Gaussian Process regression for astronomical time-series

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

[1] Department of Physics, University of Oxford, Oxford, UK, OX1 3RH; email: suzanne.aigrain@physics.ox.ac.uk
[2] Center for Computational Astrophysics, Flatiron Institute, New York, USA, NY 10010

Xxxx. Xxx. Xxx. Xxx. 2022. AA:1–26

https://doi.org/10.1146/((please add article doi))

Compiled using show your work!

Keywords

keywords, separated by comma, no full stop, lowercase

Abstract

Abstract text, approximately 150 words.

Contents

1. INTRODUCTION 2
   1.1. Brief history 3
   1.2. Motivating examples 4
   1.3. Overview of this review 6
2. BASICS OF GAUSSIAN PROCESS REGRESSION 6
   2.1. Formal definition 6
   2.2. From least-squares to GPR 7
   2.3. Inference with a GP 9
   2.4. Making predictions 9
3. GAUSSIAN PROCESS MODELING DECISIONS 10
   3.1. Gaussian Process Covariance Functions 10
   3.2. Hyperparameter Inference 13
   3.3. Model assessment, validation, and selection 14
4. GAUSSIAN PROCESSES IN TIME-DOMAIN ASTRONOMY 15
   4.1. AGN variability (DFM) 15
   4.2. Stars and exoplanets 17
   4.3. Compact objects 23
5. CHALLENGES, PITFALLS, AND SOLUTIONS 23
   5.1. Fast GP solvers 23
6. OPEN-SOURCE GAUSSIAN PROCESS SOFTWARE 23
7. CONCLUSIONS AND FUTURE DIRECTIONS 23

GP: Gaussian Process

GPR: Gaussian Process Regression

1. INTRODUCTION

Gaussian Processes (GPs) are a powerful class of statistical models, which 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 modelling. While this may seem abstract, GPs have a wide range of applications from modelling of stochastic physical processes, to high dimensional interpolation and smoothing. In particular, GP Regression (GPR) has become increasingly popular in the astronomical community over the last decade. Part of the reason for this uptake is the growing availability and importance of time-domain datasets in astronomy. These systematically contain non-trivial random or unknown signals, whether astrophysical or instrumental, that need to be modelled. In many cases, these are nuisance signals, which we need to marginalise over in order to detect or measure other signals robustly. Sometimes, we are interested in the stochastic behaviour itself, and want to infer its characteristics or predict its behaviour. GPR offers a compelling solution: statistically principled, naturally Bayesian, and extremely flexible, yet mathematically simple. To new users, however, the lack of an explicit functional form for the model can make GPR can seem a little arcane. Furthermore, the computational cost of the method, which scales cubically with the dataset 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 optimised software packages.

Add some summary text here. “In this review...”

Should we mention we are using showyourwork?

1.1. Brief history

An early use of GPR, was for spatial interpolation in geophysics (Krige 1951), and GPR has since been adopted or re-invented in a wide range of other application domains. GPs were used in simulations in a wide range of astronomical sub-fields (see e.g. Barnes et al. 1980; Constable & Parker 1988; Peebles 1997), but early mentions of GPs for modelling astronomical datasets (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 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 refereeed astronomy and astrophysics journals with the words "Gaussian Process" in full text of the article Figure 1. After an increase in popularity throughout the 1990s, the use of GPs in astrophysics remained fairly constant 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 published in the astrophysics literature.

A number of important factors have contributed to recent the democratisation 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, an additional factor has been at play: the rise of time domain surveys. Correlated noise in time-domain observations is a direct and unavoidable consequence of causation, and hence ubiquitous. Adequately modelling 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 darfs, also display complex, intrinsically or apparently stochastic behaviour, for which adequate modelling strategies are required. GPR is a natural choice to tackle these challenges. Rather than attempting to cover all applications of GPs to all of astrophysics, which would not be feasible in the space available, we have therefore opted to focus on its application to time-domain datasets in astronomy.

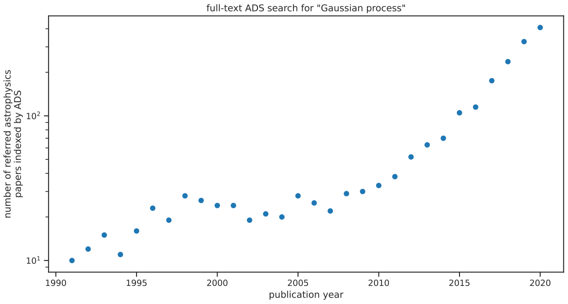

Figure 1

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

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 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 datasets since it is useful to know the "ground truth" to validate performance.

The first example is a re-implementation 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, non-parametric model for the latent (un-observed) 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, and 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 add tinygp Zenodo ref, a Python library for GPR built on top of the JAX library for numerical computing add JAX ref. Here and throughout, the probabilistic models are implemented using the NumPyro library add NumPyro reference 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 re-visit 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 modelled 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 dataset 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 dataset from Vanderriest et al. (1989). The simulated light curves are plotted in the left panel of Figure 2.

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 the left panel we show the simulated data with the median of posterior time delay applied to image A, and an arbitrary magnitude offset applied to image B for plotting purposes. Over-plotted 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.

The right panel of Figure 2 shows the posterior constraints for two of the key parameters of the model: the time delay, and the mean magnitude difference between images. Since these are simulated data, we know the true values of these parameters, and these true values are over-plotted on Figure 2b, demonstrating that our method reliably recovers the expected result.

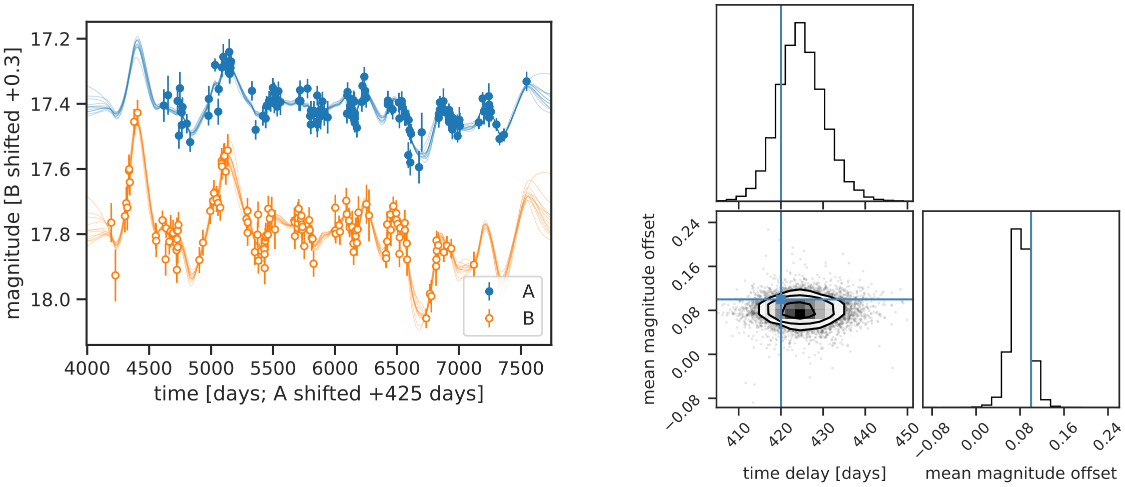

Figure 2

The results of fitting the simulated light curves of two images of a lensed quasar with a time delay between images. *Left:* 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 over-plotted on these data are posterior predictive samples of each images' predicted variability. *Right:* The posterior constraints on the time delay and mean magnitude offset between the two images, obtained using Markov chain Monte Carlo (MCMC) to fit the simulated data shown in the left panel of this figure. The true values of these parameters that were used to simulate the data are over-plotted as blue lines.

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 4.2.2 for a more detailed discussion).

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 cite Agol for light curve computation. 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 the left panel of Figure 3.

We then used the same framework to model the simulated dataset, first ignoring the correlated noise and then accounting for it explicitly. For the transit model, we fit for the planet-to-star radius ratio $R_\mathrm{p}/R_\star$, the mid-transit time T_0, the out-of-transit flux f_0, and the limb darkening parameters u_1, u_2. When also fitting for correlated noise, we also marginalise over the amplitude α and timescale λ of the GP model. The posterior constraints on the time of transit and the planet-to-star ratio are shown in the right panel of Figure 3, with the true values of these parameters over-plotted as a black line. When the correlated noise is neglected (plotted in orange in Figure 3), the inferred parameters are significantly inconsistent with the truth. Taking the correlated noise into account (blue in Figure 3) increases the uncertainties on the inferred parameters, but also shifts the results to recover the true parameters.

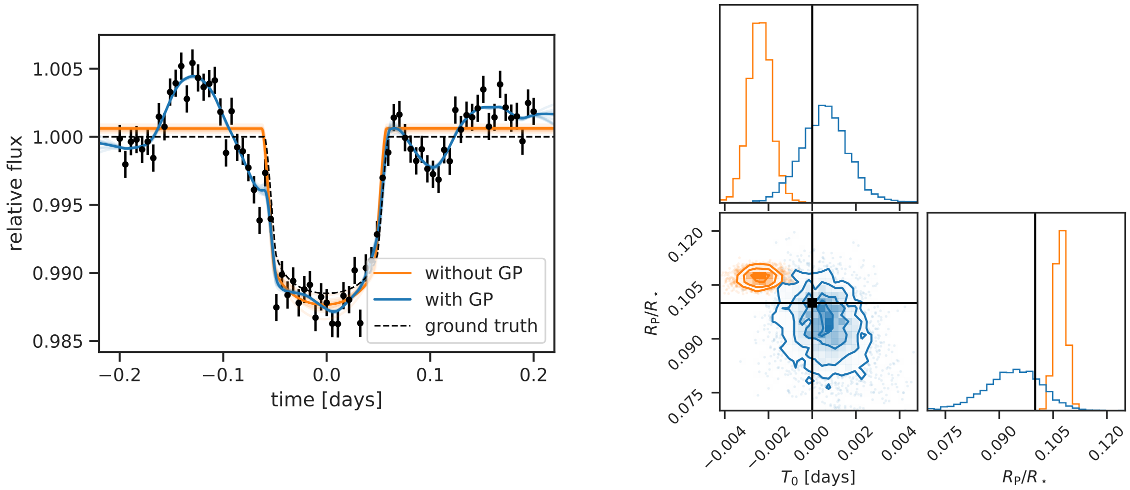

Figure 3

The results of fitting the simulated light curve of a transiting exoplanet. *Left:* The simulated dataset plotted as points with error bars. The inferred model when correlated noise is (blue) or isn't (orange) taken into account are over-plotted on the data. *Right:* The posterior constraints on the time of transit T_0 and the planet-to-star radius ratio $R_\mathrm{p}/R_\star$. The colors in this figure match those in the left panel, 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.

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 modelling 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 datasets to date. describe remaining sections

2. BASICS OF GAUSSIAN PROCESS REGRESSION

So far we have discussed how GPR has become widely used in astronomy and touched on some of their pros and cons, but we haven't 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. For more details and practical exercises, we refer the reader to Rasmussen & Williams (2006) and [Cite SA Saas Fee lecture notes if published].

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 needn't 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 $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$ from the GP is a multi-variate 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 aren't necessarily restricted to, the time t_i. The covariance function, also known as the *kernel*, 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 where 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 function are known as the *hyper-parameters* 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 marginalised 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 *Hierarchichal Bayesian hierarchical Model* (HBM).

Although is 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 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 marginalisation and conditioning, that inference with GPs (the process of marginalising over the individual functions) to be performed *analytically*, with very simple linear algebra. It is this analytic marginalisation property that sets GPR apart from other forms of HBM. ask Dan to check these last two paragraphs.

2.2. From least-squares to GPR

The formal definition above 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 will show that GPR can be thought of as a generalisation 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. 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 $\boldsymbol{\theta} = \{\boldsymbol{\theta}_j\}_{j=1,\ldots,M}$. 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 ϵ_i is the measurement error, or noise, on the i^{th} observation. Furthermore, let us assume that ϵ_i is drawn from a Gaussian distribution with mean 0 and variance σ_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

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

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

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

where $\mathbf{m} = \{m_i\}_{i=1,\ldots,N}$. From this one can readily see that $\ln \mathcal{L} = \text{constant} - 0.5\chi^2$, where the constant depends only on the σ's. Thus, if the σ's are known, maximizing $\mathcal{L}$ is equivalent to minimizing χ^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. I think we should explicitly re-write this as $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi})$ and indicate that K and m depend on these parameters. We should really highlight that this is a drop-in replacement for the likelihood defined in the previous section. Is this ok? Let us now re-write the likelihood in matrix form:

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

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 $K_{ij} = \delta_{ij}\sigma_i^2$ (δ_{ij} being the discrete Kronecker delta function). This is, of course, the *covariance matrix* of the model.

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

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

where k is a *covariance function* or *kernel*, 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 and parameters, the covariance matrix can now have non-zero 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 penalising more complex models.

The kernel 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

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

This gives rise to smooth random functions with variance α and characteristic length-scale λ. We present other commonly used kernels and discuss how to select one in Section 3.1.

2.3. Inference with a GP

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 regression is illustrated schematically in Figure **??**. For example, we can optimize the likelihood with respect to the hyper-parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$. In the machine learning literature, where GPs are commonly used, this is called *training* the GP. In doing so, we are learning the properties of the correlated signal from the data.

Given adequate priors, we can also evaluate the *posterior* distribution:

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

where $p(\mathbf{y}) = \int p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})p(\boldsymbol{\theta}, \boldsymbol{\phi})\mathrm{d}\boldsymbol{\theta}\mathrm{d}\boldsymbol{\phi}$. Usually, the posterior cannot be evaluated analytically, so we resort to sampling methods such as Markov Chain Monte Carlo (MCMC) or nested sampling.

In many astrophysical applications of GPs, we wish to take into account correlated noise, 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 thus *marginalize* (integrate over) the *nuisance parameters* $\boldsymbol{\phi}$.

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 will touch upon), 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 dataset forwards. The 'magic' of GPR is that the predictive distribution is also Gaussian, and its mean and covariance are given by simple analytic relations:

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

where we have introduced the notation $\mathbf{m}_\star \equiv m(\mathbf{t}_\star, \boldsymbol{\theta})$, $\mathbf{K}_\star \equiv k(\mathbf{t}, \mathbf{t}_\star, \boldsymbol{\phi})$ and $\mathbf{K}_{\star\star} \equiv k(\mathbf{t}_\star, \mathbf{t}_\star, \boldsymbol{\phi})$. 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. Depending on whether one wishes the predictive variance (the diagonal elements of $\mathbf{C}_\star$) to account for measurement uncertainties or not, one can include the white noise term in $\mathbf{K}_{\star\star}$ or leave it out.

2.4.2. Properties of the predictive distribution. A closer look at Equations (13) reveals some important properties. First, a GP is a *linear predictor*. 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 centred on the training points: $f_\star = \alpha^{\mathrm{T}}\,\mathbf{k}_\star$ where $\alpha = \mathbf{K}^{-1}\,\mathbf{y}$. These linearity properties are very important to understand the behaviour of GPs. They shed light on the relationship between GPR and standard linear models with very large numbers of free parameters (see e.g. Hogg & Villar 2021, for a more detailed discussion), and lead to a number of powerful extensions that are beyond scope of this review.

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 observations to optimise our predictions at a given (set of) time(s). However, in practice we rarely know the hyper-parameters *a priori*, and the predictive posterior distribution marginalised over the hyper-parameters *does* depend on the observations.

Finally, as $\mathbf{K}$ is positive semi-definite, 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 behaviour 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 will 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 $k(t_i,\, t_j;\, \boldsymbol{\phi})$. Unlike how they are typically presented in the machine learning literature, GP models in astrophysics will often—but not always—include non-trivial mean functions. For example, in the example from 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 won’t discuss the mean function in detail, focusing instead primarily on the kernel function, since that is unique to GP modeling. All this being said, in our experience, new users of GP models will 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 your choice of kernel. In this section we walk through this process in detail.

3.1. Gaussian Process Covariance Functions

Say something about to handle higher dimensional inputs and distance metrics.

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. The parameter $\boldsymbol{\phi}$ indicates that these functions are typically parameterized by a set of parameters, typically referred to as "hyper-parameters".

Hyper-parameters

In a GP model, the kernel function is typically parameterized by a set of parameters that we will label as $\boldsymbol{\phi}$ and refer to as "hyper-parameters". 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 marginalised over in closed form, leaving the hyper-parameters as the parameters of interest.

The kernel function can be any positive semi-definite scalar function, but some will be more useful than others. Given this large decision space, you may be wondering how to choose the right kernel for your specific problem. If you're very lucky, you may be able to motivate your model using physics. This can be approached from two directions. On one hand, some commonly used kernel function have a specific physical interpretation—for example, the solution to a stochastic ordinary differential equation—that can be used to motivate their use. On the other hand, if your physical model is specified by its *power spectrum*, this can recast as a GP model with a specific kernel function. Say something about the number of times a function is differentiable? That can be a useful consideration. But perhaps already mentioned later.

Even if you don't have a formal physics-based justification for your model, you may be able to identify the key scales of your 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 select a set of models that have the needed covariance structure, and then combine or select between these choices. For example, if there is only expected to only be a single non-periodic timescale in the problem of interest, you could list all the two-parameter non-periodic kernels and use a numerical model selection technique as described below.

3.1.1. Standard kernels & sums or products thereof. Some popular kernels are listed in Table 1, and some other choices are discussed in Chapter 4 of Rasmussen & Williams (2006). In any practical application, these kernel functions are not used on their own. Instead, more expressive models are designed by combining these models. In particular, any valid kernel functions can be added or multiplied to generate new valid functions. For example, the squared exponential kernel function is generally defined as

$$k(\tau;\, \boldsymbol{\phi}) = \alpha^2\, \exp\left(-\frac{\tau^2}{2\,\lambda^2}\right) \qquad 14.$$

with $\boldsymbol{\phi} = \{\alpha,\, \lambda\}$, which, in our notation, is actually the product of a constant kernel and a squared exponential kernel, as defined in Table 1. Another example that is commonly used in the astrophysics literature (Aigrain et al. 2012; Haywood et al. 2014) is the following "quasi-periodic" kernel

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

with $\boldsymbol{\phi} = \{\alpha,\, \lambda_1,\, \lambda_2,\, \gamma\}$, which has a period of λ_2, and a decoherence timescale of λ_1.

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

$$K_{i,j} = k(t_i,\, t_j;\, \boldsymbol{\phi}) \quad , \tag{16.}$$

and then generating a multivariate Gaussian sample with this covariance. As an example, Figure 4 shows several prior samples for three different kernel functions from Table 1, with a range of length scales λ and amplitudes α. In this figure, we can see some qualitative differences between the kernels—namely that the kernels become "smoother" from left to right—and we can see how the range of allowed functions change with the hyper-parameters.

Table 1 Some kernel functions commonly used in the astrophysics literature add the rational quadratic kernel

Name	Representation[a]
Constant kernel	α^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	$\left(1+\sqrt{5}\tau/\lambda + 5(\tau/\lambda)^2/3\right)\, e^{-\sqrt{5}\tau/\lambda}$
Cosine	$\cos 2\pi\tau/\lambda$
Exponential Sine Squared	$\exp\left(-\gamma \sin^2 \pi\tau/\lambda\right)$
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, τ is defined as $\tau = |t_i - t_j|$, and Greek letters indicate hyper-parameters; [b] "radial basis function"; [c] "Ornstein-Uhlenbeck", "damped random walk" or "Matérn-1/2"; [d] Foreman-Mackey et al. (2017).

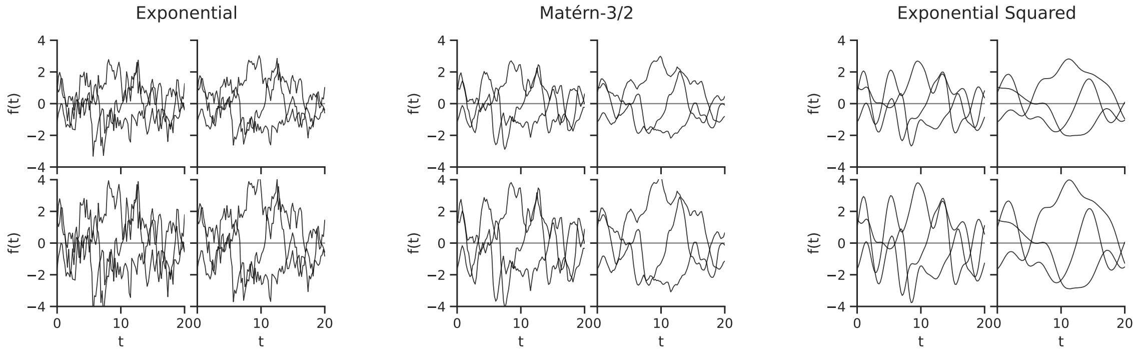

Figure 4

Prior samples for three different classes of kernel functions: (a) exponential, (b) Matérn-3/2, and (c) exponential squared. In each sub-figure, the length scale λ of the kernel increases from left to right, and the amplitude α 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, while the exponential squared kernel is infinitely differentiable. This can be seen qualitatively in this figure.

3.1.2. Operations on kernel functions. Derivatives, integrals, etc.

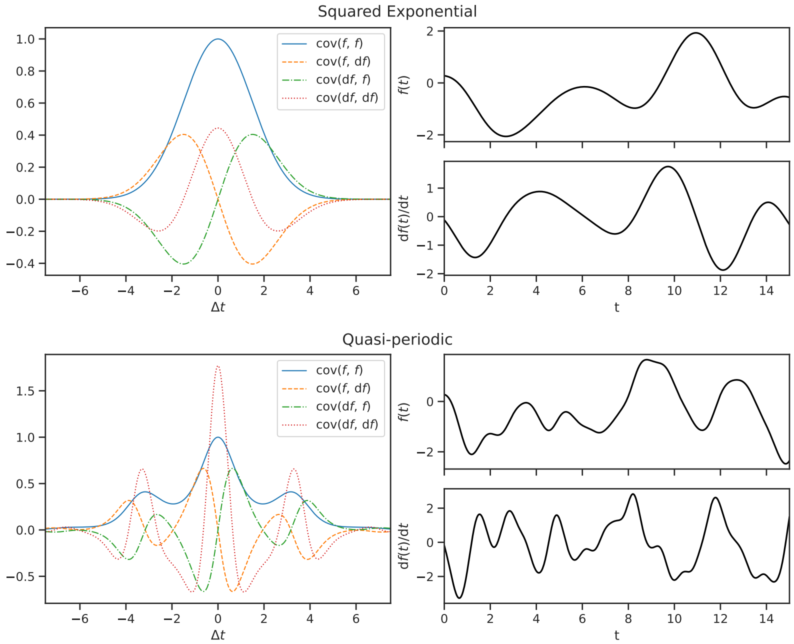

Figure 5

Derivative kernels.

3.1.3. Multi-dimensional inputs. Abs vs other norms, non-standard geometry (GPs on the sphere), parallel time series.

3.1.4. Physically motivated kernels. Cosmology, asteroseismology, etc.

Mention kernels with break points to detect sudden changes, as described in review by Roberts et al. (2012) for example?

3.2. Hyperparameter Inference

A key component of all the covariance functions discussed above is that they are all *parameterized* by a set of hyper-parameters. In most cases, you won't have *a priori* knowledge for how the values of these hyper-parameters should be set. Instead, their values will 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 hyper-parameters 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 hyper-parameters $\boldsymbol{\phi}$. 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 non-linear optimization routine to find the maximum likelihood parameter values, or in a Markov chain Monte Carlo (MCMC) procedure to marginalise over the hyper-parameters, and propagate their uncertainty to constraints on the parameters of the mean model.

An important point here is that, since many data analysis procedures in astrophysics include a step like the ones listed above, the use of a GP likelihood doesn't significantly change the processing. In fact, we like to say that the GP likelihood can be used as a drop-in replacement for anywhere you're currently using a "chi-squared" objective. There are some practical reasons why things aren't necessarily this simple (for example, computational cost, as described below), but the sentiment stands.

In this review, we won't go into too many details about the inference algorithms, but throughout the text, we will regularly use the BFGS gradient-based, non-linear optimization routine (Nocedal & Wright 1999; Virtanen et al. 2020) to find the maximum likelihood parameter values

$$\boldsymbol{\theta}_\star, \boldsymbol{\phi}_\star = \underset{\boldsymbol{\theta},\boldsymbol{\phi}}{\arg\max}\ \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi}) = \underset{\boldsymbol{\theta},\boldsymbol{\phi}}{\arg\max}\ \log p(\boldsymbol{y}\,|\,\boldsymbol{\theta}, \boldsymbol{\phi}) \quad . \tag{17.}$$

Another common inference technique used in astrophysics—and in this review!—is the use of Markov chain Monte Carlo (MCMC) to generate posterior samples

$$\boldsymbol{\theta}, \boldsymbol{\phi} \sim p(\boldsymbol{\theta}, \boldsymbol{\phi}\,|\,\boldsymbol{y}) \tag{18.}$$

that can be used to marginalise over some subset of the parameters, and estimate the uncertainty on the parameter values. All the examples in this review are implemented using the NumPyro probabilistic inference library, and all the MCMC examples use the gradient-based No U-Turn Sampler (NUTS) algorithm (Hoffman & Gelman 2014).

Section 1.2.1 and Section 1.2.2, and specifically Figure 2 and Figure 3, show the results of worked examples of an MCMC-based GPR workflow, without and with (respectively) a non-trival 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 your 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's important to note that there's nothing fundamentally different about GPs in the context when compared to other models for data, so that means that many methods that you may already use for model selection and validation in other contexts will 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 will discuss some examples of model validation and selection techniques that have been used for GPs.

Say something here about assessing Gaussianity.

When it comes to selecting between different possible kernel functions, the approach that you take may depend on your 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 of posterior predictive assessment add link to Gelman paper here can be used to evaluate different choices of kernel function.

To demonstrate these approaches, Figure 6 shows the results of performing a model comparison between three different kernel functions applied to a simulated dataset. 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: (1) the squared exponential kernel, (2) the Matérn-3/2 kernel, and (3) the rational quadratic kernel. The simulated dataset is shown in the left panel of Figure 6.

First, using the full dataset, we compute the Bayesian evidence integral for each of these model choices using a nested sampling algorithm implemented in the `jaxns` package add citation. The Bayesian evidence integral is defined as:

$$Z(H) \equiv p(\boldsymbol{y} \,|\, H) = \int p(\boldsymbol{\theta}, \boldsymbol{\phi} \,|\, H)\, p(\boldsymbol{y} \,|\, \boldsymbol{\theta}, \boldsymbol{\phi}, H)\, \mathrm{d}\boldsymbol{\theta}\, \mathrm{d}\boldsymbol{\phi} \qquad 19.$$

where H indicates the modeling choices, in this case the choice of kernel function. For better or worse[1], the Bayesian evidence is frequently used in the astrophysics literature as an ingredient in model selection procedures. 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 the middle panel of Figure 6. In this figure, it is clear that the squared exponential and rational quadratic kernels are indistinguishable under this metric, while the Matérn-3/2 kernel is somewhat disfavoured.

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

$$p(\boldsymbol{y}_\star \,|\, \boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\phi}) = \mathcal{N}(\boldsymbol{f}_\star, \boldsymbol{C}_\star) \qquad 20.$$

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 as black dots in the left panel of Figure 6, holding out the data points indicated by blue crosses. At each step in the MCMC, we evaluate the likelihood in Equation 20 for the held out data conditioned on the training data and the model parameters. In the rightmost panel of Figure 6, 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, while the Matérn-3/2 kernel is disfavoured.

4. GAUSSIAN PROCESSES IN TIME-DOMAIN ASTRONOMY

In this section we discuss a range of applications of GPR to time-domain datasets in the astronomical literature. The present-day popularity of GPs in astronomy precludes any attempt at an exhaustive review. Instead, we have selected examples which showcase the power and flexibility of the method. The order in which these applications are discussed is partly based on chronological considerations, but we have also attempted to present the simpler applications first.d

4.1. AGN variability (DFM)

This section is assigned to DFM but as I started writing something about it for the introduction then decided to move it here I have left it in – feel free to use this as a starting point or ignore it entirely!.

The Active Galactic Nuclei (AGN) at the centre of many distant galaxies produce stochastic variability in their optical and radio emission, on timescales ranging from hours to years.

[1] A discussion of the relative merits of different model selection algorithms is beyond the scope of this review.

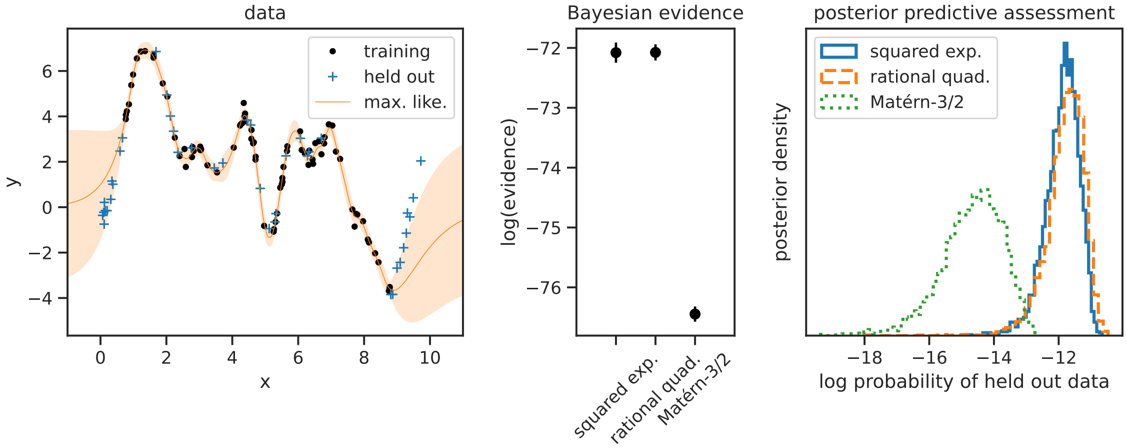

Figure 6

Model assessment.

4.1.1. Lensed quasar time-delays. As discussed in Section 1, this is –to the best of our knowledge– the first area in which GPR makes an appearance in the refereed astronomical literature.

Refer to the first simulated example in Section 1.2.1 here. Some of the text below may need to move to that Section.

When the quasar is lensed by an intervening galaxy or galaxy cluster, multiple images can be formed, whose brightness 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 weren't always consistent. Press et al. (1992a) derived a simple χ^2 metric under the assumption that the observed light curves are shifted, noisy versions of a single sample a Gaussian process with known covariance. Having proposed an empirical procedure to estimate the covariance function from the data, they then optimize the χ^2 with respect to the time-delay between images, and apply standard methods used to derive an uncertainty. While their method relied on an adhoc 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 dataets 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 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.

4.1.2. Reverberation mapping.

4.1.3. Quasar power spectrum estimation.

4.2. 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 datasets: transits in light curves, Keplerian signals in Radial Velocity (RV) or astrometry time-series. These are invariably affected by, and often buried in, correlated noise or nuisance signals, making GPR a natural choice.

4.2.1. Transit searches. Following the discovery of the first transiting exoplanet (Henry et al. 2000; Charbonneau et al. 2000), numerous ground-based photometric monitoring surveys were set up (or re-purposed) 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. While surveys such as the Optical Gravititational 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 radial velocity 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 went on to show that the SNRs for all candidates had been over-estimated due to the presence of correlated, or red, noise in the light curves. They proposed an empirical method to evaluate the red noise on transit time-scales 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 Crossfield et al. (see e.g. 2016), and are used frequently after detection to model the out-of-transit baseline alongside the transit signal itself (see next sub-section). 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 *K2* data. In that case, the primary source of out-of-transit variability is instrumental systematics (see next section), 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*, *Kepler* and *TESS*, the systematics are less prominent, so the dominant source of out-of-transit variations is intrinsic stellar variability, which occurs mainly 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 stars, the separation between transits and stellar signal becomes less clean, significantly reducing the performance irrespective of sequential detrending approaches (irrespective of the algorithm used, Hippke et al. 2019). In such cases, a simultaneous modelling 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 future missions such as *PLATO* 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 fast and scalable GP solvers are available, this becomes a more feasible proposition.

4.2.2. Transit and eclipse modelling. In practice, GPR has been much more widely applied to transit modelling than to detection *per se*. The detailed depth, timing, duration and shape of a transit depends on the planet-to-star radius ratio $R_\mathrm{p}/R_\star$, the system scale $a/R_\star$ (where a is the orbital semi-major axis), the impact parameter b, the orbital period P and the time of transit centre, 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, which 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 [ref]). The system scale is directly related to the stellar density [ref seager-mallen03], which can be used for sanity checks to ensure the transits are indeed of planetary origin [ref tingley]. 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, non-transiting planets [ref]. Wavelength-dependent measurements of the transit depth probe the effect of the planet's atmosphere on the star light that filters through it, and hence allow us to access the atmospheric composition. 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 the formulation using wavelets. Rather than modelling 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 datasets with irregular sampling or significant data gaps, for example from space telescopes in low-Earth orbit such as Hubble and Spitzer.

Colours of noise: White, pink and red noise correspond to $\gamma = 0$, 1 and 2, respectively.

This limitation is significant, as Hubble and Spitzer have been the workhorse instruments for transit and eclipse spectroscopy for over a decade. This involves measuring minute changes in the depth of the primary transit or secondary eclipse as a function of wavelength, either using successive photometric observations through different filters or using 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 favourable 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 centre 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).

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 multi-dimensional inputs to a squared exponential GP whose mean is the transit signal, allowing one to marginalise 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 has a different length scale for each input dimension, but a single, shared output scale (or amplitude), known as an Automatic Relevance Determination (ARD) kernel:

$$k_{\rm ARD}(\mathbf{x}_i, \mathbf{x}_j) = A \exp\left(- \sum_{m=1}^{M} \frac{|\mathbf{x}_i - \mathbf{x}_j|_m^2}{\lambda_m} \right), \qquad 21.$$

where M is the number of input dimensions. Combined with "shrinkage" hyper-priors favouring long λ's, this 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 marginalising over families of both parametric and GP models.

Despite these theoretical advantages and practical successes, the use of GPR to analyse low-resolution transit and eclipse spectra remains confined to a relatively small subset of the corresponding community. Parametric models (in some cases marginalising over families thereof, [ref wakeford]) remain the most widely-used approach for Hubble space telescope observations, and Pixel-Level-Decorrelation (PLD) for Spitzer observations [ref Deming]. Possible explanations for this include computational cost, as well as the fact that these other methods are perceived as easier to implement and interpret. As the recently-launched JWST begins to deliver much higher signal-to-noise observations, multiple analyses of which by different groups are actively encouraged, it will be interesting to see how this topic evolves in the next few years.

Latent variable: A random variable which is never actually observed.

One general shortcoming of current GP-based systematics models with multi-dimensional inputs is that the uncertainties on the input variables are ignored. This could in principle be remedied by treating the housekeeping variables as noisy observations of latent GP variables, and modelling them alongside the observed fluxes, 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). "Coloured" light curves, extracted in individual wavelength bins, are then divided by the best-fit systematics model derived from the white light curve, before being modelled 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 modelling 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 modelling 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

Composite GPs

A composite (additive or multiplicative) GP is a powerful way to separate different components of a time-series dataset, whether these components depend on different input variables (as in the instrumental systematics examples discussed in this section) 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 Equations (13) using that component only, while using the full covariance for $\mathbf{K}$.

that the apparent density discrepancy disappears when this granulation signal is modelled using a GP.

4.2.3. Systematics removal in the presence of stellar variability. GP-based models have also been used to correct instrumental systematics in data from the *K2* space mission, which used the *Kepler* satellite to perform an Ecliptic plane survey after the failure of two of its reaction wheels. 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 $\sim 6\,\mathrm{h}$ 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 intra-pixel variations, and to changes in the contamination of the photometric aperture by neighbouring targets as well as in aperture losses. These can be modelled 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 2-D 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. This not only improves the fit but allows the position-dependent systematics to be evaluated separately and thus removed while preserving intrinsic variability. The use of a time-depedent 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 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).

4.2.4. Quasi-periodic GP models for stellar light curves. The light curves of Sun-like stars (broadly construed) display low-amplitude quasi-periodic variations which are caused by the rotational modulation and evolution of magnetically active regions on their surfaces. These produce quasi-periodic variations in both photometry and radial velocity observations, and GPs have in recent years become one of the most popular ways of modelling them. One reason for this is that a simple quasi-periodic covariance function [refer to relevant section or introduce here], which reproduce the light curves of rotating stars with evolving active regions remarkably well.

GP models were first applied to variations by Aigrain et al. (2012), to test a new method to simulate Radial Velocity (RV) variations based space-based photometry on solar data. The context of this work was the large numbers of candidate transiting exoplanets being discovered at the time by the *CoRoT* and *Kepler* space missions. Detecting the planets signals in RV was needed to confirm the planetary nature of the candidates by measuring their masses, but was being 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 Quasi-Periodic (QP), and found the former to be preferred in the solar case. This is not entirely surprising, as the typical lifetimes of active regions on the Sun are not much longer than its rotation period [ref]. 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.

Although the QP kernel provides a phenomenological rather that 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 datasets, 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 datasets spans less than the simulated evolution timescale.

On the other hand, 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 starspot 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* or *PLATO*), composite GP models with different terms to describe activity, granulation and stellar oscillations can be useful. For example, Barros et al. (2020) analysed 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 one with a single term. They also tested a white noise only model, which provided less precise precise estimates than either GP models, albeit generally consistent with both.

4.2.5. Stellar activity in Radial Velocities. Given the remarkable instrumental precision actived by modern RV spectrographs, stellar activity is nowadays the key factor limiting the sensitivity of RV surveys low-amplitude and/or long-period planet signals. The effect of active regions on RV observations is two-fold. First, as dark spots rotate on the stellar surface, they distort the profile of spectral lines, removing a small contribution first from the red wing of each spectral line, then from the line core, and then from the blue 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 inter-granular lanes). In a facula, i.e. a region of enhanced magnetic flux density compared to the "clean" photosphere, convection is suppressed, leading to a localised reduction in this "convective blue-shift". Faculae have very small photometric contrast but can cover a much larger area than dark spots, so their RV signature can dominate over that of spots, specially for moderately slow rotators like the Sun with modest magnetic fields (Meunier et al. 2010).

Overall, active regions produce quasi-periodic RV variations which are frequently modelled with the same kind of QP GPs as 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 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 strong pressure on telescope time means that the time-sampling of these observations is not always good enough to constrain the GP parameters well.

As previously alluded to, 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 $F'(t)$ 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 co-located, 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 the case only exceptionally (see e.g. Gibson 2014).

On the other hand, 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 which trace chromospheric emission in the cores of strong lines. Rajpaul et al. (2015) developed a GP framework, schematically illustrated in Figure **??**, to model these activity indicators alongside the RVs, and to disentangle them from planetary signals. Each observed time-series is modelled 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 modelled as a (typically quasi-periodic) GP, which makes it possible to write down expressions for the covariance between any pair of observations from the 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. 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 recently (Barragán et al. 2022; Delisle et al. 2022), and extensions thereof are likely to play a significant role in the analysis of next regeneration RV surveys targeting Earth analogues, which need to achieve even better contrast between activity and planetary signals.

4.2.6. Granulation and asteroseismology. Possible addition: modelling high-dispersion spectra with GPs

4.3. Compact objects

I thought this new heading would be a neat way to introduce applications to pulsars, non-quasar QPOs and GWs?

4.3.1. Pulsars.

4.3.2. Quasi-periodic oscillations.

4.3.3. Gravitational waves.

5. CHALLENGES, PITFALLS, AND SOLUTIONS

Including when not to use GPs.

In this section we will discuss some of the main drawbacks of using Gaussian processes namely computational cost and overfitting. We will discuss how recent advances are helping overcome the former, and illustrate how the latter arises using examples from the astronomical literature. We will suggest methods for diagnosing and avoiding overfitting (e.g. cross validation). Estimated length: 5 pages

5.1. Fast GP solvers

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 analysis. 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 non-trivial 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, while cautioning 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 will focus on tools implemented in the Python programming language since it is—at the time of writing—the most popular language, both in astronomy and GP modeling, but there are tools available in all other popular languages.

Assigned: DFM

In this short section, we will provide an overview of the current ecosystem of popular open source tools for implementing Gaussian processes, including links to specific packages (mostly, but not exclusively, written in Python). Estimated length: 3 pages

7. CONCLUSIONS AND FUTURE DIRECTIONS

This section will wrap up the review by recapping the key takeaways and outline a couple of directions in which rapid progress is currently taking place, which could have a significant impact in how Gaussian processes are used in astronomy in the near term. In particular, we will discuss the relationship between scalable Gaussian processes and recent progress in applied linear algebra. We will discuss how these methodological advances will enable the use of Gaussian processes to model data from the next generation of astronomical facilities such as LSST-Rubin and SKA. Estimated length: 2 pages

SUMMARY POINTS

1. Summary point 1. These should be full sentences.
2. Summary point 2. These should be full sentences.
3. Summary point 3. These should be full sentences.
4. Summary point 4. These should be full sentences.

FUTURE ISSUES

1. Future issue 1. These should be full sentences.
2. Future issue 2. These should be full sentences.
3. Future issue 3. These should be full sentences.
4. Future issue 4. These should be full sentences.

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

Acknowledgements, general annotations, funding.

LITERATURE CITED

Aigrain S, Hodgkin ST, Irwin MJ, Lewis JR, Roberts SJ. 2015. *Monthly Notices of the Royal Astronomical Society* 447(3):2880–2893

Aigrain S, Parviainen H, Pope BJS. 2016. *Monthly Notices of the Royal Astronomical Society* 459(3):2408–2419

Aigrain S, Pont F, Zucker S. 2012. *Monthly Notices of the Royal Astronomical Society* 419(4):3147–3158

Angus R, Morton T, Aigrain S, Foreman-Mackey D, Rajpaul V. 2018. *Monthly Notices of the Royal Astronomical Society* 474(2):2094–2108

Barclay T, Endl M, Huber D, Foreman-Mackey D, Cochran WD, et al. 2015. *Astrophysical Journal* 800(1):46

Barnes JA, Tryon PV, Sargent H. H. I. 1980. *Sunspot cycle simulation using random noise.* In *The Ancient Sun: Fossil Record in the Earth, Moon and Meteorites*, eds. RO Pepin, JA Eddy, RB Merrill

Barragán O, Aigrain S, Kubyshkina D, Gandolfi D, Livingston J, et al. 2019. *Monthly Notices of the Royal Astronomical Society* 490(1):698–708

Barragán O, Aigrain S, Rajpaul VM, Zicher N. 2022. *Monthly Notices of the Royal Astronomical Society* 509(1):866–883

Barros SCC, Demangeon O, Díaz RF, Cabrera J, Santos NC, et al. 2020. *Astronomy and Astrophysics* 634:A75

Blandford RD, Narayan R. 1992. *Annual Review of Astronomy and Astrophysics* 30(1):311–358

Carter JA, Winn JN. 2009. *Astrophysical Journal* 704(1):51–67

Charbonneau D, Brown TM, Latham DW, Mayor M. 2000. *Astrophysical Journal Letters* 529(1):L45–L48

Constable CG, Parker RL. 1988. *Journal of Geophysical Research* 93(B10):11569–11581

Crossfield IJM, Ciardi DR, Petigura EA, Sinukoff E, Schlieder JE, et al. 2016. *Astrophysical Journal Supplement* 226(1):7

Delisle JB, Unger N, Hara NC, Ségransan D. 2022. *Astronomy and Astrophysics* 659:A182

Dvorak R, Edelman C. 1976. *Mitteilungen der Astronomischen Gesellschaft Hamburg* 38:192

Evans TM, Aigrain S, Gibson N, Barstow JK, Amundsen DS, et al. 2015. *Monthly Notices of the Royal Astronomical Society* 451(1):680–694

Evans TM, Pont F, Sing DK, Aigrain S, Barstow JK, et al. 2013. *Astrophysical Journal Letters* 772(2):L16

Evans TM, Sing DK, Kataria T, Goyal J, Nikolov N, et al. 2017. *Nature* 548(7665):58–61

Foreman-Mackey D, Agol E, Ambikasaran S, Angus R. 2017. *Astronomical Journal* 154(6):220

Foreman-Mackey D, Montet BT, Hogg DW, Morton TD, Wang D, Schölkopf B. 2015. *Astrophysical Journal* 806(2):215

Gibson NP. 2014. *Monthly Notices of the Royal Astronomical Society* 445(4):3401–3414

Gibson NP, Aigrain S, Roberts S, Evans TM, Osborne M, Pont F. 2012. *Monthly Notices of the Royal Astronomical Society* 419(3):2683–2694

Gibson NP, Pont F, Aigrain S. 2011. *Monthly Notices of the Royal Astronomical Society* 411(4):2199–2213

Gillon M, Pont F, Demory BO, Mallmann F, Mayor M, et al. 2007. *Astronomy and Astrophysics* 472(2):L13–L16

Haywood RD, Collier Cameron A, Queloz D, Barros SCC, Deleuil M, et al. 2014. *Monthly Notices of the Royal Astronomical Society* 443(3):2517–2531

Henry GW, Marcy GW, Butler RP, Vogt SS. 2000. *Astrophysical Journal Letters* 529(1):L41–L44

Hippke M, David TJ, Mulders GD, Heller R. 2019. *Astronomical Journal* 158(4):143

Hoffman MD, Gelman A. 2014. *Journal of Machine Learning Research* 15(47):1593–1623

Hogg DW, Villar S. 2021. *Publications of the Astronomical Society of the Pacific* 133(1027):093001

Horne K. 2003. *Status aand Prospects of Planetary Transit Searches: Hot Jupiters Galore*. In *Scientific Frontiers in Research on Extrasolar Planets*, eds. D Deming, S Seager, vol. 294 of *Astronomical Society of the Pacific Conference Series*

Jekeli C. 1991. *Manuscr. Geod.* 16(5):313–325

Kreidberg L. 2018. *Exoplanet Atmosphere Measurements from Transmission Spectroscopy and Other Planet Star Combined Light Observations*. In *Handbook of Exoplanets*, eds. HJ Deeg, JA Belmonte. SpringerLink, 100

Krige DG. 1951. *A statistical approach to some mine valuations and allied problems at the Witwatersrand*. Master's thesis, University of Witwatersrand

Luger R, Agol E, Kruse E, Barnes R, Becker A, et al. 2016. *Astronomical Journal* 152(4):100

Luger R, Foreman-Mackey D, Hedges C. 2021a. *Astronomical Journal* 162(3):124

Luger R, Foreman-Mackey D, Hedges C, Hogg DW. 2021b. *Astronomical Journal* 162(3):123

Luger R, Kruse E, Foreman-Mackey D, Agol E, Saunders N. 2018. *Astronomical Journal* 156(3):99

Mandel K, Agol E. 2002. *Astrophysical Journal Letters* 580(2):L171–L175

Meunier N, Desort M, Lagrange AM. 2010. *Astronomy and Astrophysics* 512:A39

Nicholson BA, Aigrain S. 2022. *Monthly Notices of the Royal Astronomical Society*

Nocedal J, Wright SJ. 1999. *Numerical Optimization*. Springer

Peebles PJE. 1997. *Astrophysical Journal Letters* 483(1):L1–L4

Pont F, Zucker S, Queloz D. 2006. *Monthly Notices of the Royal Astronomical Society* 373(1):231–242

Press WH, Rybicki GB, Hewitt JN. 1992a. *Astrophysical Journal* 385:404

Press WH, Rybicki GB, Hewitt JN. 1992b. *Astrophysical Journal* 385:416

Rajpaul V, Aigrain S, Osborne MA, Reece S, Roberts S. 2015. *Monthly Notices of the Royal Astronomical Society* 452(3):2269–2291

Rasmussen CE, Williams CKI. 2006. *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press

Roberts S, Osborne M, Ebden M, Reece S, Gibson N, Aigrain S. 2012. *Philosophical Transactions of the Royal Society of London Series A* 371(1984):20110550–20110550

Rybicki GB, Press WH. 1995. *Physical Review Letters* 74(7):1060–1063

Sliski DH, Kipping DM. 2014. *Astrophysical Journal* 788(2):148

Udalski A, Paczynski B, Zebrun K, Szymanski M, Kubiak M, et al. 2002. *Acta Astronomica* 52:1–37

Vanderburg A, Johnson JA. 2014. *Publications of the Astronomical Society of the Pacific* 126(944):948

Vanderriest C, Schneider J, Herpe G, Chevreton M, Moles M, Wlerick G. 1989. *Astronomy and Astrophysics* 215:1–13

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, et al. 2020. *Nature Methods* 17:261–272

von der Heide K. 1978. *Astronomy and Astrophysics* 70(6):777–784

Way MJ, Srivastava AN. 2006. *Astrophysical Journal* 647(1):102–115

Zicher N, Barragán O, Klein B, Aigrain S, Owen JE, et al. 2022. *Monthly Notices of the Royal Astronomical Society* 512(2):3060–3078