

## Characterizing Sparse Asteroid Light Curves with Gaussian Processes

CHRISTINA WILLECKE LINDBERG,<sup>1</sup> DANIELA HUPPENKOTHEN,<sup>2</sup> R. LYNNE JONES,<sup>3,4</sup> BRYCE T. BOLIN,<sup>5,6</sup> MARIO JURIC,<sup>3,4</sup>  
V. ZACH GOLKHOV,<sup>4</sup> ERIC C. BELLM,<sup>4</sup> ANDREW J. DRAKE,<sup>5</sup> MATTHEW J. GRAHAM,<sup>7</sup> RUSS R. LAHER,<sup>8</sup>  
FRANK J. MASCI,<sup>8</sup> REED RIDDLE,<sup>7</sup> AND KYUNG MIN SHIN<sup>7</sup>

<sup>1</sup>*JHU, 3400 North Charles St., 473 Bloomberg Center for Physics and Astronomy, Baltimore, MD, 21218*

<sup>2</sup>*SRON Institute for Space Research, Sorbonnelaan 3, 3584 CA Utrecht, The Netherlands*

<sup>3</sup>*Department of Astronomy, University of Washington, 3910 15th Avenue NE, Seattle, WA 98195, USA*

<sup>4</sup>*DIRAC Institute, Department of Astronomy, University of Washington, 3910 15th Avenue NE, Seattle, WA 98195, USA*

<sup>5</sup>*Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA 91125, USA*

<sup>6</sup>*IPAC, Mail Code 100-22, Caltech, 1200 E. California Blvd., Pasadena, CA 91125, USA*

<sup>7</sup>*California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125, USA*

<sup>8</sup>*IPAC, California Institute of Technology, 1200 E. California Blvd, Pasadena, CA 91125, USA*

(Received TBD; Revised TBD; Accepted TBD)

Submitted to ApJ

### ABSTRACT

In the era of wide-field surveys like the Zwicky Transient Facility and the Rubin Observatory’s Legacy Survey of Space and Time, sparse photometric measurements constitute an increasing percentage of asteroid observations, particularly for asteroids newly discovered in these large surveys. Follow-up observations to supplement these sparse data may be prohibitively expensive in many cases, so to overcome these sampling limitations, we introduce a flexible Bayesian modeling framework using Gaussian Processes for parameter inference of asteroid time series data (code can be found at [AsteroGaP](#)). We show on both simulated light curves and real observations from the Zwicky Transient Facility that the new model reliably infers rotational periods from sparsely sampled light curves, and generally provides well-constrained posterior probability densities for the model parameters.

*Keywords:* Gaussian Processes — asteroids – sparse data sets – etc.

### 1. INTRODUCTION

Asteroids are small rocky bodies known to contain information about the formation of planetary objects in our Solar System. Their shapes, sizes, spin, orbits, and compositions all help us constrain models for asteroid formation, collision history, and orbital evolution. These models influence our understanding of the protoplanetary disk along with the evolution and distribution of materials such as water in our Solar System.

Asteroids are observed in the light they reflect from the sun. The amount of light observed by a telescope on Earth at a given point in time depends on the illuminated asteroid surface visible by the telescope as well as surface properties such as composition and albedo. Because asteroids move with respect to both Earth and Sun, and also rotate around their own axis, the observed flux will change as a function of time (Barucci & Fulchignoni 1982).

Consequently, asteroid light curves can be described by parameters like rotational period, amplitude of the magnitude variations, and the variability in magnitude within a rotational period (which we denote *rotational profile* in this paper). Determining these parameter values can constrain physical asteroid characteristics such as elongation, shape, and structural properties (e.g. whether it is a rubble pile vs. a monolithic rock, Masiero et al. (2009)).

A standard ellipsoidal asteroid will produce a light curve that is roughly sinusoidal and double-peaked, which can easily be modeled with an ensemble of sine functions and characterized by a Fourier transformation (Harris et al. 2014). However, if an asteroid is not strictly tri-axial or is composed of surface materials with varying albedos, then rotational profile will become complex. While one can still derive Fourier representations of these complex light curves, their interpretation is much more difficult, since the variability within each period will

spread power across a wide range of frequencies. If the data is sparse, it then becomes difficult to derive an accurate rotational period. Recently, [Schemel & Brown \(2020\)](#) introduced a method to infer certain asteroid characteristics (e.g. color, phase, parameters, absolute magnitude, amplitude of rotation) with sparse photometric data by assuming that the light curve can be described by a single sinusoid, however, this leaves numerous other asteroid characteristics unconstrained.

Asteroid research has for the most part utilized phase dispersion minimization ([Stellingwerf 1978](#)) or Fourier transform-based methods like Fourier Analysis of Light Curves (FALC) ([Harris et al. 1989](#)) and Lomb-Scargle periodograms ([Lomb 1976](#); [Scargle 1982](#)) to characterize rotational periods and amplitudes. It is not uncommon that groups relying on differing densities of observations will report significantly different period estimates for the same object when relying on Fourier-based methods ([Fedorets et al. 2020](#); [Bolin et al. 2020](#)). These approaches have been shown to be less reliable when light curves are sparsely sampled, producing multi-peaked periodograms ([Masiero et al. 2009](#); [Warner & Harris 2011](#); [Harris et al. 2012](#)) and resulting in period estimates that can differ from the period derived through other data sets or methods. One potential solution is to gather denser time series data, which is expensive, and may be infeasible for large samples of asteroids.

Future asteroid time series data will primarily be collected from large surveys with telescopes such as the Rubin Observatory, which is expected to observe 500,000 asteroids per night with two observations per asteroid ([Jones et al. 2018](#); [Ivezić et al. 2019](#)). Likewise, the Zwicky Transit Facility (ZTF) detects 25,000 asteroids on a clear night, covering the northern sky every three nights ([Bellm et al. 2019a](#); [Graham et al. 2019](#)). Gathering dense light curves is an impractical task for these large new samples of asteroids. Hence, reliable inference of the properties of the millions of asteroids to be detected will require adequate methods to characterize these sparse light curves. Here, we propose a new model based on Gaussian Processes to overcome these obstacles and provide an alternative approach to characterizing asteroid light curves.

Gaussian Processes ([Rasmussen & Williams 2006](#)) provide flexible and generative models for time series data, and have been used to model, among others, non-sinusoidal light curves of stars ([Brewer & Stello 2009](#)) and Active Galactic Nuclei ([Kelly et al. 2014](#)). This is done by modeling the covariance between data points, rather than modeling the light curve directly. In this paper, we show that we can obtain accurate estimates of asteroid properties, in particular the rotational period,

with the use of a Gaussian Process model and sampling via Markov Chain Monte Carlo (MCMC). The flexible, generative model admits periodic, but non-sinusoidal functions as a way to encode the rapid, intra-period variability seen from many asteroids.

The development of this model was initially motivated by sparse observations of the interstellar object 1I/'Oumuamua, observed in late 2017 ([Bolin et al. 2018](#)). Using photometric time-series data from the Apache Point Observatory (APO) and the Discovery Channel Telescope (DCT), [Bolin et al. \(2018\)](#) constructed a light curve and estimated a rotational period. In this context, a Gaussian Process yielded a well-constrained estimate of the period despite the sparsity of the data. It is with this sparsity in mind that we continue to develop and validate this Gaussian Process method.

The structure of this paper is as follows: In Section 2, we briefly review the concept of Gaussian Processes, covariance functions, Bayesian priors, and MCMC samplers used in our model. In Section 3, we outline our simulations and give a motivating example of our model being used on a well-known nearby asteroid. In Section 4, we show results for asteroids recently observed with ZTF, and in Section 5, we discuss the potential of this model and describe future improvements.

## 2. METHODS

### 2.1. Bayesian Statistics

In Bayesian inference, we estimate a probability distribution of model parameters, given observed data and prior knowledge about the model and its parameters. Bayes' theorem is defined as

$$p(\theta|D, M) = \frac{p(D|\theta, M)p(\theta|M)}{p(D|M)}. \quad (1)$$

In this equation,  $p(\theta|D, M)$  denotes the *posterior probability distribution* of the model parameters  $\theta$ , given an observed set of data points  $D$  and an assumed model  $M$ . This assumed model  $M$  deviates from standard astronomical terminology in that its definition includes not only a physically motivated function to model the data, but also a number of implicit assumptions, including the type of equation used to represent the data points—in our case a Gaussian Process—as well as the model for the uncertainty on the data points (the likelihood) and other choices made during the data analysis process. The right-hand side of the equation defines two important components for inference of the parameters  $\theta$ . The *likelihood* describes the probability of observing a data set  $D$ , given some underlying model  $M$  with parameters  $\theta$ , which are assumed to be known exactly. In practice, we generally observe data and aim to infer the parameters,

which makes the likelihood a *function* of the parameters, but a *probability distribution* of the data. The likelihood includes both a parametric model of the underlying (physical) process assumed to have generated the data, as well as a model for the measurement uncertainties.

The prior probability distribution,  $p(\theta|M)$  encodes our knowledge of the parameters  $\theta$  before looking at the data. For example, asteroids cannot have arbitrary rotation periods. The possible values for the rotational period of an asteroid are constrained both by physics—for example the material the asteroid is made of—and prior knowledge about the range of periods we have observed from similar asteroids in the past. We encode that information in the form of prior probability distributions, specified in detail in Section 2.3.

The final term in Bayes’ theorem in the denominator is called the *evidence* or *marginal likelihood*. It is a normalization term ensuring that the posterior  $p(\theta|D, M)$  integrates to unity as a proper probability distribution. While important for the purposes of model selection, it is independent of the parameters  $\theta$ , and is a constant that that does not affect the shape of the posterior, only its normalization. We can therefore write Bayes’ theorem as

$$p(\theta|D, M) \propto p(D|\theta, M)p(\theta|M) \quad (2)$$

when performing inference on the parameters.

In most practical applications, the posterior defined in Equation 2 is analytically intractable and needs to be approximated numerically. One way to do this is through Markov Chain Monte Carlo (MCMC) sampling. In the ideal case of an infinitely long sampling process, MCMC guarantees convergence to the correct posterior distribution. During an initial burn-in phase, an MCMC chain—a single sampling process—should converge to a stationary distribution and ideally, any samples subsequently generated are drawn from the true posterior distribution. Since infinite MCMC runs are impossible, we run the sampler for a specified number of iterations, and check for convergence—using autocorrelation times and the visual inspection of MCMC chains—to see whether the chains have converged to the posterior distribution. For sparse asteroid observations, we expect the posterior probability distribution for the period to be multimodal, i.e. exhibit multiple peaks. This is caused by the sparsity of the data, which admits multiple periods as valid models. It is therefore difficult for any sampler to sample efficiently since chains may become constrained within their localized probability maxima and never get a chance to sample other regions of the parameter space. In this work, we use MCMC as implemented in the python package *emcee* (Foreman-Mackey et al. 2013).

## 2.2. Gaussian Processes

Gaussian Process models are a generative kernel-based framework for solving both classification and regression problems (Roberts et al. 2012); here, we focus on Gaussian Processes as a method for regression. Rather than modeling the data points directly as in standard linear regression, a Gaussian Process provides a model for the *covariance* between data points. Let us assume we have obtained  $N$  photometric measurements  $D = \{m_i, \sigma_m\}_{i=1}^N$  of an asteroid, recording magnitude and magnitude error for every data point at times  $T = \{t_i\}_{i=1}^N$ .

In standard linear regression with a Gaussian likelihood<sup>1</sup>, one defines a parametric model  $f(t_i, \phi)$  specified by a set of parameters  $\phi$  that aims to approximate the structure in the observed data points. One then compares model magnitudes derived for a specific set of values for  $\phi$  to the observations using a Gaussian likelihood, where the mean of the Gaussian is set by  $f(t_i, \phi)$ , the variance is set by measurement uncertainties,  $\sigma_m^2$ , and data points are assumed to be independent random variables, i.e. there are no covariances between data points. It is possible to write this model as a single Gaussian distribution,

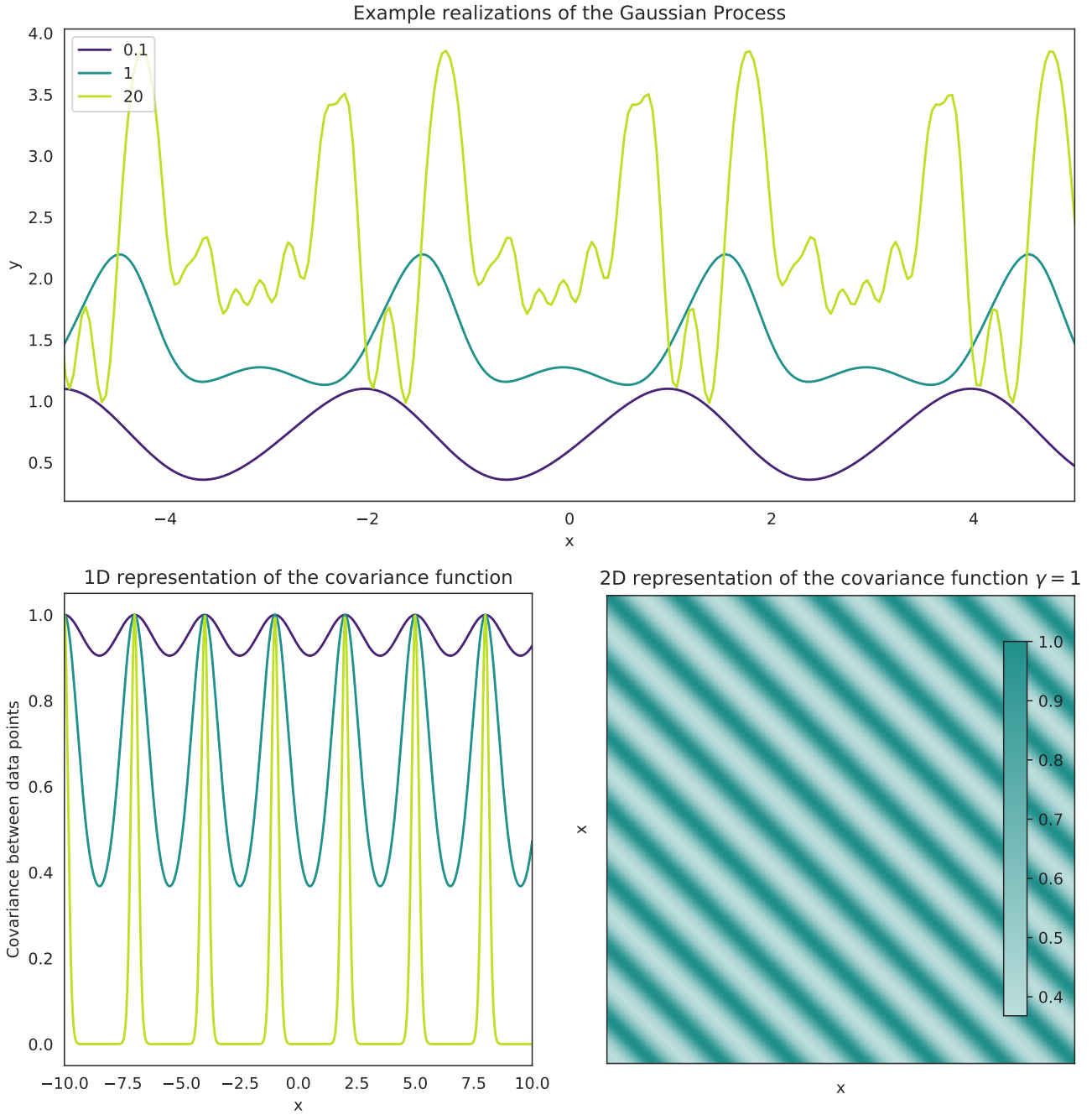
$$D \sim \mathcal{N}(f(T, \phi), \Sigma^2)$$

where  $D$  is the vector of  $N$  data points,  $\mathcal{N}$  is an  $N$ -dimensional Gaussian distribution with means set by  $f(T, \phi)$  and  $\Sigma$  is an  $N \times N$  dimensional matrix where, in this standard case, the off-diagonal elements are zero and the diagonal elements are set by  $\sigma_m^2$ .

A Gaussian Process extends this model by allowing covariances between data points to exist, and by defining a model for that covariance. It is a generalization of the standard Gaussian likelihood into a process that defines an infinite-dimensional Gaussian distribution. One may still define a mean function  $f(t, \phi)$  that governs global trends in the data set, but this model now also contains a model for the covariances, called a covariance function or a kernel,  $k(t_i, t_j, \theta)$  with its own set of parameters  $\theta$ . Instead of drawing a vector as one would from a finite Gaussian distribution, draws from a Gaussian Process are functions defined by the mean function  $f$  and covariance function  $k$ . The draws, evaluated in finite dimensions and conditioned on the observations, serve as model light curves for the observed data points.

Modeling the covariance between every pair of data points through the covariance function allows us to build flexible models in particular for stochastic processes. Here we assume that the small-scale structure

<sup>1</sup> also called  $\chi^2$  statistic in many astronomical contexts



**Figure 1.** Top: Three example realizations drawn from a Gaussian Process with a periodic covariance function. All three realizations were drawn from a Gaussian Process with a period of  $P = 3.0$  and an amplitude of  $A = 1$ , but variable intra-period lengthscale  $\Gamma$ . Bottom left: visualization of the covariance function that generated the realizations in the top panel. Given the period of  $P = 3$ , points that are integer multiples apart from each other will be exactly correlated, and thus generate a periodic pattern. The variations within each period depend on the value of  $\Gamma$ : if  $\Gamma$  is large, the correlation between points that are close together is near zero, generating rapidly varying structure within each period. If  $\Gamma$  is small, the correlation between data points is always high, generating smooth, nearly sinusoidal realizations. Bottom right: two-dimensional representation of the covariance function with  $\Gamma = 20$  from the left panel. The light yellow bands are regions of high covariance spaced at intervals of multiples of  $P$ , the dark purple bands in between show regions where the covariance between data points for a given  $\Delta x$  is near 0.



and variations in albedo of an asteroid, generating rapid magnitude variations, can be approximated by such a stochastic process.

The logarithm of the likelihood for the Gaussian Process model is defined as

$$\begin{aligned} \ln \mathcal{L}(\phi, \theta) &= \ln p(D|T, \phi, \theta) \\ &= -\frac{1}{2} r_\phi K_\theta^{-1} r_\phi - \frac{1}{2} \ln \det K_\theta - \frac{N}{2} \ln(2\pi) \end{aligned} \quad (3)$$

where

$$r_\theta = (m_1 - f(t_1, \phi) \dots m_N - f(t_N, \phi))^T \quad (4)$$

represents the vector of residuals and  $K_\theta$  represents the covariance matrix (Rasmussen & Williams 2006; see also Foreman-Mackey et al. 2017). The covariance matrix is defined as the  $N \times N$ -dimensional subset of the infinite-dimensional covariance function  $k$  for all observation time stamps  $(t_i, t_j)$ .

The covariance function defines the covariance matrix  $K_\theta$  as a function of parameters  $\theta$  and thus describes how pairs of data points are expected to covary under the assumptions of the model. There are both stationary and non-stationary covariance functions. Stationary covariance functions are often defined in terms of the distance between data points:  $k(t_i, t_j) = k(|t_i - t_j|)$  such that the statistical properties of the data-generating process do not change as a function of time. Covariance functions constructed with a symmetric function that generates periodic processes are a special case of non-stationary covariance functions.

In this work, we use the *sine-squared exponential covariance function*, which requires that the covariance between data points spaced at a certain periodic intervals  $P$  is unity. This means that for any given point in time  $t_i$ , data points at multiples of  $P$  from  $t_i$  will be exactly equal, while data points at a distance of  $t_i + \Delta t$ , where  $\Delta t < P$  may be correlated or uncorrelated. This defines a process that is strictly periodic, but not sinusoidal, allowing for rapid changes within a period. This covariance function is formally defined as

$$k_{\text{per}}(t_i, t_j) = A_{\text{periodic}} \exp(-\Gamma \sin^2[\frac{\pi}{P}|t_i - t_j|]) \quad (5)$$

The behaviour of this covariance function is governed by a number of parameters<sup>2</sup>. The amplitude  $A_{\text{periodic}}$  scales the covariance function to match the amplitude of

the light curves (see also Table 1).  $\Gamma$  is defined as the inverse of a length scale that dictates the distance between data points within a single period at which correlations tend to zero. This models the intra-period variability of our light curve, which can reveal information about the asteroid shape and albedo variations. Finally,  $P$  denotes the period and indicates the interval where the correlation between data points becomes unity. In this work, we use this parameter as a proxy for the rotational period of the asteroid. We expect that the sine-squared exponential covariance function provides an adequate representation for asteroid light curves, since it accounts for intra-periodic variability and assumes our data is periodic without necessarily being sinusoidal. However, we also note that the strict periodicity requirement it imposes may make the model unstable for light curves spanning very long intervals (e.g. years) or tumbling asteroids that rotate around non-principal axes.

In Figure 1, we show examples of time series generated by a Gaussian Process with different values of  $\Gamma$ , along with the corresponding covariance functions and behaviour. For  $\Gamma = 0.1$ , the covariance function is never zero, i.e. the covariance between data points is always high, and data points are never independent. This leads to very smoothly varying functions. In contrast, for larger values of  $\Gamma$ , the covariance function tends to 0 in between multiples of  $P$ , which allows the model to produce more complex, variable functions.

Asteroids exhibit light curves with significant intra-period variability, which we aim to model with the covariance function presented in Equation 5. For long observations, the rotational profile changes as a function of time: as the angle between Earth, Sun and asteroid changes, so does the reflecting surface visible from Earth. The covariance function generates strictly periodic function with the same rotational profile throughout. In order to model the nature of a slowly varying rotational profile, we need to add another component to the model.

A covariance function that is often used is the *squared exponential covariance function*, a stationary covariance function that models the covariance as an exponentially decreasing function of distance between data points, i.e., observations taken close together are expected to be more highly correlated than observations taken far apart. This covariance function is formally defined as

$$k_{\text{SE}} = A_{\text{long}} \exp(-\frac{|t_i - t_j|^2}{2M}) \quad (6)$$

<sup>2</sup> In the machine learning literature, the kernel parameters are often called "hyperparameters", because they kind of consider the model data points as the actual parameters. However, in Bayesian

inference, "parameters" generally refers to the parameters of the model, and hyperparameters are the parameters used to define the prior distributions. In this paper, we will consistently use Bayesian inference terminology.

The metric  $M$  determines the length scale of the correlation between data points taken at a distance of  $|t_i - t_j|$  from each other. A smaller metric leads to a correlation that drops off faster than a larger metric.

The product of this covariance function with the sine-squared exponential covariance function allows the Gaussian Process to generate functions that change the shape of their intra-period variability over time, governed by the length scale  $M$ . We use this combination of covariance functions to mimic the natural evolution of an asteroid rotational profile and its phase-angle between the Earth and the Sun changes during the course of its orbit.

Given these covariance functions, we can now compute an  $N \times N$  covariance matrix  $K_\theta$  for each of the  $N$  data points in the data set, and generate model light curves conditioned on the observed data. We include the magnitude errors into the model by adding a noise constant to the covariance matrix along the diagonal:

$$k_{\text{WN}}(t_i, t_j) = \delta_{ij} \sigma_i^2. \quad (7)$$

Our final model includes a mean function  $f(t, \phi) = c$  which we set to be constant, and the final covariance function

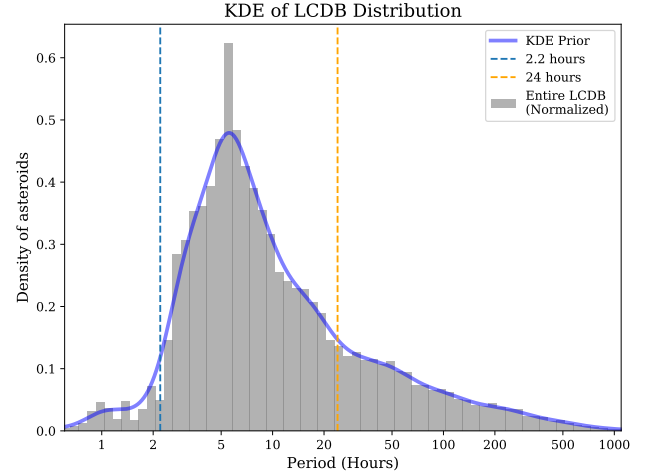
$$k(t_i, t_j) = k_{\text{SE}}(t_i, t_j) k_{\text{per}}(t_i, t_j) + k_{\text{WN}}(t_i, t_j). \quad (8)$$

We use this model to estimate the likelihood defined in Equation 3. Here, we use Gaussian Processes implemented in the Python package *george* (Ambikasaran et al. 2015).

### 2.3. Priors

Using a Gaussian Process with a constant mean function, a squared exponential covariance function, and a sine-squared exponential covariance function, we need to define prior distributions for six parameters, defined in Table 1.

We generate a prior probability density distribution for the period parameter  $P$  by approximating the observed distribution of asteroid rotational periods derived from the sample of 33969 known asteroids available in the Lightcurve Database (Warner et al. 2009a) as of March 25th 2020, shown in Figure 2. We use a Gaussian kernel density estimate (KDE) (Rosenblatt 1956; Parzen 1962) as implemented in SciPy (Virtanen et al. 2020) to approximate this distribution. We implement hard cutoffs in our prior to only allow periods longer than 1 minute and shorter than one year (8760 hours) to avoid unphysical period suggestions. We note that this prior distribution is unlikely to adequately reflect the true distribution of observed periods, but rather a convolution of



**Figure 2.** This figure shows the normalized distribution of recorded asteroid period values from the Lightcurve Database (LCDB) as of March 2020 (grey) and the kernel density estimate distribution of our period prior (blue).

that distribution with a number of observational biases (Jedicke et al. 2016; Masiero et al. 2009). We considered two alternative priors: a uniform distribution and an exponentially decaying distribution. Because the rotational periods of asteroids can span multiple orders of magnitude—from under an hour to hundreds of hours—we discarded the option of a flat (uniform) prior early on, since it would disproportionately favour very long periods, unlike what is seen in practice. Conversely, an exponentially decaying prior (often implemented as a uniform prior on the logarithm of the relevant parameter), would strongly favour extremely short periods. We arrived at the empirically derived distribution above as a compromise: while it might encode some observational biases, it allows for a wide range of possible periods, and correctly assigns a large probability where we expect most asteroid rotational periods to exist. For this work, where we infer the properties of individual asteroids, we expect that this prior will suffice. However, we caution the reader that using this model e.g. for population inference will require a more nuanced consideration of the period prior.

We impose a strict positivity requirement on both amplitude parameters,  $A_{\text{long}}$  and  $A_{\text{periodic}}$ , which by definition must not be negative. We use the logarithms of the amplitudes,  $\log(A)$  as parameters rather than  $A$  itself, and assign a uniform (for  $A_{\text{long}}$ ) and a normal distribution (for  $A_{\text{periodic}}$ ) to these parameters, respectively. Hyperparameter values for both are chosen to reflect magnitude changes broadly typical for asteroid observations. The periodic lightcurve amplitude,  $A_{\text{periodic}}$ , may increase or decrease for a given object based on its

**Table 1.** Overview of the priors used in the Bayesian models

Parameter Name	Parameter	Prior distribution	Hyperparameters
Mean constant magnitudes <sup>a</sup>	$c$	normal	$\mu_m = 1, \sigma_m = 0.5$
Periodic kernel, amplitude	$\log A_{\text{periodic}}$	normal	$\mu_{A,p} = \log(0.15), \sigma_{A,p} = \log(2)$
Periodic kernel, inverse length scale	$\log \Gamma$	normal	$\mu_\Gamma = \log(10), \sigma_\Gamma = \log(2)$
Periodic kernel, period	$\log P$	kernel density estimate	bandwidth = 0.2 / $\sigma(\text{data})^b$
Squared Exponential kernel, amplitude	$\log A_{\text{long}}$	uniform	$[-10, 10]$
Squared Exponential kernel, length scale <sup>c</sup>	$\log M$	normal	$\mu_M = \log(100), \sigma_M = (10)$

<sup>a</sup> Simulated magnitude data is normalized to 1 and ZTF magnitude data is corrected to 0.

<sup>b</sup>  $\sigma(\text{data})$  here refers to the standard deviation of the data.

<sup>c</sup> This is often also referred to as *metric*, which we will use below as the parameter name to easily distinguish it from the inverse length scale of the periodic kernel.

phase angle (Zappala et al. 1990) or its surface properties and topography (Gutiérrez et al. 2006). The dynamic range of  $A_{\text{long}}$  will depend on the class of asteroid and how much their viewing geometry evolves in time. For asteroids closer to the Earth, where their viewing geometry is more drastically and rapidly evolving in a given apparition, the dynamic range is expected to be larger compared to more distant objects like Kuiper Belt objects, which have more consistency in their apparition.

For science cases other than the case considered here, the choice of hyperparameters can be tuned to fit the physical knowledge of other light curves exhibiting periodicity. The priors with their hyperparameters are summarized in Table 1 and a visualization can be found online in the [AsteroGaP documentation](#).

#### 2.4. Assumptions

The model we describe above implicitly and explicitly makes a number of assumptions about the data we observe and the inference problem of asteroid periods. Because these assumptions are important to the validity of the model, we list them here. For cases where any of these assumptions are broken, a different model, or the addition of other model components, may be required.

1. We assume long-term trends due to the changing distance to the asteroid between the asteroid and the Sun have been removed. Our model does account for gradual changes in the rotational profile from the changes in phase-angle, but the constant mean function  $c$  implies that any major changes in the mean magnitude for our asteroids have been normalized out, e.g. by using a software that simulates the asteroid’s orbit and distance from the Sun like OpenOrb (Granvik et al. 2009).
2. We assume that our asteroid light curve data is strictly a periodic process, meaning we cannot account for tumbling asteroids which exhibit rotation

along a non-principal axis, or asteroid observations taken over many years, where the period might change.

3. The Gaussian Process, by its nature, requires measurement uncertainties to be Gaussian, which should generally be true for optical fluxes and magnitudes from astronomical surveys, but might break down for other types of data.

### 3. SIMULATED DATA

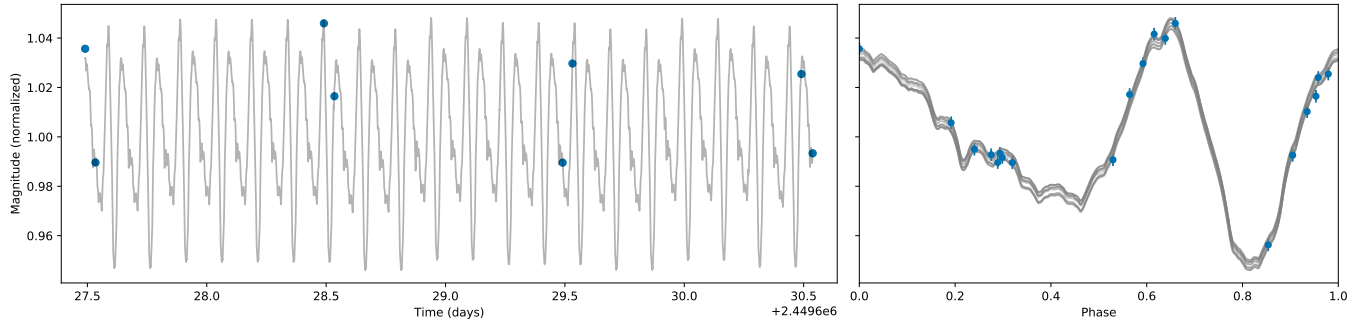
We evaluate the effectiveness of our Gaussian Process model on simulated light curves of asteroids, generated from asteroid shape models with established rotational period. With these simulations, we aim to show that the Gaussian Process can recover rotational periods in realistic simulations of sparsely sampled asteroid time series. We compare the Gaussian Process model with the more conventional Lomb-Scargle Periodogram (LSP) and test how effective each of the two methods are at rediscovering the true period in these simulations.

#### 3.1. Simulation Set-Up

We use three-dimensional asteroid shape models available in the DAMIT database<sup>3</sup> (Durech et al. 2010) to generate realistic, dense time series data spanning 5.5 months (165 days ranging from MJD 49627 to 49787) sampled every 30 seconds for two short-period asteroids, 221 Eos and 3200 Phaethon. Both asteroids have been densely observed multiple times since their discovery and have well-established rotation periods and physical models (Warner et al. 2009a).

We sub-sample our dense simulations, creating a cadence pattern similar to ZTF and proposed LSST cadences (Bellm et al. 2019b; Jones et al. 2018).

<sup>3</sup> see <http://astro.troja.mff.cuni.cz/projects/damit>



**Figure 3.** Asteroid 3200 Phaethon light curve with simulated observations over 10 days total. Grey lines denote the densely simulated light curve of 3200 Phaethon from the shape model available in the DAMIT database, and the blue markers are the simulated photometric observations obtained from the light curve. Left: 8 observations over the first 4 days of observations, shortened from 10 days for viewing clarity. Right: the light curve for all 10 days of observations, folded at the asteroid’s estimated rotational period.

A ZTF and LSST-like cadence consists of 2 observations per night spaced 30-60 minutes apart, every few nights. We simulate two observations every night for  $L$  days, such that the number of data points in each light curve  $N = 2L$ . Intervals between each set of nightly observations were randomly determined by sampling a normal distribution with a mean of 60 minutes and a standard deviation of 5 minutes. We chose  $L$  in a range from 10 to 60 days to test how well our model and the LSP will perform with different numbers of data points.

We initialize the MCMC sampler with 100 chains for a burn-in of 1,000 iterations, followed by an additional 10,000 iterations to sample. In individual cases, we increase the number of burn-in iterations when the distribution appears to not have converged. For comparison, we use a two-term multi-band periodic LSP as implemented in the *Python* package *gatspy* (VanderPlas & Ivezić 2015). For the short-period asteroids simulated here, we set the period range for the search to span 1 to 60 hours and we set our first pass coverage to 200.

Preliminary experiments reveal posterior distributions that are multimodal in the period parameter, suggesting that summaries of the distribution like the mean and standard deviation fail to capture a complete picture of the posterior inference. In order to identify and characterize modes in the marginalized posterior probability distribution, we initially generate a coarse 20-bin histogram across the full span of sampled parameters and identify the bins with the highest probability density as potential modes. We produce high-resolution histograms for each of the identified modes, and identify the highest peak in each as the putative mode. We calculate the width to half of the max of the tallest bin in our localized histograms on each side of the mode since we cannot assume our posterior modes are symmetric, and then determine the range that spans five half-widths to half maximum on either side, which preliminary tests

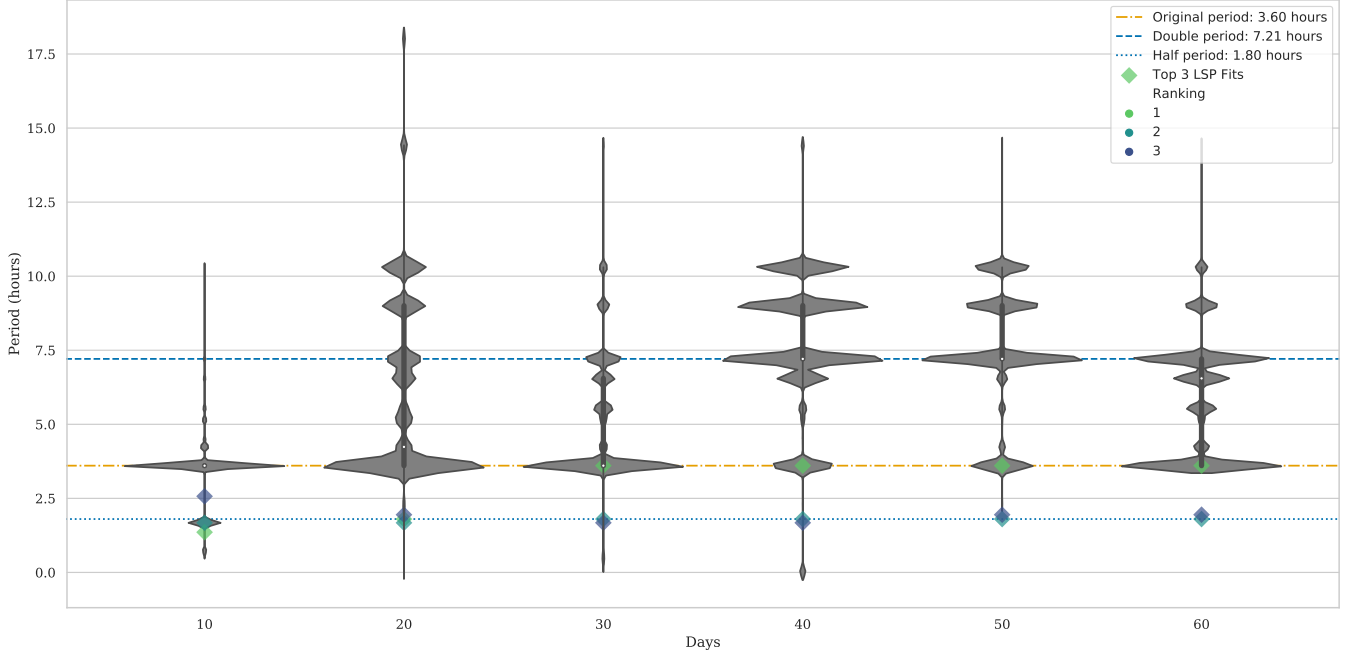
suggest is wide enough to encompass the vast majority of samples in each mode. Once this range is determined, we create a final histogram for both the entire distribution as well as each mode (as seen in Figure 8) and select the center of the tallest bin as the maximum of that mode. The total probability of a mode is determined by integrating the marginalized posterior probability density within the five standard deviation range.

### 3.2. 3200 Phaethon

We simulate light curve data of Phaethon 3200, an Apollo-type near-Earth asteroid discovered in 1983 (Green & Kowal 1983). Phaethon has an observed rotational period of 3.604 hours (Hanuš et al. 2018a) and a non-sinusoidal rotational profile (Figure 3, right panel). We sub-sample the dense light curve generated from the 3200 Phaethon model in the DAMIT database with the ZTF-LSST cadence to generate individual realizations that vary in length from 10 to 60 days. With 2 observations per night, the generated light curves have between 20 to 120 data points.

We model each light curve with the Gaussian Process as described in Section 3.1 in order to infer the period and other parameters. The posterior probability distributions for a majority of the light curves are multimodal (Figure 4), with modes at multiples of the true period, most notably twice the true period (7.208 hours) and 2.5 times the true period (9.010 hours). A significant fraction of the probability from the marginalized posterior distributions of the period for light curves with 10, 20, and 30 days of observations is located in a narrow mode around the known period (see Section 3.1), 73.3%, 46.2%, and 57.8% for 10, 20, and 30 days, respectively, while the light curves of 40 and 50 day durations have the majority of their probability mass located at double the period, 32.5% and 43.7%, respectively. The light curve with 60 days of observations has its probability primarily split





**Figure 4.** Violin plot (grey) showing posterior distributions of the different Gaussian Process results for Asteroid 3200 Phaethon, spanning 20, 60, 80, 100, and 120 data points (or 10, 30, 40, 50, and 60 days). Diamonds in green, teal and blue represent the top three best period estimates of Lomb-Scargle periodograms, respectively. Lines mark the period that was used to simulate the data (yellow, dot-dashed), double that period (blue, dashed) and half that period (blue, dotted).

between the true period and double: 39.2% and 30.4%, respectively.

We generate LSPs for all simulated light curves and identify the three highest peaks in each (Figure 4, diamonds). For light curves with 30 or more days of observations, the LSP estimates the true period and half the true period as the highest and second-highest peaks, respectively. However with 10 or 20 days of observations, the LSP fails to reliably identify the true period. The highest periodogram peaks for the 10-day light curve differ by more than 7 minutes from the true period. This is enough of a difference for a short-period asteroid like 3200 Phaethon with a period of a few hours to produce a folded light curve that appears incoherent. For the 20-day light curve, the highest peak suggests a period that is half of the true period.

For a better comparison between the LSP and the Gaussian Process posteriors, we also present the LSP for a simulated light curve over 60 days of observations in Figure 5. The periodogram characterizes the best fit period within 2 seconds of the reported period (Hanuš et al. 2018a), but also features several nearby peaks of similar heights.

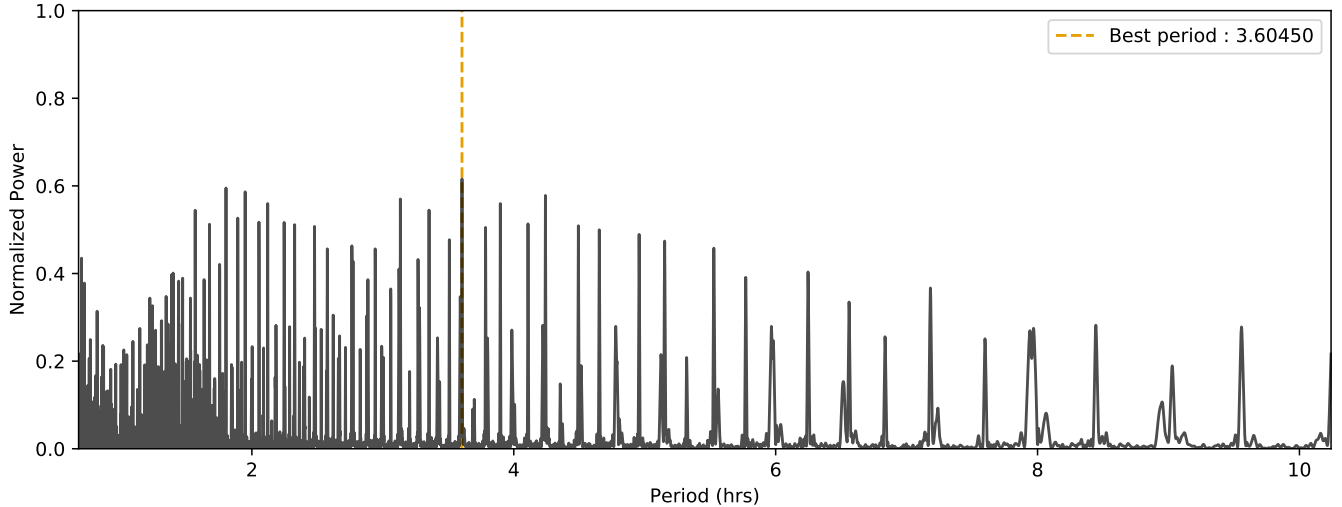
### 3.3. 221 Eos

We also test our model on 221 Eos, a well-documented main-belt asteroid with an average rotation period of

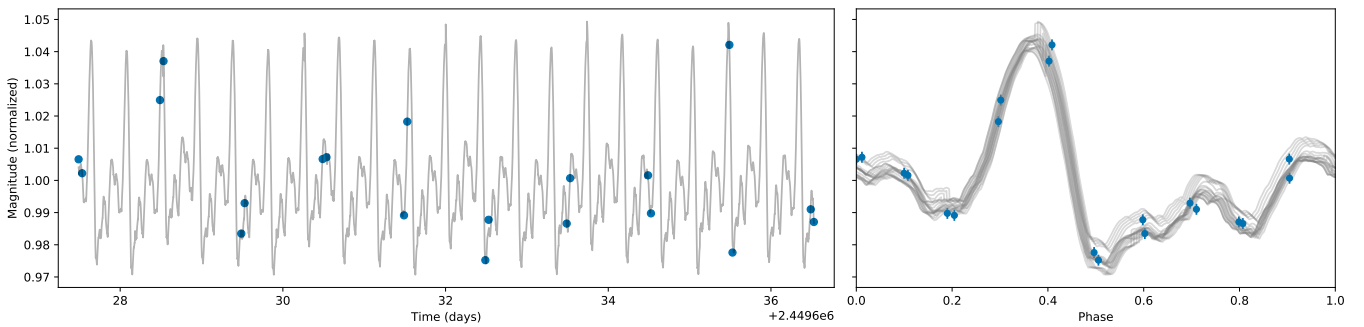
10.443 hours (Pilcher 2013). Its light curve is highly structured, with a rotational profile departing significantly from a sinusoidal shape, as shown in Figure 6. Just as for 3200 Phaethon, we sub-sample the dense light curve generated from the 221 Eos model in the DAMIT database with the ZTF-LSST cadence for 10 to 60 days.

We model each light curve with the Gaussian Process, as described in Section 3.1. As for 3200 Phaethon, the resulting posterior probability distribution for each light curve is multimodal, with most modes at multiples of the true period. The Gaussian Process is highly effective in inferring the true period across all realizations and light curve durations. A large fraction of the probability from the marginalized posterior distributions of the period for each light curve is located in a narrow mode around the true period: 74.4%, 74.8%, 76.0%, 64.4%, 84.4% and 80.5% of samples for 10, 20, 30, 40, 50, and 60 days, respectively. There is also a smaller amount of probability located at a third of true period (3.47 hours) for each of the light curves, possibly a reflection of the more complex substructure of this asteroid’s rotational profile.

Lomb-Scargle periodograms show a similar bias towards shorter periods in a more pronounced way: in many of the light curves, the highest peaks cluster around a third of the known period or less. For the light curve with only 10 days of observations and 20 data points,



**Figure 5.** LSP for asteroid 3200 Phaethon for 60 days of observations ( $N = 120$  observations). The dashed yellow line indicates the highest peak in the periodogram, similar to the period found in Hanuš et al. (2018a).



**Figure 6.** Simulated light curve of asteroid Eos 221 over 10 days. Blue markers are 20 simulated observations and the grey line is the original simulated light curve obtained from the DAMIT database.

none of the peaks were close to any obvious multiples of the true period while nearly 75% of samples from the Gaussian Process model are clustered around the known period. Only for the light curves with 20 and 30 data points does the LSP show the known period as the highest peak.

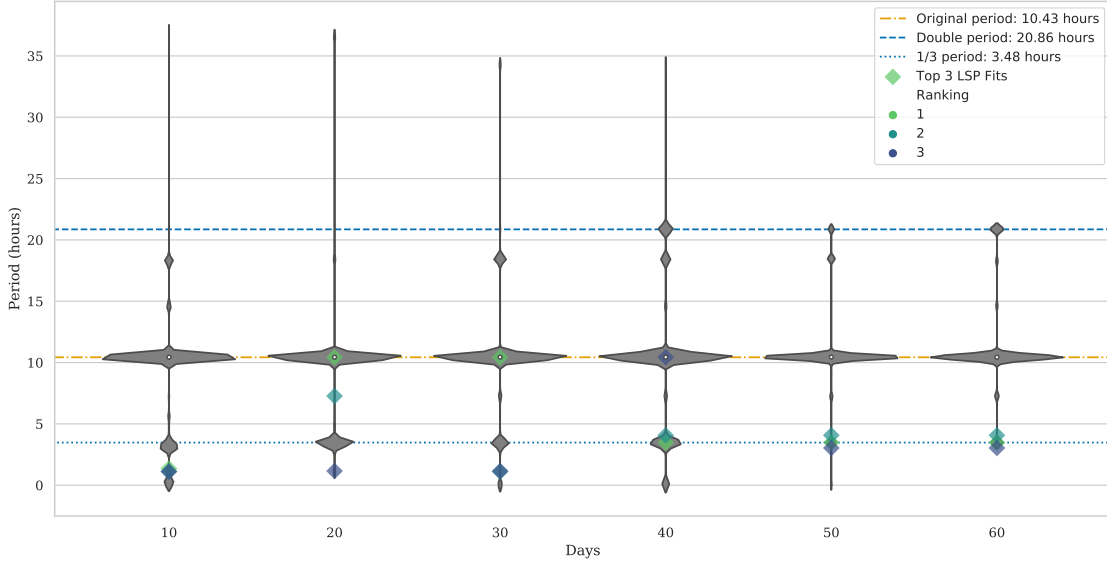
#### 4. ZTF

In order to understand the model’s performance on real data, we test our model on observations from the Zwicky Transient Facility. We first look at a short-period main-belt asteroid to provide a direct comparison to our simulation results, then we focus on long-period asteroids for which period estimation is traditionally difficult.

##### 4.1. Data Reduction

The ZTF alert stream provides aperture-corrected photometry for each object detected in difference images; detections near the predicted positions of known asteroids are tagged with this identification. We retrieved ZTF

detections of a variety of asteroids from the public and private alert stream (Prša et al. 2019), removing any observations that fell more than 10 arcseconds away from the predicted positions of these (well-known) orbits. We also retrieved orbital information for each asteroid from the JPL Small Body Database (<https://ssd.jpl.nasa.gov/sbdb.cgi>) using Astroquery (Ginsburg et al. 2019). We used the orbital information to generate expected magnitudes at each epoch, using OpenOrb (Granvik et al. 2009), thus providing a correction for distance and phase angle effects using a fairly simple  $H - G$  phase curve (Bowell et al. 1989) and assuming  $G = 0.15$ . After subtracting these corrections from the observed magnitudes, we calculate the RMS of the resulting values to look for outliers more than  $3\sigma$  away from the mean in each bandpass – these outliers are typically the result of measurements compromised by a background star or edge effect and are rejected. We then calculate the mean corrected-photometric values per bandpass to get an initial estimate of the color of the



**Figure 7.** Violin plot (grey) showing posterior distributions of the different Gaussian Process results for Asteroid 221 Eos, spanning 10, 20, 30, 40, 50, and 60 days (20, 40, 60, 80, 100, and 120 data points). Colored diamonds in green, teal and blue represent the top three best period estimates in the Lomb-Scargle periodograms. The lines show the period that was used to simulate the data (yellow, dot-dashed), double (blue, dashed) that period, and 1/3 of the true period (blue, dotted). Our posterior probability distributions for the light curves with 30 and 40 days of observations contain a small number of samples more than five standard deviations (averaged across all samples) above the mean, which, for the sake of this visualization, have been masked.

asteroid; this color is also subtracted from the relevant photometric points. The resulting set of photometry for each asteroid is well corrected for distance effects, and approximately for phase curve effects and color. These ‘corrected’ magnitudes center around zero in each and primarily reflect the effects of asteroid variability due to rotation.

Preliminary tests with the ZTF data show that the Gaussian Processes model breaks down in some cases, modeling photometric differences between bandpasses rather than finding the rotational period. This is likely driven by two problems which lead to an incorrect initial color estimate: inadequate modeling of the phase curve and uneven sampling of the rotational period in each bandpass. For some asteroids, using a single  $G$  value with the  $H - G$  phase curve correction is not sufficient, and offsets in the predicted magnitudes in different bandpasses at large phases lead to offsets in the estimated color. A more sophisticated phase curve model (and/or fitting  $G$  values in each bandpass) could improve this. For other asteroids, the sparse sampling in each bandpass may fall more frequently into faint (or bright) points in the rotation period, again biasing the initial color estimate.

In this initial study, where we primarily aim to show the use of this model on sparse observations, we proceed

to model the data from  $r$ -band only. Future work will extend the model to allow observations from multiple bandpasses to be used simultaneously.

We initialize the MCMC samplers with the same setup as our simulated asteroids (see Section 3.1 for details) and determine the probabilities and peak periods for each mode in the posterior.

#### 4.2. Asteroid 1388 Aphrodite

As a direct comparison to the asteroid light curves simulated in Section 3, we investigate a short-period main-belt asteroid from the Eos family. Asteroid 1388 was observed 121 times in the  $r$ -band over the span of 139 days and has a reported period of 11.94389 hours (Hanuš et al. 2018b).

We present our posterior inferences in Figure 8 (summaries of the marginalized posterior of the period parameter, and folded light curves), Figure 9 (a more complete representation of the folded light curve with models), and Figure 10 (a corner plot summarizing posterior inferences for all parameters). 84.9% of the probability in the posterior distribution of the period are concentrated in a narrow mode centred on a period of 11.94596 hours (Figure 8). There is a smaller mode centred on the double of the period (23.89141 hours), although it only accounts

for 4.1% of the total posterior probability. Around 6.2% of the probability mass is at 7.96627 hours, i.e. at 2/3 of the narrow posterior peak with the highest overall probability. Overall, over 95% of the probability reflects an estimate close to or at a multiple of the previous reported period.

We plot our original observations folded at the periods reflecting the peak of each of the two most probable modes (11.95 and 7.97 hours) in the lower row of Figure 8. To visualize the Gaussian Process, we use parameter values from samples from the same mode in the marginalized period posterior distribution to construct realizations of the Gaussian Process for an interval of about twenty periods (240 and 160 hours, respectively) from the start of the observing window. We fold these model lightcurves at the same respective periods as mentioned above to illustrate the behaviour of the model. For the data folded at 11.95 hours, the light curve resembles a double-peaked sinusoid, which is common for asteroids shaped like tri-axial ellipsoids (Harris et al. 2014).

The light curve folded at 7.97 hours does appear to have a visible pattern (Figure 8, bottom right panel), but the models for the first 160 hours reveal that the Gaussian Process appears to be mainly modelling a group of brighter observations from the first peak of the double-peaked sinusoid while simultaneously modelling some of the dimmer observations by varying the light curve slowly over time. Because we have opted to only plot the model light curves for the first 160 hours for clarity, we present in Figure 9 the same data and model, but spanning the full observational window rather than only the first 160 hours. The lower panel of this plot reveals a significant evolution in the intra-period variability over the 139 days of observations, where the peak shifts by nearly one half of a rotational phase in order to fit all data points. We discuss this behaviour further in Section 5.

We present a corner plot of the posterior distribution in Figure 10. As expected, the posterior probability of the constant mean flux  $c$  is consistent with zero,  $0.02 \pm 0.1$ , indicating that the subtraction of the phase angle during pre-processing largely succeeded. The two-dimensional scatter plots reveal some scatter in the distribution, and a slight bimodality for the amplitude  $A_{\text{long}}$  and the metric parameter  $M$  of the squared exponential covariance function, as well as the inverse length scale of the periodic kernel  $\Gamma$ .

We observe a negative correlation in parameter space between the log-amplitude  $A_{\text{long}}$  for the long-term squared exponential covariance function and the log-amplitude for our periodic sine squared exponential covariance function,  $A_{\text{periodic}}$ . This trend is likely a

result of both parameters contributing to the overall amplitude of the combined covariance functions: if one amplitude has a high value, the other amplitude will need to be at a lower value to produce the overall observed variability.

There is a slight positive correlation between the inverse length scale  $\Gamma$  of the periodic covariance function, which models the intra-period variability, and the period,  $P$ . This is expected: the longer the proposed period, the more complex the variations that the Gaussian Process has to model within each period. For example, a double sinusoid exhibits more complex intra-period variability when compared to a single sinusoid, so a longer period would correlate with a greater inverse length scale  $\Gamma$  value.

#### 4.3. Asteroid 299 Thora

Because our prior favours short periods, we want to explore the models performance on a long-period asteroid. Asteroid 299 Thora is a long-period main-belt asteroid with a reported period of 273 hours (Pilcher et al. 2017). It was observed 49 times by ZTF in the  $r$ -band over 153 days.

The posterior probability distribution (Figure 11) of the period indicates that over 87% of the probability is located in a mode around a peak at 269.79 hours, though the posterior is asymmetric. The posterior has a second mode at 132.57 hours (approximately half of our first period), containing 7.4% of the probability.

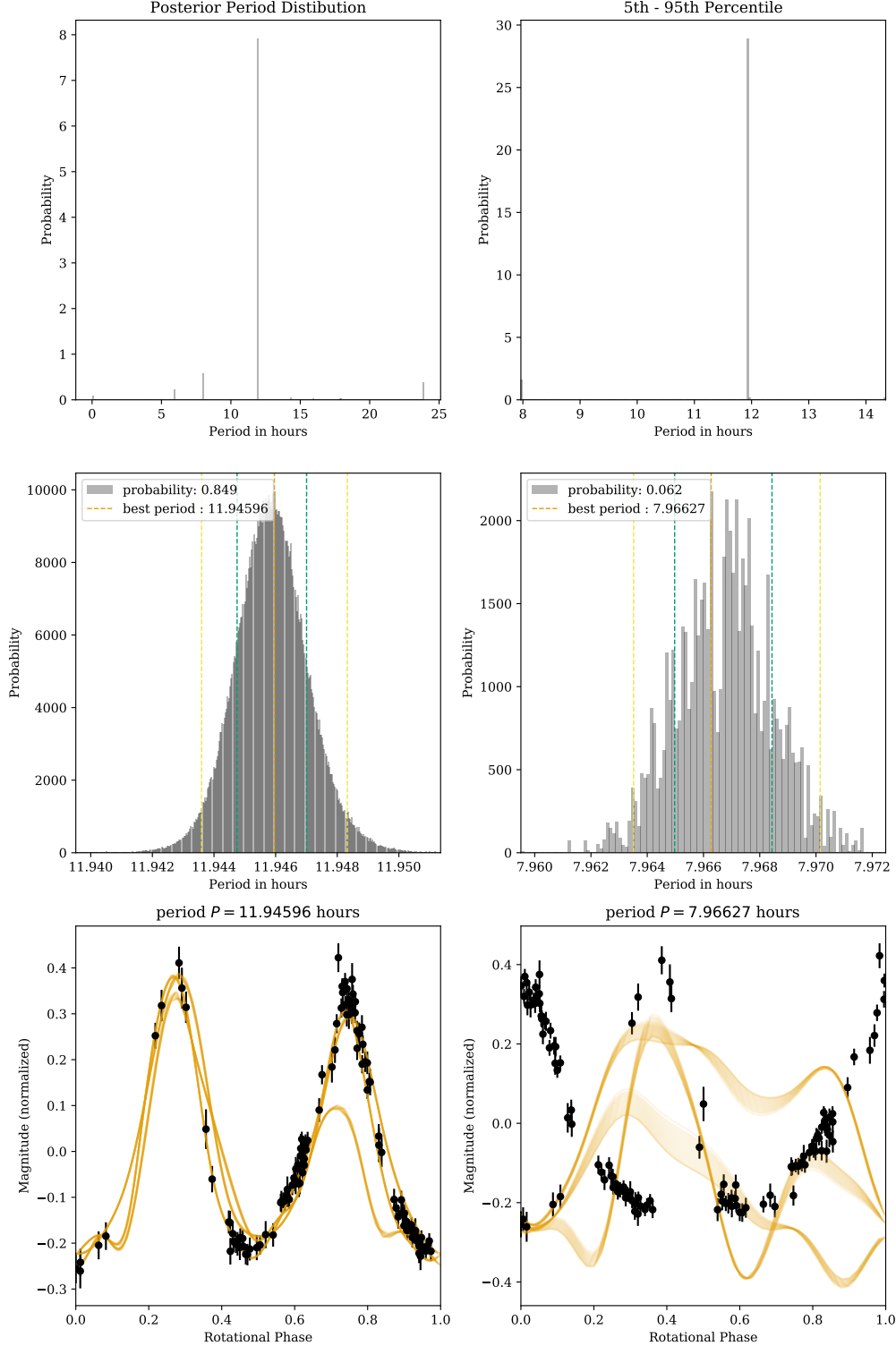
The corner plot (Figure 12) once again shows a large negative correlation between the two amplitudes,  $A_{\text{long}}$  and  $A_{\text{periodic}}$ , and a slight positive correlation between the log metric  $M$  and inverse length scale  $\Gamma$  parameters. All other parameters are well-constrained in a single mode.

#### 4.4. Asteroid 821 Fanny

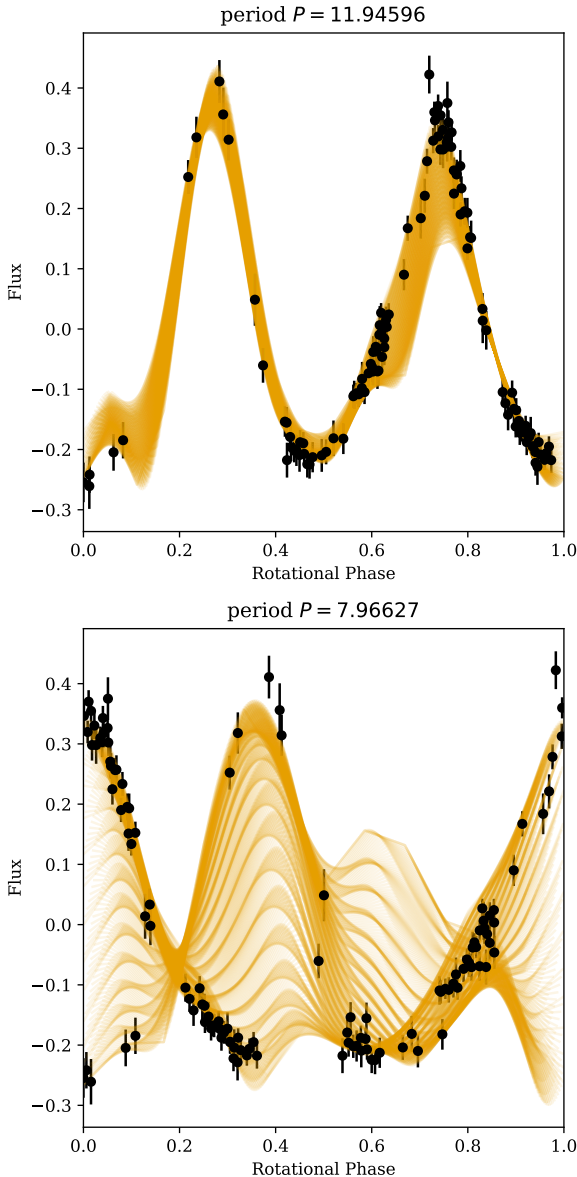
We include the asteroid 821 Fanny as an illustration of the limits of the model. Asteroid 821 is a slowly-rotating main-belt asteroid with a reported period varying from  $230.6 \pm 0.3$  hours to  $236.6 \pm 0.3$  hours in one recent analysis (Pilcher 2018) and  $238.9 \pm 0.8$  hours in another (Polakis 2018). Asteroid 821 was observed by ZTF 46 times in the  $r$ -band over 153 days. We initially sampled the model for 11,000 iterations, but convergence criteria indicated that the chains had not converged, and thus continued to sample for another 10,000 iterations.

The period parameter spans the entire allowed prior range (Figure 13) as described in Section 2.3, suggesting that either the sampler has still not converged, or that the period is entirely unconstrained given the data. The inverse length scale  $\Gamma$  also spans a very large range (values





**Figure 8.** Top left: Posterior distribution of the period for Asteroid 1388 Aphrodite. Top right: 5th-95th percentile of the posterior distribution. Note that in the top two profiles, the posterior appears as thin, sparse peaks as a result of its multimodal topology. Middle: Modes with the highest and second highest probability mass (84.9% and 6.2% of samples, respectively), with the yellow line denoting the highest bin of the histogram, from which the period for the mode is calculated (11.94594 and 7.96627 hours). Dashed lines denote the 95th (light yellow) and 68th (green) percentile ranges. Bottom: ZTF observations (black) folded at the respective periods determined from the middle plots. Yellow lines are three realizations of Gaussian Processes with periods similar (within  $\pm 0.5$  hours) to the folding periods. These models span from the start of the observational window, to 20 times their respective folding periods, allowing us to see how the rotational profile varies over time.



**Figure 9.** ZTF observations of asteroid 1388 Aphrodite folded at 11.946 hours (top) and 7.966 hours (bottom). The yellow lines are folded realizations of a Gaussian Process model conditioned on the data, and using parameters drawn from the mode at  $11.946 \pm 0.5$  hours (top), and from the mode at  $7.966 \pm 0.5$  hours. As opposed to the bottom panels in Figure 8, these models span the entire observational window of the asteroid. Covering the full observational range reveals how the rotational profile varies over time. This variation is particularly apparent in the bottom panel, where the model exploits a degeneracy between the rotational period and the metric  $M$  to generate a light curve that matches the observations despite a shorter period  $P$  by enforcing large changes in the rotational profile over the 139 days of observations.

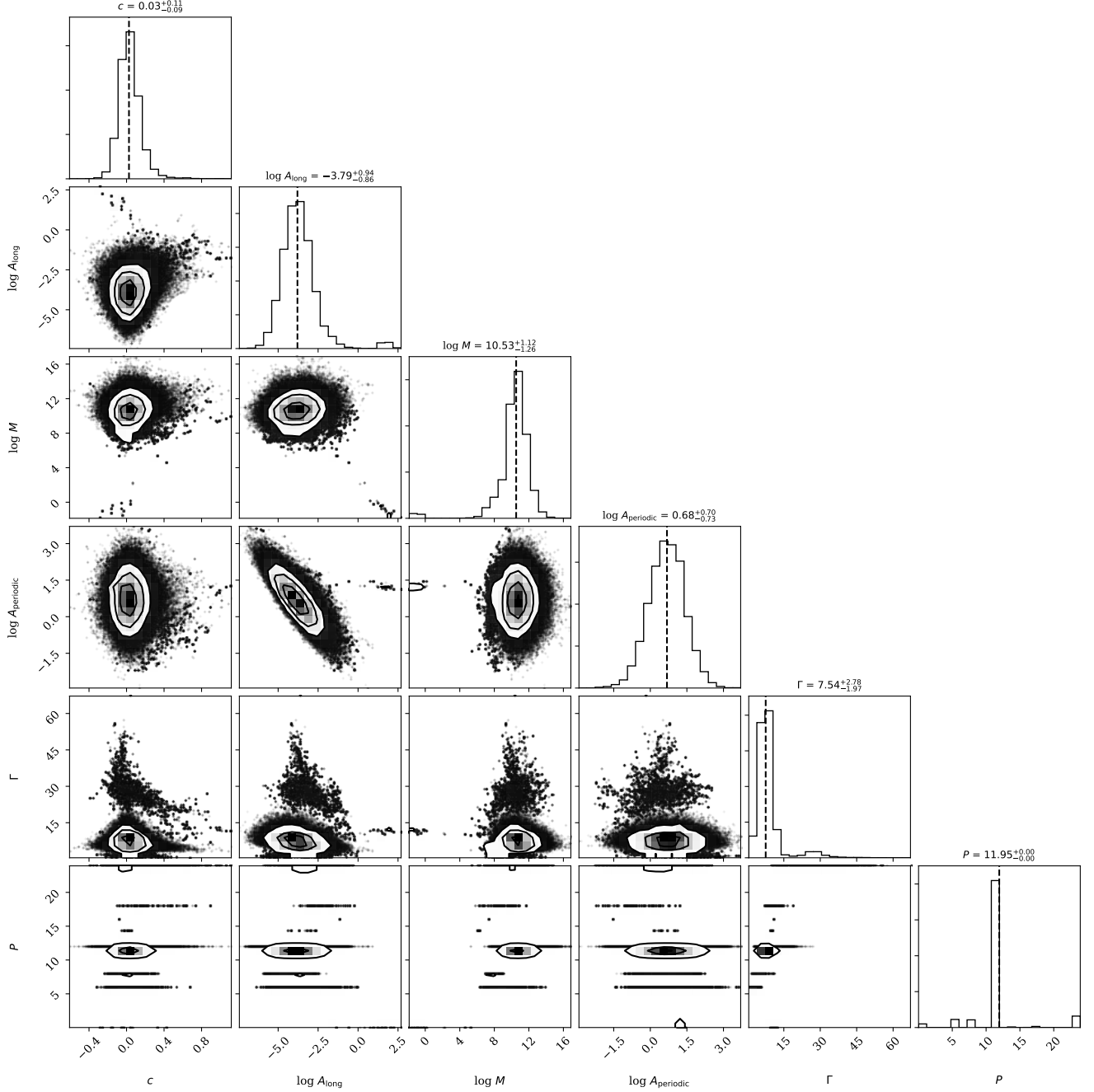
$\sim 100 - 1000$ ) when compared to better-constrained asteroids discussed earlier. This is expected, given that, if the model suggested a period of several months (much longer than our total observing time for this asteroid), then the inverse length scale  $\Gamma$  has to encapsulate all the variations in the flux, or the intra-period variability, over the span of that period.

Folding the data to 236.6 hours, one of the periods reported by Pilcher (2018), yields a noisy light curve (Figure 14) without much visible structure, but highlights the high magnitude uncertainty. In cases like this, where the amplitude due to rotation is smaller than the uncertainties on the data points and the data is sparse, any period yields an acceptable solution.

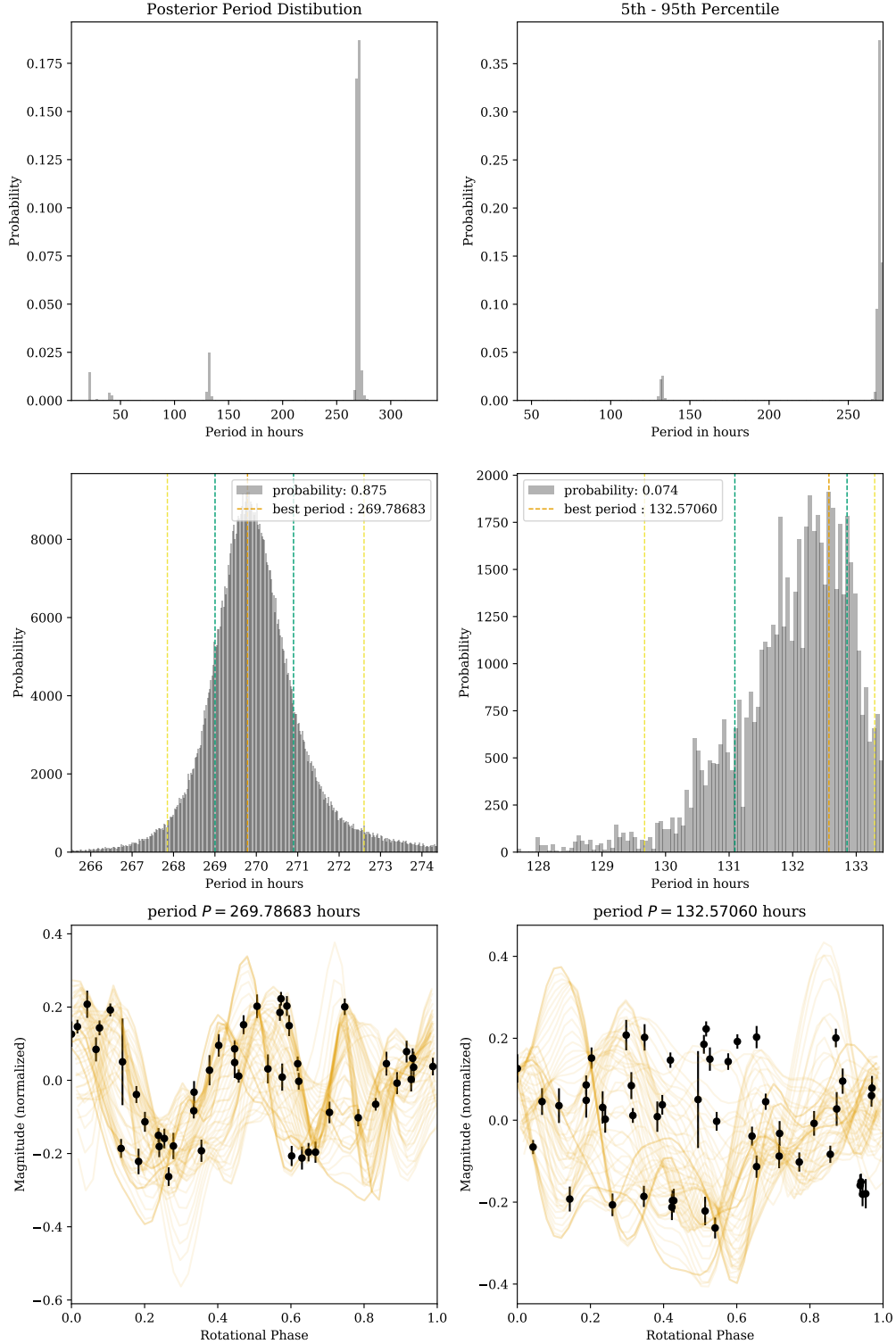
The posterior distribution (Figure 15) for the period parameter has some small peaks, but looks extremely broad overall, as expected from the corner plot (Figure 13). That the posterior probability peaks at the lower edge of the prior could in principle indicate that the solution that best represents the data sits beyond the 1-minute cutoff specified in the prior. However, smaller periods are believed to be unphysical, since they would yield an asteroid that spins itself apart (Jewitt et al. 2017; Warner et al. 2009b; Chang et al. 2015; Moreno et al. 2017). The minor peak that does exist in the probability distribution is located around 74.18 hours. We fold the observations and generated models for 74.18 hours and find model light curves with very small amplitudes, broadly consistent with a constant value. It is also possible that our model did not converge on a clear period if this asteroid is in a non-principal axis rotation state.

## 5. DISCUSSION

The standard model for an asteroid—a tri-axial ellipsoid—produces a double-peaked light curve where each peak is approximately sinusoidal. In practice, however, asteroids display a wide range of rotational profiles that deviate from that simple model. Irregular rotational profiles are observed when asteroids have either patches of surface materials with different albedos, abnormal shapes, or are relatively spherical but have surface variations that affect reflectance. The resulting light curves can be complex and generally exhibit multiple peaks as well as other substructure. In these cases, the LSP requires a large ensemble of sinusoidal functions to accurately represent the light curve, leading to power spread over many frequencies. This in turn complicates the inference of the underlying period. The behaviour is evident from the two simulated asteroids, particularly 221 Eos (Figure 7), where much of the power in the LSP is concentrated

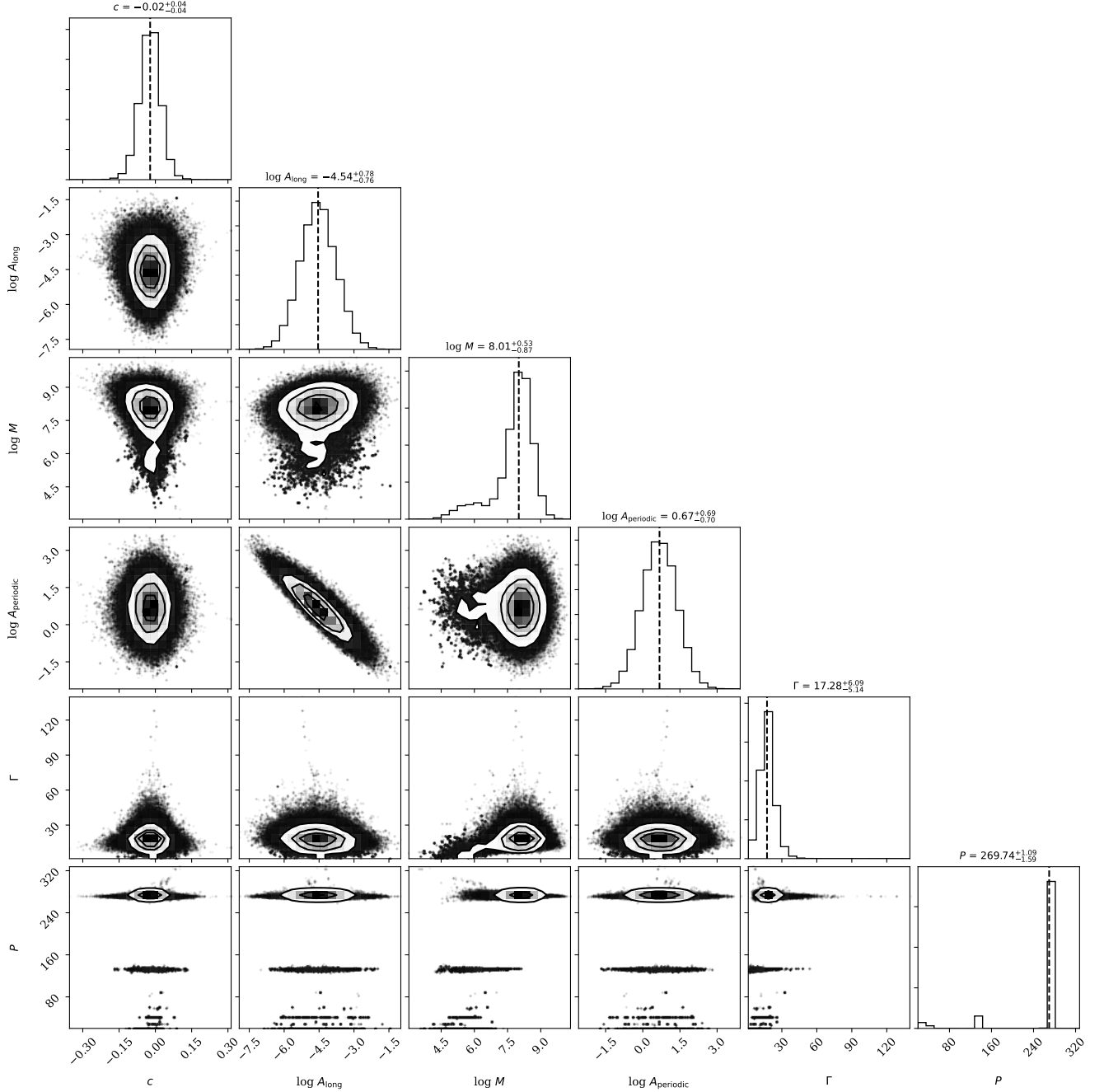


**Figure 10.** This figure, also called a *corner plot*, presents a visualization of the posterior probability distribution of model parameters for Asteroid 1388 Aphrodite. Because the posterior pdf is a six-dimensional distribution, this visualization aims to make the information contained in the posterior accessible by unfolding it into a series of 1D and 2D representations. On the diagonal, we present histograms of the marginalized posterior probabilities for each parameter. On the off-diagonal, we show scatterplots for all pairwise combinations of parameters. The median of each distribution is marked on the histograms and included in the title of each column, along with the 16th and 84th percentile differences.  $A_{\text{long}}$ ,  $M$ , and  $A_{\text{periodic}}$  are plotted on a logarithmic scale.



**Figure 11.** Top left: Posterior distribution of the period for Asteroid 299 Thora. Top right: 5th-95th percentile of the posterior distribution. Middle: Modes with the highest and second highest probability mass (87.5% and 7.4% of samples, respectively), with the yellow line denoting the highest bin of the histogram, from which the period for the mode is calculated (269.78683 and 132.57060 hours). Dashed lines denote the 95th (light yellow) and 68th (green) percentile ranges. Bottom: ZTF observations (black) folded at the respective periods determined from the middle plots. Yellow lines are three realizations of Gaussian Processes with periods similar (within  $\pm 0.5$  hours) to the folding periods. These models span from the start of the observational window, to 20 times their respective folding periods, allowing us to see how the rotational profile varies over time.



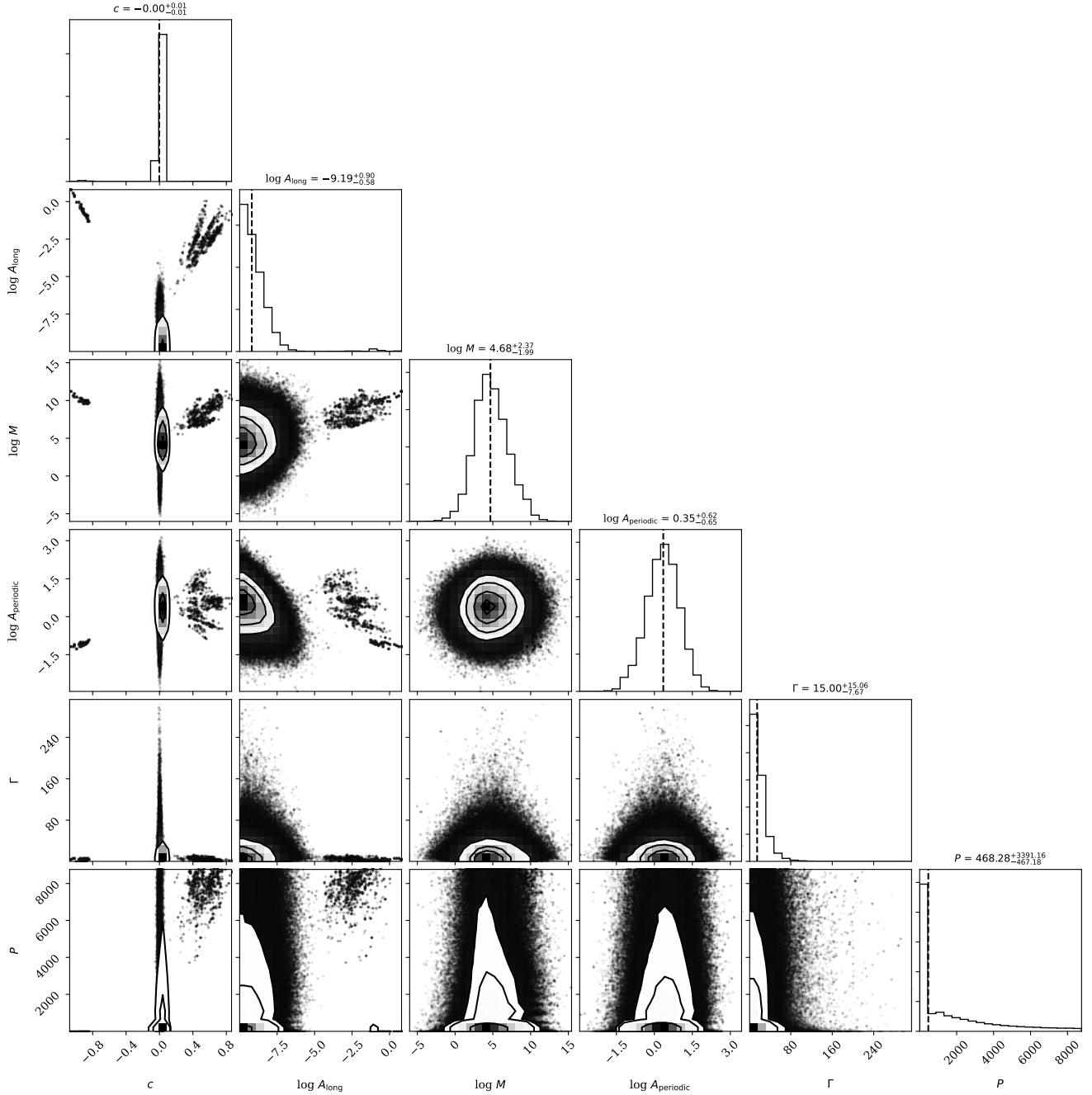


**Figure 12.** Corner plot of model parameters for Asteroid 299 Thora. The 50th percentile of each distribution is marked on the histograms and written in the title above, along with the 16th and 84th percentile differences.  $A_{\text{long}}$ ,  $M$ , and  $A_{\text{periodic}}$  are plotted on a logarithmic scale.

near 1/3 of the true rotational period. This can be explained by 221 Eos’ triple-peaked structure (Figure 6).

In contrast, the flexibility of the Gaussian Process model implemented admits a wide range of realistic light curve shapes observed in nature. The sine squared exponential covariance function only assumes an exponential decay of the covariance of data points within a period  $P$ , meaning the Gaussian Process has the flexibility to

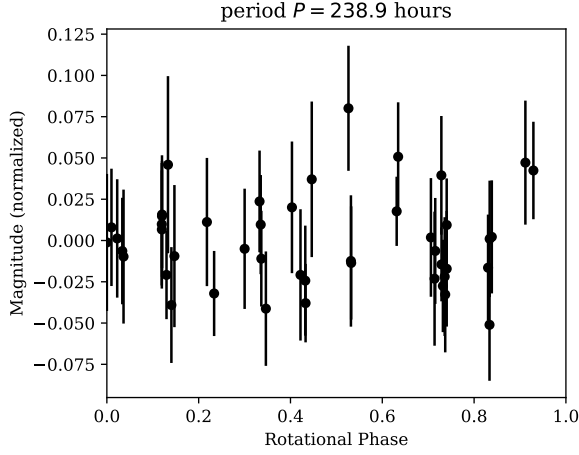
model a wide range of realistic rotational profiles as long as they remain periodic (and do not drastically change over a short time; Figure 9 shows how rotational profiles could potentially morph over time). In addition, we include another covariance term to account for long-term changes in the rotational profile as a function of time. The squared exponential covariance function enables the model to account for the natural evolution of



**Figure 13.** Corner plot of model parameters for Asteroid 821 Fanny. The 50th percentile of each distribution is marked on the histograms and written in the title above, along with the 16th and 84th percentile differences.  $A_{\text{long}}$ ,  $M$ , and  $A_{\text{periodic}}$  are plotted on a logarithmic scale.

a rotational profile using the log-metric parameter  $M$ , as the phase-angle gradually changes throughout an asteroid's orbit. We expect asteroid lightcurves to evolve over longer periods of time as the phase-angle between us, the asteroid, and the sun changes. This is especially the case for sparse asteroid light curves, where data collection routinely spans months to years, and occasionally decades.

These additional degrees of freedom allow the model to better characterize sparse data obtained over a large range of time, along with differing rates of phase-angle change. In the future, we will implement other physical processes—the phase angle via a mean function, tumbling asteroids via additional terms in the covariance function—to allow for realistic modeling of a wider range of asteroids.

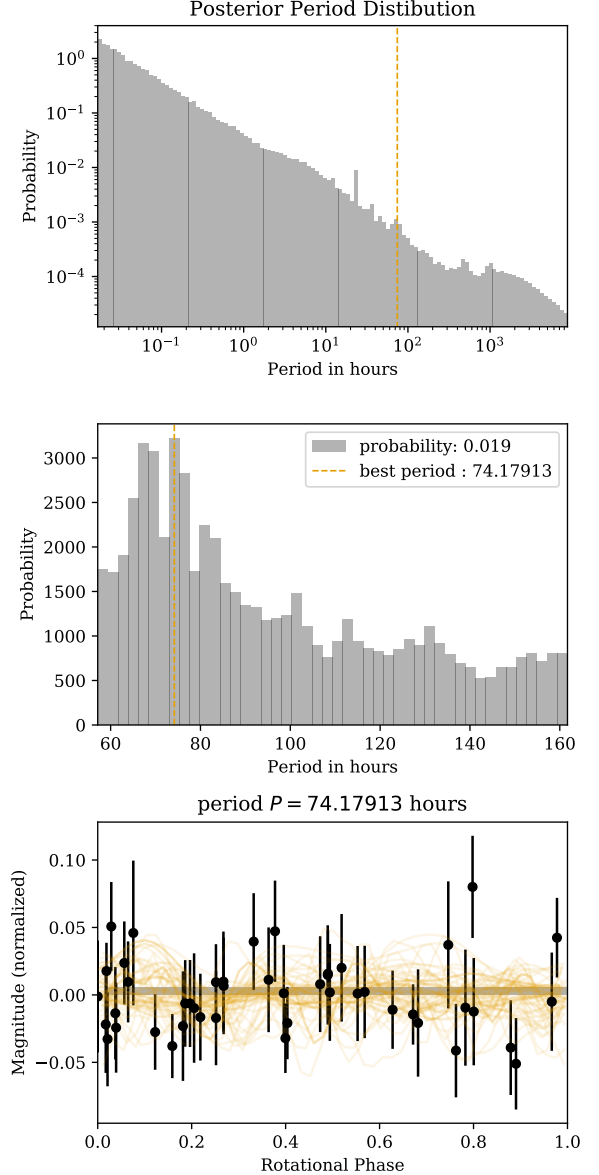


**Figure 14.** ZTF observations of Asteroid 821 Fanny folded at one of the reported periods, 238.9 hours.

The model implemented is a Bayesian model: this allows straightforward inference of relevant model parameters through the posterior distribution, as well as incorporation of known asteroid properties through the prior probability distributions. The model allows for principled *population inference* through Bayesian hierarchical modeling. While in this paper, we focused on the inference of properties of individual asteroids, hierarchical models allow for the joint inference of both individual properties as well as population parameters (e.g. the distribution of periods, amplitudes—as a proxy for shape—and intra-period variability) for physically related groups of asteroids. For example, the empirical prior for the period introduced in Section 2.3 could then be replaced by a physically motivated distribution, and the population-level parameters of this distribution be inferred.

The model performs well at characterizing simulated light curves of asteroids, especially for sparse observations with a small number of data points, as long as the intrinsic variance in the magnitude due to rotation exceeds the measurement uncertainties. The posterior probability for both simulated and observed light curves is multimodal, but concentrated in a small number of narrow peaks, compared to the wider spread in power across frequencies in the Lomb-Scargle periodogram (Figure 5). The majority of the probability density is concentrated for each simulated light curve at either the true period or at twice that value.

The posterior probability for the 3200 Phaethon simulation with the fewest observations (10 nights and 20 data points, see Figure 4) has its highest mode at the true period (73%). This simulation is also the only 3200 Phaethon light curve for which the LSP fails to find



**Figure 15.** Top: Posterior distribution of the period for Asteroid 821 Fanny, with logarithmic scales on both axes with a yellow line denoting 74.18 hours. Middle: Mode with the highest probability mass (1.9% of samples) with a yellow line denoting the highest bin of the histogram, from which the period for the mode is calculated (74.18 hours). Bottom: ZTF observations (black) folded at the period determined from the middle plot. Yellow lines are three realizations of the Gaussian Process with periods similar (within  $\pm 0.5$  hours) to the folding period. These models span from the start of the observational window, to 20 times the folding period (1500 hours).

either the true period or any of its multiples within its three highest peaks.

Overall, the model performs well when data sets span multiple rotational periods, and the magnitude uncertainty is relatively small. Failure to meet these criteria results in incorrect parameter inference (see also Section 5.1). We report to a high degree of confidence that the rotational period for Asteroid 1388 Aphrodite is 11.94597 hours with a 68% credible interval ranging from 11.94475 to 11.94700 hours. The previously observed rotational period of 11.94389 (Hanuš et al. 2018b) falls within a 95% credible interval. Hanuš et al. (2018b) uses 16 dense light curves along with an additional 220 sparse individual measurements to calculate their period. We use 121 observations spaced over 139 days to arrive at a compatible result.

Asteroid 1388 Aphrodite has a secondary mode in its posterior probability distribution of the period with a peak  $P$  of 7.97 hours and an  $M$  value of 1754.6 days based on Figure 10 (as opposed to an  $M$  value of 45251.9 days for the highest probability mode with  $P$  11.95 hours). This means the Gaussian Process models variations in the light curve and phase-angle on a much shorter timescale at 7.95 hours than at 11.95 hours. In practice, while both periods presented are possible in theory, the posterior distribution for the period  $P$  favors the previously established period of 11.94389 hours. The evolution of the rotational profile can be seen directly in Figure 9, where a realization of the Gaussian Process model with a period near 7.97 hours (yellow) has been plotted and folded for the entire observational window.

While this secondary mode is less probable than the primary mode at  $P$  11.97 hours (and therefore likely is not the correct intrinsic period, especially when compared to previously-reported periods), it illustrates how in the limit of sparse observations, there is an intrinsic degeneracy between the period  $P$  and the metric  $M$  of the squared exponential covariance function: in order to accommodate a period that is shorter than the likely underlying period, the model adjusts to incorporate rapid changes in the rotational profile in order to adequately represent all data points. Whether this is realistic or not depends on the actual location and movement of the asteroid, something we currently do not directly take into consideration. For distinct populations of asteroids with some known physical properties (e.i. near-Earth asteroids versus Trojan asteroids), it may be advisable to modify the prior so that the metric parameter  $M$  better reflects the timescales over which we might expect the shape of the rotational profile to change. Here, we use a relatively broad prior for  $M$  to model asteroids since we are not fitting any distinct population. In practical use

scenarios, the Gaussian Process can be guided towards realistic rotational profiles through physically motivated priors for the metric parameter  $M$ .

Asteroid 299 Thora has a sparse data set, 49 observations over 153 days, and is characterized as having a rotational period of 269.79 hour with a 68% credible interval ranging from 269.00300003 to 270.8985369 hours. The 273-hour rotational period reported by Pilcher et al. (2017) falls within three standard deviations of the peak and upon visual inspection, ZTF observations folded at 273 hours produce a more tightly-constrained rotational profile with less scatter. This could imply that 273 hours is a more accurate rotational period for Asteroid 299, and that the current selection of observations don't lend themselves to this interpretation. Another possible explanation could be that the phase-angle was not fully normalized with OpenOrb. ZTF observations for Asteroid 299 show a slight average increase in magnitude over time (about a 0.1 magnitude rise over 153 days) which could be indicative of an inaccurate phase-angle model. Our model does not account for any changes in the average light curve magnitude. That aspect of the light curve is modeled with the mean function, which at the moment we assume is constant (see Section 5.3 for further discussion). Thus, if our phase-angle correction was calculated incorrectly, then, our assumption that the mean function is constant is false, and it is possible that the metric  $M$  of the squared exponential covariance function is compensating for our error by modeling the biases in the light curve introduced by an incorrect phase-angle correction. In this case, the metric and amplitude of the squared exponential covariance function should be interpreted with caution.

### 5.1. Limitations

We note some of the inherent limitations to the Gaussian Process model. There are occasions where the model fails to characterize an asteroid, either because (1) large outliers are present in the data, (2) uncertainties associated with the magnitude measurements are large, or (3) the span of observations is too short.

Outliers in data sets, several standard deviations removed from the mean, can cause Gaussian Processes to fail. In these cases, we often see marginalized parameter posteriors that reproduce their prior distributions, spanning several orders of magnitude or their entire range as specified by their prior (e.g. the period parameter cannot exceed one year, or 8000 hours, set by its prior). If we remove these outliers, the model is able to successfully characterize the light curve.

The Gaussian Process model fails to generate a good model for light curves if an asteroid's magnitude uncer-



tainty constitutes more than 10% of the intrinsic magnitude variability. 821 Fanny is a prime example of this limitation. It has a 25% relative magnitude uncertainty when compared to its intrinsic variability, suggesting the observations are not constraining the Gaussian Process model, allowing it to fit almost any period to the data. It is possible that the current viewing angle of 821 Fanny is such that the light curve amplitude is minimized, maximizing the relative magnitude uncertainty. There exists a degeneracy between light curve amplitude and viewing angle, where observing an asteroid pole-on will result in a very small light curve amplitude, whereas observing an asteroid equatorially will allow for the full light curve amplitude to be observed, in the case of an oblate tri-axial ellipsoid. Knowing an asteroid’s pole orientation will reveal when an asteroid is being observed equatorially versus pole-on (Vokrouhlický et al. 2017). Regardless, the current ratio between 821 Fanny’s magnitude uncertainty and the intrinsic magnitude variability does not constrain the model enough to provide valuable results.

The model also does not perform well when given data that does not span multiple rotational periods. However, if the rotational period of an asteroid is undetermined, it is impossible to know if this criterion has been met. If the model suggests a rotational period close to the length of the observational window (typically within an hour), it is a strong indication that the observational window of the data is too small (e.g. nightly observations consistently produce rotational period estimates 24 or 48 hours). These results should not be trusted, and we suggest adding a secondary source of data to increase the length of the observational window. In most cases, this occurs when data only spans a couple of hours from one night of observations. In this situation, the model is unable to find a suitable rotational period shorter than the length of the observations, indicating that no smaller period exists and that no other method would be able to find an appropriate period.

### 5.2. Recommendations for Usage

Gaussian Processes implemented with the covariance functions used here are generally computationally expensive and scale with the cube of the number of data points,  $N^3$ , since they require computation of both determinant and inverse of the  $N \times N$  covariance matrix. This model should ideally be used in scenarios where faster methods like the LSP either fail to find a believable period or show multiple peaks with similar heights in the periodogram. In both situations, the Gaussian Process model adds probabilistic information and, as we’ve shown, might be able to better constrain the asteroid’s properties. These

scenarios are most likely to occur with sparse data sets, which makes inference computationally feasible.

We envision an implementation in a two-step process, where initial screening of a large survey of asteroid light curves is initially analyzed using fast, traditional methods of period detection. An automatic process then generates candidates for further studies: either asteroids that are of particular interest, cases where the inferred period deviates from previous measurements for that asteroid, or cases where traditional methods produce ambiguous results. These sources may then be followed up with the approach laid out here to produce better inferences for periods, as well as other other parameters such as amplitudes and measures of the variation of the rotational profile that traditional methods cannot provide. In cases where the Gaussian Process produces ambiguous results, too, its probabilistic forecasting capabilities can be employed to make predictions for when potential follow-up observations should be optimally scheduled in order to rule out certain modes in the posterior probability distribution.

### 5.3. Future Work

While the Gaussian Process model constrains both the periodicity and the profile evolution as a result of the changing phase-angle, there is currently no term explicitly included in the Gaussian Process for modeling the overall phase-angle change of the light curve. Here, we normalized any variation in magnitudes resulting from the phase-angle out of the light curve. We accomplish this by modeling the expected ZTF-observed asteroid light curves using OpenOrb and then subtracting this model from the observed data. Future work would see an overall phase-angle term added to the model through a time-dependent mean function. This allows the model to correctly incorporate uncertainties in the phase angle model and help avoid situations where the period determination is strongly affected by biases introduced in an earlier phase angle correction step.

If we observe an asteroid in multiple bands, it is necessary to expand the model to simultaneously fit multiple light curves of the same period with differing profiles. Future work would see the addition of multiple bands to the model, similar to how the LSP from Gatspy can fit multiple bands at once.

Computational improvements to the model may include a sampler more tailored towards multimodal posterior distributions, as well as a more computationally efficient covariance functions. Because the affine-invariant sampler implemented in *emcee* uses a subset of chains to generate proposals for new parameters in other chains, it is generally not well-suited to the problem considered

here. A future exploration and comparison of sampling algorithms and implementations may significantly improve both inferences and computational efficiency. The covariance functions chosen in this work have a computational scaling of  $\mathcal{O}(N^3)$ . While reasonable for sparse light curves, this scaling quickly makes the model prohibitively expensive when aiming to provide comparisons on densely sampled light curves. Other periodic or quasi-periodic covariance functions with more favourable computational complexity exist (e.g. Foreman-Mackey et al. 2017), and may provide a similarly appropriate model for asteroid light curves.

In this paper, we design the prior probability for the period based on the distribution of known asteroid periods in the LCDB. While this distribution is potentially biased towards lower periods because of observational biases, we also show that it can nevertheless infer the period for long-period asteroids. In the future, implementing more realistic priors for both the period  $P$  and metric  $M$  parameters, potentially tailored towards certain asteroid populations of interest, should lead to improved inferences for these asteroids, and will allow better population studies.

Additionally, while we primarily focus on the inference of the rotational period in this paper, we can also explore the physical meaning of other parameters. In particular the length-scale parameter  $\Gamma$  provides a proxy for the overall reflective properties of the asteroid which could be exploited to explore the composition and albedo of different populations.

## 6. CONCLUSION

In the era of wide-field surveys like the Rubin Observatory’s Legacy Survey of Space and Time and the Zwicky Transient Facility, sparse photometric measurements will constitute an increasing percentage of asteroid observations, particularly for currently unknown asteroids that have yet to be detected. These surveys open the door for large-scale population studies, but require inferring physical properties from individual asteroids within these vast data sets. At the same time, follow-up observations to supplement existing data may be prohibitively expensive in many cases.

In this paper, we set out to add to the toolbox of methods that can infer periods and other relevant asteroid properties from sparse light curves, and show that the Gaussian Process model can infer the correct period for both short-period and long-period asteroids. By

implementing both a periodic sine-squared exponential covariance function and a squared exponential covariance function, we are able to model complex rotational profiles with only a few dozen observations. We test our Gaussian Process model on sparse simulated asteroid observations as well as ZTF observations. We find that our model produces reliable parameter estimates as long as the data do not include the concerns put forth in Section 5.1. We hope that the addition of this new model will allow astronomers to characterize asteroids whose properties were previously unconstrained.

## ACKNOWLEDGMENTS

The authors wish to thank Josef Ďurech for helpful discussion in accessing the DAMIT asteroid model database in generating the synthetic asteroid lightcurves used in this study as well as help with the DAMIT synthetic asteroid lightcurve code.

Based on observations obtained with the Samuel Oschin Telescope 48-inch and the 60-inch Telescope at the Palomar Observatory as part of the Zwicky Transient Facility project. ZTF is supported by the National Science Foundation under Grant No. AST-1440341 and a collaboration including Caltech, IPAC, the Weizmann Institute for Science, the Oskar Klein Center at Stockholm University, the University of Maryland, the University of Washington, Deutsches Elektronen-Synchrotron and Humboldt University, Los Alamos National Laboratories, the TANGO Consortium of Taiwan, the University of Wisconsin at Milwaukee, and Lawrence Berkeley National Laboratories. Operations are conducted by COO, IPAC, and UW.

D.H. is supported by the Women In Science Excel (WISE) programme of the Netherlands Organisation for Scientific Research (NWO). D.H. acknowledges support from the DIRAC Institute in the Department of Astronomy at the University of Washington. The DIRAC Institute is supported through generous gifts from the Charles and Lisa Simonyi Fund for Arts and Sciences, and the Washington Research Foundation. This work was supported by a Data Science Environments project award from the Gordon and Betty Moore Foundation (Award #2013-10-29) and the Alfred P. Sloan Foundation (Award #3835) to the University of Washington eScience Institute and by the eScience Institute.

This research has made use of NASA’s Astrophysics Data System Bibliographic Services.

## REFERENCES

- Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg, D. W., & O’Neil, M. 2015, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38, 252
- Barucci, M. A., & Fulchignoni, M. 1982, *Moon and Planets*, 27, 47

- Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019a, *PASP*, 131, 018002
- Bellm, E. C., Kulkarni, S. R., Barlow, T., et al. 2019b, *PASP*, 131, 068003
- Bolin, B. T., Weaver, H. A., Fernandez, Y. R., et al. 2018, *ApJ*, 852, L2
- Bolin, B. T., Fremling, C., Holt, T. R., et al. 2020, *ApJL*, 900, L45
- Bowell, E., Hapke, B., Domingue, D., et al. 1989, in *Asteroids II*, ed. R. P. Binzel, T. Gehrels, & M. S. Matthews, 524–556
- Brewer, B. J., & Stello, D. 2009, *MNRAS*, 395, 2226
- Chang, C.-K., Ip, W.-H., Lin, H.-W., et al. 2015, *ApJS*, 219, 27
- Durech, J., Sidorin, V., & Kaasalainen, M. 2010, *A&A*, 513, A46
- Fedorets, G., Micheli, M., Jedicke, R., et al. 2020, *AJ*, 160, 277
- Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, *AJ*, 154, 220
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, 125, 306
- Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., et al. 2019, *AJ*, 157, 98
- Graham, M. J., Kulkarni, S. R., Bellm, E. C., et al. 2019, *PASP*, 131, 078001
- Granvik, M., Virtanen, J., Oszkiewicz, D., & Muinonen, K. 2009, *Meteoritics and Planetary Science*, 44, 1853
- Green, S., & Kowal, C. 1983, *IAUC*, 3878, 1
- Gutiérrez, P. J., Davidsson, B. J. R., Ortiz, J. L., Rodrigo, R., & Vidal-Nuñez, M. J. 2006, *A&A*, 454, 367
- Hanuš, J., Vokrouhlický, D., Delbo', M., et al. 2018a, *A&A*, 620, L8
- Hanuš, J., Delbo', M., Alí-Lagoa, V., et al. 2018b, *Icarus*, 299, 84
- Harris, A. W., Pravec, P., & Warner, B. D. 2012, *Icarus*, 221, 226
- Harris, A. W., Young, J. W., Bowell, E., et al. 1989, *Icarus*, 77, 171
- Harris, A. W., Pravec, P., Galád, A., et al. 2014, *Icarus*, 235, 55
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, *ApJ*, 873, 111
- Jedicke, R., Bolin, B., Granvik, M., & Beshore, E. 2016, *Icarus*, 266, 173
- Jewitt, D., Agarwal, J., Li, J., et al. 2017, *AJ*, 153, 223
- Jones, R. L., Slater, C. T., Moeyens, J., et al. 2018, *Icarus*, 303, 181
- Kelly, B. C., Becker, A. C., Sobolewska, M., Siemiginowska, A., & Uttley, P. 2014, *ApJ*, 788, 33
- Lomb, N. R. 1976, *Ap&SS*, 39, 447
- Masiero, J., Jedicke, R., Ďurech, J., et al. 2009, *Icarus*, 204, 145
- Moreno, F., Pozuelos, F. J., Novaković, B., et al. 2017, *ApJL*, 837, L3
- Parzen, E. 1962, *Ann. Math. Statist.*, 33, 1065.  
<https://doi.org/10.1214/aoms/1177704472>
- Pilcher, F. 2013, *Minor Planet Bulletin*, 40, 158
- . 2018, *Minor Planet Bulletin*, 45, 356
- Pilcher, F., Franco, L., & Pravec, P. 2017, *Minor Planet Bulletin*, 44, 270
- Polakis, T. 2018, *Minor Planet Bulletin*, 45, 347
- Prša, A., Zhang, M., & Wells, M. 2019, *PASP*, 131, 068001
- Rasmussen, C. E., & Williams, C. K. I. 2006, *Gaussian Processes for Machine Learning*
- Roberts, S., Osborne, M., Ebdon, M., et al. 2012, *Philosophical Transactions of the Royal Society of London Series A*, 371, 20110550
- Rosenblatt, M. 1956, *Ann. Math. Statist.*, 27, 832.  
<https://doi.org/10.1214/aoms/1177728190>
- Scargle, J. D. 1982, *ApJ*, 263, 835
- Schemel, M., & Brown, M. E. 2020, *arXiv e-prints*, arXiv:2011.01329
- Stellingwerf, R. F. 1978, *ApJ*, 224, 953
- VanderPlas, J. T., & Ivezić, Ž. 2015, *ApJ*, 812, 18
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261
- Vokrouhlický, D., Pravec, P., Durech, J., et al. 2017, *A&A*, 598, A91
- Warner, B. D., & Harris, A. W. 2011, *Icarus*, 216, 610
- Warner, B. D., Harris, A. W., & Pravec, P. 2009a, *Icarus*, 202, 134
- . 2009b, *Icarus*, 202, 134
- Zappala, V., Cellino, A., Barucci, A. M., Fulchignoni, M., & Lupishko, D. F. 1990, *A&A*, 231, 548