ANNUAL REVIEWS

*Annual Review of Astronomy and Astrophysics*

# Gaussian Process Regression for Astronomical Time Series

Suzanne Aigrain[1] and Daniel Foreman-Mackey[2]

[1]Department of Physics, University of Oxford, Oxford, United Kingdom; email: suzanne.aigrain@physics.ox.ac.uk

[2]Center for Computational Astrophysics, Flatiron Institute, New York, NY, USA; email: dforeman-mackey@flatironinstitute.org

Annu. Rev. Astron. Astrophys. 2023. 61:329–71

First published as a Review in Advance on June 13, 2023

The *Annual Review of Astronomy and Astrophysics* is online at astro.annualreviews.org

https://doi.org/10.1146/annurev-astro-052920-103508





## Keywords

astronomy data analysis, time-series analysis, time-domain astronomy, astrostatistics techniques, computational methods

## Abstract

The past two decades have seen a major expansion in the availability, size, and precision of time-domain data sets in astronomy. Owing to their unique combination of flexibility, mathematical simplicity, and comparative robustness, Gaussian processes (GPs) have emerged recently as the solution of choice to model stochastic signals in such data sets. In this review, we provide a brief introduction to the emergence of GPs in astronomy, present the underlying mathematical theory, and give practical advice considering the key modeling choices involved in GP regression. We then review applications of GPs to time-domain data sets in the astrophysical literature so far, from exoplanets to active galactic nuclei, showcasing the power and flexibility of the method. We provide worked examples using simulated data, with links to the source code; discuss the problem of computational cost and scalability; and give a snapshot of the current ecosystem of open-source GP software packages. In summary:

- GP regression is a conceptually simple but statistically principled and powerful tool for the analysis of astronomical time series.
- It is already widely used in some subfields, such as exoplanets, and gaining traction in many others, such as optical transients.
- Driven by further algorithmic and conceptual advances, we expect that GPs will continue to be an important tool for robust and interpretable time-domain astronomy for many years to come.

## Contents



## 1. INTRODUCTION

Gaussian processes (GPs) are a powerful class of statistical models that allow us to define a probability distribution over random functions. Rather than write down an explicit mathematical formula for the function from which some observations are generated, we model the covariance between pairs of samples from the process, using our physical domain knowledge and/or available data to guide our modeling. Although this may seem abstract, GPs have a wide range of applications from modeling of stochastic physical processes, to high dimensional interpolation and smoothing. In particular, Gaussian process regression (GPR) has become increasingly popular in the astronomical community over the past decade. Part of the reason for this uptake is the growing availability and importance of time-domain data sets in astronomy. These systematically contain nontrivial random or unknown signals, whether astrophysical or instrumental, that need to be modeled. In many cases, these are nuisance signals, which we need to marginalize over in order to detect or measure other signals robustly. Sometimes, we are interested in the stochastic behavior itself, and want to infer its characteristics or predict its behavior. GPR offers a compelling solution: statistically principled, readily incorporated in a Bayesian inference framework, and extremely flexible, yet mathematically simple. To new users, however, the lack of an explicit functional form for the model can make GPR seem a little arcane. Furthermore, the computational

**GP:** Gaussian process

**GPR:** Gaussian process regression

cost of the method, which naively scales cubically with the data set size, can be an obstacle. These factors initially impeded its dissemination in the astronomical community, but have been largely overcome in recent years thanks to the availability of user-friendly, computationally optimized software packages.

In this review, we aim to give the reader a concise introduction to the basic theoretical framework of GPR and to illustrate the range, strengths, and limitations of the method by discussing some of its applications to time-domain data sets in astronomy so far. We made no attempt to provide a complete theoretical description or to discuss the relevant literature exhaustively: There are excellent textbooks that already do the former, whereas the latter would have been impossible in the space available. Instead, our goal was to provide an accessible introduction that might give interested readers starting points for further reading, and help them determine whether GPR might be useful for their own data sets. With this in mind, we took pains to require minimal prior knowledge, to provide practical advice regarding the modeling choices one needs to make when using GPR, and to review relevant open-source software.

As a technical note, this manuscript was prepared using the showyourwork package[1] and the source code used to generate each figure is available in a public GitHub repository.[2]

## 1.1. Brief History

An early use of GPR was for spatial interpolation in geophysics (Krige 1951), and GPR has since been adopted or reinvented in a wide range of other application domains. GPs were used in simulations in a wide range of astronomical subfields (see, e.g., Barnes et al. 1980, Constable & Parker 1988, Peebles 1997), but early mentions of GPs for modeling astronomical data sets (see, e.g., Dvorak & Edelman 1976, von der Heide 1978, Jekeli 1991) received limited attention.

Perhaps the earliest use of GPR in the refereed astronomical literature that will be familiar to a modern reader was published by Press et al. (1992a) in the context of quasar variability, and for a long time this remained its main application domain in astronomy. GPR then gradually appeared in other areas, starting with photometric redshift estimation (Way & Srivastava 2006), then exoplanet transit observations (Carter & Winn 2009, Gibson et al. 2012) and radial velocity (RV) planet searches (Aigrain et al. 2012, Haywood et al. 2014). Nonetheless, GPR remained relatively niche, and few astronomers had heard of it until a few years ago. To illustrate this point, we searched on the NASA Astrophysics Data System (ADS) for articles published in refereed astronomy and astrophysics journals with the words "Gaussian process" in the full text of the article (**Figure 1**). After an increase in popularity throughout the 1990s, the use of GPs in astrophysics remained fairly constant at around ∼20 publications per year until 2010. Since 2010, the popularity of GPs has grown significantly, and in 2021, more than 500 refereed papers referencing GPs were

[1]Available at **https://show-your.work**.

[2]Available at **https://github.com/dfm/araa-gps**. Specific versions of the Jupyter notebook that was executed to generate each figure are available here: **Figure 1**, **https://github.com/dfm/araa-gps/blob/8a5b2a4c642e8ad0f562a2bbd1056135fac8da94/src/scripts/literature.ipynb**; **Figure 2**, **https://github.com/dfm/araa-gps/blob/8a5b2a4c642e8ad0f562a2bbd1056135fac8da94/src/scripts/quasar.ipynb**; **Figure 3**, **https://github.com/dfm/araa-gps/blob/8a5b2a4c642e8ad0f562a2bbd1056135fac8da94/src/scripts/transit.ipynb**; **Figure 5**, **https://github.com/dfm/araa-gps/blob/8a5b2a4c642e8ad0f562a2bbd1056135fac8da94/src/scripts/samples.ipynb**; **Figure 6**, **https://github.com/dfm/araa-gps/blob/8a5b2a4c642e8ad0f562a2bbd1056135fac8da94/src/scripts/kernel-ops.ipynb**; **Figure 7**, **https://github.com/dfm/araa-gps/blob/8a5b2a4c642e8ad0f562a2bbd1056135fac8da94/src/scripts/assessment.ipynb**; **Figure 9**, **https://github.com/dfm/araa-gps/blob/8a5b2a4c642e8ad0f562a2bbd1056135fac8da94/src/scripts/scaling.ipynb**.

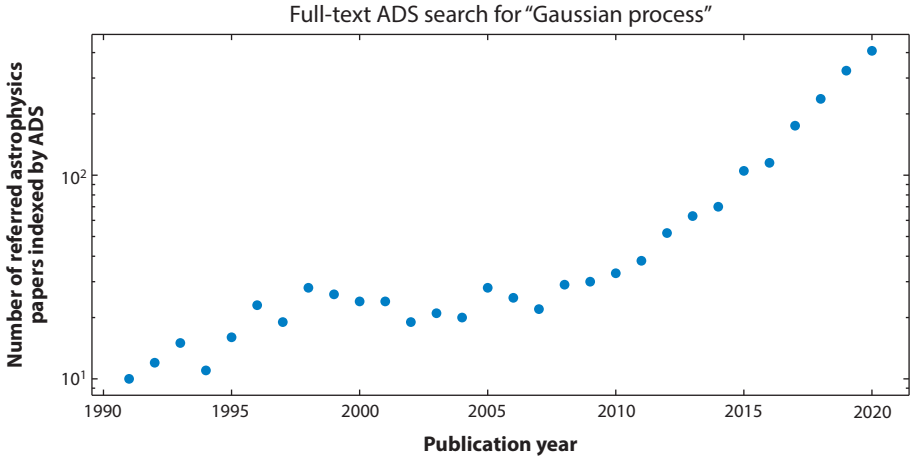


**Figure 1**

The number of refereed publications in the astronomy and astrophysics literature that include the text "Gaussian process" as indexed by the NASA Astrophysics Data System (ADS).

published in the astrophysics literature. A snapshot of the areas in which GPs are currently used in astronomy can be obtained readily from the keywords of those papers published in the past two years (using the NASA ADS analytics tools). This shows that approximately a third of these are cosmology papers, where GPs are primarily used as emulators; this falls outside the scope of this review because it does not involve time-series data. The second and largest grouping covers exoplanets and, by extension, stellar variability (because most exoplanet discoveries rely on indirect detection of the planetary signals in observations of their host stars). Finally, the remaining 30% or so of the papers relate to other kinds of astrophysical variability including active galactic nuclei (AGNs), compact objects, transients, and gravitational waves.

A number of important factors have contributed to the recent democratization of GPR across a wide range of scientific disciplines, including the publication of a dedicated textbook (Rasmussen & Williams 2006), as well as the availability of cheap computing power and user-friendly, open-source GPR software. In astronomy, specifically, packages such as george (Ambikasaran et al. 2015) and celerite (Foreman-Mackey et al. 2017), which were developed by astronomers working with applied mathematicians, have been instrumental in broadening the user base for GPR. An additional factor has also been at play in this application domain: the rise of time-domain surveys. Correlated noise in time-domain observations is a direct and unavoidable consequence of causation, and hence ubiquitous. Adequately modeling this correlated noise is vital when searching for faint signals, for example, from exoplanets. Astrophysical sources, from accretion disks on all scales to magnetically active stars or cloudy brown dwarfs, also display complex, intrinsically or apparently stochastic behavior, for which adequate modeling strategies are required. GPR is a natural choice to tackle these challenges. Rather than attempting to cover all applications of GPs in astrophysics, which would not be feasible in the space available, we have therefore opted to focus on its application to time-domain data sets in astronomy.

## 1.2. Motivating Examples

By way of motivation, before delving any further into the methodology, we have produced two illustrative examples that highlight some typical use cases for GPR in astrophysics. These examples are not meant to comprehensively summarize the space of use cases. Similarly, the goal of this

section is not to formally compare the performance of GPR to that of other methods with the same goals. Instead, these examples are meant to identify some common, but qualitatively different applications. In both examples, we use simulated data sets because it is useful to know the "ground truth" to validate performance.

The first example is a reimplementation of one of the earliest uses of GPR in time-domain astronomy (Press et al. 1992a), using modern language and techniques. In this example, we measure the time delay of a lensed quasar, using a GP as a flexible, nonparametric model for the latent (unobserved) variability of the unlensed quasar system.

The second example demonstrates the use of GPR to account for stellar variability in the light curve of a transiting exoplanet, when inferring its parameters. In this case, the parameters of interest are the parameters of the mean model, the GP is a nuisance model, and our goal is to propagate uncertainty introduced by the stellar variability to our constraints on the physical parameters of interest.

These examples—and all the examples throughout this review—are implemented using tinygp,[3] a Python library for GPR built on top of the JAX[4] library for numerical computing. Here and throughout, the probabilistic models are implemented using the NumPyro[5] library (Phan et al. 2019) and the Markov chain Monte Carlo (MCMC) inference is performed using the No U-Turn Sampling (NUTS) algorithm (Hoffman & Gelman 2014).

**1.2.1. Example 1: The time delay of a gravitational lensed quasar.** In this example, we revisit the method developed by Press et al. (1992a) to measure the time delay of the gravitationally lensed quasar 0957+561, one of the earliest applications of GPR for time-domain astronomy. The underlying model here is that the unobserved latent variability of the source is modeled using a GP, in this case we use a Matérn-3/2 covariance function as discussed and defined in Section 3.1. The images sample this time series at lagged times and with different mean magnitudes and variability amplitudes.

Under this assumed model, we simulate a pair of light curves with the same cadence and uncertainties as the data set from Vanderriest et al. (1989) that was analyzed by Press et al. (1992a). In this simulation, the parameters of the covariance model, the mean magnitudes, and the time delay are all set to known values, designed to produce qualitatively similar features to the data set from Vanderriest et al. (1989). The simulated light curves are plotted in **Figure 2*a***.

Using these simulated data, we fit a GP model using MCMC, varying the time delay, the mean magnitude of each image, the variability amplitude of each image, and the timescale of the covariance. The results of this inference are shown in **Figure 2**. In **Figure 2*a***, we show the simulated data with the median of posterior time delay applied to the light curve of image A, and an arbitrary magnitude offset applied to the light curve of image B for plotting purposes. Overplotted on these data are 12 posterior samples of the GP model predictions for the noise-free photometry for each image. This figure captures how a GP can be used to flexibly capture a stochastically variable process under certain smoothness constraints, and how the uncertainties on the interpolated and extrapolated predictions increase away from the observed data.

**Figure 2*b*** shows the posterior constraints for two of the key parameters of the model: the time delay, and the mean magnitude difference between images. Because these are simulated data,

[3]Available at **https://tinygp.readthedocs.io**; dfm/tinygp: tinygp v0.2.2, from D. Foreman-Mackey, S. Yadav, R. Tronsgaard, S. Schmerler, and T. Rashid.
[4]Available at **https://jax.readthedocs.io**; composable transformations of Python+NumPy programs, from J. Bradbury, R. Frostig, P. Hawkins, et al.
[5]Available at **https://num.pyro.ai**.

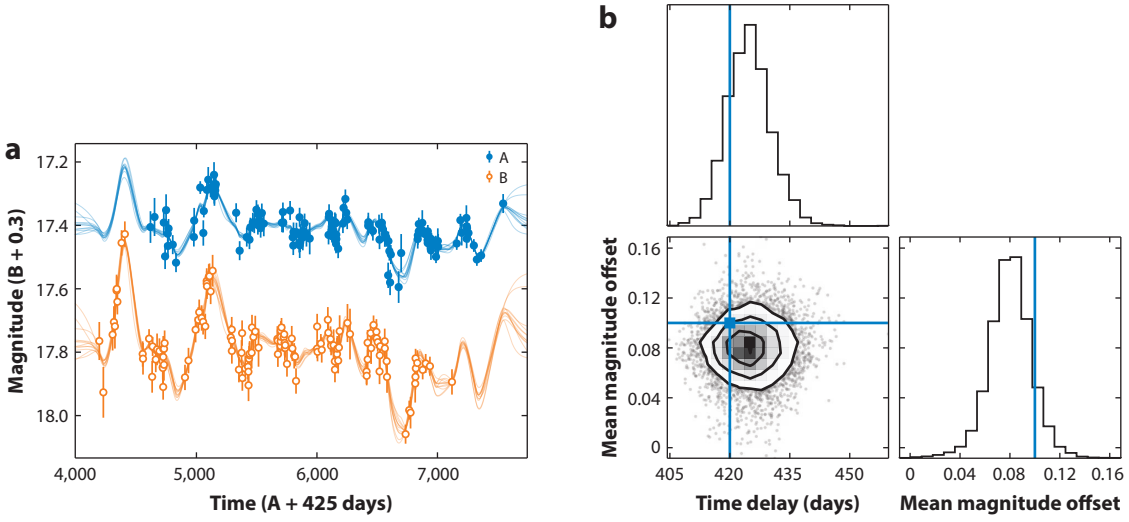


**Figure 2**

The results of fitting the simulated light curves of two images of a lensed quasar with a time delay between images. (*a*) The simulated light curves for each image with an arbitrary magnitude offset applied to image B, and the median of posterior time delay applied to image A, bringing the light curves into a common frame. The data are plotted as points with error bars, and overplotted on these data are posterior predictive samples of each image's predicted variability. (*b*) The posterior constraints on the time delay and mean magnitude offset between the two images, obtained using Markov chain Monte Carlo to fit the simulated data shown in panel *a* of this figure. The contours correspond to the 2-dimensional 0.5-, 1-, 1.5-, and 2-sigma credible intervals, and the true values of these parameters that were used to simulate the data are overplotted as blue lines.

we know the true values of these parameters, and these true values are overplotted in **Figure 2*b***, demonstrating that our method reliably recovers the expected result.

It is worth noting that the determination of the time delay for 0957+561 attracted considerable controversy at the time (see Kundić et al. 1995, and references therein) and that later observations (Kundić et al. 1997) favor a significantly shorter time delay than that obtained by Press et al. (1992a). Whatever the statistical method used, the likelihood function for translation problems of this type can be highly multimodal on large scales, making identification of the global optimum challenging. Furthermore, the assumption that one quasar image's light curve is merely a scaled and shifted version of the other is probably too simplistic. This is another reason why we used simulated data in this example, rather than work with the data set originally analyzed by Press et al. (1992a).

**1.2.2. Example 2: Fitting an exoplanet transit with stellar variability.** In this second example, we demonstrate another common application of GPR in astrophysics, as a flexible model for nuisances and correlated noise, where we want to correctly capture uncertainty introduced by this noise model into our constraints on the parameters of interest. To this end, we simulate the light curve of a transiting exoplanet and aim to infer the physical parameters of the system, taking correlated noise into account. The correlated noise in transit light curves is typically caused by the variability of the host star and by instrumental effects like focus or pointing changes. It has been demonstrated that neglecting to account for these variations can cause significant errors when inferring the properties of the planet (Pont et al. 2006, Gillon et al. 2007). This correlated noise can often be well modeled by a GP, and the use of GPs for transit modeling has been a fruitful area of research in the astrophysics literature (see Section 4.1.2 for a more detailed discussion).

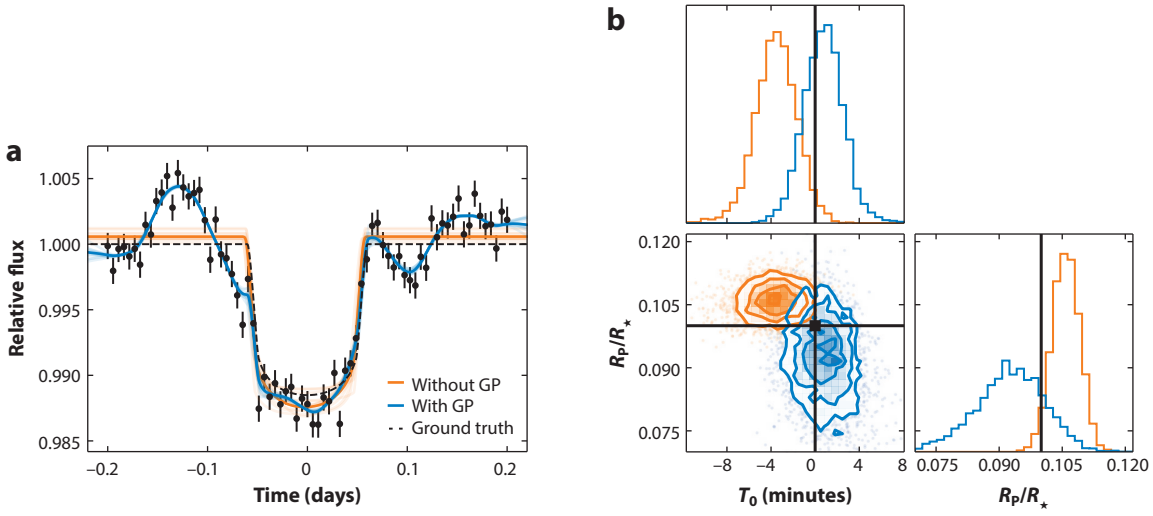


**Figure 3**

The results of fitting the simulated light curve of a transiting exoplanet. (*a*) The simulated data set plotted as points with error bars. The inferred model when correlated noise is (*blue*) or is not (*orange*) taken into account are overplotted on the data. (*b*) The posterior constraints on the time of transit $T_0$ and the planet-to-star radius ratio $R_p/R_\star$. The colors in this figure match those in panel *a*, and the true values of these parameters are indicated with black lines. The results when neglecting correlated noise (*orange*) are significantly inconsistent with the true value, but when the correlated noise is modeled using a GP (*blue*), the correct parameters are recovered, albeit with larger uncertainty. As in **Figure 2**, the contours indicate the 2D 0.5-, 1-, 1.5-, and 2-sigma credible intervals. Abbreviation: GP, Gaussian process.

For this example, the transit was simulated with known physical properties such as the planet-to-star radius ratio, the impact parameter, and quadratic limb-darkening parameters (Agol et al. 2020). We then add correlated and white noise by sampling from a GP model with a Matérn-3/2 covariance function as defined in Section 3.1, with a known amplitude, timescale, and white noise amplitude. These simulated data are shown in **Figure 3*a***.

We then used the same framework to model the simulated data set, first fitting only for an excess variance or "jitter" term (a method commonly used to account for model misspecification) and then accounting for the correlated nature of the noise explicitly using a GP. For the transit model, we fit for the planet-to-star radius ratio $R_p/R_\star$, the mid-transit time $T_0$, the out-of-transit flux $f_0$, and the limb-darkening parameters $u_1$ and $u_2$. We then marginalize over either the variance of the jitter term or the amplitude $\alpha$ and timescale $\lambda$ of the GP noise model. The posterior constraints on the time of transit and the planet-to-star ratio are shown in **Figure 3*b***, with the true values of these parameters overplotted. When the correlated noise is absorbed into a jitter term (plotted in **Figure 3**), the inferred parameters are significantly inconsistent with the truth. Taking the correlated noise into account increases the uncertainties on the inferred parameters but also shifts the results to recover the true parameters.

## 1.3. Overview of This Review

The remainder of this review is structured as follows. Section 2 provides a brief, accessible introduction to the basic theory of GPR, and Section 3 discusses the key modeling choices one needs to make when using a GP to model data, with practical advice for how to make these choices. Section 4 gives an overview of applications of GPR to time-domain astronomical data sets to

date, from stars and exoplanets to AGNs, pulsars, and gravitational waves. In Section 5, we discuss some challenges of using GPR in practice, and how they may be overcome, whereas Section 6 reviews the ecosystem of open-source GPR software available today. Finally, we summarize the main takeaway points and outline future directions in Section 7.

## 2. BASICS OF GAUSSIAN PROCESS REGRESSION

So far, we have discussed how GPR has become widely used in astronomy and touched on some of its pros and cons, but we have not explicitly defined what a GP is or explained how GPR works. This section gives a brief introduction to the theory of GPR, written in a way that we hope will be intelligible to most astronomers. A much more in-depth treatment is provided by Rasmussen & Williams (2006), whose textbook to this day remains the main reference on GPs.

### 2.1. Formal Definition

A GP is a type of stochastic process based on the Gaussian probability distribution. A probability distribution describes a random variable with a finite number of dimensions. A stochastic process extends this concept to an infinite number of dimensions, allowing us to define a probability distribution over functions. Just what do we mean by "extending to an infinite number of dimensions"? Well, this can be a little problematic mathematically, but we need not worry about it, because in practice we only ever deal with finite samples from the stochastic process.

The formal definition of a GP is that the joint probability distribution over any finite sample $\mathbf{y} = \{y_i\}_{i=1,\ldots,N}$ from the GP is a multivariate Gaussian:

$$p(\mathbf{y}) = \mathcal{N}(\mathbf{m}, \mathbf{K}), \tag{1.}$$

where $\mathbf{m}$ is the mean vector and $\mathbf{K}$ the covariance matrix.

The elements of the mean vector and covariance matrix are given by the mean function $m$ and the covariance function $k$, respectively:

$$m_i = m(\boldsymbol{x}_i, \boldsymbol{\theta}), \tag{2.}$$

$$K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{\phi}), \tag{3.}$$

where $\boldsymbol{x}_i$ is the set of inputs (independent variables) corresponding to the $i$th sample. For time-series data, the inputs usually include, but are not necessarily restricted to, the time $t_i$. The covariance function, also known as the kernel function, is the fundamental ingredient of a GP model, and considerable care must be taken to select it adequately (or to test different possibilities). Sometimes the mean function is assumed to be constant, or even zero, everywhere; this is often done in the wider GPR literature to keep derivations uncluttered. However, for many astrophysical applications in which a GP is used to model a nuisance signal, the mean function contains the signal of interest and is important.

The parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ of the mean and covariance functions are known as the hyperparameters of the GP. Strictly speaking, the parameters of the GP are the (infinitely many) unknown functions that share the specified mean vector and covariance matrix and could have given rise to the observations. However, these parameters are always marginalized over: We never explicitly deal with the individual functions, except when drawing samples for illustrative purposes (as we did in **Figures 2** and **3**). GPs are therefore a type of hierarchical Bayesian model (HBM). We never observe the unknown function that gave rise to the data directly, but infer a probability distribution for it from our noisy observations.

Although it is possible to construct and use stochastic process models based on other distributions, GPs are by far the most popular, for two main reasons. The first is the Central Limit

theorem: It implies that the assumption of Gaussianity is often at least approximately correct. The second is that Gaussian distributions obey simple mathematical identities for marginalization and conditioning that enable inference with GPs (the process of marginalizing over the individual functions) to be performed analytically, with very simple linear algebra. It is this analytic marginalization property that sets GPR apart from other forms of HBM.

## 2.2. From Least-Squares Regression to Gaussian Process Regression

The formal definition given in Equation 1 does not necessarily provide an intuitive understanding of how GPs work. Most readers of this review will, however, be more familiar with least-squares regression. In this section, we show that GPR can be thought of as a generalization of least-squares regression, allowing for correlated noise (or signals) in the data. Conversely, least-squares regression, as traditionally presented, is a special case of GPR, where the covariance matrix is assumed to be purely diagonal, and the variances associated with each observation are known a priori.

**2.2.1. Revisiting least squares.** Once again, we consider $N$ observations of a variable $\mathbf{y} = \{y_i\}_{i=1,\ldots,N}$, taken at times $\mathbf{t} = \{t_i\}$, with associated measurement uncertainties $\boldsymbol{\sigma} = \{\sigma_i\}$. We wish to compare these to a model function $m(t, \boldsymbol{\theta})$ controlled by parameters $\boldsymbol{\theta} = \{\theta_j\}_{j=1,\ldots,M}$. In most astrophysical applications, we are interested in estimating some or all of those parameters. In least-squares regression, we minimize the quantity

$$\chi^2 \equiv \sum_{i=1}^{N} (y_i - m_i)^2 / \sigma_i^2, \tag{4.}$$

where $m_i \equiv m(t_i, \boldsymbol{\theta})$, with respect to $\boldsymbol{\theta}$. Where does this come from?

Let us assume that the observations are given by

$$y_i = m(t_i, \boldsymbol{\theta}) + \epsilon_i, \tag{5.}$$

where $\epsilon_i$ is the measurement error, or noise, on the $i$th observation. Furthermore, let us assume that $\epsilon_i$ is drawn from a Gaussian distribution with mean 0 and variance $\sigma_i^2$:

$$p(\epsilon_i) = \mathcal{N}(0, \sigma_i^2) \equiv \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{\epsilon_i^2}{2\sigma_i^2}\right), \tag{6.}$$

then the likelihood for the $i$th observation is simply

$$\mathcal{L}_i(\boldsymbol{\theta}) \equiv p(y_i|\boldsymbol{\theta}) = \mathcal{N}(m_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(y_i - m_i)^2}{2\sigma_i^2}\right]. \tag{7.}$$

We also assume that the noise is uncorrelated, or white, meaning that the $\epsilon_i$'s are drawn independently of each other from their respective distributions. Then, the likelihood for the whole data set $\mathbf{y}$ is merely the product of the likelihoods for the individual observations:

$$\mathcal{L}(\boldsymbol{\theta}) \equiv p(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^{N} \mathcal{L}_i = \prod_{i=1}^{N} \left\{ \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(y_i - m_i)^2}{2\sigma_i^2}\right] \right\}. \tag{8.}$$

From this one can readily see that $\ln \mathcal{L} = \text{constant} - 0.5\chi^2$, where the constant depends only on the $\sigma$'s. Thus, if the $\sigma$'s are known, maximizing $\mathcal{L}$ is equivalent to minimizing $\chi^2$. In other words, least-squares regression yields the maximum likelihood estimate (MLE) of the parameters under the assumption of white, Gaussian noise with known variance.

**2.2.2. Link to Gaussian process regression.** Let us now rewrite the likelihood in matrix form:

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \frac{1}{\sqrt{|2\,\pi\,\mathbf{K}|}} \exp\left(-\tfrac{1}{2}(\mathbf{y}-\mathbf{m})^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{y}-\mathbf{m})\right), \quad 9.$$

where, as before, the mean vector $\mathbf{m}$ has elements $m_i = m(t_i, \boldsymbol{\theta})$, and $\mathbf{K}$ is a purely diagonal $(N, N)$ matrix with elements $K_{ij} = \delta_{ij}\sigma_i^2$ ($\delta_{ij}$ being the discrete Kronecker delta function). This is, of course, the covariance matrix of the model. The notation $|\mathbf{A}|$ refers to the determinant of the matrix $\mathbf{A}$, and $\mathbf{A}^{-1}$ to its matrix inverse.

Now, instead of assuming that the covariance matrix takes this very specific form, let us allow a more flexible covariance model:

$$K_{ij} = k(t_i, t_j, \boldsymbol{\phi}) + \delta_{ij}\sigma_i^2, \quad 10.$$

where $k$ is a covariance function, or kernel function, controlled by parameters $\boldsymbol{\phi}$. The result is a GP, and its likelihood is still given by Equation 9. Depending on the choice of kernel function and parameters, the covariance matrix can now have nonzero off-diagonal elements, allowing us to explicitly model correlated noise or stochastic signals in the data. Note that the likelihood depends not only on the argument of the exponential but also on the determinant of the covariance matrix, which therefore needs to be evaluated explicitly. This term acts as a built-in Occam's razor, automatically penalizing more complex models.

The kernel function $k$ encodes our beliefs about the stochastic, or random, element of the model, in just the same way as the mean function $m$ encodes our beliefs about the deterministic component of the model. For example, in many circumstances, we would expect that two observations taken close together in time should be more strongly correlated than observations taken further apart, and we would use a decreasing function of the time interval $\tau = |t_i - t_j|$ to represent that. One of the most widely used covariance functions is the squared exponential

$$k(\tau;\, \boldsymbol{\phi}) = \alpha^2 \exp\left(-\frac{\tau^2}{2\,\lambda^2}\right). \quad 11.$$

This gives rise to smooth (infinitely differentiable) random functions with variance $\alpha^2$ and characteristic length scale $\lambda$. We present other commonly used kernel functions and discuss how to select one in Section 3.1.

## 2.3. Inference with a Gaussian Process

Now that we know how to evaluate the likelihood (Equation 9), we are ready to perform inference, that is, to use observations to update our prior beliefs about the system we are observing. The overall Bayesian inference workflow for GPR is illustrated schematically in **Figure 4**. Given a data set (**Figure 4*a***) our first task is typically to select a kernel function, along with an educated initial guess for its hyperparameters (**Figure 4*b***). These choices are discussed in more detail in Section 3 and should be guided by a combination of domain knowledge (what we know about the process that generated the data), practical considerations (e.g., ease of implementation and computational cost), and diagnostics based on the data themselves (e.g., examining the autocorrelation function and/or the power spectral density of the data). Given a kernel function and hyperparameters, we can draw samples from the GP prior (**Figure 4*f***), which can serve as sanity check for the choice of kernel function (in the sense that, measurement errors aside, samples from the GP prior should ideally look qualitatively similar to the observations).

We are now in a position to compute the GP covariance matrix for the data (**Figure 4*c***), and to evaluate the likelihood of the model (**Figure 4*d***), which we can then optimize with respect to

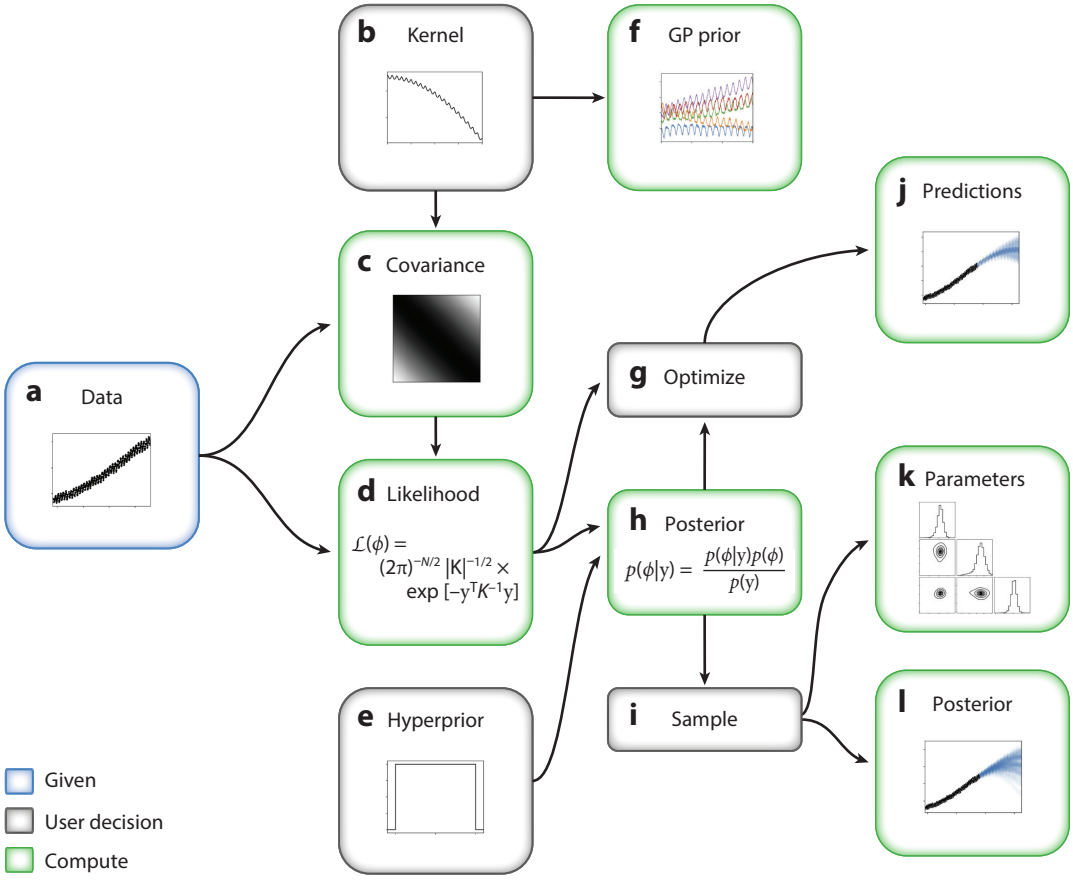


**Figure 4**

Schematic representation of a typical GPR workflow. Given a data set (*blue box*) and some modeling choices (*gray boxes*; Section 3), the mathematical framework presented in Section 2 can be used (*green boxes*) to evaluate the likelihood and posterior distribution over the hyperparameters that can be optimized or sampled. Note that, for simplicity, we have assumed a zero mean function in this figure. We can also condition the GP on the data to predict observations at locations where we do not yet have observations. Several of the plots in this figure are taken from the tutorial on hyperparameter optimization included in the george package documentation (**https://george.readthedocs.io**), and show the composite GP model proposed by Rasmussen & Williams (2006) for the Mauna Loa $CO_2$ data set of Keeling & Whorf (2004). Abbreviations: GP, Gaussian process; GPR, Gaussian process regression.

the hyperparameters (**Figure 4*g***) using standard optimization methods; this is called training the GP. Often, however, we would incorporate priors on the hyperparameters (**Figure 4*e***), which we select once again based on domain knowledge and practical considerations. We can then evaluate the posterior distribution (**Figure 4*h***),

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}, \boldsymbol{\phi})}{p(\mathbf{y})}, \qquad 12.$$

where the denominator $p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}, \boldsymbol{\phi})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\phi}$ is the marginal likelihood, or model evidence, and is often omitted from calculations because it is independent of the hyperparameters but becomes relevant for model comparison. Once again, we can optimize this posterior with respect to the hyperparameters (**Figure 4*g***). For a given set of hyperparameters (for example, the set that maximizes either the likelihood or the posterior), the GP prior can be conditioned on the

observations to make predictions (**Figure 4*j***); this is called conditioning the GP and is discussed in Section 2.4.

However, in astronomical applications we often need to estimate the full posterior distribution over the hyperparameters. Generally we cannot do this analytically, so we resort to sampling methods such as MCMC or nested sampling (**Figure 4*i***). The samples can be used to estimate posterior distributions for individual hyperparameters of interest (**Figure 4*k***). For example, we might be using a GP to take into account correlated noise when fitting an otherwise deterministic model to some data, but we are not interested in the noise per se, only on the impact it has on the parameters of the mean function, $\boldsymbol{\theta}$. We would then marginalize (integrate over, or project) the posterior samples over the nuisance parameters $\boldsymbol{\phi}$. Some samplers also enable us to compute the model evidence, which can be used to compare different kernel functions (see Section 3.3 for a discussion). Finally, posterior samples can also be used in a Monte Carlo fashion to compute predictions that incorporate the uncertainties on the hyperparameters (**Figure 4*l***).

## 2.4. Making Predictions

An important use case for GPR is as a statistically principled interpolation method. We wish to "learn" an unknown function that gave rise to some data, in order to make predictions for some new set of inputs. Importantly, a GP model provides not a point estimate but a full probability distribution for the function at any desired location(s) in the input domain. This allows for robust uncertainty propagation (though there are some important caveats we touch upon in Section 2.4.3), and can also motivate strategies for active sampling (deciding when or where to make new observations).

**2.4.1. The predictive equations.** Given some existing observations $\mathbf{y}$, taken at times $\mathbf{t}$, how do we make predictions at some new set of times $\mathbf{t}_\star$? We are after $p(\mathbf{y}_\star|\mathbf{y})$, the conditional probability distribution for $\mathbf{y}_\star$ given $\mathbf{y}$. In GPR, this is also known as the predictive distribution, because it is often used to extrapolate a time-series data set forward. The magic of GPR is that the predictive distribution is also Gaussian, and its mean and covariance are given by simple analytic relations:

$$\mathbf{f}_\star = \mathbf{m}_\star + \mathbf{K}_\star^{\mathrm{T}}\,\mathbf{K}^{-1}\,(\mathbf{y}-\mathbf{m}) \quad \text{and} \quad \mathbf{C}_\star = \mathbf{K}_{\star\star} - \mathbf{K}_\star^{\mathrm{T}}\,\mathbf{K}^{-1}\,\mathbf{K}_\star, \tag{13.}$$

where we have introduced the vector $\mathbf{m}_\star \equiv m(\mathbf{t}_\star;\boldsymbol{\theta})$ and the matrices $\mathbf{K}_\star$ and $\mathbf{K}_{\star\star}$, with elements

$$[\mathbf{K}_\star]_{ij} = k(t_i, \mathbf{t}_{\star,j};\boldsymbol{\phi}) \quad \text{and} \quad [\mathbf{K}_{\star\star}]_{ij} = k(\mathbf{t}_{\star,i}, \mathbf{t}_{\star,j};\boldsymbol{\phi}). \tag{14.}$$

Note that, as real observations are always noisy, $\mathbf{K}$ generally includes a white noise term ($\delta_{ij}\sigma_i^2$), but $\mathbf{K}_\star$ does not. The white noise term is usually omitted from (the diagonal elements of) $\mathbf{K}_{\star\star}$, in which case the predictive covariance $\mathbf{C}_\star$ accounts only for the uncertainties in the function inferred with the GP model. When making predictions that represent hypothetical additional observations taken with the same setup as the training set, however, the white noise term should be included in $\mathbf{K}_{\star\star}$.

**2.4.2. Properties of the predictive distribution.** A closer look at Equation 13 reveals some important properties. First, a GP is a linear predictor.[6] The predictive mean for a specific time $t_\star$ can be written as a linear combination of the observations: $f_\star = \mathbf{w}^{\mathrm{T}}\,\mathbf{y}$, where $\mathbf{w} = \mathbf{K}^{-1}\,\mathbf{k}_\star$ and $\mathbf{k}_\star = k(\mathbf{t}, t_\star, \boldsymbol{\phi})$. It can also be written as a linear combination of covariance functions centered on the training points: $f_\star = \boldsymbol{\alpha}^{\mathrm{T}}\,\mathbf{k}_\star$ where $\boldsymbol{\alpha} = \mathbf{K}^{-1}\,\mathbf{y}$. These linearity properties are very important to

[6]There is some subtlety with respect to the nonlinear mean function so, in this section, we set the mean to zero or, equivalently, consider the residuals away from the mean model as the data.

understand the behavior of GPs, as they shed light on the relationship between GPR and other types of models, including standard linear models with very large numbers of free parameters (see, e.g., Hogg & Villar 2021, for a more detailed discussion).

Second, the predictive covariance is independent of the data. $\mathbf{C}_\star$ depends only on the locations $\mathbf{t}$ of the observations, not on their values $\mathbf{y}$. This has important consequences for observation planning: If we know the covariance function $k$ and its parameters, we can decide when to make observations to optimize our predictions at a given (set of) time(s). However, in practice we rarely know the hyperparameters a priori, and the predictive posterior distribution marginalized over the hyperparameters does depend on the observations.

Finally, as $\mathbf{K}$ is positive semidefinite, so is $\mathbf{K}^{-1}$. Therefore, $\mathbf{k}_\star^{\mathrm{T}}\,\mathbf{K}^{-1}\,\mathbf{k}_\star \geq 0$, and $\mathrm{Var}(\mathbf{f}_\star) \leq k(t_\star, t_\star)$. This is as we would expect: Obtaining more data should only ever improve the accuracy of our predictions.

**2.4.3. Cautionary notes.** It is important to note that the behavior of $\mathbf{f}_\star$ is not the same as that of individual samples from the predictive distribution. Typically, $\mathbf{f}_\star$ tends to be smoother than individual samples. This should be borne in mind when displaying GPR results or using them in subsequent analysis.

It is also important to note that the predictive variance accounts for the imperfect ability of the specific model under consideration to explain the data but not for the choice of model (i.e., the choice of mean and kernel functions and their parameters). We discuss how to choose a kernel function and fit for its parameters in the next section.

## 3. GAUSSIAN PROCESS MODELING DECISIONS

In the previous section, we presented an overview of GP methods, and the key mathematical details. In this section, we dive deeper into some of the practical decisions that arise when using GPs. The two core elements of a GP model are the mean function $m(t;\, \boldsymbol{\theta})$, and the kernel or covariance function $k(t_i, t_j;\, \boldsymbol{\phi})$. In contrast to the way GP models are typically presented in the machine learning literature, in astrophysics they often—but not always—include nontrivial mean functions. For example, in the case study discussed in Section 1.2.2, the mean function $m(t;\, \boldsymbol{\theta})$ is a physical transit model that is a function of the orbital parameters of the system and includes a realistic limb-darkening model (Mandel & Agol 2002). However, in this review we do not discuss the mean function in detail, focusing instead primarily on the kernel function, because that is unique to GP modeling. All this being said, in our experience, new users of GP models often focus and worry more than necessary about the choice of kernel function for their problem. As with any probabilistic modeling problem, there are several well-defined workflows for motivating, selecting, and validating the choice of kernel function. In this section, we walk through this process in detail. (For definitions, see the sidebar titled Kernel Function and Hyperparameters.)

### 3.1. Gaussian Process Covariance Functions

The kernel function can be any positive scalar function that gives rise to a positive semidefinite covariance matrix over the input domain, but some are more useful than others. Given this large decision space, the reader may be wondering how to choose the right kernel function for their specific problem. If one is very lucky, one may be able to motivate the choice of model using physics. This can be approached from two directions. On one hand, some commonly used kernel functions have a specific physical interpretation—for example, the solution to a stochastic ordinary differential equation—that would naturally motivate their use in certain circumstances. On the other hand, if the physical model is specified by its power spectrum, this can be recast as a GP model with a specific kernel function.

## KERNEL FUNCTION AND HYPERPARAMETERS

### Kernel Function

The kernel or covariance function, typeset here as $k(t_i, t_j; \boldsymbol{\phi})$, is a parametric description of the covariance between two data points $t_i$ and $t_j$. This function takes a pair of inputs (here $t_i$ and $t_j$) as arguments and is controlled by a set of parameters $\boldsymbol{\phi}$, typically referred to as hyperparameters.

### Hyperparameters

In a GP model, the kernel function is typically parameterized by a set of parameters that we label as $\boldsymbol{\phi}$ and refer to as hyperparameters. Formally, the parameters of a GP are the (unknown) true values of the process at the observed times, but the magic of Gaussians means that these parameters can be marginalized over in closed form, leaving the hyperparameters as the parameters of interest.



Even in the absence of a formal physics-based justification for the choice of model, it may be possible to identify the key scales of the problem and design a kernel function that captures these features. The usual approach to this problem is to take sums and products of commonly used kernel functions to construct a set of models that have the desired covariance structure, and then combine or select between these choices. For example, if the problem of interest is expected to involve only a single nonperiodic timescale, one could list all the two-parameter nonperiodic kernel functions and use a numerical model selection technique as described below.

**3.1.1. Standard kernel functions.** Some popular kernel functions are listed in **Table 1**, and some other choices are discussed by Rasmussen & Williams (2006, see their chapter 4). We already met the squared exponential kernel function, defined in Equation 11. Its hyperparameters are $\boldsymbol{\phi} = \{\alpha, \lambda\}$, where $\alpha$ controls the amplitude (output scale) and $\lambda$ the length scale (input scale) of the functions. Due to its simplicity, this is the most widely used kernel for modeling smooth functions of unknown shape.

**Table 1 Some kernel functions commonly used in the astrophysics literature**

| Name | Representation[a] |
|---|---|
| Constant | $\alpha^2$ |
| Squared exponential[b] | $e^{-(\tau/\lambda)^2/2}$ |
| Exponential[c] | $e^{-\tau/\lambda}$ |
| Matérn-3/2 | $\left(1+\sqrt{3}\tau/\lambda\right) e^{-\sqrt{3}\tau/\lambda}$ |
| Matérn-5/2 | $\left(1+\sqrt{5}\tau/\lambda+5(\tau/\lambda)^2/3\right) e^{-\sqrt{5}\tau/\lambda}$ |
| Rational quadratic | $\left(1+\frac{\tau^2}{2\gamma\lambda^2}\right)^{-\gamma}$ |
| Cosine | $\cos 2\pi\tau/\lambda$ |
| Sine squared exponential | $\exp\left(-\Gamma\sin^2\pi\tau/\lambda\right)$ |
| Stochastic harmonic oscillator[d] | $\cos\left(\sqrt{1-\beta^2}\frac{\tau}{\lambda}\right)+\frac{\beta}{\sqrt{1-\beta^2}}\sin\left(\sqrt{1-\beta^2}\frac{\tau}{\lambda}\right)$ |

[a] In each case, $\tau$ is defined as $\tau = |t_i - t_j|$, and Greek letters indicate hyperparameters.
[b] Radial basis function.
[c] Ornstein–Uhlenbeck, damped random walk, or Matérn-1/2.
[d] Stochastically driven, damped, simple harmonic oscillator (Foreman-Mackey et al. 2017).

Another example that is commonly used in the astrophysics literature (Aigrain et al. 2012, Haywood et al. 2014) is the following quasi-periodic (QP) kernel function:

$$k(\tau;\, \boldsymbol{\phi}) = \alpha^2 \exp\left(-\frac{\tau^2}{2\,\lambda_1{}^2} - \Gamma \sin^2\left[\frac{\pi\,\tau}{\lambda_2}\right]\right), \qquad 15.$$

with $\boldsymbol{\phi} = \{\alpha, \lambda_1, \lambda_2, \Gamma\}$, which has a period of $\lambda_2$, and a decoherence timescale of $\lambda_1$. The hyperparameter $\Gamma$ controls the extent to which the periodic component of the signal resembles a simple sinusoid or is more complex.

It is worth noting that, in any practical application, these kernel functions are not generally used on their own. Instead, more expressive models are designed by combining these models (as is discussed in Section 3.1.2).

When selecting a kernel function, it can be useful to generate samples from this implied prior distribution over functions to get a qualitative sense of the properties of the kernel. In practice, this is done by choosing a grid of times $\boldsymbol{t} = \{t_i\}$, evaluating the elements of the covariance matrix

$$K_{ij} = k(t_i, t_j;\, \boldsymbol{\phi}), \qquad 16.$$

and then generating a multivariate Gaussian sample with this covariance. As an example, **Figure 5** shows several prior samples for three different kernel functions from **Table 1**, with a range of length scales $\lambda$ and amplitudes $\alpha$. In this figure, we can see some qualitative differences between the kernel functions—namely that the samples become smoother from left to right—and we can see how the characteristic input and output scales of the allowed functions change with the hyperparameters.

A key feature of a kernel function that defines its qualitative behavior is its mean square (MS) differentiability. We do not formally define MS differentiability—interested readers are instead referred to Rasmussen & Williams (2006, their section 4.1.1) for a more detailed discussion—but we want to highlight that MS differentiability is important because it defines the smoothness of the functions that can be modeled by the GP. It can be seen in **Figure 5** that the squared exponential kernel function, which is infinitely MS differentiable, is significantly smoother than the exponential kernel function, which is not MS differentiable.

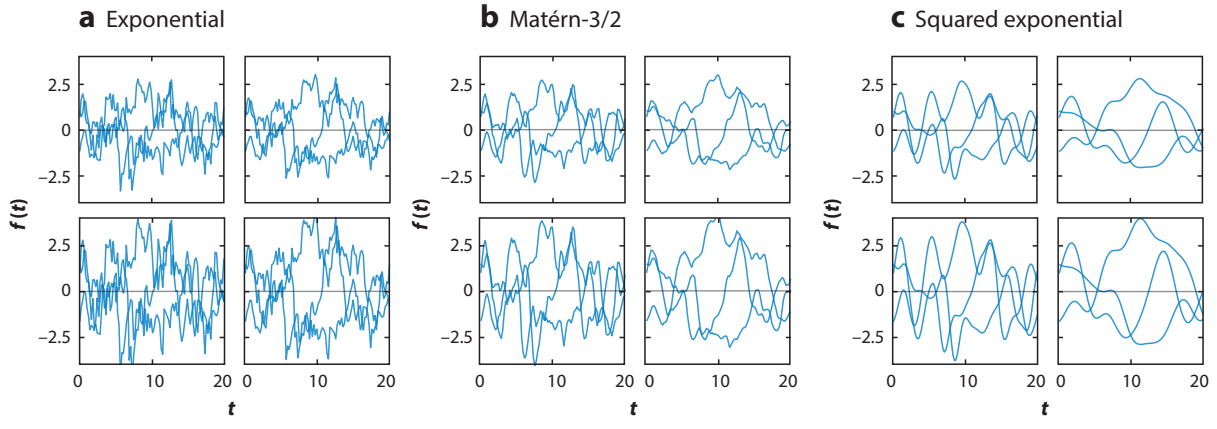


**Figure 5**

Prior samples for three different classes of kernel functions: (*a*) exponential, (*b*) Matérn-3/2, and (*c*) squared exponential. In each subfigure, the length scale $\lambda$ of the kernel increases from left to right, and the amplitude $\alpha$ increases from top to bottom. One thing to notice in this figure is that these kernel functions differ in their smoothness properties. Specifically, the exponential kernel is not mean-square differentiable, whereas the squared exponential kernel is infinitely differentiable. This can be seen qualitatively in this figure.

**3.1.2. Operations on kernel functions.** As discussed above, more flexible kernel functions are often constructed using products and sums of the standard kernels listed in **Table 1**. Besides addition and multiplication, other operations can be used to impose structure on the standard kernels. For example, linear operations like scalar multiplication, more general affine transformations, differentiation, or integration can all be used to develop new kernel functions.

For example, there are good arguments (Aigrain et al. 2012) to expect that the RV time series of a spotted rotating star is related to the photometric time series via its time derivative. More formally, assume we have two time series—one with observations of some function $f(t)$ and another with observations of $\dot{f}(t)$, where the dot indicates the derivative with respect to time—and the function $f(t)$ is modeled as a GP with some covariance function $k(t_i, t_j; \boldsymbol{\phi})$. The covariance between an observation of $f(t)$ and one of $\dot{f}(t)$ is

$$\mathrm{cov}\left[f(t_i), \dot{f}(t_j)\right] = \left.\frac{\partial k(t_i, t; \boldsymbol{\phi})}{\partial t}\right|_{t=t_j}, \tag{17.}$$

and the covariance between two observations of $\dot{f}(t)$ is

$$\mathrm{cov}\left[\dot{f}(t_i), \dot{f}(t_j)\right] = \left.\frac{\partial^2 k(t, t'; \boldsymbol{\phi})}{\partial t\, \partial t'}\right|_{t=t_i, t'=t_j}. \tag{18.}$$

This means that we can compute the GP model for observations of both sets of observations and their covariances. There is one technical point that is worth noting here: For Equation 18 to define a valid covariance function, the base kernel $k(t, t'; \boldsymbol{\phi})$ must be mean square differentiable, as mentioned briefly in the previous section. Therefore, the exponential kernel, for example, should not be used for modeling derivative observations because it does not have the appropriate properties.

**Figure 6** illustrates the elements of GP models with different kernel functions for observations of a time series and its time derivative. The left panels show the covariance functions for all pairwise permutations of $f(t)$ and $\dot{f}(t)$, and the right panels show samples of the function and its derivative. The use of this type of GP kernel is discussed in more detail in Section 4.1.5, with applications to RV time-series observations of exoplanets.

Integration or convolution also has similar properties. For example, real time-series observations are made with finite exposure time $\Delta$, and therefore, if the underlying stochastic process $f(t)$ is modeled as a GP with covariance $k(t_i, t_j; \boldsymbol{\phi})$, the observations are actually of

$$f_{\mathrm{int}}(t) = \int_{t-\Delta/2}^{t+\Delta/2} f(t')\, \mathrm{d}t', \tag{19.}$$

which has a covariance of

$$k_{\mathrm{int}}(t_i, t_j; \boldsymbol{\phi}) = \int_{t_i-\Delta/2}^{t_i+\Delta/2} \int_{t_j-\Delta/2}^{t_j+\Delta/2} k(t, t'; \boldsymbol{\phi})\, \mathrm{d}t\, \mathrm{d}t', \tag{20.}$$

which can be evaluated in closed form for some standard kernel functions. Another context in which integrated GPs have recently been used for astrophysics (albeit not for time series) is as a model of the Galactic dust distribution (Miller et al. 2022).

**3.1.3. Multivariate Gaussian Processes.** Because this review is focused on GPs for time-domain astronomy, we primarily consider univariate GPs with just time as the input coordinate. However, we would be remiss to neglect discussing multivariate GPs altogether. Most of our discussion so far applies to multivariate data sets without any change, but there are some subtleties that are worth mentioning here.

When working with univariate GPs, it is relatively straightforward and unambiguous to compute the distance between two points $t_i$ and $t_j$: $\rho(t_i, t_j) = \tau = |t_i - t_j|$. For multivariate inputs $\boldsymbol{x}$, more

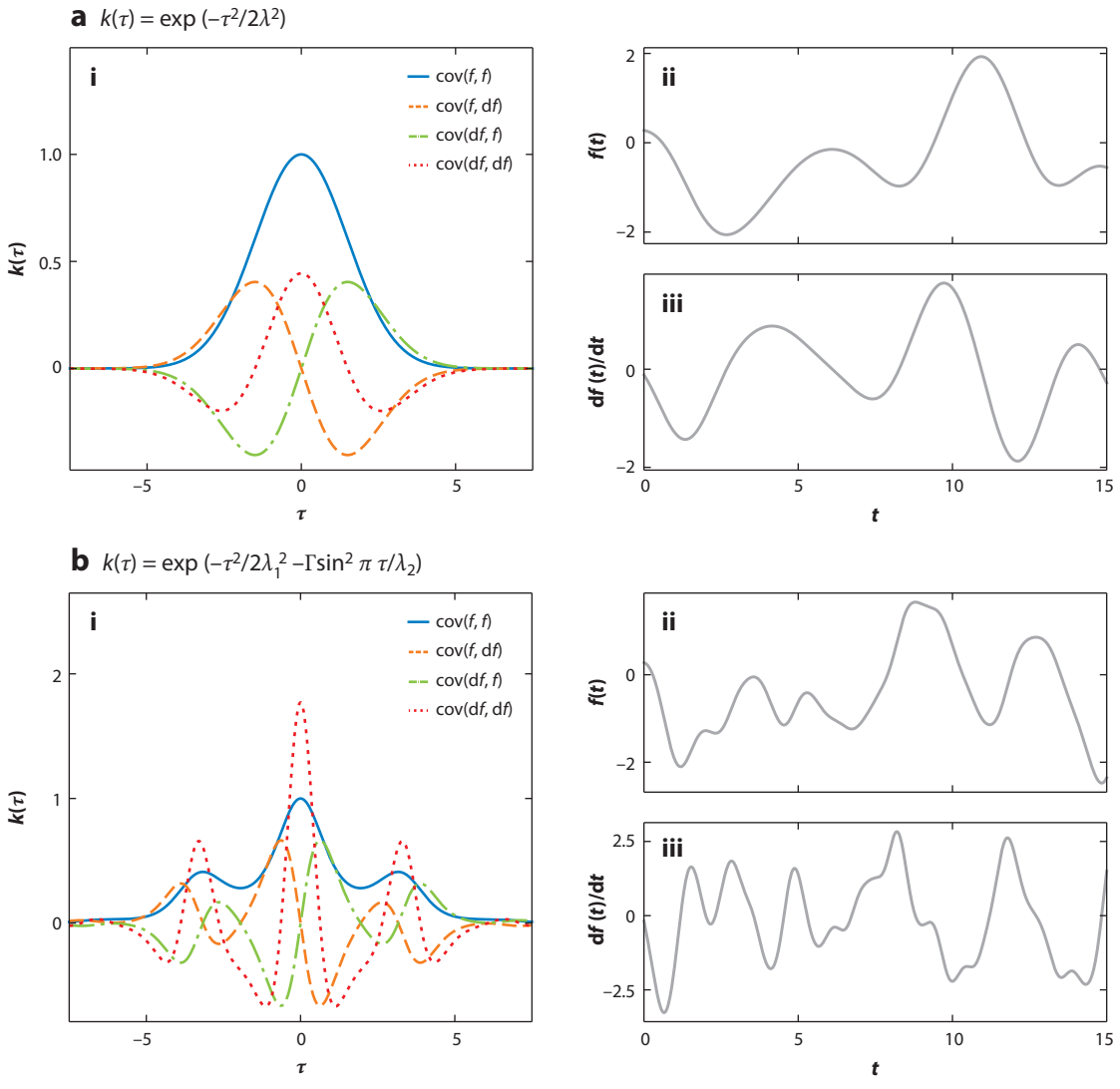


**Figure 6**

A demonstration of the effects of differentiation on a GP model. Each row shows the same panels for a different base kernel: (*a*) a squared exponential kernel ($\lambda = 1$), and (*b*) the product of a squared exponential kernel ($\lambda_1 = 2.5$) with a sine squared exponential kernel ($\Gamma = 1$; $\lambda_2 = 3.5$). In each row, subpanel *i* on the leftmost side shows the kernel function and its first and second time derivatives, which are used to evaluate the covariance matrix for a GP model for a process and its derivative. The right-hand panels show a realization from the process [$f(t)$; subpanels *ii*] and its time derivative [$\dot{f}(t)$; subpanels *iii*] generated by jointly sampling from the GP including the covariance between $f(t)$ and $\dot{f}(t)$. Abbreviation: GP, Gaussian process.

care must be taken to define a sensible distance metric $\rho(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Common choices include the $d$th norm $\rho_d(\boldsymbol{x}_i, \boldsymbol{x}_j) = ||\boldsymbol{x}_i - \boldsymbol{x}_j||_d$ for $d = 1$ or $d = 2$, but other metrics—for example, the Haversine distance when the elements of $\boldsymbol{x}$ are coordinates on a sphere—might be better motivated for a particular use case.

When the input coordinates have some geometric interpretation (e.g., spatial coordinates), simple isotropic distance metrics, like those presented so far, are generally well justified. However,

when the input coordinates are heterogeneous with, for example, different units, it is often more appropriate to use a more general distance metric. In particular, it can be useful to fit for a different length scale parameter $\lambda_d$ for each input dimension $d$:

**ARD:** automatic relevance determination

$$\rho(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{\sum_{d=1}^{D} \left( \frac{x_{id} - x_{jd}}{\lambda_d} \right)^2}, \quad 21.$$

or even include support for covariance between different input dimensions:

$$\rho(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sqrt{\boldsymbol{x}_i \boldsymbol{M}^{-1} \boldsymbol{x}_j}, \quad 22.$$

where $\boldsymbol{M}$ is a $D \times D$ positive definite matrix of hyperparameters.

Kernels with a different length scale parameter for each dimension (Equation 21) are commonly referred to as automatic relevance determination (ARD; Rasmussen & Williams 2006) kernels, the idea being that the data can inform the importance of each input dimension when describing the observations. In practice, such a kernel is often used in conjunction with a shrinkage prior on the hyperparameters, favoring large values of $\lambda_d$, as discussed further in Section 4.1.2.

A common class of multivariate GP model for time-domain astronomy are data sets with multiple parallel covariant time series. For example, multiband time series produced by surveys like Pan-STARRS (Panoramic Survey Telescope & Rapid Response System) or LSST (Large Synoptic Survey Telescope), or RV time series with parallel activity indicators. One useful way to specify these data sets is to define the inputs as $\boldsymbol{x_i} \equiv (t_i, \ell_i)$, where $t_i$ is the time of the $i$th observation, and $\ell_i$ is a label for which time series the $i$th observation is drawn from. In this case, one common choice of kernel function is

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j; \boldsymbol{\phi}) = \boldsymbol{a}_{\ell_i}^{\mathrm{T}} \boldsymbol{a}_{\ell_j} k_0(t_i, t_j; \boldsymbol{\phi}), \quad 23.$$

where $k_0(t_i, t_j; \boldsymbol{\phi})$ is a standard one-dimensional kernel, and the set of $\{\boldsymbol{a}_d\}_{d=1}^{D}$ are also hyperparameters of the model. This class of kernels includes the models commonly used for interpolating multiband photometric observations of transients (e.g., Villar et al. 2020, 2021). In the case in which the data are rectangular (i.e., all input dimensions are observed at all times), the covariance matrix defined by Equation 23 has Kronecker structure (e.g., Wilson et al. 2015, Gordon et al. 2020), which can be exploited to improve the computational performance of the GP model.

**3.1.4. Physically motivated kernels.** Physical processes that are expected to be GPs appear frequently in astrophysics. For example, a physical model that is specified by a parametric form for its power spectrum in the Fourier domain (for example, in the context of the cosmic microwave background or solar-like asteroseismic oscillations) could equivalently be specified by its covariance function (or kernel function) in real space. It is common practice in astrophysics to analyze data sets with these models by first Fourier transforming the data, and then fitting these transformed data using a physical model for the power spectrum. There is a rich literature on this topic, discussing the motivation for this approach, and the rigorous propagation of measurement uncertainty for these data sets. These analyses could also be performed using GPR without first transforming the data, but this has typically been avoided since naïve implementation of GPR would be significantly more computationally expensive or even completely intractable.

There are, however, many cases in which analysis in the time domain is preferred—at least in principle—over a Fourier analysis. This is especially true when working with short, significantly unevenly sampled or otherwise heterogeneous data sets, or when combined with nonstationary functions that cannot be compactly represented in the frequency domain.

Another class of GPs that are commonly used for astrophysics are stochastic differential equations (SDEs; see Särkkä & Solin 2019 for a detailed discussion of these models), although this

terminology is not standard in the literature. The most widely used model in this class is the Ornstein–Uhlenbeck (Uhlenbeck & Ornstein 1930), or damped random walk process (e.g., Kelly et al. 2009, Kozłowski et al. 2010, MacLeod et al. 2010, Bailer-Jones 2012, Hu & Tak 2020). These models—or the generalized continuous-time autoregressive moving average (CARMA) process (e.g., Kelly et al. 2014, Foreman-Mackey et al. 2017, Yu et al. 2022)—are commonly used as a physically motivated model for the stochastic time variability of AGNs. Although this fact is not always noted in the literature, these models are examples of GPs with specific kernel function choices. Besides their physical interpretability, these models can also be evaluated with computationally efficient and scalable algorithms (e.g., Kelly et al. 2014, Foreman-Mackey et al. 2017), which can be crucial for practical applications, as we discuss in Section 5.1.

Asteroseismology is another domain in which Fourier space models are common, and recently there has been some work using time-domain GPR to analyze these data sets (e.g., Grunblatt et al. 2017, Farr et al. 2018). The kernel functions used for these analyses are typically parameterized by the frequency and amplitude of the oscillation modes, but the likelihood is evaluated using Equation 9 in the time domain. Time-domain GP models for asteroseismic oscillations can be particularly useful when combined with a mean model that is compactly specified in the time domain (e.g., a transiting planet; Grunblatt et al. 2017), or when the time sampling of the data is strongly nonuniform with complicated windowing effects (Farr et al. 2018). Like the AGN models discussed above, these asteroseismic GP models can typically be evaluated efficiently using scalable algorithms. So far, the applications of GPR to asteroseismology have been generally restricted to simple effective models with a single wide mode used to capture the power spectrum of the oscillations, but it is conceivable that more complex models could be applied to fully capture the structure of the power spectrum.

Although the models discussed so far in this section have strong physical motivations, it is also common to apply GPR using effective kernel models with physically motived parameters. For example, GPs have been used to measure the rotation periods of stars in photometric time series using a QP kernel function, where one parameter of the model can be interpreted as the stellar rotation period (Angus et al. 2018). These models perform well and can be used to formally propagate the observational uncertainty to the inferred rotation period. Recently, some more physically interpretable kernel functions have been developed for measuring stellar rotation, and mapping stellar surfaces using photometric time series (Luger et al. 2021a), but these models have so far seen limited use because of their computational cost.

## 3.2. Hyperparameter Inference

A key component of all the covariance functions discussed above is that they are all parameterized by a set of hyperparameters. In most cases, one would not have a priori knowledge of how the values of these hyperparameters should be set. Instead, their values need to be numerically tuned or incorporated into a larger inference scheme.

In the astrophysics literature, the most common approach for taking this uncertainty into account—and the method that we advocate for here—is to treat the hyperparameters directly as parameters of the model. In other words, instead of just fitting for the parameters of the mean model $\boldsymbol{\theta}$, we can simultaneously fit for both $\boldsymbol{\theta}$ and the hyperparameters $\boldsymbol{\phi}$. In Section 2.2.2, we defined the likelihood for a GP model and, in Section 2.3, we sketched the procedure used for Bayesian inference with such a model. The likelihood function defined in Equation 9 is a function of both $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, and we can use that function as an objective for a nonlinear optimization routine to find the maximum likelihood parameter values, or in an MCMC (e.g., Hogg & Foreman-Mackey 2018) procedure to marginalize over the hyperparameters and propagate their uncertainty to constraints on the parameters of the mean model.

An important point here is that, because many data analysis procedures in astrophysics include a step like the ones listed above, the use of a GP likelihood does not significantly change the processing. In fact, we like to say that the GP likelihood can be used as a drop-in replacement for any situation where a chi-squared objective is currently used.[7] There are some practical reasons why things are not necessarily this simple (for example, computational cost, as described below), but the sentiment stands.

In this review, we do not go into too many details about the inference algorithms, but throughout the text, we regularly use the Broyden–Fletcher–Goldfarb–Shanno (BFGS) gradient-based, nonlinear optimization routine (Nocedal & Wright 1999, Virtanen et al. 2020) to find the maximum likelihood parameter values,

$$\boldsymbol{\theta}_\star, \boldsymbol{\phi}_\star = \arg\max_{\boldsymbol{\theta},\boldsymbol{\phi}} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \arg\max_{\boldsymbol{\theta},\boldsymbol{\phi}} \log p(\boldsymbol{y} \,|\, \boldsymbol{\theta}, \boldsymbol{\phi}). \quad 24.$$

Another common inference technique used in astrophysics—and in this review!—is the use of MCMC or nested sampling (Skilling 2006) to generate posterior samples,

$$\boldsymbol{\theta}, \boldsymbol{\phi} \sim p(\boldsymbol{\theta}, \boldsymbol{\phi} \,|\, \boldsymbol{y}), \quad 25.$$

that can be used to marginalize over some subset of the parameters and estimate the uncertainty on the parameter values. Sections 1.2.1 and 1.2.2, and specifically **Figures 2** and **3**, show the results of worked examples of an MCMC-based GPR workflow, without and with (respectively) a nontrivial mean function.

### 3.3. Model Assessment, Validation, and Selection

Given the wide array of possible kernel functions described in Section 3.1, it can be important to assess the performance of one's model and the relevant choices. This includes both assessing the choice to use a GP in the first place, and the specific choice of kernel function. It is important to note that there is nothing fundamentally different about GPs in this context when compared to other models for data. Therefore, many model selection and validation methods, which the reader may already have encountered in other contexts (see, e.g., Claeskens & Hjort 2008, for a useful textbook reference), also apply when using GPs. That being said, within the astrophysics literature formal probabilistic model checking has had limited use, and GPs do come with some specific technical complications, therefore we discuss some examples of model validation and selection techniques that have been used for GPs.

When it comes to selecting between different possible kernel functions, the approach to take may depend on the specific research goals. For example, in many cases, including the transiting exoplanet example in Section 1.2.2, the main parameters of physical interest may be the parameters of the mean model, and the GP is simply an effective model for stochastic nuisances. In this case, it may be sufficient to demonstrate that the inferred results are not significantly inconsistent for different choices of kernel function.

Other common use cases, like the time-delay example in Section 1.2.1, primarily require good predictive performance for the GP model. In these cases, methods like cross validation (e.g., Gelfand et al. 1992) or posterior predictive assessment (e.g., Gelman et al. 1996) can be used to evaluate different choices of kernel function.

[7] There may be cases in which a model including correlated noise is not required to explain the data but, because the independent noise model is a special case of most GP models, using a GP isn't conceptually problematic even in those situations.

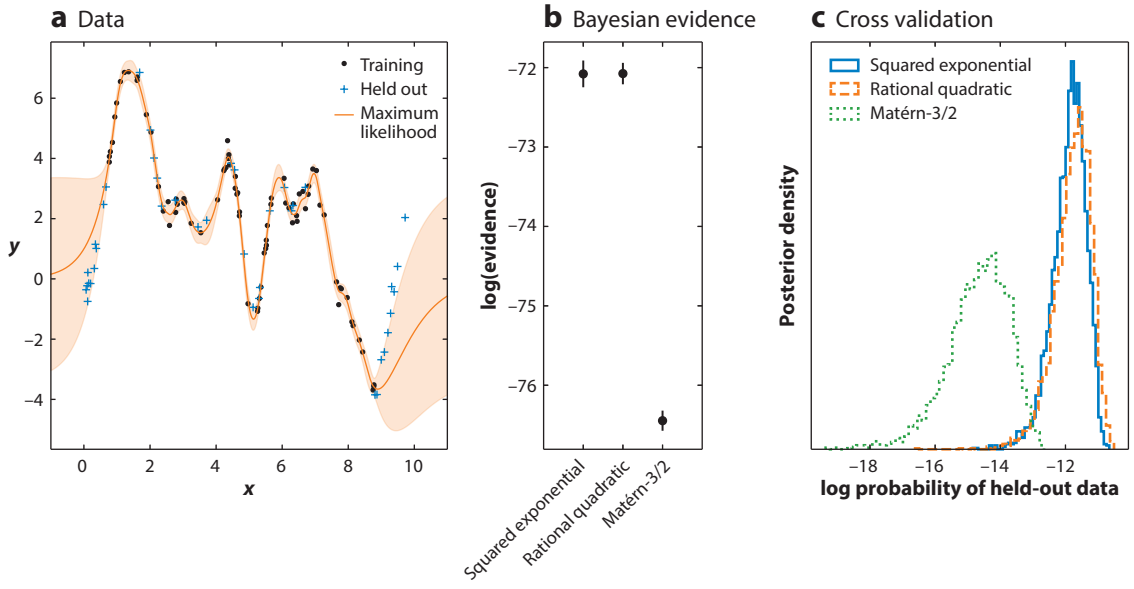


**Figure 7**

A demonstration of model assessment using data simulated from a GP model using a squared exponential kernel with known parameters. (*a*) The simulated data are shown as black dots, with some held out for cross validation indicated by blue crosses. The maximum likelihood estimate of the GP's predictive distribution based only on the black points is indicated with the orange $1-\sigma$ contours. (*b*) The value of the Bayesian evidence integral numerically estimated using nested sampling and all the data in panel *a* (including the data held out for cross validation) is shown for each kernel function. (*c*) The posterior distribution of the held-out log probability is plotted for each kernel function. Abbreviation: GP, Gaussian process.

To demonstrate these approaches, **Figure 7** shows the results of performing a model comparison among three different kernel functions applied to a simulated data set. The simulated data were generated from a GP model with a squared exponential kernel with known parameters, and we aim to compare the performance of three kernels: (*a*) the squared exponential kernel, (*b*) the Matérn-3/2 kernel, and (*c*) the rational quadratic kernel. The simulated data set is shown in **Figure 7*a***.

First, using the full data set, we compute the Bayesian evidence integral for each of these model choices using a nested sampling algorithm implemented in the jaxns package (Albert 2020). The Bayesian evidence integral is defined as

$$Z(H) \equiv p(\boldsymbol{y} \mid H) = \int p(\boldsymbol{\theta}, \boldsymbol{\phi} \mid H)\, p(\boldsymbol{y} \mid \boldsymbol{\theta}, \boldsymbol{\phi}, H)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\boldsymbol{\phi}, \qquad 26.$$

where $H$ indicates the modeling choices, in this case the choice of kernel function. The Bayesian evidence is frequently used in the astrophysics literature as an ingredient in model selection procedures (see Trotta 2008 for a more detailed discussion). Nested sampling is an algorithm for numerically estimating this integral, and we apply it to produce estimates of $Z$ for each choice of kernel $H$, and plot these results in **Figure 7*b***. In this figure, it is clear that the squared exponential and rational quadratic kernels are indistinguishable under this metric, whereas the Matérn-3/2 kernel is somewhat disfavored.

Another popular method for model selection that is less commonly used in the astrophysics literature is cross validation. Cross validation is designed to assess the predictive performance of the model, and it proceeds by holding out some data, fitting the rest of the data, and then

computing the likelihood of the held-out data conditioned on the fit results. These steps can then be repeated for different held-out samples. For a GP model, the likelihood of the held-out data can be computed using the predictive distribution discussed in Section 2.4. In particular, the likelihood is the following multivariate Gaussian,

$$p(\boldsymbol{y}_\star \,|\, \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\phi}) = \mathcal{N}(\boldsymbol{f}_\star, \boldsymbol{C}_\star), \tag{27.}$$

where $\boldsymbol{f}_\star$ and $\boldsymbol{C}_\star$ are defined in Equation 13, noting (importantly!) that the (squared) observational uncertainties on the held-out data should be included on the diagonal of $\boldsymbol{C}_\star$.

In this example, we use MCMC to fit the data shown in **Figure 7*a***, holding out the data points as indicated. At each step in the MCMC, we evaluate the likelihood in Equation 27 for the held-out data conditioned on the training data and the model parameters. In **Figure 7*c***, we plot the posterior distribution of the held-out log probability to show that, like with the evidence integral, the squared exponential and rational quadratic kernels are indistinguishable, whereas the Matérn-3/2 kernel is disfavored. In this case, we hold out a single fixed subset of the data, but it is common practice to repeat this process for several unique held-out subsets, and average these results. There is some freedom in the choice of held-out data, and it is not trivial to select the best procedure for correlated time-series data. In this example, we held out some contiguous data points at the beginning and end of the time series to assess the model's performance when extrapolating, and then held out a random subset of the remaining data points to validate the bulk statistics of the process. For GP models, simple or other uniform sampling schemes are unlikely to be sufficient for assessing the correlation structure of the data, and care should be taken to select validation data sets that capture the details of interest.

Besides the model comparison tests discussed so far, it can also be useful to assess the quality of fit in a more absolute sense. Like with the other assessment tools, there is not anything particularly unique about GP models when it comes to model checking, although it is generally nontrivial to quantify goodness of fit metrics in all but the simplest cases (e.g., Gelman et al. 1995). Our preferred approach to model checking is to use graphical posterior predictive checks (see Gabry et al. 2019, for example). The idea behind posterior predictive checks is that the observed data $\boldsymbol{y}$ should not be an extreme outlier with respect to synthetic data $\boldsymbol{y}_\star$ sampled from the posterior:

$$\boldsymbol{y}_\star \sim p(\boldsymbol{y}_\star \,|\, \boldsymbol{y}) = \int p(\boldsymbol{y}_\star \,|\, \boldsymbol{\theta}, \boldsymbol{\phi})\, p(\boldsymbol{\theta}, \boldsymbol{\phi} \,|\, \boldsymbol{y})\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\boldsymbol{\phi}. \tag{28.}$$

Operationally, a single sample of $\boldsymbol{y}_\star$ is generated by drawing a sample of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ from the posterior, and then drawing $\boldsymbol{y}_\star$ from the sampling distribution $p(\boldsymbol{y}_\star \,|\, \boldsymbol{\theta}, \boldsymbol{\phi})$ conditioned on the sampled values of $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. Data sets in astrophysics are typically sufficiently high dimensional that it is not feasible to visualize the full distribution of $\boldsymbol{y}_\star$, and we instead plot scalar statistics of the data, such as the mean, variance, or skew. If the model is appropriate for the problem at hand, the same statistic computed on the observed data should be consistent with the posterior predictive distribution.

## 4. GAUSSIAN PROCESSES IN TIME-DOMAIN ASTRONOMY

In this section, we discuss a range of applications of GPR to time-domain data sets in the astronomical literature. The present-day popularity of GPs in astronomy precludes any attempt at an exhaustive review. Instead, we have selected examples that showcase the power and flexibility of the method. No special meaning should be ascribed to the order in which these applications are discussed, beyond the fact that we started with the application domains we are most familiar with.

## 4.1. Stars and Exoplanets

It is no coincidence that the recent uptake of GPR in astronomy coincides with the meteoritic rise of exoplanet studies. Almost all exoplanet detections and related discoveries are made indirectly and involve detecting small and/or short-lived signals in time-domain data sets: transits in light curves, Keplerian signals in RV, or astrometry time series. These are invariably affected by, and often buried in, correlated noise or nuisance signals, making GPR a natural choice.

**4.1.1. Transit searches.** Following the discovery of the first transiting exoplanet (Charbonneau et al. 2000, Henry et al. 2000), numerous ground-based photometric monitoring surveys were set up (or repurposed) to search for planetary transits. Early estimates of their yield (see, e.g., Horne 2003) ran into the hundreds of planets per year, but these soon turned out to be highly optimistic. Although surveys such as the Optical Gravitational Lensing Experiment (OGLE) did indeed discover numerous transit-like events (see, e.g., Udalski et al. 2002), only a handful were ultimately confirmed as bona fide transiting planets, and the process took years. The vast majority of early transit candidates were later diagnosed as diluted or grazing stellar eclipses based on RV follow-up. Intriguingly, a small fraction turned out to be spurious detections, despite having relatively high signal-to-noise ratio (SNR). Indeed, Pont et al. (2006) noted that all the confirmed OGLE planets had very high SNRs, and they went on to show that the SNRs for all candidates had been overestimated due to the presence of correlated or red noise[8] in the light curves. They proposed an empirical method to evaluate the red noise on transit timescales from the individual light curves (analogous to the method proposed by Press et al. (1992a) to evaluate the correlation matrix of a quasar light curve), and then a prescription for adjusting the detection threshold accordingly. In effect, they were searching for a way to model the covariance of the data. This is precisely what GPR allows, but the methodology was not known in the exoplanet community at the time.

GP models have been used on occasion to detrend light curves from transit surveys prior to running transit searches, i.e., to remove variability on timescales significantly longer than a transit (see, e.g., Crossfield et al. 2016), and are used frequently after detection to model the out-of-transit baseline alongside the transit signal itself (see Section 4.1.2). In principle, a GP could also be used to model out-of-transit variations (intrinsic or instrumental) alongside transits as part of a detection pipeline. Doing this simultaneously rather than sequentially allows more flexible models to be used for the out-of-transit variations and avoids the risk of corrupting the transit signal by using an excessively aggressive filter at the detection stage, as demonstrated by Foreman-Mackey et al. (2015) for data from the *Kepler 2* space mission (Howell et al. 2014), which used the *Kepler* satellite to perform an Ecliptic plane survey after the failure of two of its reaction wheels. In that case, the primary source of out-of-transit variability is instrumental systematics (see Section 4.1.2), which are to some degree common to all light curves in a given observing run. Foreman-Mackey et al. (2015) model these systematics as a linear combination of the first 150 principal components of the ensemble of light curves, alongside a simple box-shaped transit model, varying the latter's period, phase and duration.

In other space-based transit surveys such as CoRoT (*Convection, Rotation & Planetary Transits*; Baglin et al. 2006), *Kepler* (Borucki et al. 2010), and TESS (*Transiting Exoplanet Survey Satellite;* Ricker et al. 2015), the systematics are less prominent, so the dominant source of out-of-transit variations is intrinsic stellar variability, which occurs mainly (for main sequence stars) on timescales longer than transits, and can therefore be filtered out quite effectively without removing the transits. However, in some cases, for example for rapidly rotating and magnetically active

[8]White, pink, and red noise are colors of noise, i.e., power spectral densities with power-law indices $\gamma = 0$, 1, and 2, respectively.

stars, the separation between transits and stellar signal becomes less clean, significantly reducing the performance of sequential detrending approaches (irrespective of the algorithm used; Hippke et al. 2019). In such cases, a simultaneous modeling approach could improve detection performance significantly. Since the nuisance signal is specific to each star, it cannot be modeled using other light curves, but a GP is a credible alternative. Such an approach could also offer a small but significant enhancement in the sensitivity of future missions such as PLATO (Rauer et al. 2014) to transits of habitable planets around Sun-like stars (which are very shallow and relatively long). The main issue would be computational cost: transit surveys monitor $10^4$–$10^5$ stars at a time, with $10^3$–$10^5$ observations per run, and the transit search must be run over a fine grid of periods and phases. Doing this with standard GPR methods would be impractical. However, now that fast and scalable GP solvers are available, this becomes a more feasible proposition.

**4.1.2. Transit and eclipse modeling.** In practice, GPR has been much more widely applied to transit modeling than to detection per se. The detailed depth, timing, duration, and shape of a transit depends on the planet-to-star radius ratio $R_p/R_\star$, the system scale $a/R_\star$ (where $a$ is the orbital semimajor axis), the impact parameter $b$, the orbital period $P$, and the time of transit center $T_0$, as well as the limb-darkening profile of the star. High-precision observations of one or more transits of a given planet can be used to infer these parameters, and this has a wide range of scientific applications. The most obvious is to estimate the planet size (and hence, given a mass estimate, its bulk density and composition). The system scale is directly related to the stellar density (Seager & Mallén-Ornelas 2003), which can be used for sanity checks to ensure the transits are indeed of planetary origin. Small departures from strictly periodic timing, known as Transit Timing Variations (TTVs), probe dynamical interactions between multiple planets in a given system, and can be used measure the planets' masses and eccentricities and to reveal the presence of additional, nontransiting planets (Holman & Murray 2005). Wavelength-dependent measurements of the transit depth probe the effect of the planet's atmosphere on the starlight that filters through it, and hence allow us to access the atmospheric composition (Seager & Sasselov 2000). The depth of any secondary eclipse (when the planet passes behind the star) depends on the planet-to-star flux ratio, and can therefore be used to measure an emission or reflection spectrum, while its timing yields strong constraints on the orbital eccentricity. All of this requires precise and accurate measurements of the transit parameters, for which any correlated noise must be accounted for explicitly, as we demonstrated using a simulated example in Section 1.2.2.

To address this, Carter & Winn (2009) proposed a wavelet-based method to model correlated noise in transit light curves, which can be seen as a special case of GPR with a covariance function belonging to the exponential family, but formulated in terms of wavelets. Rather than modeling the covariance in the time domain, the Power Spectral Density (PSD) of the noise is assumed to be of the form $P(f) \propto f^{-\gamma}$. Noting that the wavelet transform of such a process gives rise to a nearly diagonal covariance matrix, Carter & Winn (2009) derived an expression for the likelihood that can be computed in $\mathcal{O}(N)$ operations. This method has been used widely since its publication to model transit observations from space missions which stare continuously at a given field, such as *Kepler* or TESS. However, the use of a wavelet transform requires regular time sampling, which precludes its application to data sets with irregular sampling or significant data gaps, for example, from ground-based, space telescopes in low-Earth orbit such as *Hubble*.

This limitation is significant, as most data sets for transit and eclipse spectroscopy are unevenly sampled with significant data gaps. Transit spectroscopy involves measuring minute changes in the depth of the primary transit or secondary eclipse as a function of wavelength, using either successive photometric observations through different filters or a spectrograph (see Kreidberg 2018 for a review). The signal of interest is the wavelength dependence of the transit depth, or the

depth of the eclipse, which are of order $10^{-4}$ and $10^{-3}$ respectively, in the most favorable cases. Even when observed from space, these signals are typically dwarfed by instrumental systematics. The pointing jitter and thermal relaxation of the telescope as it orbits the Earth causes the target star to move on the detector, typically by a fraction of a pixel. As the sensitivity of the detector varies from pixel to pixel (the "flat-field") and between the center and edges of each pixel, this motion causes spurious variations in the recorded flux from the target. Although we expect the measured flux to depend on "housekeeping" variables such as the satellite orbital phase, telescope attitude, the centroid of the image or the locus and angle of the spectrum of the detector, or the temperatures of various parts of the instrument, a physically motivated model for the form of this dependence is generally lacking. Ad hoc parametric (e.g., polynomial) models are problematic, as the choice of model inputs and functional form is arbitrary, yet can drastically alter the resulting exoplanet spectrum (Gibson et al. 2011).

**Latent variable:** a random variable that is never actually observed

To address this, Gibson et al. (2012) proposed a GPR framework to model the systematics in space-based transit observations. The aforementioned housekeeping variables are treated as multidimensional inputs to a squared exponential GP whose mean is the transit signal, allowing one to marginalize over broad families of systematic models without assuming a specific functional form for them, all the while propagating the resulting uncertainties on the physical parameter of interest, namely the (wavelength-dependent) planet-to-star radius ratio. The kernel used was a squared exponential kernel with an automatic relevance determination (ARD) distance metric (Equation 21), which allows one to try including a wide range of housekeeping data into the fit; in principle only those which are genuinely relevant will have an effect on the result. GP-based systematics models of this kind were later extended to eclipse spectroscopy, where the parameter of interest is the planet-to-star flux ratio, which controls the eclipse depth. Among other successes, this approach has enabled the first measurement of the wavelength-dependent albedo of a hot Jupiter (Evans et al. 2013), helped resolve early controversies surrounding the treatment of systematics in Spitzer observations (Evans et al. 2015) and led to the first unambiguous detection of a thermal inversion in an exoplanet emission spectrum (Evans et al. 2017). Using simulated data, Gibson (2014) showed that GPs outperform parametric models when the true form of the systematics is unknown, but also that the most robust results overall are obtained by marginalizing over families of both parametric and GP models.

Despite these theoretical advantages and practical successes, the use of GPR to analyze low-resolution transit and eclipse spectra remains confined to a relatively small subset of the corresponding community. Parametric models (in some cases marginalizing over families thereof; Wakeford et al. 2016) remain the most widely used approach for HST observations, and pixel-level-decorrelation (PLD) for Spitzer observations (Deming et al. 2015). Possible explanations for this include computational cost, as well as the fact that these other methods are perceived as easier to implement and interpret. Another likely reason, at least in the case of HST transmission spectra, is that GP systematics models tend to produce exoplanet spectra with larger uncertainties than parametric ones. This does not mean that they are less accurate, however, merely that they make a less restrictive set of assumptions about the functional form of the systematics. As discussed by Gibson (2014) and demonstrated in the appendix of Evans et al. (2018), when the true functional form of the systematics is unknown, choosing the "wrong" systematic model can lead to overconfident estimates of the transit depth, which is more problematic for the robust interpretation of the results than the enlarged uncertainties produced by a more agnostic GP model.

One general shortcoming of current GP-based systematics models with multidimensional inputs is that the uncertainties on the input variables are ignored. This could be remedied by treating the housekeeping variables as noisy observations of GP latent variables, and modeling

them alongside the observed fluxes, as done by Almosallam et al. (2016) in the context of photometric redshift determination, but we are not aware of published attempts to do this in practice to date. In spectroscopic observations, whether using GPs or parametric models, it has become standard practice to model the "white" light curve first (obtained by integrating the spectrum over the full wavelength range at each time step). "Colored" light curves, extracted in individual wavelength bins, are then divided by the best-fit systematics model derived from the white light curve, before being modeled further. To our knowledge, no attempt has been made to model the wavelength dependence of the systematics directly.

On the other hand, GPR has become quite widely used for modeling correlated noise in single-band transit observations. One striking example is the case of transiting planet candidates discovered by *Kepler* around giant stars. By comparing the stellar densities derived from transit modeling to those expected from independent estimates, Sliski & Kipping (2014) argued that many of these candidates, including the confirmed planet Kepler-91, might be false positives. However, red giant light curves contain stochastic variability on timescales of hours due to granulation. In the case of Kepler-91, Barclay et al. (2015) showed that the apparent density discrepancy disappears when this granulation signal is modeled using a GP.

**4.1.3. Systematics removal in the presence of stellar variability.** GP-based models have also been used to correct instrumental systematics in data from the *Kepler 2* (K2) space mission (Howell et al. 2014). During the K2 observations, the satellite underwent significant roll-angle variations, causing the stars to move on the detector by more than a pixel over a timescale of hours. The satellite thrusters were fired every ∼6 h to return the spacecraft to its nominal attitude, but the resulting drift caused significant changes in the measured stellar fluxes, due to the detector inter- and intrapixel variations, and to changes in the contamination of the photometric aperture by neighboring targets as well as in aperture losses. These can be modeled effectively using a GP with a squared exponential covariance function depending on the roll angle (Aigrain et al. 2015, Crossfield et al. 2016) or the star's 2D position on the detector (Aigrain et al. 2016). In these approaches, a second, time-dependent term is added to the GP covariance function to represent the target star's intrinsic variability (see the sidebar titled Composite Gaussian Processes). This not only improves the fit but also allows the position-dependent systematics to be evaluated separately and, thus, removed while preserving intrinsic variability. The use of a time-dependent GP term improves the photometric performance over other widely used methods for correcting K2 systematics (e.g., Vanderburg & Johnson 2014) when the stellar variability is significant and/or occurs on a timescale similar to the roll-angle variations. Ultimately, the best overall photometric precision for K2 was achieved by combining PLD to model the position-dependent systematics with a time-dependent GP to model intrinsic variability (Luger et al. 2016, 2018).

## COMPOSITE GAUSSIAN PROCESSES

A composite (additive or multiplicative) GP is a powerful way to separate different components of a time-series data set, whether these components depend on different input variables (as in the instrumental systematics examples discussed in Section 4.1.3) or not. To evaluate the predictive distribution for a particular component of a composite GP, one simply evaluates $\mathbf{K}_\star$ and $\mathbf{K}_{\star\star}$ in Equation 13, using that component only, while using the full covariance for $\mathbf{K}$.

**4.1.4. Quasi-periodic Gaussian process models for stellar light curves.** The light curves of Sun-like stars (broadly construed) display low-amplitude QP variations, which are caused by the rotational modulation and evolution of magnetically active regions on their surfaces. These produce QP variations in both photometry and RV observations, and GPs have in recent years become one of the most popular ways of modeling them. One reason for this is that a simple QP covariance function (Equation 15) gives rise to functions, which reproduce the light curves of rotating stars with evolving active regions remarkably well.

GP models were first applied to the simulated observations of the Sun's total irradiance variations by Aigrain et al. (2012), to test a new method to simulate RV variations based on space-based photometry. The context of this work was the large numbers of candidate transiting exoplanets being discovered at the time by the CoRoT (*Convection, Rotation & Planetary Transits*) and *Kepler* space missions. Detecting the planet signals in RV was needed to confirm the planetary nature of the candidates by measuring their masses, but this was hampered by the apparent RV variations caused by active regions. Using simple geometric considerations, Aigrain et al. (2012) derived a simple (approximate) relationship between the flux perturbation caused by active regions, $F$, and their RV signature, which depends on both $F$ and its time derivative $F'$. They tested this data-driven "$FF'$ method" on simulated photometric and RV observations of the Sun-as-a-star produced using resolved magnetograms, using a GP as a principled smoothing tool to evaluate $F$ and $F'$ from the photometry, and comparing the results to the simulated RVs. Aigrain et al. (2012) tested a number of covariance functions, both aperiodic and QP, and found the former to be preferred in the solar case, as expected if most active regions on the Sun do not persist for much longer than its rotation period. More active and/or more rapidly rotating stars display variability that is coherent over multiple rotation periods, making QP GPs the model of choice. For example, Haywood et al. (2014) used a QP kernel when applying the aforementioned $FF'$ method to the active planet–host star CoRoT-7.

Today, using GPs to model light curves containing stellar variability is standard practice. The kernel most frequently used for this purpose is the aforementioned QP kernel, whose simplicity and flexibility make it a popular choice. Angus et al. (2018) implemented a Bayesian inference framework based on this kernel to measure accurate rotation periods from *Kepler* light curves. The very complex dependence of the likelihood surface on the parameters, particularly the period, makes this challenging, and careful tuning of the posterior sampling strategy is required. These methods have since been adapted and applied to ground-based photometric surveys (Gillen et al. 2020) and the shorter baseline K2 light curves (Gordon et al. 2021).

Although the QP kernel provides a phenomenological rather than physically motivated description of the variability, some of its parameters lend themselves to a physical interpretation in terms of the star's rotation period and the evolution timescale of active regions. Nicholson & Aigrain (2022) tested this using simulated light and RV curves based on physical star-spot models and confirm that, for moderately well-sampled data sets, the period of the GP does indeed provide a precise and accurate measure of the stellar rotation period. The same is true, albeit to a lesser extent, for the evolution timescale: The correlation between simulated and recovered values is more scattered and breaks down when the data set spans less than the simulated evolution timescale.

By contrast, the QP kernel (or any covariance function of time only) does not give access to the physical properties of the spots, such as their size, latitude, or contrast. The problem of inferring these properties from light curves is fundamentally ill-posed (Luger et al. 2021b), making direct inference of individual spot properties or brightness maps highly degenerate (whatever the methodology used). However, Luger et al. (2021a) 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 star spot sizes, contrasts, and latitudes. This can be used in a hierarchical Bayesian framework to infer the distribution in question from ensembles of light curves.

Most studies focusing on the photometric signatures of stellar activity ignore variations on short timescales, either working with binned data or using a jitter term to absorb them. This keeps the model simple and, when binning, speeds up computing time. However, when the time sampling and precision of the observations allow it [e.g., for data from CHEOPS (*CHaracterising ExOPlanets Satellite*; Benz et al. 2021), or PLATO (i.e., Planetary Transits and Oscillations of stars)], composite GP models with different terms to describe activity, granulation, and stellar oscillations can be useful. For example, Barros et al. (2020) analyzed light curves of stars observed at high cadence and high precision in the CoRoT asteroseismology field. They showed that the parameters of planetary transits injected into these light curves were recovered more accurately when using a composite GP to model the stellar variability than when using one with a single term. They also tested a white noise only model, which provided less precise estimates than either of the GP models, albeit generally consistent with both.

**4.1.5. Stellar activity in radial velocities.** Given the remarkable instrumental precision achieved by modern RV spectrographs, stellar activity is nowadays the key factor limiting the sensitivity of RV surveys to low-amplitude and/or long-period planet signals. The effect of active regions on RV observations is twofold. First, as dark spots rotate on the stellar surface, they distort the profile of spectral lines, removing a small contribution first from the blue wing of each spectral line, then from the line core, and then from the red wing. The second effect is more subtle: In the absence of active regions, spectral lines of Sun-like stars display a net blue-shift due to granulation (the emission from the hot, up-welling material in the granules, which is blue-shifted, dominates over that from the cooler material falling back down in the intergranular lanes). In a facula, i.e., a region of enhanced magnetic flux density compared to the clean photosphere, convection is suppressed, leading to a localized reduction in this convective blue-shift. Faculae have very small photometric contrasts but can cover a much larger area than dark spots, so their RV signature can dominate over that of spots, especially for moderately slow rotators like the Sun with modest magnetic fields (Meunier et al. 2010).

Overall, active regions produce QP RV variations that are frequently modeled with the same kind of QP GPs as that used for light curves. However, unlike transits, planetary signatures in RV occur on timescales similar to activity signals, meaning that the risk of overfitting is significant. This is partly mitigated for RV follow-up of transiting planets by strong priors on the stellar rotation period and on the period and phase of the planetary orbit, but it is a severe problem for blind RV searches. The issue is compounded by the fact that RV observations are invariably ground-based and often have sparse time sampling.

As previously mentioned, Aigrain et al. (2012) showed that an approximate relationship should exist between the photometric and RV signatures of active regions. Specifically, if $F(t)$ is the photometric signature and $F'(t)$ is its time derivative, the RV variations are expected to scale as $AF(t)F'(t) + BF^2(t)$, where the first term corresponds to spots and the second to faculae (assuming that the two are generally colocated, although the latter covers a larger area), and $A$ and $B$ are tunable free parameters. In its original form, however, this method was of limited applicability, as it requires the RV observations to be quasi-simultaneous with high-precision, tightly sampled light curves, which is the case only exceptionally (see, e.g., Haywood et al. 2014).

However, RV extraction pipelines routinely provide additional indicators measuring the characteristic width and asymmetry of the spectral lines, which are also affected by active regions, as well as spectral activity indicators that trace chromospheric emission in the cores of strong lines. Rajpaul et al. (2015) developed a GP framework, schematically illustrated in **Figure 8**, to

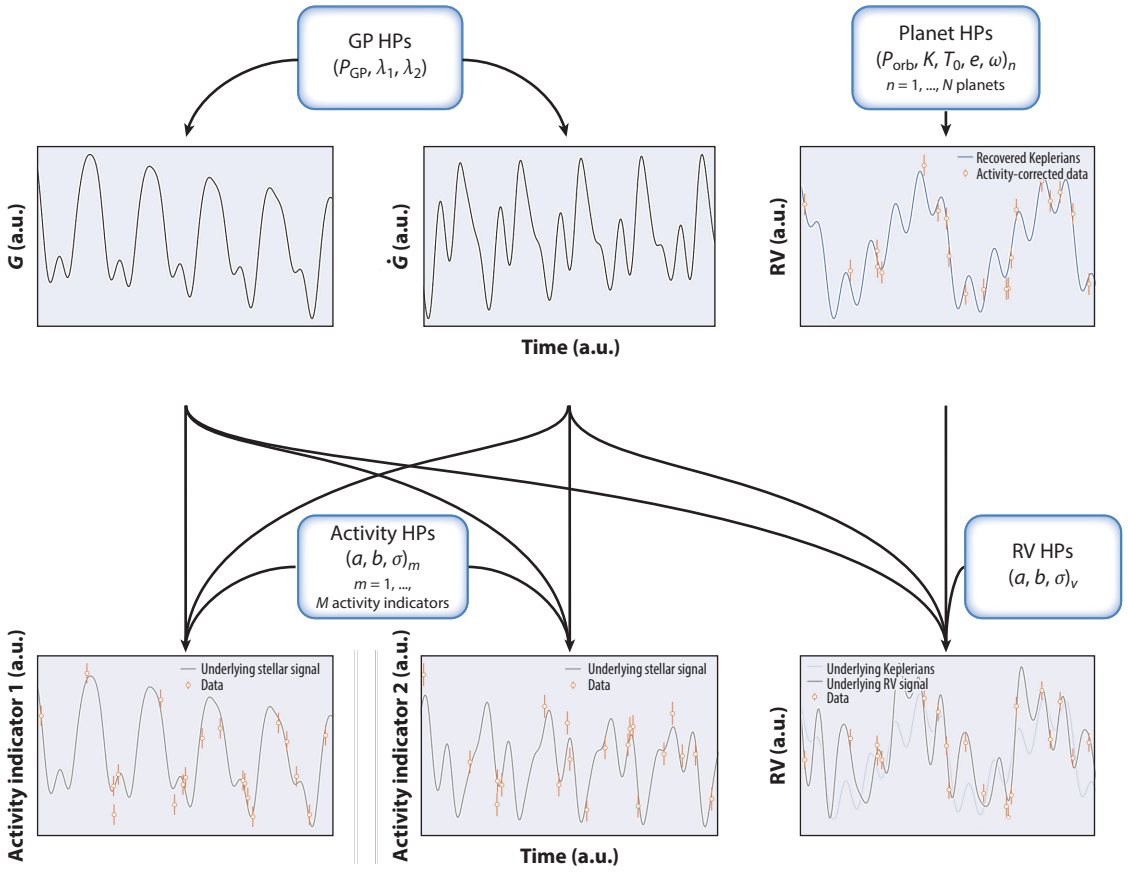


**Figure 8**

Schematic of the multidimensional GP framework of Rajpaul et al. (2015) for modeling activity signals in RV data. The observed RVs (shown as *orange points with error bars* in the *bottom right* panel) are modeled alongside observations of activity indicators extracted from the spectra or line profiles (*orange points with error bars* in the *bottom left* and *bottom middle* panels). In all these time series, the activity signal (shown by the *black line* in each panel) is modeled as a linear combination of an unobserved (latent) GP, shown in the top left panel, and its first time derivative, shown in the top middle panel. In addition, the Keplerian signal(s) of putative planet(s) (the sum of which is shown by the *blue lines* in the *top right* and *bottom right* panels) contribute to the RVs but not to the activity indicators. This hierarchical model is implemented in a Bayesian framework, sampling over all the hyperparameters listed in the blue boxes (subject to appropriate priors) to derive posteriors over the hyperparameters of interest and/or estimate the evidence for models with different numbers of planets. Models with different numbers of planets (or without planets) are then compared using the methods discussed in Section 3.3. Abbreviations: GP, Gaussian process; HP, hyperparameter; RV, radial velocity.

model these activity indicators alongside the RVs and to disentangle them from planetary signals. Each observed time series is modeled as a linear combination of some latent variable $G(t)$, its time derivative $G'(t)$, white noise, and (in the case of the RVs only) one or more Keplerian signals. $G(t)$ does not have a direct physical interpretation; it loosely corresponds to $F^2$ in the $FF'$ framework of Aigrain et al. (2012), so that $G'(t) \propto f(t)F'(t)$, but it is modeled as a (typically QP) GP, which makes it possible to write down expressions for the covariance between any pair of observations from any of the time series included in the analysis. This results in a global covariance matrix of dimensions ($NM$, $NM$), where $N$ is the number of observations and $M$ the number of activity indicators included in the analysis (including the RVs themselves). This framework has proved

particularly successful for RV confirmation of transiting planets around young stars (Barragán et al. 2019, Zicher et al. 2022), enabling the detection of planetary signals almost 50 times smaller than the activity signals. Open-source implementations of this framework have been published (Barragán et al. 2022, Delisle et al. 2022), and extensions thereof are likely to play a significant role in the analysis of next-generation RV surveys targeting Earth analogs, which need to achieve even better contrast between activity and planetary signals. Camacho et al. (2022) proposed a GPR network model for stellar activity signals in RV, showing that it can be used to model RVs of the Sun-as-a-star simultaneously with other activity indicators.

**4.1.6. Granulation and asteroseismology.** As discussed in Section 3.1.4, asteroseismology is an area in which physically motivated GP models arise quite naturally. This has motivated studies in which GPs have been used to model stellar oscillations in the time domain (Grunblatt et al. 2017, Farr et al. 2018), but these remain very much an exception to the much more widespread practice of transforming the data into the Fourier domain and identifying and modeling oscillation modes therein. Using a GP in this context makes particular sense when the data are irregularly sampled or inhomogeneous, or when the GP is combined with a mean function that is much easier to specify in the time domain (for example, a planetary transit). For example, the PLATO mission (Rauer et al. 2014) aims to characterize most of the host stars of the transiting planets it will discover up to magnitude $V \sim 11$ through asteroseismology (except for late K and M stars for which the oscillation amplitudes are too small and the frequencies too high). Current plans for the PLATO pipeline involve either subtracting the best-fit transit model from the light curve or cutting and interpolating over the transits, before performing the seismic analysis. However, a GP model would enable a simultaneous analysis, which should in principle be more robust. Conversely, Barros et al. (2020) demonstrated the importance of explicitly accounting for short-term stellar variability signals, including both granulation and oscillations, to obtain robust and accurate values of the transiting planets' parameters, and they compared a range of GP models with different kernels for this purpose.

## 4.2. Active Galactic Nucleus Variability

The AGNs at the center of many distant galaxies produce stochastic variability in their optical and radio emission on timescales ranging from hours to years. GPs can be useful descriptive models for these processes, and research in this domain has driven the development of new methods for GPR with large data sets. In particular, as discussed in Section 1, a GP was used by Press et al. (1992a) to model the underlying variability of gravitationally lensed quasar 0957+561 to measure its time delay, in one of the first explicit uses of GPR in the refereed astronomical literature.

When the quasar is lensed by an intervening galaxy or galaxy cluster, multiple images can be formed, whose brightnesses can be monitored individually. Measurements of the time delays between the resulting light curves can be used to constrain the Hubble constant, $H_0$ (Blandford & Narayan 1992). Early measurements of these time delays in the decade following the discovery of the first multiply imaged quasar, the double quasar 0957+561, were hampered by the lack of a generative model for the light curves. Heuristic methods based on interpolation and cross correlation were developed, but uncertainties were difficult to derive, and results obtained with different instruments and different bandpasses were not always consistent. Press et al. (1992a) derived a simple $\chi^2$ metric under the assumption that the observed light curves are shifted, noisy versions of a single sample from a GP with known covariance. Having proposed an empirical procedure to estimate the covariance function from the data, they then optimized the $\chi^2$ with respect to the time delay between images and applied standard methods to derive an uncertainty. Although their method relied on an ad hoc method to estimate the covariance matrix, and did not propagate

the corresponding errors into the final result, it did resolve the prior discrepancy between the estimates of the time delay from optical and radio data sets (Press et al. 1992b). Even then, the requirement to invert the covariance matrix was a barrier to the wider application of this method. This was addressed a few years later by Rybicki & Press (1995), who introduced a fast algorithm to invert the covariance matrix for exponential kernels. Updated versions of this algorithm now form the basis for fast GP solvers discussed in Section 5.1. A closely related application domain in which GPR is used to model stochastic AGN variability in order to infer a time delay is reverberation mapping (see, e.g., Zu et al. 2011). In this case, the time delay of interest is between the continuum (which traces the emission from the central engine) and emission lines (which trace emission by nearby gas clouds that are photoionized by the continuum radiation).

GPR is also used to study the population-level properties of AGN samples by characterizing and interpreting the statistics of the light curve variability for large samples. The idea is that the properties of the stochastic variations of an AGN light curve are driven by an accretion disk near the central supermassive black hole and that this variability provides an observational probe of the small scale astrophysical processes. The standard approach is to fit a GP model to the light curve—typically using a damped random walk kernel (e.g., Kozłowski et al. 2010; MacLeod et al. 2010, 2012), or the more general CARMA kernel (e.g., Kelly et al. 2014, Zhang et al. 2022, Yu et al. 2022), because these permit efficient computations even with large data sets—and then to use the inferred hyperparameter values to constrain the variability amplitudes and timescales of the system (e.g., Kozłowski et al. 2010; MacLeod et al. 2010, 2012; Kelly et al. 2014; Kasliwal et al. 2017; Moreno et al. 2019; Stone et al. 2022; Zhang et al. 2022; Yu et al. 2022). These amplitudes and timescales have been shown to empirically correlate with the fundamental properties of the AGN system, such as the mass of the central black hole, Eddington ratio, and bolometric luminosity.

Driven by the success of this approach, and the scale of upcoming survey data sets from, for example, the Vera C. Rubin Observatory and the LSST survey, there has been a concerted effort within this field to develop GP methods that can scale to large data sets. This began, perhaps, with Rybicki & Press (1995) and the development of a fast GP solver for exponential kernels. Although widely used, this kernel is very restrictive, and more flexible models are required to capture the full range of variability observed in AGN light curves. To this end, Kelly et al. (2014) developed a scalable GP solver for the CARMA kernel, which uses a Kalman filtering approach to achieve linear scaling of the computational cost with the number of data points, which we discuss further in Section 5.1.

### 4.3. Compact Objects, Gravitational Waves, and Transients

GPs have also been used for the analysis of a broad range of other astrophysical objects, including pulsars, gravitational wave sources, and other transients. In this section, we highlight some of these remaining applications, in somewhat less detail than the other applications because we have less familiarity with these topics. In many of these domains, it is standard practice to work in the Fourier domain by computing the fast Fourier transform (FFT) of the data and then assume that the noise process is uncorrelated between frequency bins (e.g., Romano & Cornish 2017). In other words, the data covariance matrix in the Fourier domain is assumed to be diagonal, and this significantly reduces the computational cost of likelihood evaluations compared to a GP model with a dense covariance matrix. This is a reasonable assumption for data sets with long observational baselines relative to the signals of interest and stationary noise processes (e.g., Unser 1984). Although these requirements are not met by many of the analyses that we have discussed, large data sets from pulsar timing surveys and gravitational wave interferometers do often satisfy these constraints, or they can be transformed in such a way that they do.

**4.3.1. Pulsar timing.** GPs are widely used in the analysis of pulsar observations to model timing residuals, pulse shape variations, and the stochastic signals of interest. In early work, van Haasteren & Vallisneri (2014) describe a comprehensive set of GP models, and computational techniques for their implementation, for pulsar timing studies aimed at detecting gravitational waves. These models include the effects of timing noise, dispersion-measure variations, and the stochastic gravitational wave background. Such GP models are now routinely used for pulsar timing studies (e.g., Parthasarathy et al. 2019, Champion et al. 2020, Antoniadis et al. 2022).

GPs have also been used to model pulse shape variations in decades-long pulsar observations (Brook et al. 2016, 2019; Shaw et al. 2022). In this case, the GP is used to model the pulses themselves and to track long-term changes in pulse periodicity and shape. This is done by evaluating (analytically) the first and second derivatives of the fitted GP and using those to find extrema and inflection points of the pulse profile along with the associated uncertainties. Finally, Posselt et al. (2021) used GPs to characterize the off-pulse noise background in observations of pulsars with large duty cycles and no obvious flat baseline.

**4.3.2. Gravitational waves.** Within the LIGO–Virgo project, GP models are used to model the detector noise (e.g., Abbott et al. 2020) and waveform modeling uncertainties (Moore et al. 2016). For these data sets, the power spectrum has been empirically calibrated, and it includes nontrivial structure at all frequencies. As discussed above, under some generally reasonable conditions, this noise process diagonalizes in the Fourier domain, and this empirically calibrated model can be used to compute the likelihood of a signal model also represented in the Fourier domain. This type of model becomes significantly more complicated for ring-down analyses in which edge effects invalidate the assumption that the data are periodic (Isi & Farr 2021).

**4.3.3. Transient classification.** In the context of transient light curve classification, GPs have been widely used as a tool for interpolating sparsely sampled and noisy photometric time series onto uniformly sampled grids that are more amenable to standard machine learning classification methods (e.g., Lochner et al. 2016, Boone 2019, Pruzhinskaya et al. 2019, Villar et al. 2020). These pipelines differ in how they treat the covariances between photometric bands, with most learning a correlation length in effective wavelength. Recently, Villar et al. (2021) proposed a wavelength covariance matrix that uses a distance metric defined by the throughput overlap between each pair of filters. So far in this domain, GPs have primarily been used in a preprocessing step to produce data correctly structured for standard machine learning methods, but recent developments that integrate GPs with deep learning frameworks could permit these GPs to be embedded within the larger classification pipeline.

**4.3.4. Quasi-periodic oscillations.** One very active area of research in which GPs have not been widely applied to date is the characterization of quasi-periodic oscillations (QPOs). QPOs are seen in astrophysical transients right across the electromagnetic spectrum, including gamma-ray bursts, X-ray binaries, flares in Sun-like stars, magnetized pulsars, and fast radio bursts. Most QPOs are believed to arise from the interaction between one or more compact objects and an accretion disk, but a detailed understanding of the origin of most classes of QPOs remains elusive. Their phenomenology is typically studied in the time domain, by fitting the power spectrum of individual, contiguous blocks of observations with empirical models, most frequently a sum of Lorentzian functions (see Ingram & Motta 2019, and references therein). Very recently, two studies (Yang et al. 2021, Hübner et al. 2022) used GPs to model QPOs directly in the time domain, with encouraging results for both the detection of QPOs in the presence of correlated noise and the characterization of their properties such as frequency.

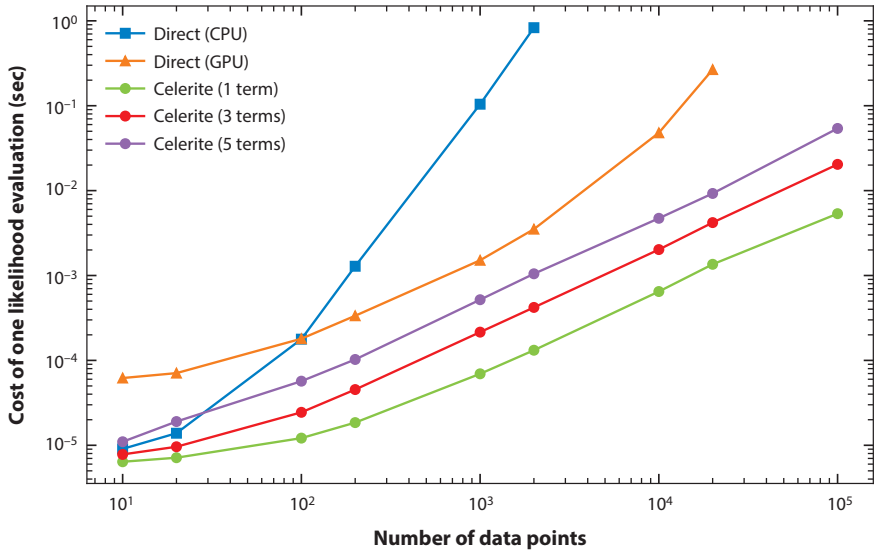


**Figure 9**

Computational cost of a single evaluation of the GP likelihood function (Equation 9) versus the size of the data set using different algorithms. In each case, the kernel function is a one-dimensional Matérn-3/2 with randomly generated data. The top two lines labeled "direct (CPU)" and "direct (GPU)" show the cost of evaluating Equation 9 directly using the Cholesky factorization of **K** provided by a tuned and optimized linear algebra library. The CPU solver was run using a single-threaded implementation, and the GPU version was executed on an A100 NVIDIA GPU. In both cases, the asymptotic scaling of the algorithm is approximately $\mathcal{O}(N^3)$ for $N$ data points. The three lines labeled "celerite ($J$ terms)" show the performance of the celerite algorithm (Foreman-Mackey et al. 2017; as implemented by the tinygp package) for a sum of $J$ Matérn-3/2 kernels, because the computational cost of that method scales with the complexity of the kernel. The asymptotic performance of each of these curves is linear in the number of data points, $\mathcal{O}(N)$.
Abbreviations: CPU, central processing unit; GP, Gaussian process; GPU, graphics processing unit.

## 5. CHALLENGES, PITFALLS, AND SOLUTIONS

So far, we have primarily discussed the benefits of GP models for astrophysics, but in this section we aim to highlight some challenges faced when using GPs in practice. The most significant challenge that has so far limited broader adoption of GPs for astrophysics is their computational cost and scaling, but as we discuss below, several techniques have been developed recently that overcome this limitation, in particular for one-dimensional data sets that are common in time-domain astronomy. We also briefly discuss some contexts in which GP models may overfit or be otherwise misspecified for specific problems.

### 5.1. Scalable Gaussian Process Inference

The core computations required for GPR are the likelihood function (Equation 9) and predictive distributions (Equation 13). All of these computations require the solution to a linear system, whose dimension scales with the number of data points $N$. Although this can be solved using optimized linear algebra libraries, possibly including acceleration with a GPU or other highly parallel hardware, the raw computational cost of evaluating a GP model typically scales as the cube of the number of data points $\mathcal{O}(N^3)$, as shown in **Figure 9**. This makes it intractable to use GPs for large data sets, practically restricting their use to data sets with less than a few hundred data points. General approaches have been developed for scaling GP models to larger data sets, and there are some methods that have seen significant impact within the astronomical literature.

Scalable methods for GPR typically fall into two broad—but sometimes conceptually overlapping—categories: (*a*) approximate, or (*b*) exact on a restricted function space. Approximate methods typically rely on algorithms from computational linear algebra to construct sparse, low-rank, or otherwise structured approximations to the covariance matrix that can be used to approximately solve the required systems to within a target numerical accuracy. However, there exist scalable exact methods that are restricted to a particular class of covariance functions or data with a specific structure (for example, one-dimensional evenly sampled data). Approximate methods (e.g., Quinonero-Candela & Rasmussen 2005, Ambikasaran et al. 2015, Wilson & Nickisch 2015, Wilson et al. 2015) have been widely used in the astrophysics literature thanks, in part, to the existence of publicly available software (e.g., GPy;[9] george, Ambikasaran et al. 2015; GPyTorch, Gardner et al. 2018). However, within time-domain astronomy, scalable exact methods have been more widely used, because time-series data often have more amenable structure.

One simple example of a scalable exact method is applicable to one-dimensional evenly sampled data, where the covariance matrix has "Toeplitz" structure (e.g., Cunningham et al. 2008), which permits more efficient computations using an FFT. The use of Toeplitz structure for scalable Gaussian noise modeling has primarily been applied for gravitational wave data analysis (e.g., Talbot & Thrane 2020, Isi & Farr 2021); however, most data sets in astrophysics are not evenly sampled, limiting the applicability of this method.

The most widely used algorithms for scalable exact GPR in time-domain astrophysics are based—often not explicitly—on SDEs (Särkkä & Solin 2019), and these typically require that the data be sortable. As an early example in the astrophysics literature, Rybicki & Press (1995) presented an algorithm for GPR with linear scaling with the size of the data set, when using an exponential kernel with unevenly sampled time-series data. This method was later generalized to a wider range of covariance functions (Ambikasaran 2015; Foreman-Mackey et al. 2017; Gordon et al. 2020; Delisle et al. 2020, 2022). The real-world performance of these generalized methods (as implemented by the tinygp package) is shown in **Figure 9** by the curves labeled "Celerite", and indexed by the number of "terms" (in this case Matérn-3/2 functions) in the kernel [see Foreman-Mackey et al. (2017) for a more detailed discussion]. The celerite[10] and S+LEAF[11] implementations of these methods have been widely used for a wide range of applications, including measuring stellar rotation periods (e.g., Gillen et al. 2020, Gordon et al. 2021, Newton et al. 2022), characterizing AGN light curves (e.g., Yu et al. 2022), and modeling the light curves of transiting planets (e.g., Kostov et al. 2019) and eclipsing binaries (e.g., Conroy et al. 2020). A qualitatively different, and independently developed, approach to implementing these models uses a state–space representation for the underlying SDE to enable scalable inference using a Kalman filter (Kelly et al. 2014, Jordán et al. 2021, Meyer et al. 2022). Of particular interest, though the core algorithms discussed here are inherently serial and cannot benefit from parallel hardware, this filtering approach to GPR has recently been formally parallelized (Särkkä & García-Fernández 2020).

## 5.2. Overfitting

In our experience, a common concern that astrophysicists have when working with GPs is that these flexible models may overfit the observations, and decrease the significance of their results, especially when a GP is used as an effective model for noise sources. Although this is not an unjustified concern, we find that the risk is generally overestimated. As discussed in Section 2.2.2,

[9] Available at **https://github.com/SheffieldML/GPy**.
[10] Available at **https://github.com/dfm/celerite**.
[11] Available at **https://obswww.unige.ch/~delisle/spleaf/doc**.

the determinant in Equation 9 penalizes GP models with unspecific hyperparameters, instead preferring to explain as much of the data as possible with the mean model. In other (but equally informal) terms, this means that a GP's modeling capacity will only be used to describe residuals that are not captured within the mean model's support.

All this being said, there do exist cases in which the use of a GP can lead to overfitting. One such example is RV planet searches when the ephemeris of the planet(s) is unknown (i.e., excluding transiting planets). Unless the time sampling of the data is truly excellent, the hyperparameters of the GP activity model are poorly constrained, and the highly multimodal nature of the posterior surface means that standard Bayesian inference and model comparison techniques can be misleading. One example of such a situation is the short-period, Earth-mass planet reported by Dumusque et al. (2012) around $\alpha$ Centauri B, which was shown by Rajpaul et al. (2016) to be an artifact of the time sampling: A small peak in the window function of the observations was serendipitously boosted by the complex model necessary to account for the star's activity signal. This is a concern with any flexible model applied to a sparse data set, but it can be less obvious for GPs, which can be very flexible with comparatively few hyperparameters. Useful strategies to deal with this kind of problem include cross validation (as discussed in Section 3.3) and performing exhaustive injection-recovery tests to build (or undermine, as the case may be!) confidence in one's results.

### 5.3. Model Misspecification

Like any other model for data, GPs are subject to the risk of model misspecification. A particularly common example of model misspecification in astronomy is the use of a Gaussian observation model when there are outliers in the data set with extremely large residuals. These outliers can be caused by a wide range of phenomena (e.g., instrumental artifacts, cosmic rays, etc.), but they cannot be properly modeled as generated by Gaussian noise. Model misspecification when faced with outliers does not uniquely affect GP models, and approaches have been developed to handle these problems in general. In the absence of correlated noise, a probabilistic approach that is commonly used in astrophysics is to build a two-component mixture model for the observations (e.g., Press 1997, Hogg et al. 2010), which can be used to infer which data points are outliers. This method does not generalize to GP models because when there is correlated noise, the observation model is no longer independent over data points: Every observation "cares" about whether or not every other measurement is an outlier, because we must compute the covariance between every pair of points.

In the astrophysics literature, the most commonly used approach to handling outliers in GP models is to perform an initial iterative sigma clipping procedure to identify and remove outliers (e.g., David et al. 2019). There are numerous sigma clipping procedures used in the literature, but a common approach involves the following steps: (1) fit the GP model (including the mean model) to the full data set using a numerical maximum likelihood procedure (Section 3.2), (2) compute the predictive distribution (Section 2.4) at the observed times, (3) remove data points with significantly low likelihood under this predictive distribution, (4) iterate from step 1 with this clipped data set until no more outliers are removed. This procedure can work well when the outliers are rare and extreme, but it may not be appropriate for all use cases.

More generally, outside of astronomy, there has been some work on developing a framework for Student's-*t* processes as generalizations of GPs, at no additional computational cost (e.g., Shah et al. 2014, Tracey & Wolpert 2018). A Student's-*t* distribution can be used to model processes with heavier tails than a GP, so a Student's-*t* process may be more robust to model misspecification. However, the Student's-*t* process in and of itself is not a good model for a data set with outliers because the underlying functions are still required to be smooth. Instead, a

more promising approach might be to use a GP model for the true function, combined with a per-observation independent Student's-*t* noise model (e.g., Vanhatalo et al. 2009). The practical application of this method is nontrivial and potentially extremely computationally expensive, because the marginal likelihood of such a model can no longer be evaluated in closed form.

Given the current state of the field, our current recommendation for assessing and mitigating model misspecification is to use the methods discussed in Section 3.3 to determine the suitability of a GP for a specific application. In the case in which there are significant outliers in the data set, the most mature and commonly used approach is to perform iterative sigma clipping on the data as a preprocessing step. As an alternative, Czekala et al. (2015) propose a method for handling nonstationary and structured model misspecification in the context of stellar spectral modeling by designing a basis of local kernel functions tuned to capture line shape residuals. Such domain-specific methods can be developed and validated using the methods discussed in Section 3.3.

## 6. OPEN-SOURCE GAUSSIAN PROCESS SOFTWARE

It is reasonably straightforward to implement a simple GP model in code, and within astrophysics it has been common for authors to implement custom GPs for their analyses. However, things get significantly more complicated when implementing the scalable or approximate methods discussed in the previous section. Similarly, it can be tedious to experiment with different kernel functions and inference methods without building a nontrivial modeling infrastructure, something that has typically been ad hoc in astrophysics research.

Luckily, driven by the immense popularity of GPR in machine learning, the physical sciences, and other fields, there are a plethora of open-source tools that have been developed to simplify this process for a wide range of applications. Many of these tools are designed for scalability, flexibility, and ease of use. In this section, we describe some popular libraries in this space but caution the reader that this is not a comprehensive list, and this domain changes quickly, so the discussion may become outdated more quickly than the rest of this document. Most of our discussion focuses on tools implemented in the Python programming language because it is—at the time of writing—the most popular language, in both astronomy and GP modeling, but there are tools available in all other popular languages.

Alongside this review, we have released all the source code to generate the figures in this document using the tinygp library,[12] which is a relatively new open-source Python library for GP modeling. Although this library has not yet been widely used in the astrophysics literature, we believe[13] that it is a good choice for many applications. tinygp uses the JAX[14] library as its computational backend, giving it good computational performance (including support for GPU hardware) with a pure-Python user interface. JAX also provides support for automatic differentiation (e.g., Margossian 2019), which can be used to accelerate all the inference schemes described in Sections 3.2 and 3.3.

Perhaps the most widely used Python library for GP inference in astrophysics is george[15] (Ambikasaran et al. 2015), which is a relatively simple library that implements most of the standard kernel functions described in Section 3.1 and an approximate solver that can be used to scale to

[12] Available at **https://tinygp.readthedocs.io**; dfm/tinygp: tinygp v0.2.2, from D. Foreman-Mackey, S. Yadav, R. Tronsgaard, S. Schmerler, and T. Rashid.
[13] Note that the development of the tinygp library is led by D. Foreman-Mackey, an author of this review.
[14] Available at **https://jax.readthedocs.io**; composable transformations of Python+NumPy programs, from J. Bradbury, R. Frostig, P. Hawkins, et al.
[15] Available at **https://george.readthedocs.io**.

moderately large ($N \sim 1{,}000$) data sets. More recently, the scalable methods implemented by celerite[16] (Foreman-Mackey et al. 2017) and S+LEAF[17] (Delisle et al. 2020, 2022) have become popular for time-domain GP models in astrophysics. These libraries require one-dimensional—or appropriately structured (Gordon et al. 2020, Delisle et al. 2022)—input data, and a specific set of kernel functions, although these kernels form a complete (Loper et al. 2021) and physically interpretable (Foreman-Mackey et al. 2017) basis, and it has recently been demonstrated that the Matérn class of kernels can be exactly represented (Jordán et al. 2021).

Outside of astrophysics, the GPy,[18] GPyTorch[19] (Gardner et al. 2018), and GPFlow[20] (Matthews et al. 2017, van der Wilk et al. 2020) projects provide mature and widely used GP implementations. GPyTorch, in particular, is designed for scaling GP modeling to large data sets, using approximate methods (e.g., Wilson & Nickisch 2015, Wilson et al. 2015) that do not have the same restrictions as celerite and S+LEAF. GPyTorch has been used for a variety of applications in astrophysics (e.g., Dharmawardena et al. 2022, Li et al. 2022), but not yet as widely as the other libraries discussed so far, perhaps because it does not integrate into the astronomical workflow as easily as the projects that were designed with astrophysicists in mind.

## 7. CONCLUSIONS

GPR has become a key element in the toolkit of modern astronomers, particularly those working with time-domain data sets. It offers a principled approach to modeling stochastic signals and nuisances, particularly correlated noise, whose importance has grown significantly in the past decade with the development of major precision time-domain monitoring surveys (e.g., planetary transit surveys). GPR is explicitly rooted in probability theory, making it readily interpretable and enabling astronomers to quantify their confidence (or lack thereof) in the results of their analysis. Despite the simplicity of the underlying mathematical framework, even basic GP models are extremely flexible, and a small handful of commonly used building blocks have been used successfully to explain the observed variability of astronomical objects on all scales. In contrast, this flexibility comes at a price: A good fit with a particular model is no longer a fail-safe indication that this particular model is the correct one, so users must proceed with extra caution.

### 7.1. Summary of the Review

After a very brief historical introduction, we started this review by introducing two simulated examples, which demonstrate the two main classes of use for GPR in astronomy, namely to account for a nuisance signal (in the case of a planetary transit) or to model a stochastic signal of interest (in the quasar time-delay example). The source code for these and other worked examples in the review is available alongside this review in the form of Jupyter notebooks that can be accessed via the links provided in footnote 1, and interested users are encouraged to use these to further their understanding of the methodology and potentially adapt the code to their own data sets. Next, we gave a succinct introduction to the key equations underlying GPR and sketched out a typical workflow for applying GPR to a new data set. Like any data analysis method, GPR requires the user to make a number of choices, the most important of which is the choice of covariance function. We listed some of the most widely used covariance functions, outlined how to construct

[16] Available at **https://github.com/dfm/celerite**.
[17] Available at **https://obswww.unige.ch/~delisle/spleaf/doc**.
[18] Available at **https://github.com/SheffieldML/GPy**.
[19] Available at **https://gpytorch.ai**.
[20] Available at **https://www.gpflow.org**.

more complex ones, and gave practical advice on how to assess the validity of one's modeling choices. These strategies include standard Bayesian model comparison but also cross validation, which can be more appropriate in cases in which predictive power is important.

We then reviewed the growing literature on applications of GPR to time-domain data sets, from exoplanet detection and characterization, where GPs are mostly used to account for instrumental or stellar nuisance signals, to the characterization of stellar variability itself, where the GP is now used to model a stochastic signal for its own sake. These examples are drawn from the exoplanet and stellar astrophysics literature, which we are most familiar with, but we also attempted to give an overview of recent applications of GPR to two other classes of astronomical objects where it is increasingly important, namely AGNs and compact objects. Altogether, these applications include a very diverse array of GP models with varying levels of complexity, including multidimensional inputs and outputs, derivative and integral observations, and applications in which the GP model is physically motivated versus cases in which it is merely an effective model. We noted how GPR can be used to tease apart the deterministic and stochastic components of a model but also different components of a stochastic process.

We discussed the challenges and pitfalls of GPR, noting recent algorithmic and technical developments that have brought down its computational cost, which was once prohibitive for large data sets, and outlining some strategies to limit the risk of overfitting. Nonetheless, the size of the data set remains an important consideration when deciding whether or not to use GPR for a particular problem. If too small, learning the kernel hyperparameters and flagging outliers can be difficult, whereas computing time can still be problematic for certain types of kernels and/or computations on very large data sets. Finally, we reviewed the ecosystem of free software resources for GPR, which make it easy to test a GP model for a new problem in just a few lines of code, and to implement a computationally efficient version suitable for a large data set with minimal additional effort.

## 7.2. Future Perspectives

In certain application domains, such as the analysis of photometric and RV time series from exoplanet surveys, GPR is already used fairly widely. Its usefulness as an effective model for astrophysical or instrumental nuisance signals is well established, and it is seen as a fairly standard element of the astronomer's data analysis toolkit. In these existing applications, we foresee three main routes for further progress in the near future. One is the development of more physically motivated (as opposed to merely effective) GP models for stellar variability, so that GPR can be used to characterize and understand these stellar signals for their own sake. The other is a more systematic use of GPR as part of the detrending and detection pipelines for exoplanet transit and RV surveys. Computational cost considerations would have made this impractical just a few years ago, but this obstacle has been lifted thanks to the recent development of scalable GP models. Finally, as *the James Webb Space Telescope* (Gardner et al. 2006) begins to deliver numerous exoplanet atmosphere observations with much higher SNR (including, importantly, more precise and relevant housekeeping parameters), a new generation of systematics models will no doubt be developed to get the most out of these observations, and it seems likely that GPR would have a role to play in that.

In other applications, such as pulsars, QPOs and astrophysical transients, GPR remains more of a niche pursuit so far, but promising early results should motivate further work in the next few years. The next decade will see a vast increase in the scale of the relevant data sets, with major surveys such as the Vera C. Rubin Observatory's LSST or the Square Kilometer Array (SKA) routinely delivering tens of terabytes per night. Even with the recent development of scalable GP models, it is unclear whether GPR can be useful as a first-contact analysis tool for these very large surveys, but it has clear potential for the analysis of specific subsets of data.

One area of active development concerns the problem of non-Gaussianity. Although the assumption of Gaussianity implicit to GPR is often good enough in practice, it is almost never strictly correct, and this can become problematic when extreme precision is required on the model hyperparameters or when the data contain (a relatively small proportion of) significant outliers. Stochastic process models can be based on distributions with higher kurtosis than the Gaussian; for example, Student's-*t* processes are an interesting alternative, though they have not been extensively used in astronomy.

Another area where GPR could become more widely used in the future is for intelligent observation planning: If a set of variable objects is being monitored as part of a survey, using a GP to model existing observations would enable the observer to make predictions (including uncertainties) for future observations and, therefore, to prioritize among the target list in real time according to some metric of choice.

GPs are an extremely powerful and general data analysis method, which explains their rapidly growing popularity in a wide range of application domains, including time-domain astronomy. However, their very flexibility, together with their built-in property of automatic marginalization over random functions, can make them appear arcane, occasionally leading to lukewarm reception from the wider community. We hope this review has demystified the subject somewhat and provided our readers with enough information to make informed decisions about when and how to use GPs for their data sets.

## DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

## ACKNOWLEDGMENTS

S.A. acknowledges support from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant agreement No. 865624), and useful discussions with Steve Roberts, Mike Osborne, Oscar Barragán, Vinesh Rajpaul, and Aris Karastergiou.

D.F.M. acknowledges useful conversations with Eric Agol, Will Farr, Alex Gagliano, Tyler Gordon, and Maximiliano Isi.

The authors thank the community members who sent feedback on the public draft of this review: Eric Agol, Nestor Espinoza, Christopher Kochanek, Konstantin Malanchev, Maria Pruzhinskaya, Steve Roberts, Fergus Simpson, Joshua Winn, Dahai Yan, and Ying Zu.

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

## LITERATURE CITED

Abbott BP, Abbott R, Abbott TD, et al. 2020. *Class. Quantum Gravity* 37(5):055002
Agol E, Luger R, Foreman-Mackey D. 2020. *Astron. J.* 159(3):123
Aigrain S, Hodgkin ST, Irwin MJ, Lewis JR, Roberts SJ. 2015. *MNRAS* 447(3):2880–93
Aigrain S, Parviainen H, Pope BJS. 2016. *MNRAS* 459(3):2408–19
Aigrain S, Pont F, Zucker S. 2012. *MNRAS* 419(4):3147–58
Albert JG. 2020. arXiv:2012.15286
Almosallam IA, Jarvis MJ, Roberts SJ. 2016. *MNRAS* 462:726–39
Ambikasaran S. 2015. *Numer. Linear Algebra Appl.* 22(6):1102–14
Ambikasaran S, Foreman-Mackey D, Greengard L, Hogg DW, O'Neil M. 2015. *IEEE Trans. Pattern Anal. Mach. Intell.* 38:252

Angus R, Morton T, Aigrain S, Foreman-Mackey D, Rajpaul V. 2018. *MNRAS* 474(2):2094–108
Antoniadis J, Arzoumanian Z, Babak S, et al. 2022. *MNRAS* 510(4):4873–87
Baglin A, Auvergne M, Barge P, et al. 2006. In *Proceedings of The CoRoT Mission Pre-Launch Status—Stellar Seismology and Planet Finding (ESA SP-1306)*, ed. M Fridlund, A Baglin, J Lochard, L Conroy, pp. 33–38. Paris: Eur. Space Agency
Bailer-Jones CAL. 2012. *Astron. Astrophys.* 546:A89
Barclay T, Endl M, Huber D, et al. 2015. *Ap. J.* 800:46
Barnes JA, Sargent HH III, Tryon PV. 1980. In *Proceedings of the Conference on the Ancient Sun: Fossil Record in the Earth, Moon and Meteorites*, ed. RO Pepin, JA Eddy, RB Merrill, pp. 159–63. New York/Oxford: Pergamon
Barragán O, Aigrain S, Kubyshkina D, et al. 2019. *MNRAS* 490:698–708
Barragán O, Aigrain S, Rajpaul VM, Zicher N. 2022. *MNRAS* 509:866–83
Barros SCC, Demangeon O, Díaz RF, et al. 2020. *Astron. Astrophys.* 634:A75
Benz W, Broeg C, Fortier A, et al. 2021. *Exp. Astron.* 51:109–51
Blandford RD, Narayan R. 1992. *Annu. Rev. Astron. Astrophys.* 30:311–58
Boone K. 2019. *Astron. J.* 158(6):257
Borucki WJ, Koch D, Basri G, et al. 2010. *Science* 327(5968):977–80
Brook PR, Karastergiou A, Johnston S, et al. 2016. *MNRAS* 456(2):1374–93
Brook PR, Karastergiou A, Johnston S. 2019. *MNRAS* 488(4):5702–12
Camacho JD, Faria JP, Viana PTP. 2022. *MNRAS* 519(4):5439–53
Carter JA, Winn JN. 2009. *Ap. J.* 704:51–67
Champion D, Cognard I, Cruces M, et al. 2020. *MNRAS* 498(4):6044–56
Charbonneau D, Brown TM, Latham DW, Mayor M. 2000. *Ap. J. Lett.* 529:L45–48
Claeskens G, Hjort NL. 2008. *Model Selection and Model Averaging*. Cambridge Ser. Statistical Probabilistic Math. Cambridge, UK: Cambridge Univ. Press
Conroy KE, Kochoska A, Hey D, Pablo H, Hambleton KM, et al. 2020. *Ap. J. Suppl.* 250(2):34
Constable CG, Parker RL. 1988. *J. Geophys. Res.* 93(B10):11569–81
Crossfield IJM, Ciardi DR, Petigura EA, et al. 2016. *Ap. J. Suppl.* 226:7
Cunningham JP, Shenoy KV, Sahani M. 2008. In *Proc. 25th Intl. Conf. Mach. Learn., Helsinki, Finland*, ed. A McCallum, S Roweis, pp. 192–99. **https://icml.cc/Conferences/2008/papers/icml2008proceedings.pdf**
Czekala I, Andrews SM, Mandel KS, Hogg DW, Green GM. 2015. *Ap. J.* 812(2):128
David TJ, Petigura EA, Luger R, et al. 2019. *Ap. J. Lett.* 885:L12
Delisle JB, Hara N, Ségransan D. 2020. *Astron. Astrophys.* 638:A95
Delisle JB, Unger N, Hara NC, Ségransan D. 2022. *Astron. Astrophys.* 659:A182
Deming D, Knutson H, Kammer J, et al. 2015. *Ap. J.* 805(2):132
Dharmawardena TE, Bailer-Jones CAL, Fouesneau M, Foreman-Mackey D. 2022. *Astron. Astrophys.* 658:A166
Dumusque X, Pepe F, Lovis C, et al. 2012. *Nature* 491(7423):207–11
Dvorak R, Edelman C. 1976. *Mitt. Astron. Ges. Hambg.* 38:192
Evans TM, Aigrain S, Gibson N, et al. 2015. *MNRAS* 451:680–94
Evans TM, Pont F, Sing DK, et al. 2013. *Ap. J. Lett.* 772(2):L16
Evans TM, Sing DK, Goyal JM, et al. 2018. *Astron. J.* 156(6):283
Evans TM, Sing DK, Kataria T, et al. 2017. *Nature* 548(7665):58–61
Farr WM, Pope BJS, Davies GR, et al. 2018. *Ap. J. Lett.* 865(2):L20
Foreman-Mackey D, Agol E, Ambikasaran S, Angus R. 2017. *Astron. J.* 154(6):220
Foreman-Mackey D, Montet BT, Hogg DW, et al. 2015. *Ap. J.* 806(2):215
Gabry J, Simpson D, Vehtari A, Betancourt M, Gelman A. 2019. *J. R. Stat. Soc.: Ser. A* 182(2):389–402
Gardner JP, Mather JC, Clampin M, et al. 2006. *Space Sci. Rev.* 123(4):485–606
Gardner JR, Pleiss G, Bindel D, Weinberger KQ, Wilson AG. 2018. In *Proc. 32nd Int. Conf. Neural Inf. Process. Syst., 2018*, ed. S Bengio, pp. 7587–97. Red Hook, NY: Curran Assoc.
Gelfand AE, Dey DK, Chang H. 1992. *Model determination using predictive distributions with implementation via sampling-based methods*. Tech. Rep., Stanford Univ CA Dept of Statistics

Gelman A, Carlin JB, Stern HS, Rubin DB. 1995. *Bayesian Data Analysis*. Boca Raton, FL: Chapman and Hall/CRC
Gelman A, Meng XL, Stern H. 1996. *Stat. Sinica* 6:733–60
Gibson NP. 2014. *MNRAS* 445(4):3401–14
Gibson NP, Aigrain S, Roberts S, et al. 2012. *MNRAS* 419(3):2683–94
Gibson NP, Pont F, Aigrain S. 2011. *MNRAS* 411(4):2199–213
Gillen E, Briegal JT, Hodgkin ST, et al. 2020. *MNRAS* 492:1008–24
Gillon M, Pont F, Demory BO, et al. 2007. *Astron. Astrophys.* 472(2):L13–16
Gordon TA, Agol E, Foreman-Mackey D. 2020. *Astron. J.* 160(5):240
Gordon TA, Davenport JRA, Angus R, et al. 2021. *Ap. J.* 913:70
Grunblatt SK, Huber D, Gaidos E, et al. 2017. *Astron. J.* 154(6):254
Haywood RD, Collier Cameron A, Queloz D, et al. 2014. *MNRAS* 443(3):2517–31
Henry GW, Marcy GW, Butler RP, Vogt SS. 2000. *Ap. J. Lett.* 529:L41–44
Hippke M, David TJ, Mulders GD, Heller R. 2019. *Astron. J.* 158(4):143
Hoffman MD, Gelman A. 2014. *J. Mach. Learn. Res.* 15(47):1593–623
Hogg DW, Bovy J, Lang D. 2010. arXiv:1008.4686
Hogg DW, Foreman-Mackey D. 2018. *Ap. J. Suppl.* 236:11
Hogg DW, Villar S. 2021. *Publ. Astron. Soc. Pac.* 133(1027):093001
Holman MJ, Murray NW. 2005. *Science* 307(5713):1288–91
Horne K. 2003. In *Scientific Frontiers in Research on Extrasolar Planets*, Vol. 294, *Astron. Soc. Pac. Conf. Ser.*, ed. D Deming, S Seager, pp. 361–70. San Francisco: ASP
Howell SB, Sobeck C, Haas M, et al. 2014. *Publ. Astron. Soc. Pac.* 126(938):398–408
Hu Z, Tak H. 2020. *Astron. J.* 160(6):265
Hübner M, Huppenkothen D, Lasky PD, et al. 2022. *Ap. J.* 936:17
Ingram A, Motta S. 2019. *New Astron. Rev.* 85:101524
Isi M, Farr WM. 2021. arXiv:2107.05609
Jekeli C. 1991. *Manuscr. Geod.* 16(5):313–25
Jordán A, Eyheramendy S, Buchner J. 2021. *Res. Notes Am. Astron. Soc.* 5(5):107
Kasliwal VP, Vogeley MS, Richards GT. 2017. *MNRAS* 470(3):3027–48
Keeling CD, Whorf TP. 2004. *Atmospheric $CO_2$ Records from Sites in the SIO Air Sampling Network*. In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, US Department of Energy, Oak Ridge. **https://data.ess-dive.lbl.gov/datasets/doi:10.3334/CDIAC/ATG.012**
Kelly BC, Bechtold J, Siemiginowska A. 2009. *Ap. J.* 698:895–910
Kelly BC, Becker AC, Sobolewska M, Siemiginowska A, Uttley P. 2014. *Ap. J.* 788:33
Kostov VB, Schlieder JE, Barclay T, et al. 2019. *Astron. J.* 158:32
Kozłowski S, Kochanek CS, Udalski A, et al. 2010. *Ap. J.* 708(2):927–45
Kreidberg L. 2018. In *Handbook of Exoplanets*, ed. HJ Deeg, JA Belmonte, pp. 2083–105. Cham, Switz.: Springer
Krige DG. 1951. *A statistical approach to some mine valuations and allied problems at the Witwatersrand*. Master's Thesis, Univ. Witwatersrand, Johannesburg
Kundić T, Colley WN, Gott J. Richard I, et al. 1995. *Ap. J. Lett.* 455:L5–8
Kundić T, Turner EL, Colley WN, et al. 1997. *Ap. J.* 482:75–82
Li T, Davies GR, Lyttle AJ, et al. 2022. *MNRAS* 511(4):5597–610
Lochner M, McEwen JD, Peiris HV, Lahav O, Winter MK. 2016. *Ap. J. Suppl.* 225(2):31
Loper J, Blei D, Cunningham JP, Paninski L. 2021. *J. Mach. Learn. Res.* 22(234):1–36
Luger R, Agol E, Kruse E, et al. 2016. *Astron. J.* 152(4):100
Luger R, Foreman-Mackey D, Hedges C. 2021a. *Astron. J.* 162(3):124
Luger R, Foreman-Mackey D, Hedges C, Hogg DW. 2021b. *Astron. J.* 162(3):123
Luger R, Kruse E, Foreman-Mackey D, Agol E, Saunders N. 2018. *Astron. J.* 156(3):99
MacLeod CL, Ivezić Ž, Kochanek CS, et al. 2010. *Ap. J.* 721(2):1014–33
MacLeod CL, Ivezić Ž, Sesar B, et al. 2012. *Ap. J.* 753(2):106
Mandel K, Agol E. 2002. *Ap. J. Lett.* 580(2):L171–75

Margossian CC. 2019. *Wiley Interdiscip. Rev.: Data Min. Knowledge Discov.* 9(4):e1305
Matthews AGdG, van der Wilk M, Nickson T, et al. 2017. *J. Mach. Learn. Res.* 18(40):1–6
Meunier N, Desort M, Lagrange AM. 2010. *Astron. Astrophys.* 512:A39
Meyer AD, van Dyk DA, Tak H, Siemiginowska A. 2022. arXiv:2207.09327
Miller AC, Anderson L, Leistedt B, et al. 2022. arXiv:2202.06797
Moore CJ, Berry CPL, Chua AJK, Gair JR. 2016. *Phys. Rev. D* 93(6):064001
Moreno J, Vogeley MS, Richards GT, Yu W. 2019. *Publ. Astron. Soc. Pac.* 131(1000):063001
Newton ER, Rampalli R, Kraus AL, et al. 2022. *Astron. J.* 164(3):115
Nicholson BA, Aigrain S. 2022. *MNRAS* 515:5251–66
Nocedal J, Wright SJ. 1999. *Numerical Optimization*. New York: Springer
Parthasarathy A, Shannon RM, Johnston S, et al. 2019. *MNRAS* 489(3):3810–26
Peebles PJE. 1997. *Ap. J. Lett.* 483:L1–4
Phan D, Pradhan N, Jankowiak M. 2019. arXiv:1912.11554
Pont F, Zucker S, Queloz D. 2006. *MNRAS* 373:231–42
Posselt B, Karastergiou A, Johnston S, et al. 2021. *MNRAS* 508(3):4249–68
Press WH. 1997. *Understanding data better with Bayesian and global statistical methods*. Paper presented at Unsolved Problems in Astrophysics Conference, Princeton, April 1995. astro-ph/9604126
Press WH, Rybicki GB, Hewitt JN. 1992a. *Ap. J.* 385:404–15
Press WH, Rybicki GB, Hewitt JN. 1992b. *Ap. J.* 385:416–20
Pruzhinskaya MV, Malanchev KL, Kornilov MV, et al. 2019. *MNRAS* 489(3):3591–608
Quinonero-Candela J, Rasmussen CE. 2005. *J. Mach. Learn. Res.* 6:1939–59
Rajpaul V, Aigrain S, Osborne MA, Reece S, Roberts S. 2015. *MNRAS* 452(3):2269–91
Rajpaul V, Aigrain S, Roberts S. 2016. *MNRAS* 456:L6–10
Rasmussen CE, Williams CKI. 2006. *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press
Rauer H, Catala C, Aerts C, et al. 2014. *Exp. Astron.* 38(1–2):249–330
Ricker GR, Winn JN, Vanderspek R, et al. 2015. *J. Astron. Telesc. Instrum. Syst.* 1:014003
Romano JD, Cornish NJ. 2017. *Living Rev. Relativ.* 20:2
Rybicki GB, Press WH. 1995. *Phys. Rev. Lett.* 74(7):1060–63
Särkkä S, García-Fernández ÁF. 2020. *IEEE Trans. Autom. Control* 66:299–306
Särkkä S, Solin A. 2019. *Applied Stochastic Differential Equations*, Vol. 10. Cambridge, UK: Cambridge Univ. Press
Seager S, Mallén-Ornelas G. 2003. *Ap. J.* 585(2):1038–55
Seager S, Sasselov DD. 2000. *Ap. J.* 537(2):916–21
Shah A, Wilson A, Ghahramani Z. 2014. In *Proceedings of the Seventeenth Conference on Artificial Intelligence and Statistics*, Vol. 33, *Proc. Mach. Learn. Res.*, ed. S Kaski, J Corander, pp. 877–85. Reykjavik, Icel.: PMLR
Shaw B, Stappers BW, Weltevrede P, et al. 2022. *MNRAS* 513(4):5861–80
Skilling J. 2006. *Bayesian Anal.* 1(4):833–59
Sliski DH, Kipping DM. 2014. *Ap. J.* 788(2):148
Stone Z, Shen Y, Burke CJ, et al. 2022. *MNRAS* 514:164–84
Talbot C, Thrane E. 2020. *Phys. Rev. Res.* 2(4):043298
Tracey BD, Wolpert D. 2018. In *2018 AIAA Non-Deterministic Approaches Conference*, *Kissimmee*, *FL*, *Jan. 8–12*, AIAA 2018-1659
Trotta R. 2008. *Contemp. Phys.* 49(2):71–104
Udalski A, Paczynski B, Zebrun K, et al. 2002. *Acta Astron.* 52:1–37
Uhlenbeck GE, Ornstein LS. 1930. *Phys. Rev.* 36(5):823–41
Unser M. 1984. *Signal Proc.* 7(3):231–49
van der Wilk M, Dutordoir V, John S, et al. 2020. arXiv:2003.01115
van Haasteren R, Vallisneri M. 2014. *Phys. Rev. D* 90(10):104012
Vanderburg A, Johnson JA. 2014. *Publ. Astron. Soc. Pac.* 126(944):948–58
Vanderriest C, Schneider J, Herpe G, et al. 1989. *Astron. Astrophys.* 215:1–13
Vanhatalo J, Jylänki P, Vehtari A. 2009. Gaussian process regression with Student-t likelihood. In *Advances in Neural Information Processing Systems*, Vol. 22, ed. Y Bengio, D Schuurmans, J Lafferty, C Williams, A Culotta, pp. 3227–57. Red Hook, NY: Curran Associates, Inc.

Villar VA, Cranmer M, Berger E, et al. 2021. *Ap. J. Suppl.* 255(2):24
Villar VA, Hosseinzadeh G, Berger E, et al. 2020. *Ap. J.* 905(2):94
Virtanen P, Gommers R, Oliphant TE, et al. 2020. *Nat. Methods* 17:261–72
von der Heide K. 1978. *Astron. Astrophys.* 70(6):777–84
Wakeford HR, Sing DK, Evans T, Deming D, Mandell A. 2016. *Ap. J.* 819:10
Way MJ, Srivastava AN. 2006. *Ap. J.* 647:102–15
Wilson A, Nickisch H. 2015. In *Proc. 32nd Int. Conf. Mach. Learn.*, pp. 1775–84. Lille, Fr.: PMLR
Wilson AG, Dann C, Nickisch H. 2015. arXiv:1511.01870
Yang S, Yan D, Zhang P, Dai B, Zhang L. 2021. *Ap. J.* 907(2):105
Yu W, Richards GT, Vogeley MS, Moreno J, Graham MJ. 2022. *Ap. J.* 936(2):132
Zhang H, Yan D, Zhang L. 2022. *Ap. J.* 930(2):157
Zicher N, Barragán O, Klein B, et al. 2022. *MNRAS* 512(2):3060–78
Zu Y, Kochanek CS, Peterson BM. 2011. *Ap. J.* 735(2):80

## RELATED RESOURCES

S.A. gave a set of lectures on GPR for time-domain astronomy as part of the 2021 *Saas Fee Advanced Course in Astrophysics*, for which PDF lecture notes are available at **shorturl.at/hrstz**, along with Python exercises in the form of Jupyter notebooks.