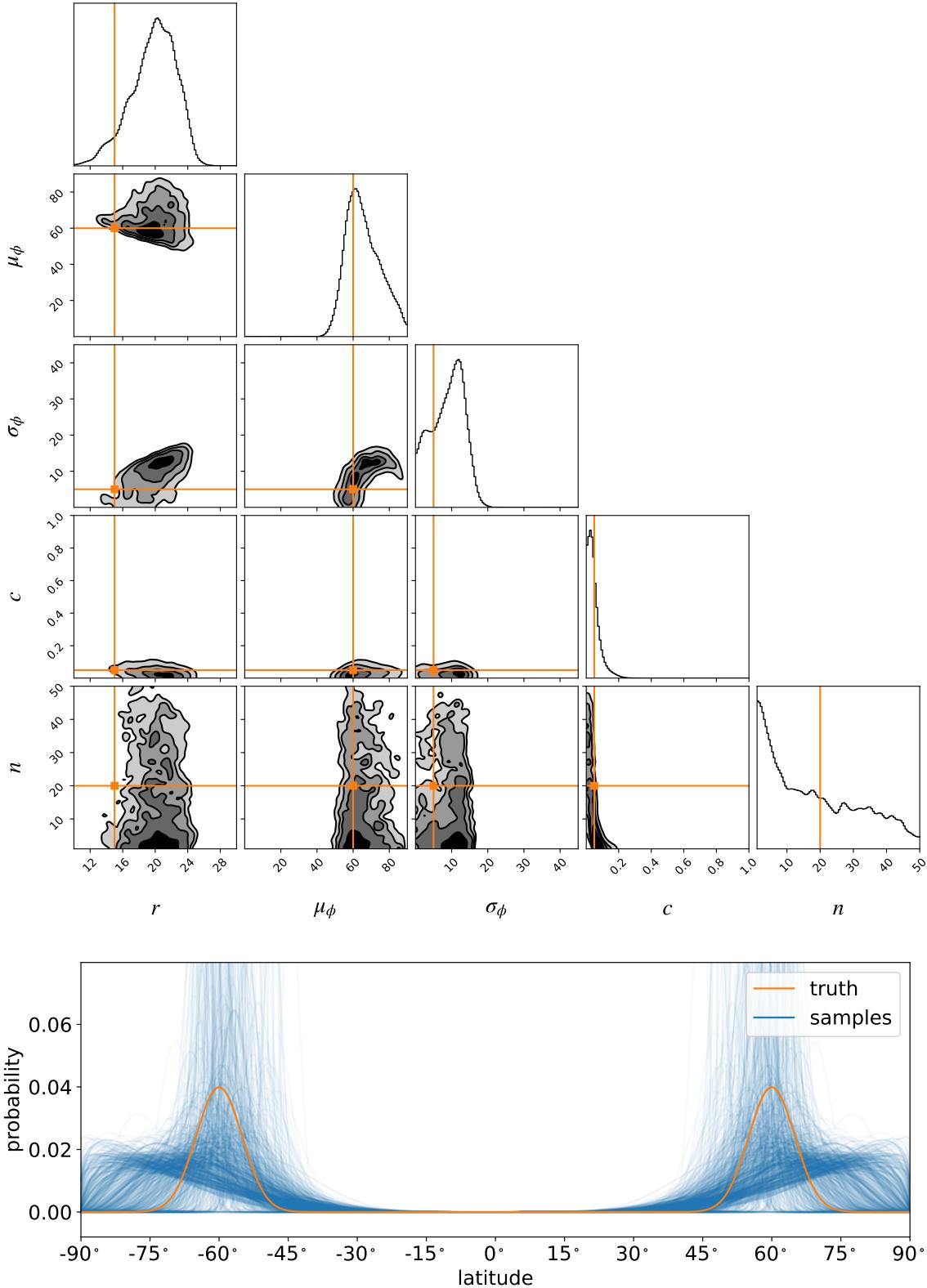
1957

Figures S17-S30 below are referenced in §3.5 discussing additional calibration runs for our GP.

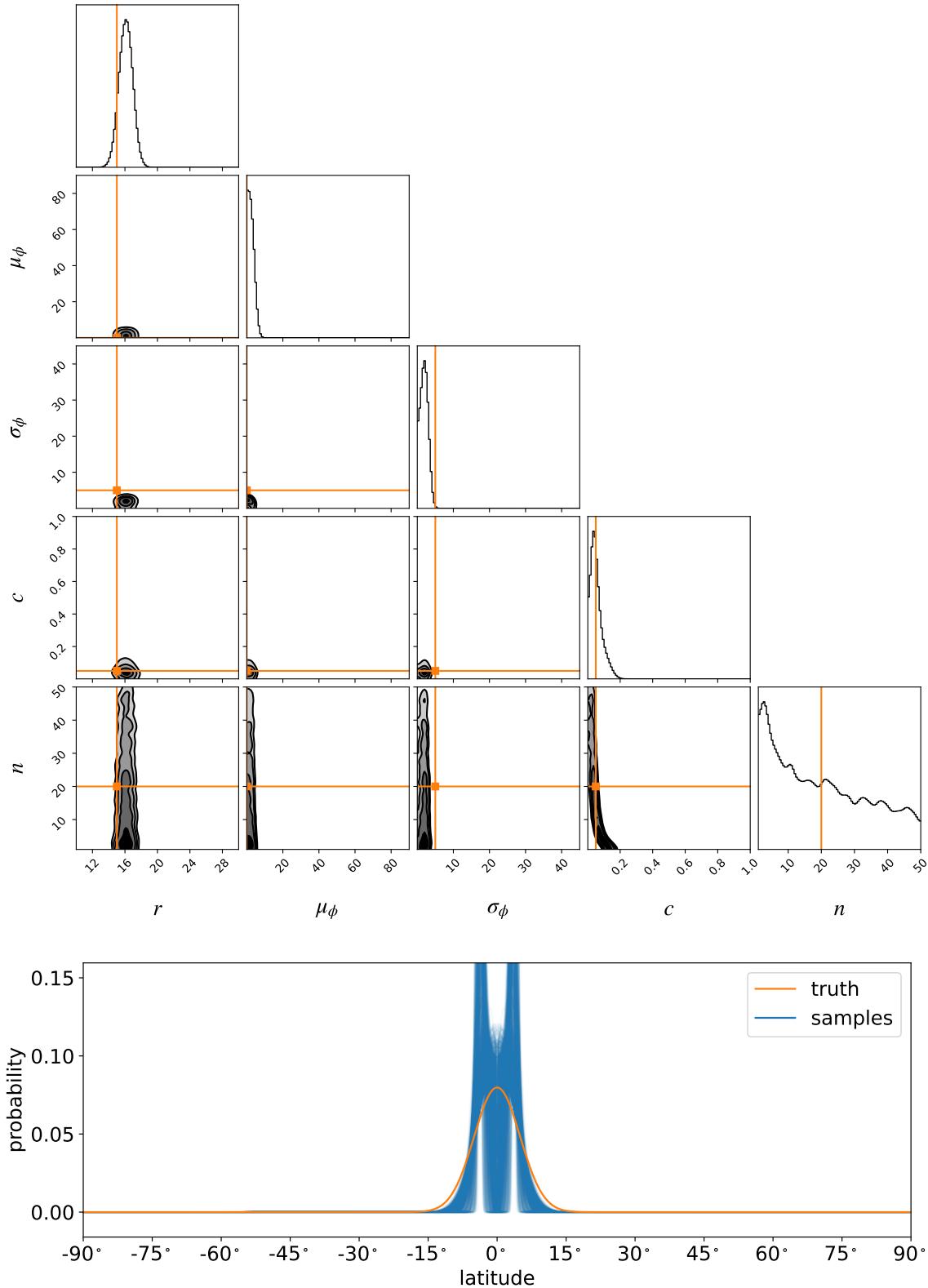
1958



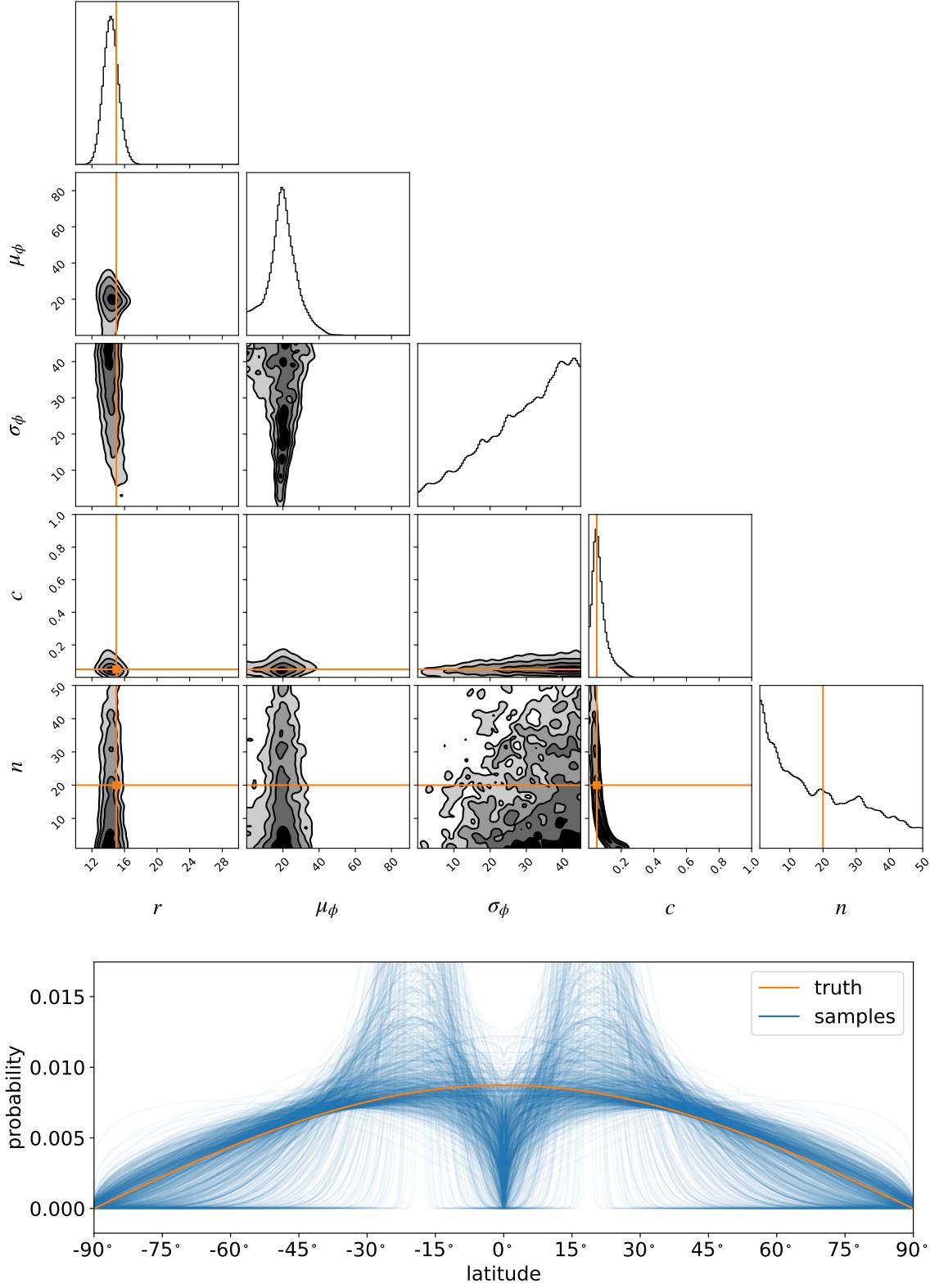
**Figure S17.** Same as Figures 6 and 7, but for mid-latitude spots with  $\mu_\phi = 45^\circ$ . The radius and latitude parameters are again inferred correctly.



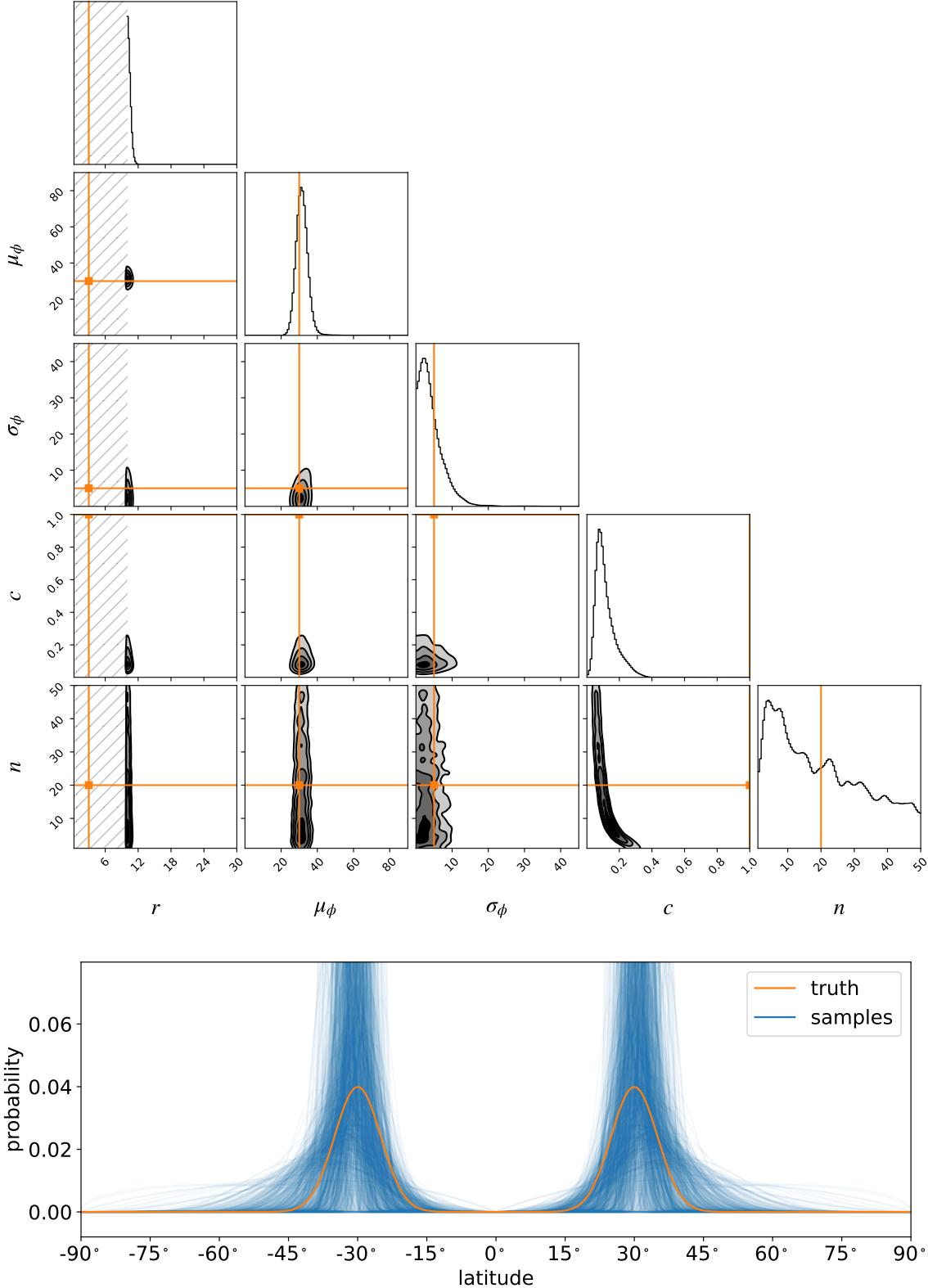
**Figure S18.** Same as Figure S17, but for high-latitude spots with  $\mu_\phi = 60^\circ$ . The radius and latitude parameters are again inferred correctly, although the model cannot rule out the presence of polar spots.



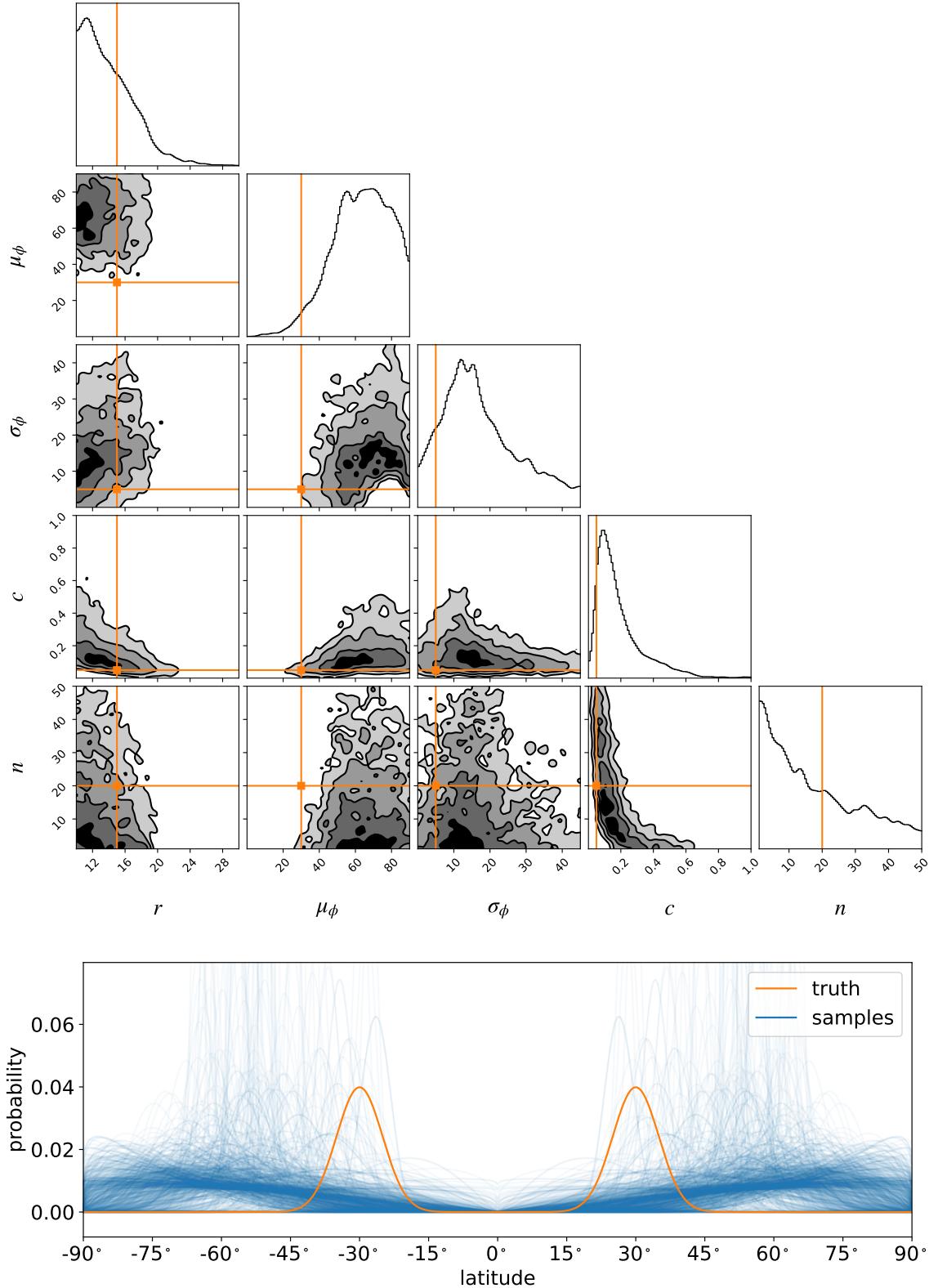
**Figure S19.** Same as Figure S17, but for equatorial spots with  $\mu_\phi = 0^\circ$ . Even though the model favors a bimodal distribution at low latitudes, the posterior strongly supports the presence of equatorial spots.



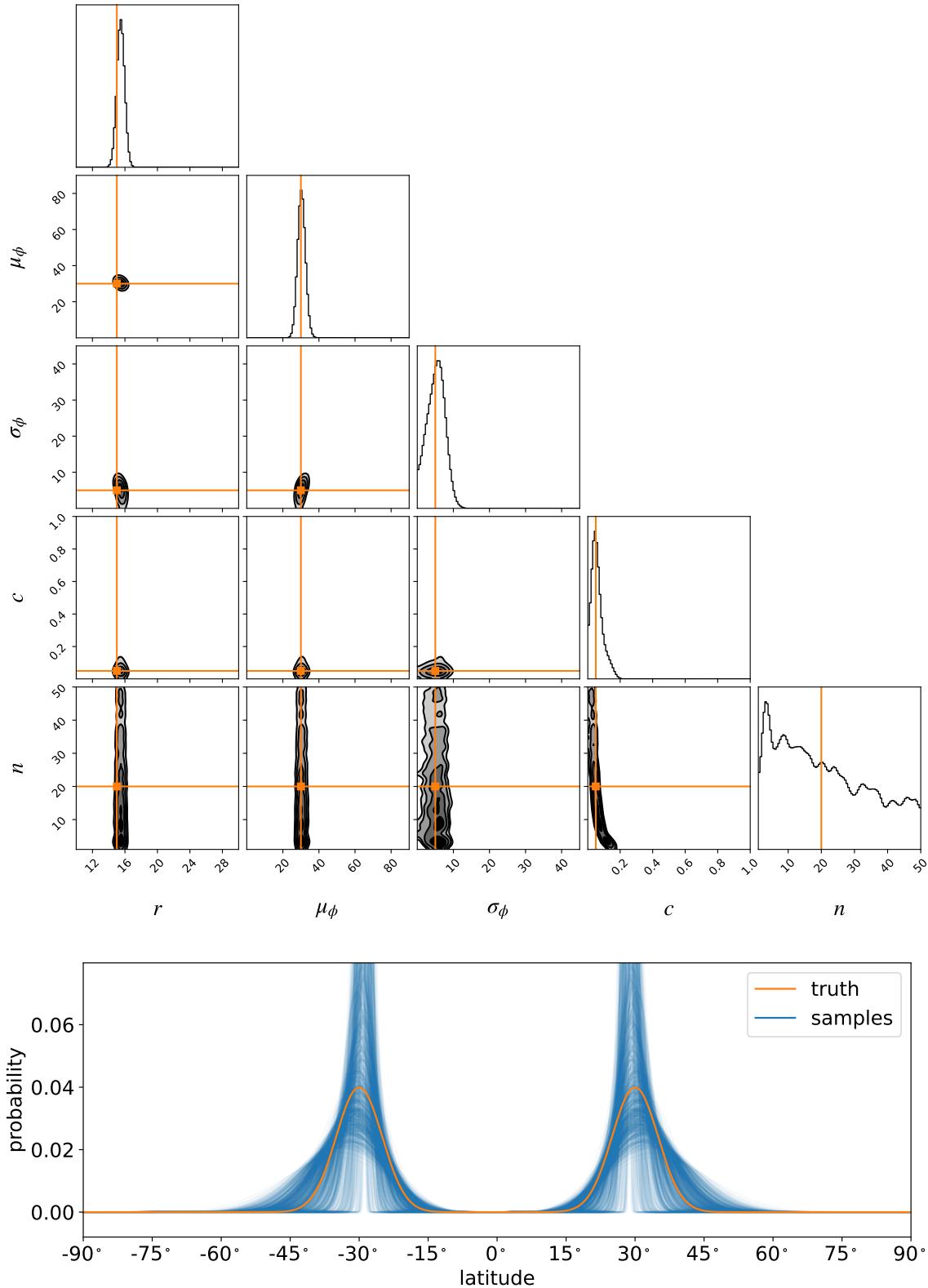
**Figure S20.** Same as Figure S17, but for isotropically-distributed spots. The posterior accurately captures the cosine-like distribution of spot latitudes.



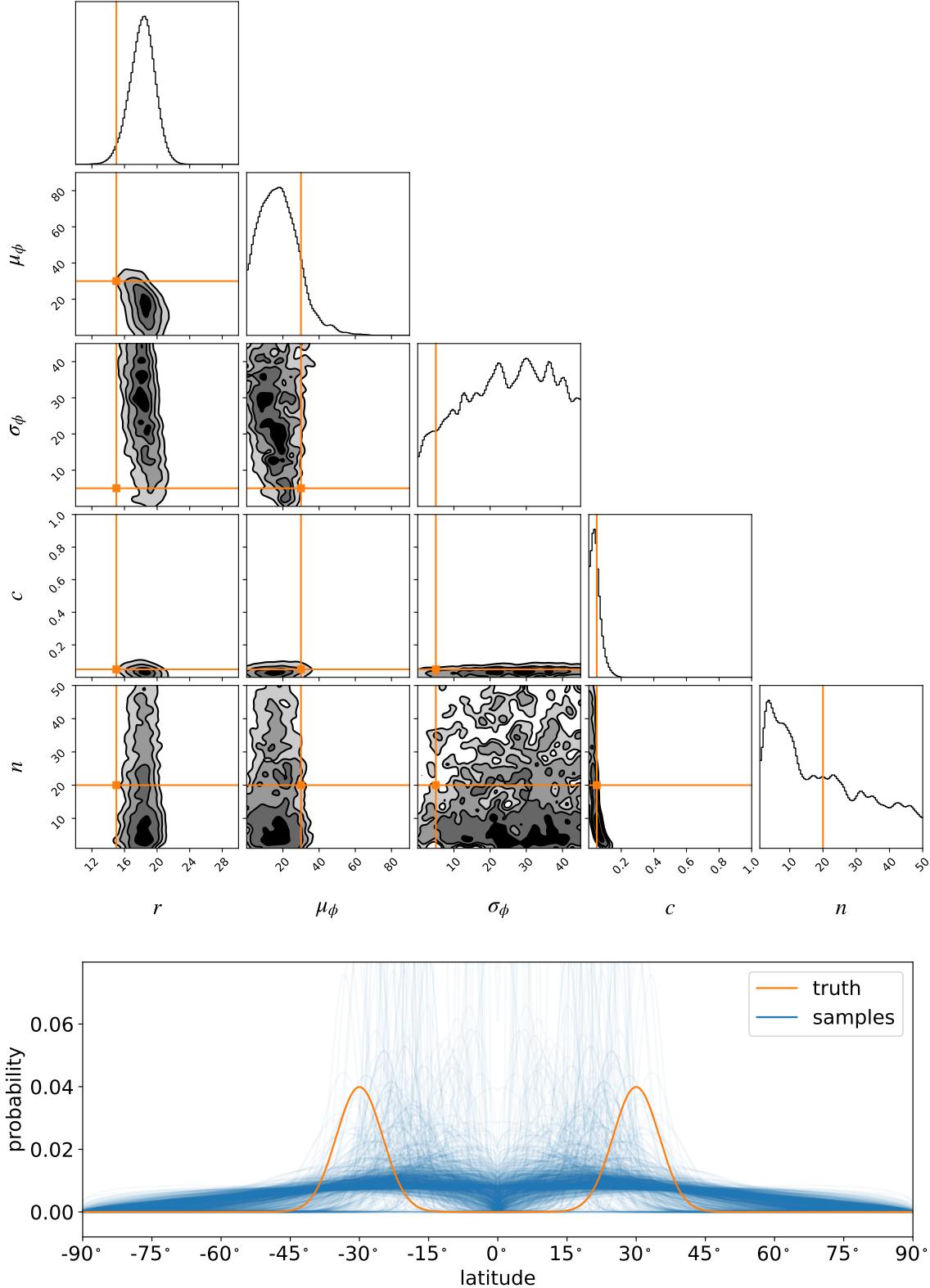
**Figure S21.** Same as Figures 6 and 7, but for high contrast ( $c = 1$ ) small ( $r = 3^\circ$ ) spots (significantly lower than the effective resolution of the model). The hatched regions in the posterior plots for the radius ( $r < 10^\circ$ ) are excluded by the prior, since the model cannot capture features that small. Despite this, the spot latitude distribution is still inferred correctly, although the spot contrast is off by more than  $10\sigma$ .



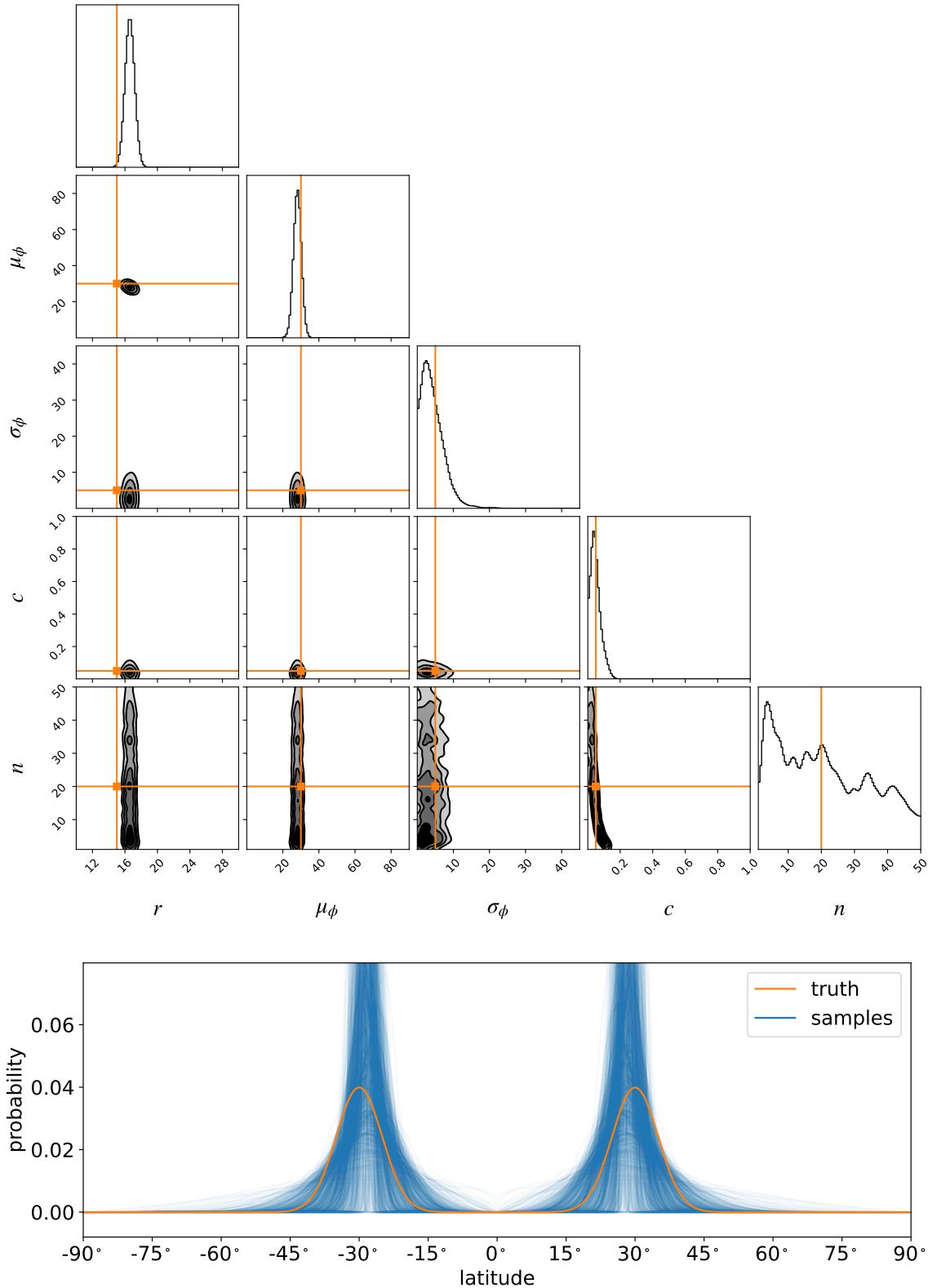
**Figure S22.** Same as Figures 6 and 7, but for inference based on a single light curve ( $M = 1$ ). The constraints on all of the parameters are dramatically weaker.



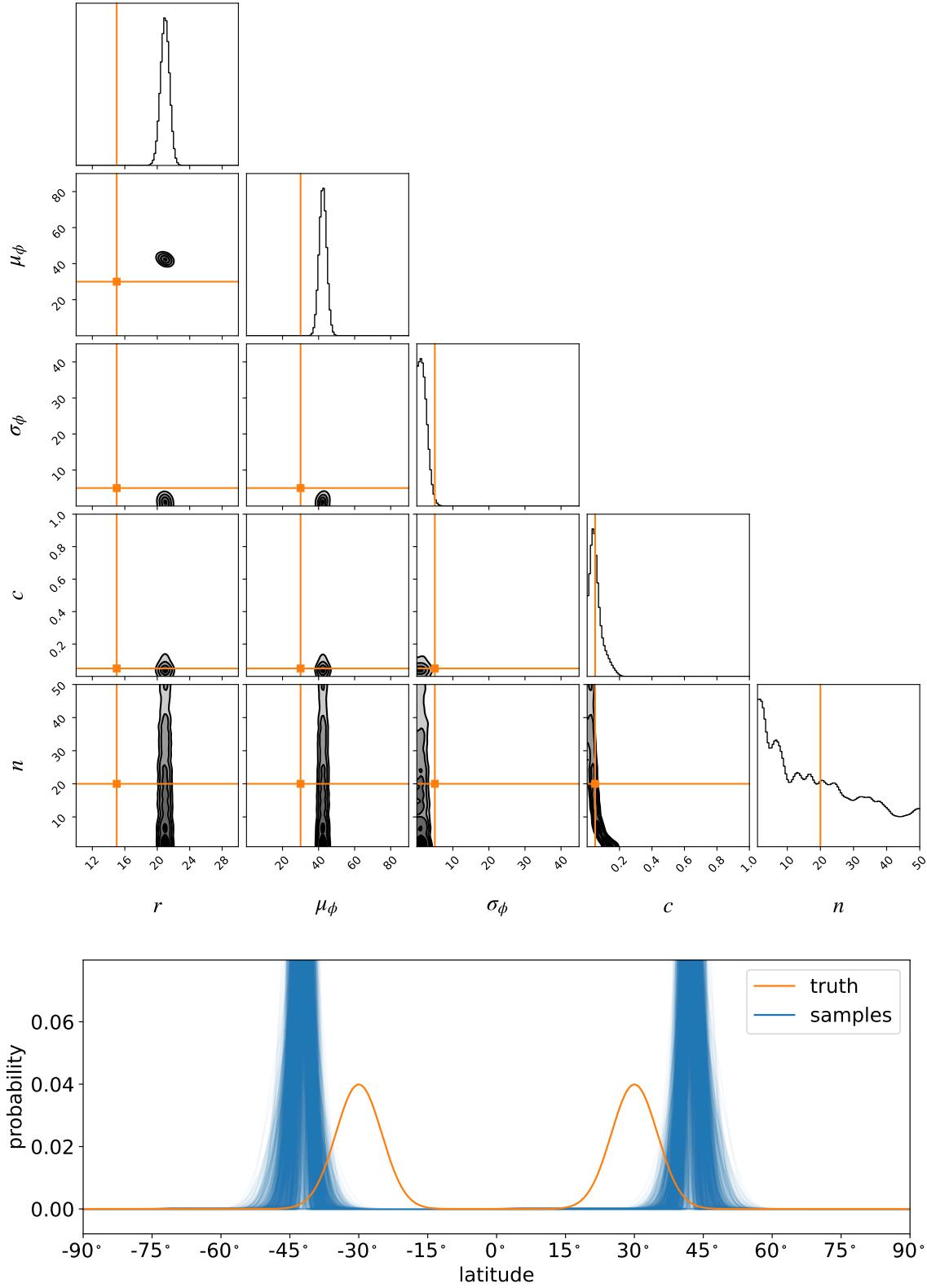
**Figure S23.** Same as Figures 6 and 7, but for inference based on one thousand light curves ( $M = 1,000$ ). The constraints on the radius and the latitude parameters are dramatically tighter. 



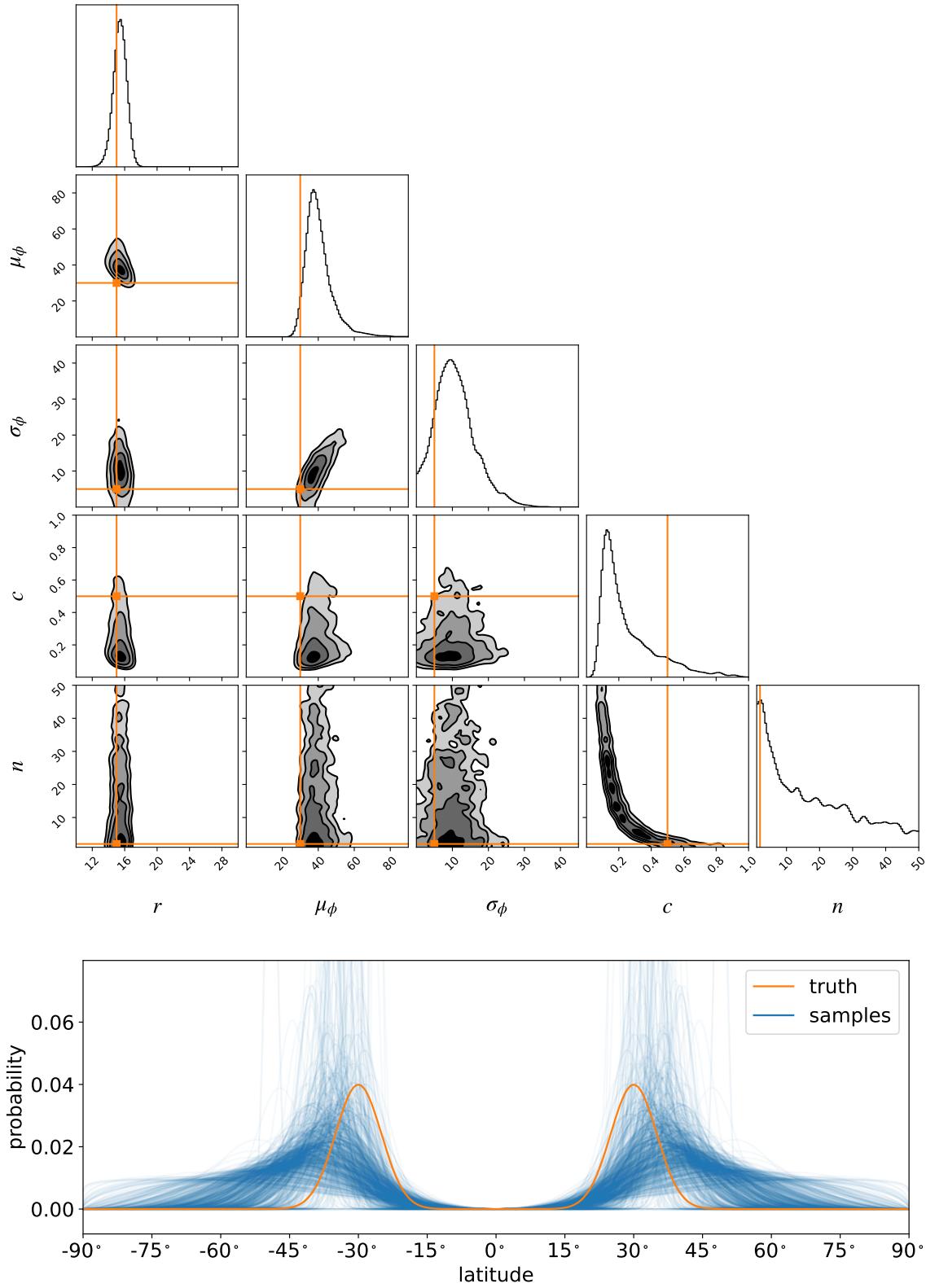
**Figure S24.** Same as Figures 6 and 7, but for stars with (assumed known) limb darkening coefficients  $u_1 = 0.50$  and  $u_2 = 0.25$ . Limb darkening makes it much harder to infer the variance of the distribution of starspot latitudes.



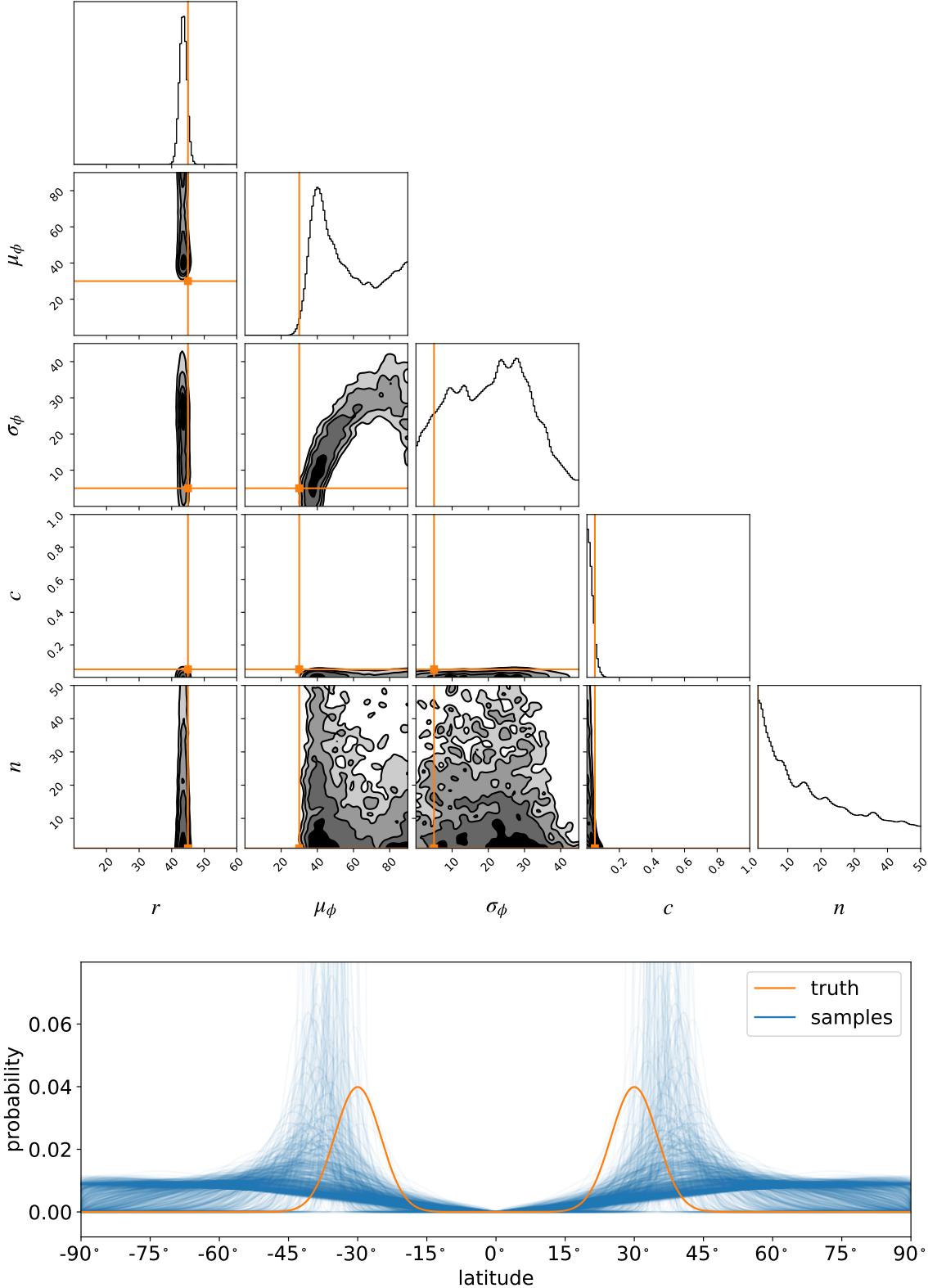
**Figure S25.** Same as Figure S24, but for an ensemble consisting of  $M = 1,000$  light curves. For a sufficiently large ensemble, it is possible to correctly infer the spot radii and latitudes in the presence of limb darkening.



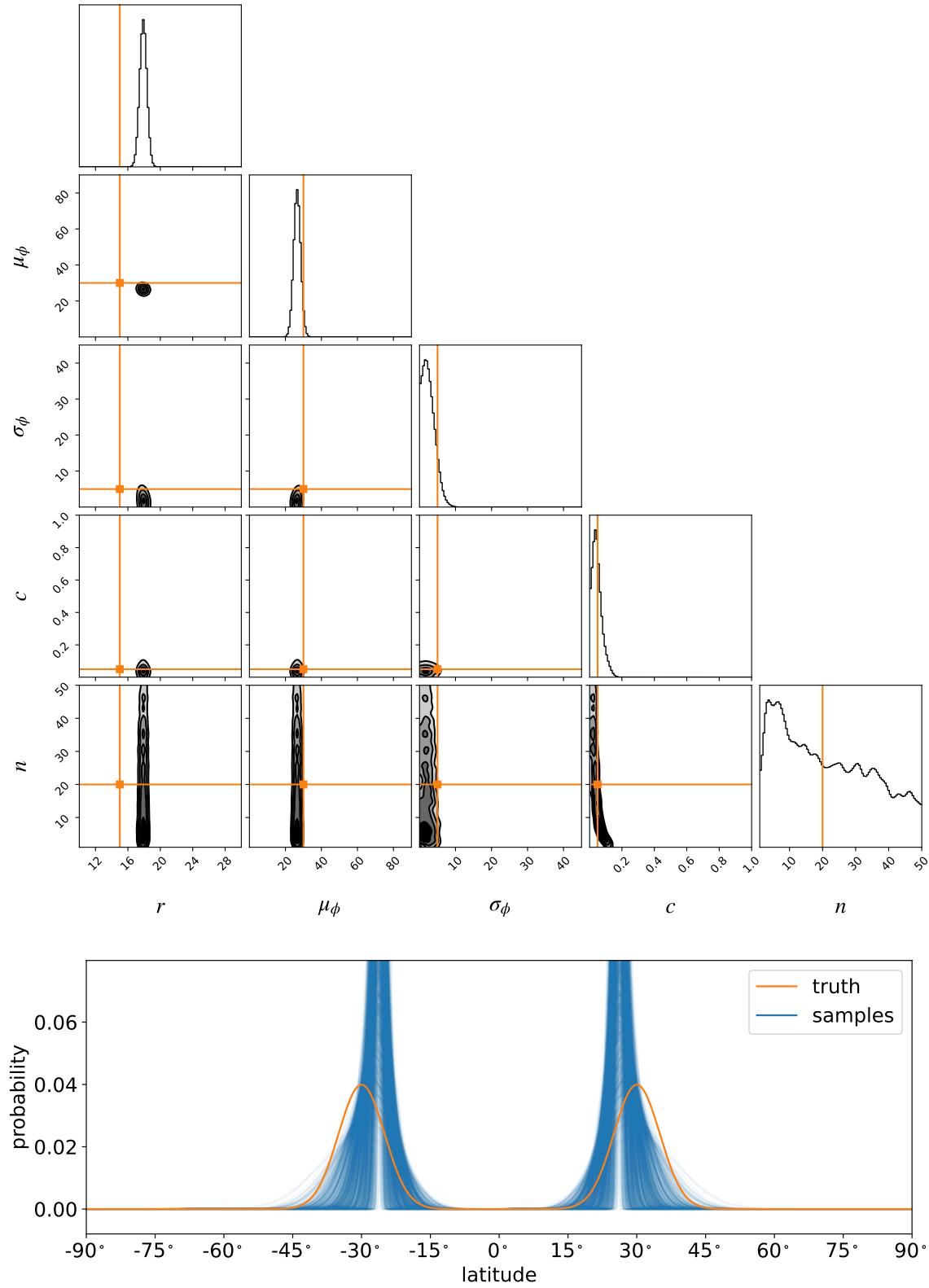
**Figure S26.** Same as Figure S24, but for  $M = 500$  light curves and assuming no limb darkening when doing inference. Neglecting limb darkening leads to significant bias in the inferred spot radii and to a lesser extent in the mean spot latitude.



**Figure S27.** Same as Figures 6 and 7, but for stars with  $n = 2$  spots with high contrast  $c = 0.5$ . Our model correctly captures the increased contrast, but it is still strongly degenerate with the number of spots.

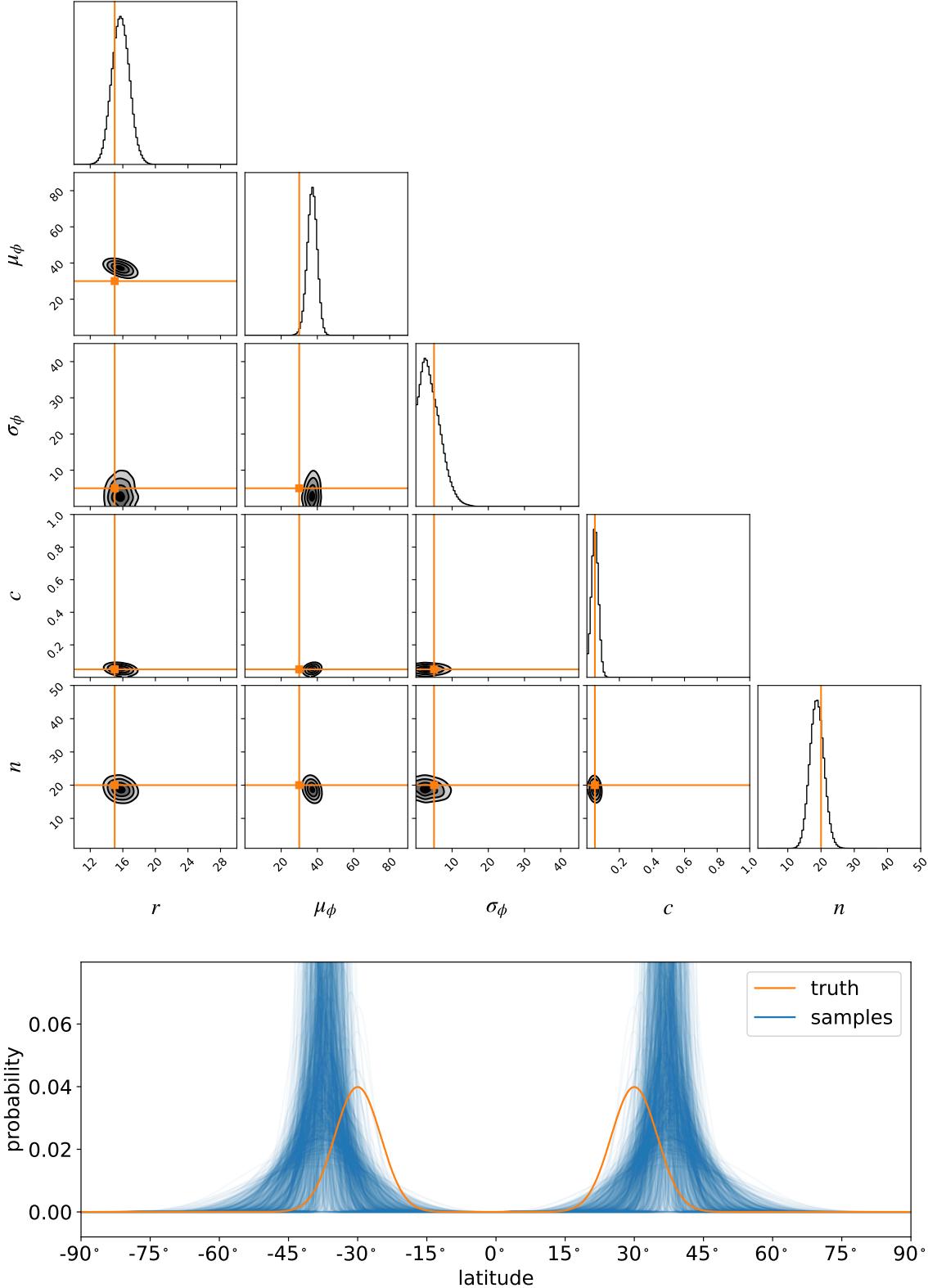


**Figure S28.** Same as Figures 6 and 7, but for a single ( $n = 1$ ) large ( $r \sim \mathcal{N}(45^\circ, 5^\circ)^2$ ) spot on each star. Our model correctly infers the larger radius, and infers the latitude within  $3\sigma$ , albeit with large uncertainty.



**Figure S29.** Same as Figures 6 and 7, 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 S30.** Same as Figures 6 and 7, but assuming the light curves are not normalized and the true amplitude is known. Knowledge of the normalization breaks the  $c - n$  degeneracy and allows us to infer the total number of spots.