New Metrics for Identifying Variables and Transients in Large Astronomical Surveys

Shih Ching Fu (符史青)[1] , Arash Bahramian[1] , Aloke Phatak[2] , James C. A. Miller-Jones[1] , Suman Rakshit[3,4] ,
Alexander Andersson[5] , Robert Fender[5] , and Patrick A. Woudt[6]

[1] International Centre for Radio Astronomy Research—Curtin University, GPO Box U1987, Perth, WA 6845, Australia; shihching.fu@postgrad.curtin.edu.au
[2] ARC Centre for Transforming Maintenance through Data Science—Curtin University, GPO Box U1987, Perth, WA 6845, Australia
[3] School of Electrical Engineering, Computing and Mathematical Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia
[4] Curtin Biometry and Agricultural Data Analytics, Centre for Crop and Disease Management, Curtin University, GPO Box U1987, Perth, WA 6845, Australia
[5] Astrophysics, Department of Physics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK
[6] Department of Astronomy, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa

Abstract

A key science goal of large sky surveys such as those conducted by the Vera C. Rubin Observatory and precursors to the Square Kilometre Array is the identification of variable and transient objects. One approach is analyzing time series of the changing brightness of sources, namely, light curves. However, finding adequate statistical representations of light curves is challenging because of the sparsity of observations, irregular sampling, and nuisance factors inherent in astronomical data collection. The wide diversity of objects that a large-scale survey will observe also means that making parametric assumptions about the shape of light curves is problematic. We present a Gaussian process (GP) regression approach for characterizing light-curve variability that addresses these challenges. Our approach makes no assumptions about the shape of a light curve and, therefore, is general enough to detect a range of variable and transient source types. In particular, we propose using the joint distribution of GP amplitude hyperparameters to distinguish variable and transient candidates from nominally stable ones and apply this approach to 6394 radio light curves from the ThunderKAT survey. We compare our results with two variability metrics commonly used in radio astronomy, namely η_ν and V_ν, and show that our approach has better discriminatory power and interpretability. Finally, we conduct a rudimentary search for transient sources in the ThunderKAT data set to demonstrate how our approach might be used as an initial screening tool. Computational notebooks in PYTHON and R are available to help deploy this framework to other surveys.

1. Introduction

Advancing our understanding of exotic astrophysical phenomena depends on accessing sufficiently large samples of such phenomena for analysis. As large survey projects such as the Rubin Observatory Legacy Survey of Space and Time (LSST; Ž. Ivezić et al. 2019) and those planned for the Square Kilometre Array (P. E. Dewdney et al. 2009) get commissioned and start monitoring vast portions of the sky, a key task is searching huge data sets for objects of interest. Indeed, surveys such as LSST and The HUNt for Dynamic and Explosive Radio transients with MeerKAT (ThunderKAT;[7] R. Fender et al. 2017) have science goals contingent on finding transient and variable phenomena. The investigation of high-energy transient phenomena, detected in X-ray, gamma-ray, radio, or other bands, helps us explore new physics under extreme conditions that cannot be reproduced in a laboratory. One discovery method is the statistical time series analysis of the changing brightness of sources, that is, their light curves.

To identify transient and variable sources, the statistical characterization of their light curves must be descriptive and scalable: descriptive, in that we want to encode the properties of sources that best facilitate their identification and classification, and scalable, such that we can use these encodings across many surveys monitoring thousands to millions of sources (A. Andersson et al. 2025). This paper proposes a novel representation of light curves based on the hyperparameters derived by Gaussian process (GP) regression. We describe a new method for characterizing sources with an emphasis on capturing the time-domain variability of a source, intending to identify transient and variable sources.

There are many existing approaches for identifying and discovering specific subtypes of transient and variable objects. These typically take a template or matched filter approach, such as by L. Feng et al. (2017) in radio and B. Allen et al. (2012) in gravitational wave signals, which search for a specific shape or pattern of curve. Further complications exist, such as particular source types varying on particular timescales. For example, subsecond for fast radio bursts and pulsars, minutes for ultralong period transients, hours to days for X-ray binaries (XRBs) and gamma-ray bursts, and weeks to months for tidal disruption events. However, in the context of commensal and serendipitous discovery in large surveys, no predetermined shape or timescale is available, so a general-purpose characterization is needed.

---

[7] https://science.uct.ac.za/thunderkat

Our approach is therefore not tuned specifically to any particular source type but can fit models and quantify the variability of various light-curve shapes. We envisage its use as an initial single-pass screening tool for rapidly reducing the search space of candidate sources without overly inflating the false negative rate. We demonstrate our approach on radio light curves and will apply it to survey data at other wavelengths in future work.

GPs have grown in popularity in the astrophysics literature, largely because they can be fitted to the sparse and unevenly sampled data typical of astronomical light curves. However, there are additional uses of GPs beyond resampling sparse data or fitting smooth curves. The parameters of GP regression models are physically interpretable, and their estimation easily fits into a Bayesian hierarchical framework such that observational uncertainties can be incorporated into every analysis stage. Therefore, we present our characterization of light curves using GPs, striking a balance between over-simplified and overspecified models such that the results are statistically justified and astrophysically meaningful.

1.1. Variability Statistics in Radio Surveys

Numerous methods have been used to identify transient and variable sources among survey observations, but here we focus on those that calculate descriptive statistics for a light curve and proceed to flag sources that have extreme values of those statistics. For example, the Caltech-NRAO Stripe 82 Survey (K. P. Mooley et al. 2016) identified variable sources by applying the Student t-test to flux densities measured at different epochs with the null hypothesis that the source has not varied significantly between epochs. However, any selection bias in the choice of epochs for that comparison would have a strong effect on the success of this approach. The Murchison Widefield Array Transients Survey (M. E. Bell et al. 2019) used two metrics for identifying candidate variable objects: the time-averaged modulation index, $\overline{M}$, and the slope parameter of a linear model fitted to the time-averaged light curve, ∇_S. A source was identified as a variable candidate if both metrics exceeded a threshold. However, since typical light curves are not adequately modeled as straight lines, model misspecification bias will cause an underreporting of candidates.

Other large surveys conducted by radio telescopes such as the Australian Square Kilometre Array Pathfinder (T. Murphy et al. 2021), Murchison Widefield Array (M. E. Bell et al. 2014), Low-Frequency Array (J. D. Swinbank et al. 2015), and MeerKAT (A. Andersson et al. 2023) have also used two measures: the flux density coefficient of variation, V_ν, and the reduced χ^2 statistic of variability, η_ν, to quantify the significance of the epoch-to-epoch variability of sources. The former is mathematically equivalent to the modulation index, M, and the latter is similar to a Student t-test approach except that all data points are simultaneously compared with a reference value to produce a single statistic for the entire light curve. Since V_ν and η_ν are commonly used in radio astronomical surveys, we shall restate their formulations and use them to benchmark our approach.

For a light curve that consists of N time-ordered flux density measurements $S_{i,\nu}$, $i = 1, \ldots, N$, at frequency ν with uncertainties $\sigma_{i,\nu}$, V_ν is defined as the sample standard deviation s divided by the overall mean $\overline{S_\nu}$:

$$V_\nu = \frac{s}{\overline{S_\nu}} = \frac{1}{\overline{S_\nu}} \sqrt{\frac{N}{N-1}\left(\overline{S_\nu^2} - \overline{S_\nu}^2\right)}. \tag{1}$$

The statistic η_ν is defined as

$$\eta_\nu = \frac{1}{N-1}\sum_{i=1}^{N}\frac{(S_{i,\nu} - \bar{S_\nu}^*)^2}{\sigma_{i,\nu}^2} \sim \chi^2_{N-1}, \tag{2}$$

where $\overline{S_\nu}^*$ denotes the weighted mean such that for weights $w_{i,\nu} = 1/\sigma_{i,\nu}^2$,

$$\bar{S_\nu}^* = \frac{\sum_{i=1}^{N}w_{i,\nu}\,S_{i,\nu}}{\sum_{i=1}^{N}w_{i,\nu}}. \tag{3}$$

Following the guidance of M. E. Bell et al. (2014), sources with high values of both V_ν and η_ν are considered promising transient or variable candidates. A high value of V_ν suggests that the source was observed to brighten (or dim) to a large multiple (or fraction) of its usual level, and a high value of η_ν suggests that the observed size and number of variations in flux density are inconsistent with a source that is nonvarying.

The main advantages of these two metrics are that they are easy to calculate and straightforward to deploy into the data-processing pipeline of large surveys. However, as is evident from Equations (1)–(2), and in common with the methods mentioned earlier, none of these metrics incorporates any time information and, therefore, amount to summarizing light curves into a coarse single-number statistic. This information loss dilutes their discriminatory power.

In contrast, some studies have used statistical learning for the automated clustering and classification of light curves (e.g., A. Rowlinson et al. 2019). Such methods have proven effective when confined to the specific survey from which the training data were obtained; however, they are susceptible to overfitting and cannot be easily translated to new surveys without being retrained.

To avoid both oversimplification and overspecification, we propose using GPs to characterize light curves. A GP representation of light curves will implicitly model the covariance between observations, the temporal information that a (η_ν, V_ν) parameterization discards. Furthermore, careful choice of covariance kernel functions will mitigate overfitting. Finally, light-curve fits from a Bayesian hierarchical modeling (BHM) approach using GP priors will fully incorporate both observational and parameter estimation uncertainties into any posterior predictions.

1.2. GPs in Time-domain Astronomy

Developed by D. G. Krige (1951) for geostatistics, the interested reader is directed to C. E. Rasmussen & C. K. I. Williams (2006, Chapter 2) and G. Matheron (2019, Chapter 4) for a full mathematical treatment of GP regression. Recently, GPs have become popular for analyzing astronomical time series (S. Aigrain & D. Foreman-Mackey 2023) with applications in estimating the value of astrophysical parameters, modeling components of variability in a source, and detecting specific phenomena. Examples of the latter are the numerous studies searching for exoplanets and stellar companions that have used GPs to model the variability of stellar light curves for determining rotation periods and radial velocities (I. Czekala et al. 2017; R. Angus et al. 2018; A. Bortle et al. 2021). Similarly, GPs have been used to model
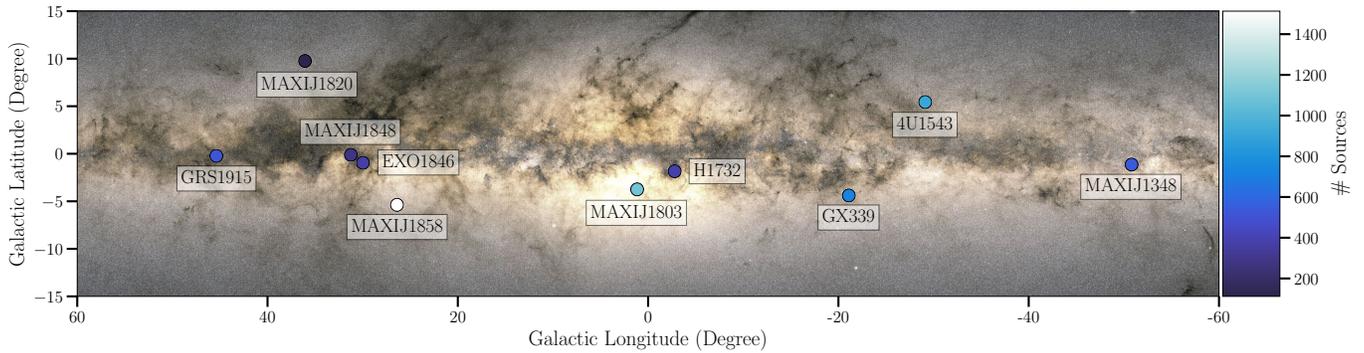
Figure 1. Projection onto the Galactic plane of the ThunderKAT fields of observation included in this work. The color bar indicates the number of sources for which we have light curves in each field (A. Andersson et al. 2023). Fields are labeled according to their observation target. Background image from the Gaia mission (A. Moitinho; ESA/Gaia/DPAC).

and detect (quasi-)periodicity in active galactic nuclei (AGN; S. Covino et al. 2020, 2022; A. B. Kovacevic et al. 2020). In gravitational wave research, C. J. Moore et al. (2016) and V. D'Emilio et al. (2021) used GP modeling to estimate parameters such as chirp mass. Another common application of GPs is in data preprocessing before further scientific analysis. This includes the disentangling of "nuisance" signals, that is, signal components attributed to factors not of immediate scientific interest, such as in pulsar timing studies (R. van Haasteren & M. Vallisneri 2014; J. Antoniadis et al. 2022), or correcting for systematic oscillations due to mechanical faults (B. J. S. Pope et al. 2016). Concerning sparse data, interpolating light curves using GP regression has improved the accuracy of automated supernovae classifiers (V. A. Villar et al. 2020; H. F. Stevance & A. Lee 2023).

However, in each of these applications, the GP model has been tuned to a specific type of signal or waveform. These models are therefore biased toward detecting particular source types at the expense of others. In this work, we propose a general-purpose model that can characterize variable sources that exhibit various behaviors, such as slowly evolving, bursting, persistent, and ultralong periods.

Another contrast with the studies mentioned above is that we use GP hyperparameters as direct descriptors of light curves, instead of using GPs as only a means to interpolate poorly sampled observations. Similar to the approach of S. A. J. McLaughlin et al. (2024), who used raw hyperparameter values to characterize and classify AGN flares, we focus on using estimated hyperparameter values as the primary descriptors of light curves.

2. Data and Methodology

2.1. ThunderKAT Data

We used the radio light curves of 6394 sources from commensal survey data captured by ThunderKAT (R. Fender et al. 2017). Between 2018 and 2022, ThunderKAT was scheduled on MeerKAT to observe XRBs, cataclysmic variables, short gamma-ray bursts, and Type Ia supernovae. For each of the targeted sources, multiple epochs of data were collected on the hundreds of radio sources found in the 1 deg^2 fields of view centered on these targets. The data used in this study were made available through the citizen science project by A. Andersson et al. (2023) and are publicly accessible online (A. Andersson 2024). As a commensal survey, all light curves are labeled with the name of the XRB on which the

field of view of the observation was centered (Figure 1): 4U 1543−47 (4U1543), EXO 1846−031 (EXO1846), GRS 1915 +105 (GRS1915), GX 339−4 (GX339), MAXI J1848−015 (J1848G), Swift J1858.6−0814 (J1858), MAXI J1348−630 (MAXIJ1348), MAXI J1803−298 (MAXIJ1803), H 1743-322 (H1732), and MAXI J1820+070 (MAXIJ1820).

Data for each source consists of flux densities at 1.28 GHz (*L* band) and their corresponding standard errors. These flux densities are peak values calculated within the synthesized beam of the source in radio images at each epoch, and the uncertainties were derived from fitting elliptical Gaussians to each island of pixels that constituted a source (J. D. Swinbank et al. 2015). Example light curves are shown in Figure 2, and cutouts of typical radio images may be found in A. Rowlinson et al. (2019), A. Andersson et al. (2022), and L. N. Driessen et al. (2022). The number of points in the light curves ranged from one to 166 data points, and only detections are used in this analysis. Observations ranged from 1 to 1296 days. Furthermore, most of these light curves had multiple uneven gaps in their observations, lasting from 1 to 226 days. Recorded flux densities spanned from the noise limit of the telescope up to 1.24 Jy, with a similar range for the observational uncertainties.[8]

We restricted our analysis to light curves with at least five data points, leaving $n = 5371$ light curves for analysis. This ensured sufficient information was available to fit a model and eliminated convergence problems in the fitting procedure. Before analysis, the flux density values were standardized by subtracting a light curve's sample mean and dividing it by its sample standard deviation. Similarly, the observational standard errors were standardized by dividing by the sample standard deviation. Consequently, all model fitting was performed in standardized units of source brightness before being transformed back into flux densities to report results. As described in Appendix A, standardization helps to improve the numerical stability for computation and simplifies the specification of prior distributions.

2.2. Bayesian Hierarchical Model

This section briefly describes the BHM we have used to characterize light curves. We adopted the framework of C. Wikle et al. (2019), which organizes model hierarchies

---

[8] A small fraction of flux densities were ≲50 μJy—below the approximate sensitivity limit of individual ThunderKAT observations—and therefore likely to be spurious. Our analysis retained these since spurious measurements are typical of large surveys.
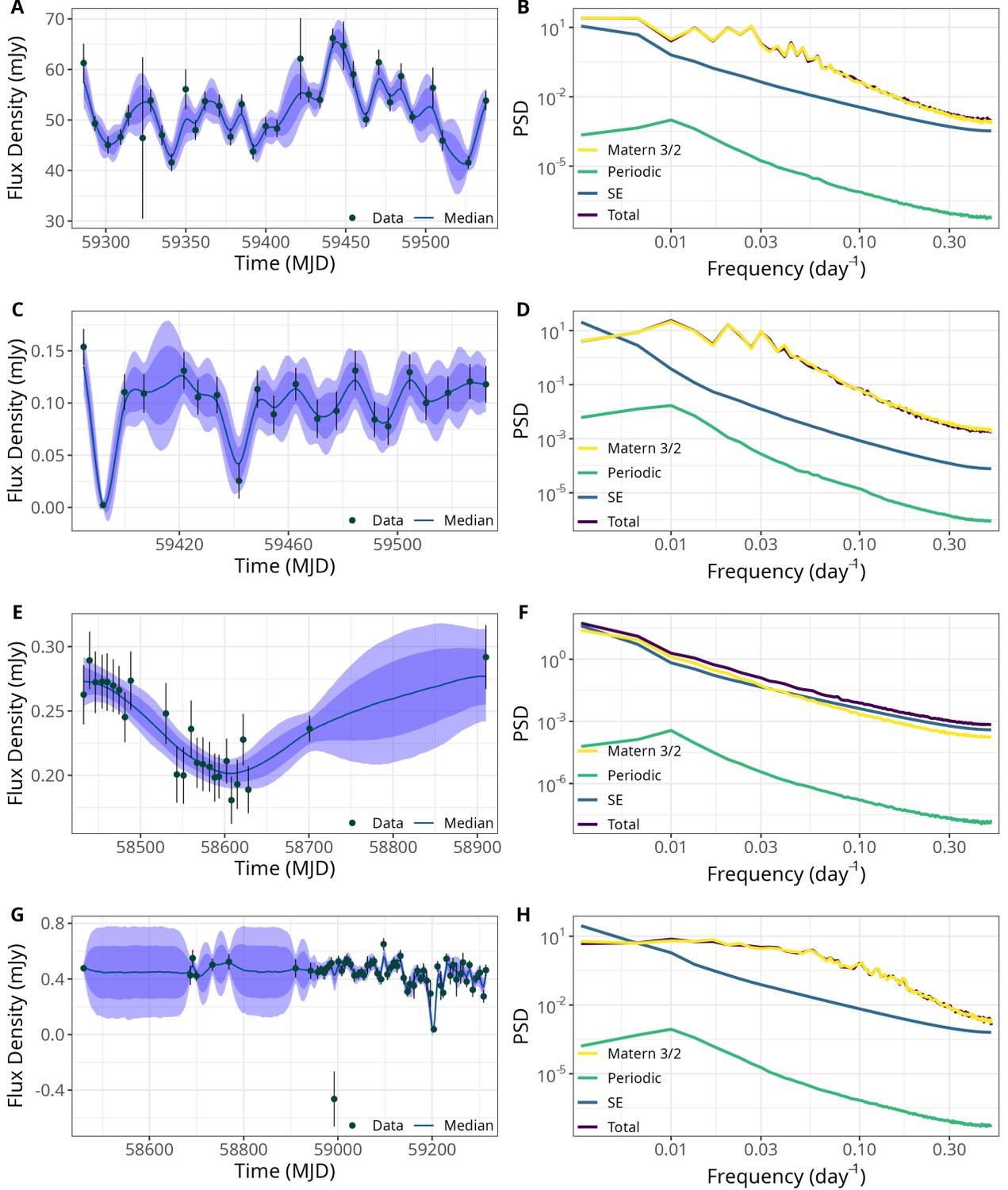
Figure 2. Left: posterior predictive samples of light curves in the fields around (A) J1848G, (C) 4U1543, (E) J1858, and (G) GRS1915. Shaded regions are 68% and 90% quantile intervals, respectively. Standardized results have been transformed back to their original flux density scale. The inconsistent size of uncertainties is due to poor data quality in certain epochs. Right: power spectral density (PSD) estimates of the posterior predictive curves in the left panel. Note that, in D and H, numerical approximation artifacts misleadingly show the squared exponential kernel exceeding the total PSD. See Table 1 for the statistics of each source.

into three levels: a *data model*, specifying the probabilistic model for the data; a *process model*, which sets out the unobserved, or latent, process of interest; and a *parameter model*, where the prior and hyperprior distributions are specified. These are described in Equations (4)–(11).

In Bayesian inference, we update prior distributions of the model parameters by using observed data to produce posterior distributions for each parameter. We then calculate the *posterior predictive distribution* of fitted light curves from these posterior distributions. Curves sampled from this

posterior predictive distribution may be compared to the originally observed light curve, providing a way to assess the plausibility of the fitted model.

2.2.1. Data Model

For each source in the ThunderKAT survey, we assumed its flux density S_i at time t_i, $i = 1, \ldots, N$ to be a Gaussian distributed random variable with a (latent) mean f_i and variance $\hat{e}_i{}^2$ (Equation (4)). Note that the variance of each flux density measurement is not an estimated parameter but the reported standard error $\hat{e}_i$ from the observed light curve.

Data model

$$S_i \sim \mathcal{N}(f_i, \hat{e}_i{}^2), \quad i = 1, \ldots, N. \tag{4}$$

Process model

$$\boldsymbol{f} \sim \mathcal{GP}(\boldsymbol{0}, \boldsymbol{K}_{N \times N}) \tag{5}$$

$$
\begin{aligned}
{[K]}_{r,c} = \kappa(t_r, t_c | \boldsymbol{\theta}) &= \sigma_{\mathrm{SE}}^2 \exp\left\{ -\frac{1}{2} \frac{(t_r - t_c)^2}{\ell_{\mathrm{SE}}^2} \right\} \\
&+ \sigma_{\mathrm{M32}}^2 \left(1 + \sqrt{3} \, \frac{|t_r - t_c|}{\ell_{\mathrm{M32}}} \right) \exp\left\{ -\sqrt{3} \, \frac{|t_r - t_c|}{\ell_{\mathrm{M32}}} \right\} \\
&+ \sigma_{\mathrm{P}}^2 \exp\left\{ -\frac{2}{\ell_{\mathrm{P}}^2} \sin^2\left(\pi \frac{|t_r - t_c|}{T} \right) \right\}, \quad r, c = 1, \ldots, N.
\end{aligned}
\tag{6}
$$

Parameter model

$$\sigma_{\mathrm{SE}}, \sigma_{\mathrm{M32}}, \sigma_{\mathrm{P}} \sim \mathcal{N}^+(0, 1) \tag{7}$$

$$\ell_{\mathrm{SE}}, \ell_{\mathrm{M32}}, \ell_{\mathrm{P}} \sim \mathrm{InvGamma}\left(\alpha = 3, \beta = \frac{1}{2} \times \lceil \mathrm{range}(t) \rceil \right) \tag{8}$$

$$T \sim \mathrm{Uniform}\left(2 \times \min(\Delta t), \frac{1}{2} \times \mathrm{range}(t) \right) \tag{9}$$

$$\ell_{\mathrm{SE}}, \ell_{\mathrm{M32}}, \ell_{\mathrm{P}} > \min(\Delta t) \tag{10}$$

$$\ell_{\mathrm{SE}} > \ell_{\mathrm{M32}} \tag{11}$$

BHM used in this study. S_i is the observed flux density of the ith data point in a light curve whose true (but unobserved) flux density is f_i. The uncertainty in that data point is $\hat{e}_i$. $[K]_{r,c}$ is the covariance kernel of the GP prior for f, which is parameterized by $\boldsymbol{\theta}$, which is itself decomposed into three kernel terms: squared exponential (SE), Matern 3/2 (M32), and periodic (P). Each term has its own hyperparameters: σ_{SE}, σ_{M32}, and σ_{P}, are amplitude hyperparameters; ℓ_{SE}, ℓ_{M32}, and ℓ_{P}, are length-scale hyperparameters; and T is the period hyperparameter. The quantity range (t) is the total duration of a light curve, and $\min(\Delta t)$ is the shortest time gap between points in a light curve.

2.2.2. Process Model

Since the latent f is an unknown time-varying function, we model it as a stochastic process. More specifically, as shown in Equation (5), we model it as a GP defined with a zero mean function and an $N \times N$ covariance matrix $\boldsymbol{K}_{N \times N}$. Each (r, c)th matrix element $[K]_{r,c}$ is specified by the kernel function $\kappa(t_r, t_c | \boldsymbol{\theta})$ in Equation (6). GPs are often used with a zero mean function, though prior knowledge of the underlying phenomena may justify specifying a parametric or nonparametric function instead (C. E. Rasmussen & C. K. I. Williams 2006,

p. 27). Depending on the choice of mean and kernel functions, a GP can generate flexible curves to fit various light-curve characteristics, such as high-frequency fluctuations, smooth monotonic trends, and quasiperiodicity. Our choices for the mean and kernel functions are justified below.

Mean function. Surveys such as ThunderKAT contain thousands of sources whose light curves exhibit a wide range of behaviors (G. J. Babu & A. Mahabal 2016). These include long-term trends, broad peaks or troughs, sharp bursts, repeating patterns, and high-frequency oscillations. Examples of these are illustrated in Figure 2. Furthermore, different fields of view have different numbers of pointings and data points. This means that no single parametric shape or pattern can be reliably assumed for an arbitrarily selected light curve. Consequently, in our GP regression model, we chose a zero mean function to avoid biasing our model toward any particular type of astrophysical phenomenon. Specifying a mean function of zero is consistent with having no specific knowledge about the shape or pattern of the modeled light curve[9] but does assume that the light curve has a stationary mean; that is, it neither trends upward nor downward. However, as shown in the third panel of Figure 2, some ThunderKAT light curves appear to brighten or dim during their observation. Nevertheless, some nonstationary behavior can be addressed by using kernel functions tuned to appropriately long length scales. This study uses the SE kernel term described below for this purpose.

In analyses of phenomena whose typical light-curve shapes are known a priori, a GP model can be employed to model the deviations from some underlying nonzero mean function. Examples include modeling the systematics in exoplanet studies, such as that by B. J. S. Pope et al. (2016), or the characterization of quasiperiodic oscillations in X-ray light curves of gamma-ray bursts, as done by F.-F. Song & J. Mao (2024). In this study, by specifying a zero mean function, we are directly modeling the observed flux densities and not the residuals.

Kernel function. The kernel function specifies the covariance relationship between data points in a GP. Properties of the kernel function, such as smoothness or periodicity, affect the final fit of the model, so the choice of kernel function strongly influences the results (W. T. Stephenson et al. 2022). Following C. E. Rasmussen & C. K. I. Williams (2006), we chose a composite kernel function comprising three distinct terms. Each constituent kernel term reflects a behavior we expect to observe in light curves, such as smooth long-timescale trends, jagged short-timescale fluctuations, and, sometimes, periodicity. The covariance function κ used in this analysis is described in Equation (6) and consists of the sum of three kernel terms: SE, M32, and P. These terms are a function of seven hyperparameters $\boldsymbol{\theta} = (\sigma_{\mathrm{SE}}, \ell_{\mathrm{SE}}, \sigma_{\mathrm{M32}}, \ell_{\mathrm{M32}}, \sigma_{\mathrm{P}}, \ell_{\mathrm{P}}, T)$.

Trends and smooth variations at scales comparable to the total length of observation are modeled by the SE kernel, whose length-scale hyperparameter is constrained to favor that scale. This is a straightforward way to address nonstationarity in the mean, which cannot be accommodated by the zero mean function mentioned earlier. Jagged fluctuations are modeled by the M32 kernel constrained to short length scales. Finally, P effects are encapsulated by the P kernel.

---

[9] Standardized to have a mean of zero.

Including both an SE kernel and an M32 kernel term may introduce mathematical identifiability problems. This is partly due to the close functional relationship between these two kernels: the SE kernel can be imagined as a Matern kernel of "infinite order," and the SE kernel can be likened to a Matern kernel with extra smoothness constraints. Consequently, these kernels may recognize the same behaviors in a light curve, resulting in convergence problems or inflated uncertainty intervals. Details on how we mitigated this problem by imposing constraints on the hyperparameters are found in Appendix A.

Lastly, we assumed that light curves could be treated as (second-order) stationary stochastic processes. That is, their covariance structures are only a function of the difference in time $(t - t')$ between data points. More intuitively, this means that, regardless of which segment of a light curve is examined, its statistical properties are approximately the same as any other segment. However, verifying this assumption without knowledge of the underlying phenomenon proves challenging in astronomical contexts because data are often so sparse that metrics of nonstationarity are unreliable. The use of nonstationary GP kernels is a subject for future work.

2.3. Parameter Model

Since a GP can be thought of as a prior distribution over functions, we will refer to the parameters of the kernel function as *hyper*parameters and the prior distributions assigned to these as *hyper*priors. These hyperpriors are shown in Equations (7)–(9). The notation range (t) indicates the total duration of a light curve (in days), and $\min(\Delta t)$ is the shortest time gap between points in a light curve (in days).

In addition to the choice of kernel function, the choice of hyperpriors will also influence the results of GP regression modeling. For example, if the data are sparse and the hyperpriors are strongly informative, these hyperpriors dominate the hyperparameter posterior distribution. Fortunately, it is straightforward to assess the influence of hyperpriors by comparing them with the posterior distribution of the hyperparameter, and a detailed justification of our choice of hyperpriors is provided in Appendix A.

2.4. Posterior Distributions

The output of Bayesian inference is the joint distribution of all the parameters of interest. From this, we can derive estimates of particular hyperparameters expressed as posterior distributions. These distributions are useful and interpretable in their own right, but can also be used to generate so-called posterior predictive samples. These posterior predictive samples are curves that conform to the constraints of both the parameter model and the observed data and are useful for assessing the plausibility of model fit.

2.4.1. Hyperparameter Posteriors

The posterior distributions of the seven hyperparameters in the GP regression model can be interpreted as descriptors of particular aspects of the original light curve. For example, if the posterior distribution of the length-scale parameter for the SE kernel $p(\ell_{SE}|y)$ has large values, it suggests that the light curve varies smoothly because of the strong autocorrelation between distant points. This can be extended to exploring the relationships between hyperparameter posteriors, which may

reveal novel formulations for characterizing astronomical sources.[10] For example, in the bivariate case, GP regression modeling amounts to a dimension reduction procedure that situates every light curve in a two-dimensional subspace of the seven-dimensional hyperparameter space where statistically (and perhaps astrophysically) similar sources might form clusters. In Section 4.1.2, we describe a two-dimensional subspace that partitions out variable sources and provides an alternative to using the η_ν and V_ν statistics discussed in Section 1.1.

2.4.2. Posterior Predictive Curves

After hyperparameter posterior distributions have been obtained for each fitted light curve, new ("predicted") curves y^* can be generated from the posterior predictive distribution $p(y^*|y)$. These predicted curves are conditioned on the observed data, and their distribution accounts for both the uncertainty in the data and in the estimated hyperparameters. C. E. Rasmussen & C. K. I. Williams (2006, p. 17) derive an analytical expression for the posterior predictive distribution $p(y^*|y)$, which makes drawing these samples both straightforward and efficient.

Posterior predictive curves are useful for assessing the fit of a GP regression model to the original data. This is done by inspecting the envelope of curves formed by repeated sampling of curves from the posterior predictive distribution, where the envelope's width reflects the uncertainty of the fit at any point along the curve. For example, in the left panels of Figure 2, we observe that the 68% quantile intervals of the posterior predictive samples encapsulate most of the original data (aside from the points with very large uncertainties), which suggests that the model is a plausible description of the data.

2.4.3. Power Spectral Densities

In addition to providing a visual diagnostic for goodness of fit in the time domain, curves sampled from the posterior predictive distribution can be used to calculate power spectral densities (PSDs). Furthermore, we can estimate both the overall PSD and the individual contribution of each term in the kernel function. The relative contributions of the individual components can be compared by plotting them alongside the overall PSD, as in the right-hand panels of Figure 2. In these examples, the M32 curve typically dominates the other two terms, and the very weak P kernel responses suggest an absence of periodicity.

Thus, inspecting the full and component-wise PSDs is another way to validate the model fit and scrutinize anomalies that might be difficult to identify in time-domain visualizations. This is particularly true for P patterns. However, care must be taken because the finite duration of light curves introduces fringing and other sampling artifacts into PSDs.

2.5. Implementation

The model fitting in this study was implemented in both PYTHON and R to allow validation and facilitate wider deployment. It was implemented in PYTHON (v3.10) with the PYMC library (v5.6.1; T. Wiecki et al. 2023) using the default No U-turn Sampler (M. D. Hoffman & A. Gelman 2011). It

---

[10] In this work, we have used the median of the hyperparameter posterior distributions.

Table 1
Fitted Hyperparameter Values for the Example Light Curves Shown in Figure 2

Figure 2	R.A. (deg)	Decl. (deg)	Field	η_ν	V_ν	Posterior Median						
						σ_{M32}	ℓ_{M32}	σ_{SE}	ℓ_{SE}	σ_{P}	ℓ_{P}	T
A	281.9044	−2.0325	J1848G	2.91	0.12	**1.20**	12.5	0.35	48.5	0.45	37.6	85.6
C	236.3573	−48.1665	4U1543	0.52	0.27	**1.38**	7.74	0.36	28.8	0.47	22.0	50.1
E	284.4313	−8.0519	J1858	0.94	0.15	0.56	73.4	**0.82**	117.3	0.53	72.39	165.0
G	288.3110	11.4117	GRS1915	20.24	0.36	**1.21**	17.3	0.31	171.7	0.42	128.6	292.2

Note. Medians of the hyperparameter posteriors are shown beside the variability statistics of η_ν and V_ν. Boldface indicates median amplitudes greater than 0.67, a rule-of-thumb threshold above which a source is nominally labeled as a variable or transient.

was implemented in R (v4.4.1; R Core Team 2023) and CMDSTAN (v2.35.0; Stan Development Team 2023) using the CMDSTANR package (v0.8.1; J. Gabry et al. 2024) and the default Hamiltonian Monte Carlo sampler.

For both implementations, hyperparameter estimation was performed using chains of 2000 samples after discarding the first 2000 samples. Known as "burn-in" samples, these initial samples are discarded because they are likely biased due to the time it takes Monte Carlo chains to reach and explore the high-density regions of the parameter space. The rank-normalized $\hat{R}$ statistic (A. Gelman & D. B. Rubin 1992; A. Vehtari et al. 2021) was used as a convergence diagnostic, and the effective sample size (ESS) was used to measure sampling efficiency. These metrics are provided by default in both PYMC and CMDSTAN.

The computational cost of GP regression is known to scale at order $\mathcal{O}(n^3)$, where n is the number of input data points. In practice, we observed that ThunderKAT light curves are sparse enough that it is not the regression computation that is the most time consuming; rather, it is sampling the posterior predictive distribution that is the most onerous. This work has sped up this latter step by implementing a *marginal* GP regression model, as described in Appendix B.

3. Results

In fitting the GP regression model described in Section 2, we segment the total variability exhibited by a light curve into components related to the amplitude and timescale of change. In the following sections, we present the results of light-curve fitting, particularly the medians of posterior estimates of individual hyperparameters. All hyperparameters were estimated using Markov Chain Monte Carlo sampling, and, for 99% of cases, we achieved reasonable convergence and efficiency measures of $\hat{R} \leqslant 1.003$ and bulk ESS > 1753, respectively.

3.1. Light-curve Fits

Before using the hyperparameter posteriors, confirming the plausibility of the model fits is important. This is accomplished by visually inspecting plots of curves sampled from the posterior predictive distribution. Conveniently, the additive nature of our GP model means that we can also plot the PSD of any posterior predictive curve as the sum of its constituent kernel terms. Seeing the relative contributions of each term in the frequency domain complements the interpretation of the fitted model in the time domain.

Below, we describe the results of fitting the GP regression model specified in Equations (4)–(11) to four ThunderKAT light curves. These were chosen because they differ by

sampling rate, nonstationarity, presence of outliers, and distribution of observational errors. Each example highlights how our GP model accommodates a diversity of light-curve features that may arise from astrophysical processes, data quality issues, or both.

Table 1 lists the medians of the fitted posteriors for the seven fitted hyperparameters of each light curve, namely, {σ_{SE}, ℓ_{SE}, σ_{M32}, ℓ_{M32}, σ_{P}, ℓ_{P}, T}. The table also lists their respective η_ν and V_ν statistics for comparison. As discussed in Appendix A, amplitude posterior estimates with medians above 0.67 warrant further interest and are highlighted in boldface in Table 1. Further examples of model fits to known XRB sources are included in Appendix C.

The light curve in panel (A) of Figure 2 features numerous undulations on the scale of $15-20$ days and is certainly a variable source. This is reflected in the estimated posterior median of the amplitude hyperparameter for the M32 kernel, $\sigma_{\mathrm{M32}} = 1.20$, which is greater than our nominal threshold of significance of 0.67. Notice that the model has closely fitted these undulations and that most of the original observations lie within the 90% uncertainty interval. Of note is the outlier near MJD 59324, which has a large observational uncertainty and is consequently ignored by the model in favor of its more precise neighbors. A similar relaxation of the model fit is observed around MJD 59424 and reflects how GP regression in a Bayesian framework directly incorporates uncertainties. Indeed, this is in common with other modeling approaches that weight the contribution of individual observations by their respective standard errors. These uncertainties are further propagated into the posterior estimates.

The PSD in panel (B) of Figure 2 confirms that most of the variability is explained by the M32 kernel term. In contrast, the P component is very weak. This is consistent with the estimated amplitude posterior medians: $\sigma_{\mathrm{M32}} = 1.20$, $\sigma_{\mathrm{SE}} = 0.35$, $\sigma_{\mathrm{P}} = 0.45$, where the former is above 0.67 and the latter are below. This implies that the model has detected a nontrivial yet nonperiodic correlation structure in the light curve.

The light curve in Figure 2(C) is also regularly sampled but has fewer observations than the curve in panel (A). This example shows that fitting to fewer observations results in wider uncertainty intervals. Again, there are two outliers, but these have smaller standard errors compared with the remaining data points than the outliers in panel (A). Consequently, the model has deflected sharply toward these observations rather than ignoring them. Inspection of the radio image for that first apparent outlier revealed it as intrinsic variation in the source and not an image artifact to be ignored. The large median value of $\sigma_{\mathrm{M32}} = 1.38$ again implies the presence of nonnegligible variability, which contrasts with its
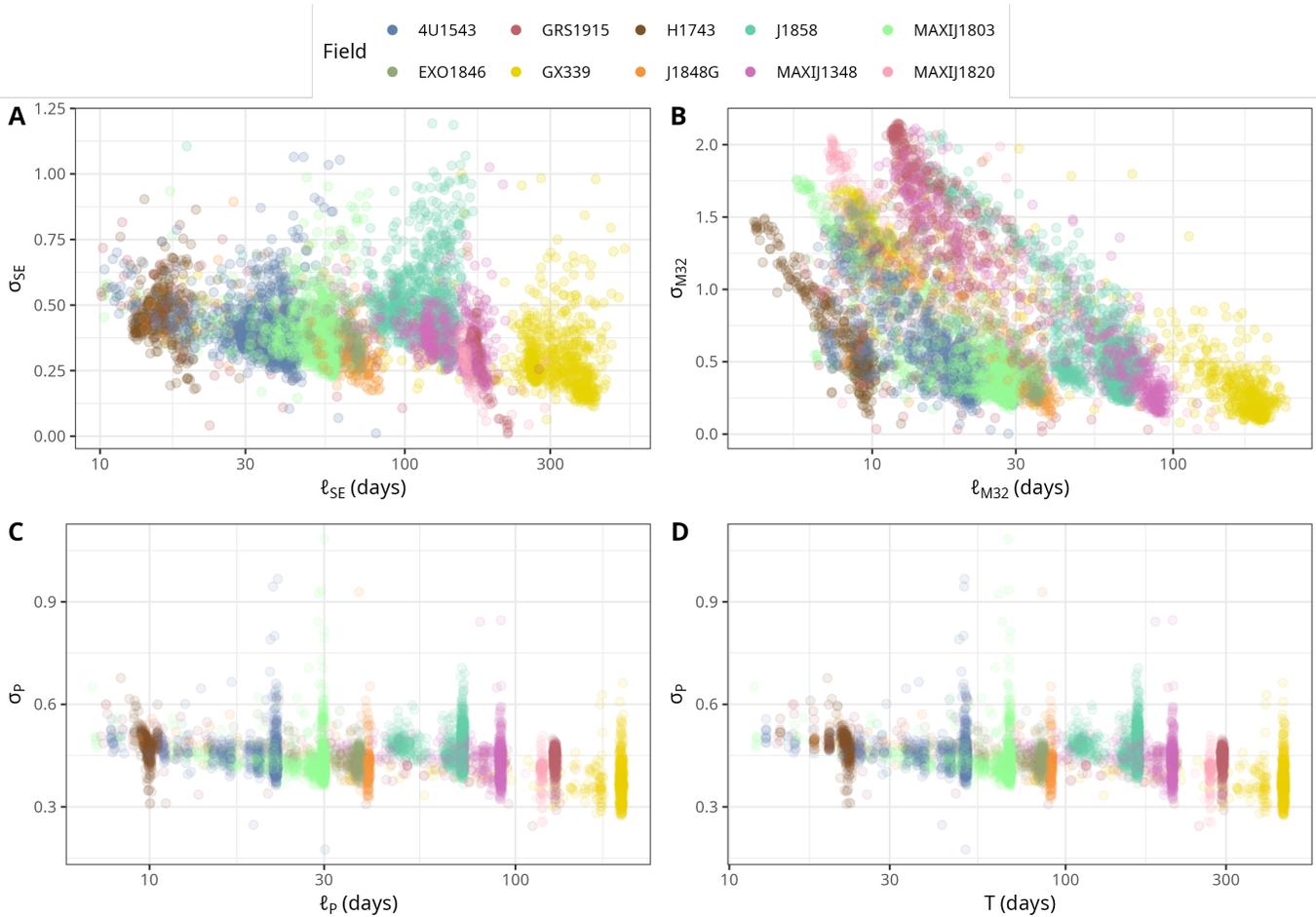
Figure 3. Posterior medians of our GP regression model's hyperparameters from each kernel term. (A): SE kernel, (B): M32 kernel, (C) and (D): P kernel. Each point corresponds to one light curve. Amplitudes, σ, are in standardized units, and length scales ℓ and period T are in days on a logarithmic scale. Colors indicate the field in which the light curve was observed. Notice the distinct clustering associated with the field of observation.

small and unremarkable η_ν and V_ν statistics of 0.52 and 0.27, respectively.

Figure 2(E) shows a light curve that is unevenly sampled with widely spaced observations past MJD 58630. The model interpolates within these sparse data regions, but the uncertainty interval is wide. Overall, this light curve's first and second halves have a smooth down-then-upward trend. The moderately large estimates for amplitude ($\sigma_{\mathrm{SE}} = 0.70$) and length-scale ($\ell_{\mathrm{SE}} = 124.6$ days) hyperparameters encapsulate this behavior. This is also reflected in the left-to-right downward slope in its PSD. Again, the globally calculated variability statistics of ($\eta_\nu = 0.94$, $V_\nu = 0.15$) failed to detect the presence of this strong flux density inflection.

The light curve in Figure 2(G) is of another unevenly sampled source with sparse regions where the uncertainty intervals are again very wide. Of particular interest are two outlying observations. The value observed just before MJD 59000 is clearly erroneous because flux densities must not be negative. The model mostly ignores this point because of its large standard error. In contrast, the other outlying point around MJD 59200, with its very small reported standard error, drags the model into a trough shape. The effect of this second outlier on inflating $\eta_\nu = 20.24$ and $V_\nu = 0.36$ is modest, making this light curve unremarkable according to these measures. The short-timescale, jagged behavior exhibited in

the latter half of this light curve is reflected in the hyperparameters of the M32 kernel, namely a large $\sigma_{\mathrm{M32}} = 1.21$ and short $\ell_{\mathrm{M32}} = 17.3$ days. This behavior is also captured in the PSD in panel (H), where the M32 kernel response dominates the SE and P kernels.

3.2. Hyperparameter Posteriors

3.2.1. SE Kernel

Figure 3(A) shows the scatter plot of the two estimated hyperparameters of the SE kernel, namely ℓ_{SE} and σ_{SE}. Recall that these are from the first term of the kernel function specified in Equation (6). Each plotted point corresponds to the median of the posterior distribution for each hyperparameter of a single ThunderKAT light curve. The different colors indicate the XRB in whose field the light curve was observed. An obvious pattern is the clustering by field, as evidenced by identically colored points appearing close to each other. Sources observed in the same field tend to have similar hyperparameter posterior medians. This is particularly true for the length-scale hyperparameter ℓ_{SE}, where clusters usually span distinct ranges in that dimension. Some fields, such as GX339 and J1858, also exhibit some internal clustering, which suggests that they, in turn, are composed of a mixture of subpopulations.
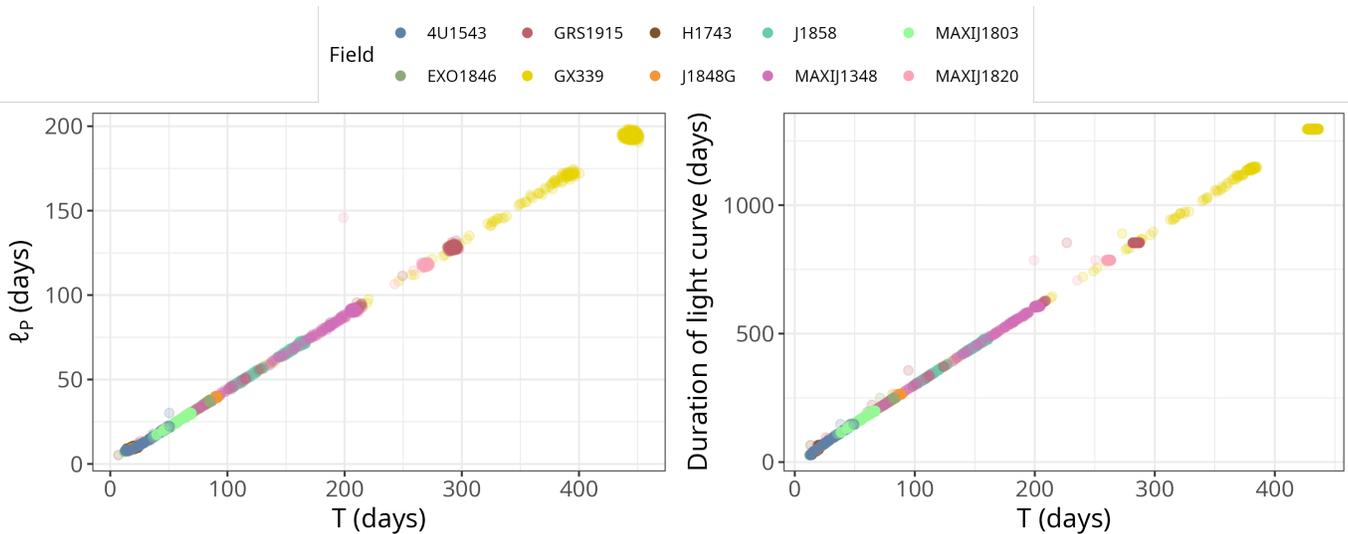
Figure 4. Left: scatter plot of the posterior medians of the period (T) and length-scale (ℓ_P) hyperparameters of the P kernel. These two quantities have a strong positive correlation ($R^2 = 0.996$). Right: scatter plot of the posterior median of T against the total duration of each light curve.

We further observe that the clusters mentioned above are not as well separated in the vertical σ_{SE} direction as in the horizontal ℓ_{SE} direction. Indeed, many different fields have overlapping ranges of amplitudes. In contrast, fields are much better delineated by length scale. For example, fields 4U1543 and J1858 have clear separation in ℓ_{SE}, but have ranges in σ_{SE} that heavily overlap.

Visual inspection of the bivariate relationship between ℓ_{SE} and σ_{SE} reveals no obvious correlation. Changes in ℓ_{SE} are not accompanied by a consistent change in σ_{SE}. However, with increasing ℓ_{SE}, there does appear to be an increase in the variance of σ_{SE} with a right skew—amplitudes below 0.5 units are favored across all fields and especially in GX339. However, within each field, the lack of strong patterns suggests that these two hyperparameters are mostly uncorrelated when conditioned by the field of observation.

One explanation for the clustering by field is the presence of confounding effects attributable to the field. For example, each field has its particular duration and cadence of observation. Factors such as observational cadence are nuisance contaminants in the characterization of light curves since they are artifacts of the data collection process rather than properties of the source itself. These effects would bias any source identification results and should be removed. In Section 4.2, we discuss our finding that two hyperparameters (σ_{SE} and σ_{M32}) appear largely unaffected by field confounding.

3.2.2. M32 Kernel

The distribution of the hyperparameters for the M32 kernel has some distinct differences from that of the SE kernel. In Figure 3(B), there is a strong negative correlation between ℓ_{M32} and σ_{M32}. Increases in the length-scale hyperparameter usually accompany decreases in the amplitude hyperparameter. This suggests that, concerning the variability captured by the M32 kernel term, there is a continuum of light curves ranging from smoothly undulating (small σ_{M32}, long ℓ_{M32}) to steep and jagged (large σ_{M32}, short ℓ_{M32}). Light curves at the latter extreme may constitute good candidates for transient sources.

As with the SE hyperparameters, there is some clustering by field for the M32 kernel hyperparameters, with similar proportions of overlap. This is again more associated with the length scale rather than the amplitude, which is better mixed. Indeed, even if the coloring in Figure 3 were removed, distinguishing different fields horizontally by clusters would still be possible visually. This suggests that, like the SE kernel, the M32 kernel term is still biased by field-specific effects.

3.2.3. P Kernel

Visual inspection of the two bivariate scatter plots between the posterior medians of the three hyperparameters of the P kernel (Figures 3(C) and (D)) does not reveal any associations between the length scale and amplitude, or the period and amplitude hyperparameters. However, the similarity between these two plots suggests a strong correlation between length scale (ℓ_P) and period (T). This is confirmed in the left panel of Figure 4 by the almost perfect positive correlation ($R^2 = 0.996$) between these two hyperparameters.

One explanation for this strong correlation is again a confounding factor. The right panel of Figure 4 is a plot of T against the total duration of each light curve, the latter being a property of the data, not of the underlying source. This duration information has leaked into the hyperparameter posteriors through their respective hyperpriors (Equations (8) and (9)) and fully accounts for the estimated values of T and ℓ_P. This implies no periodicity was detected; otherwise, shorter periods or length scales would have been found.

4. Discussion

4.1. Amplitude Hyperparameters as Indicators of Transience and Variability

The novel aspect of this work is that we use the posterior medians of GP regression *hyperparameters* as the primary descriptors of light curves instead of employing a GP model as an intermediary step in a preprocessing pipeline. In this section, we argue that, of the seven hyperparameters estimated for each light curve, the amplitude hyperparameters of the
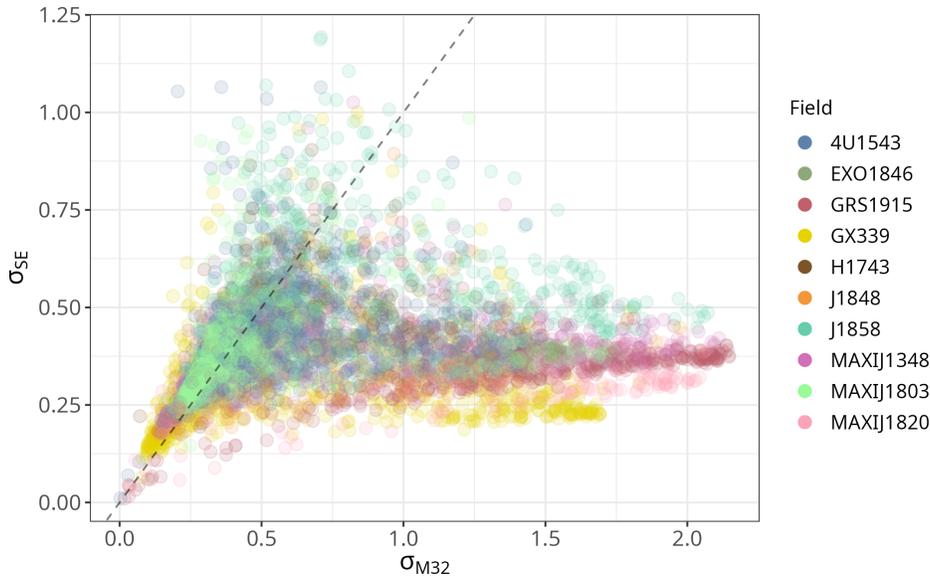
Figure 5. Scatter plot of the posterior medians of the amplitude hyperparameters of the SE (σ_{SE}) and M32 (σ_{M32}) kernels for each fitted light curve. Amplitudes are in standardized units, and colors indicate the field where the light curve was observed. The dashed line indicates the line of equality between the two hyperparameters.
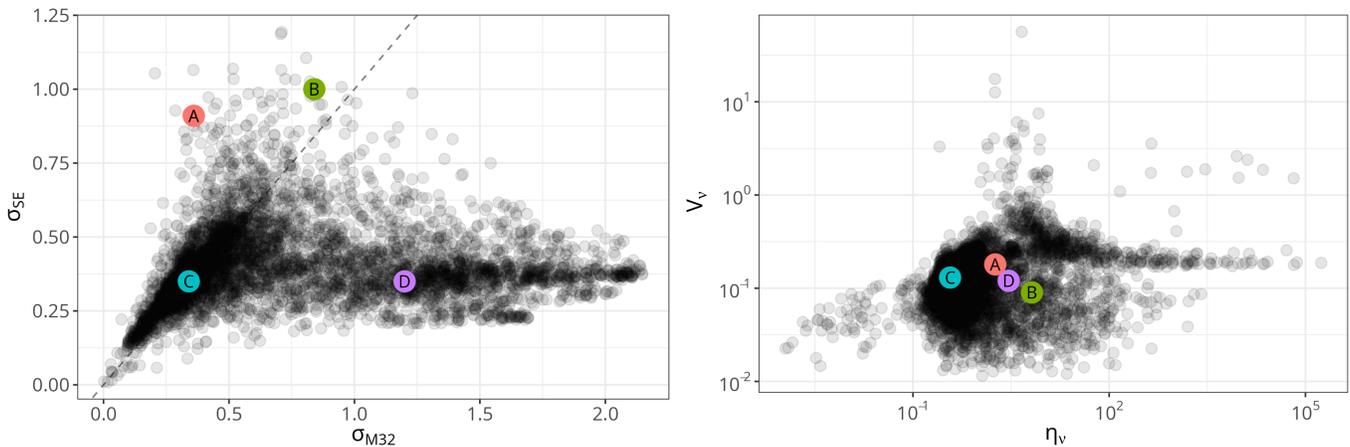

Figure 6. Left: scatter plot of the posterior medians of the amplitude hyperparameter of the SE and M32 kernels, where each point corresponds to a ThunderKAT light curve. Amplitudes are in standardized units, and the dashed line indicates equality between these two hyperparameters. Right: scatter plot of the (η_ν, V_ν) statistics for each ThunderKAT light curve. The labeled points indicate where in the parameter space the four light curves shown in Figure 7 reside. We observe that the four labeled light curves are better separated in (σ_{M32}, σ_{SE}) space than in (η_ν, V_ν) space. See Table 2 for metrics.

M32 and SE kernels show the most promise as robust indicators of variability. The location of sources in the two-dimensional subspace of (σ_{M32}, σ_{SE}) is a useful cue for identifying variable and transient candidates from among other sources whose observations do not exhibit much evidence of change over time. In Appendix D, we justify why using only two of the seven hyperparameters is sufficient to discriminate between variable and nonvariable sources.

4.1.1. Region of Stable Candidate Sources

The scatter plot of σ_{M32} against σ_{SE} in Figure 5 reveals a fan-shaped distribution with conspicuous clustering about the line of equality between these two hyperparameters. A representative light curve chosen from this elongated cluster, marked by the (C) in Figure 6, is plotted in panel (C) of Figure 7 and whose metrics are shown in Table 2. It shows a

very flat fitted curve. Indeed, the inspection of light curves located in this region of the hyperparameter space reveals that they are all similarly nonvarying or have observational uncertainties that are too large to classify them as variable or transient. Therefore, we have designated this region of the hyperparameter space as containing nominally "stable" sources.

Using the rule-of-thumb value of 0.67 as a threshold, the lower-left quadrant of this hyperparameter space where σ_{SE}, $\sigma_{M32} \leqslant 0.67$ covers approximately 58% of the total number of light curves analyzed in this study. We conjecture that this is a conservative lower bound on the proportion of sources observed in ThunderKAT that are neither variable nor transient. Sources in this quadrant should be deprioritized in any search for transients. The true boundary of this region requires a more precise definition and could be
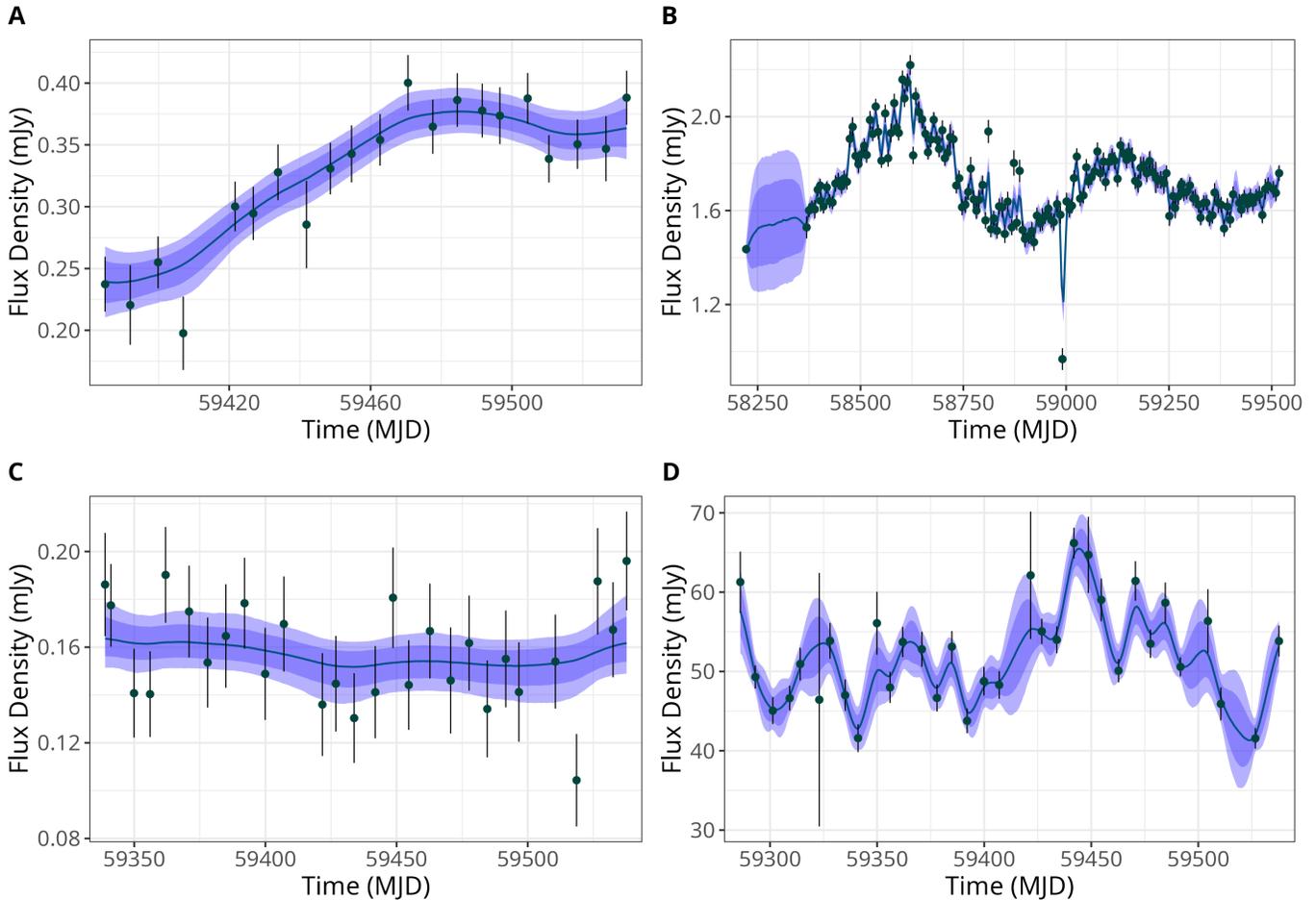
Figure 7. Posterior predictive samples of four light curves from the ThunderKAT survey. The dark curve traces the median of 2000 posterior predictive samples, and the shaded regions form the 68% and 90% intervals. The points and vertical bars indicate the original observations and their respective uncertainties. Each light curve is drawn from a distinct region of the (σ_{M32}, σ_{SE}) hyperparameter space, as labeled in Figure 6. (A): σ_{SE} is dominant. (B): Both hyperparameters are comparably large valued. (C): Both hyperparameters are small valued. (D): σ_{M32} is dominant. Specific coordinates and amplitude estimates are shown in Table 2.

Table 2
Amplitude Hyperparameter Values for the Light Curves Shown in Figure 7

Figure 7	R.A. (deg)	Decl. (deg)	Field	η_ν	V_ν	σ_{M32}	σ_{SE}
A	236.8389	−47.2731	4U1543	1.81	0.18	0.36	**0.91**
B	255.1565	−48.9462	GX339	6.59	0.09	**0.84**	**1.00**
C	270.3887	−29.7998	MAXIJ1803	0.37	0.13	0.34	0.35
D	281.9044	−2.0325	J1848G	2.91	0.12	**1.20**	0.35

Note. Medians of the amplitude hyperparameter posteriors are shown alongside the commonly reported variability statistics of η_ν and V_ν. Boldface indicates median amplitudes greater than 0.67, a rule-of-thumb threshold above which a source might be considered variable or transient. All panels except for (C) would be potential transient or variable candidates.

approximated by bootstrapping the joint distribution of (σ_{M32}, σ_{SE}) from light curves of sources that are known to be nonvarying.

4.1.2. σ_{M32} and σ_{SE} as Variability Heuristics

If nominally stable sources populate the lower-left quadrant of the (σ_{M32}, σ_{SE}) hyperparameter space, light curves of (statistically) varying sources should populate the other three quadrants. Visual inspection of light curves from those quadrants confirms this supposition, as shown in the remaining panels of Figure 7. In Figure 7(A), the light curve shows a long-timescale, smooth

upward trend with few undulations, and is characterized by large $\sigma_{SE} = 0.91$ and small $\sigma_{M32} = 0.36$. In Figure 7(D), this light curve with shorter timescale variations has small $\sigma_{SE} = 0.35$ and large $\sigma_{M32} = 1.20$. Finally, in Figure 7(B), where both hyperparameters are large, the light curve is a composite of very high-frequency jittering superimposed onto intermediate timescale trends. These examples demonstrate that, in addition to screening out nonvarying (or noisily indeterminate) sources, the magnitudes of the σ_{SE} and σ_{M32} amplitude hyperparameters are a useful heuristic for distinguishing between qualitatively different types of source behaviors.
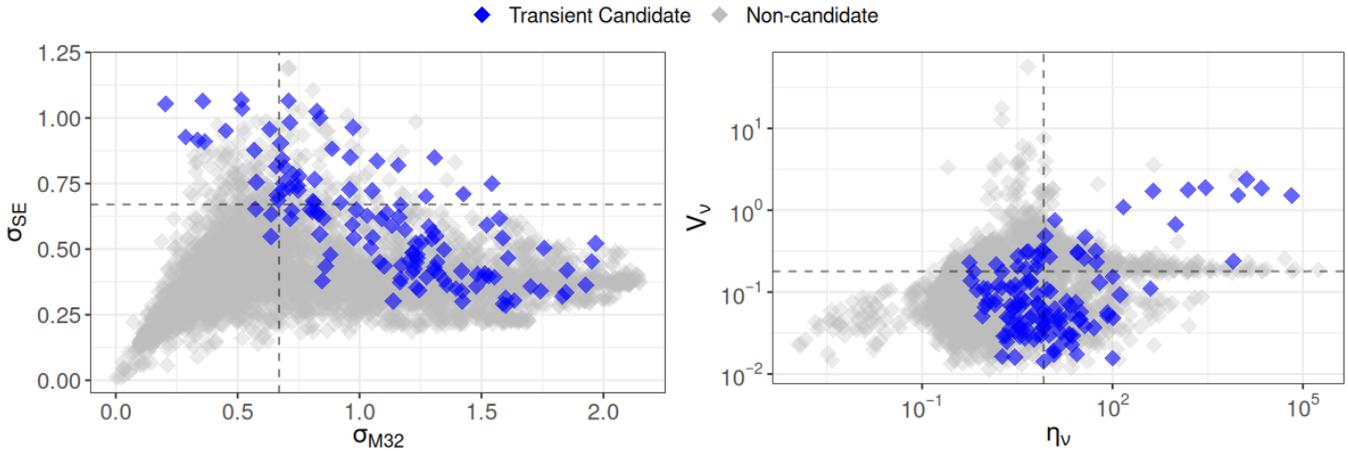
Figure 8. Left: scatter plot of the posterior medians of the amplitude hyperparameter of the SE and M32 kernels, where each point corresponds to a ThunderKAT light curve. Amplitudes are in standardized units, with dashed lines drawn at $\sigma. = 0.67$. Right: scatter plot of the (η_ν, V_ν) statistics for each ThunderKAT light curve with dashed lines at 8.22 and 0.179, respectively. Blue points are sources voted as probable variable or transient candidates by citizen scientists in A. Andersson et al. (2023). Gray points are sources that were not voted as variable or transient. Searches for transient or variable sources would typically avoid starting in the lower-left quadrants of these parameter spaces.

Table 3
Comparison of Detection Rates of Variable/Transient Sources, Treating the 126 Candidates Identified by Citizen Scientists in A. Andersson et al. (2023) as Ground Truth (Second Column from the Left)

Classification	A. Andersson et al. (2023)	$\sigma_{M32} > 0.67$ or $\sigma_{SE} > 0.67$	$\eta_\nu > 8.22$ or $V_\nu > 0.179$
Transient/variable	126 (100%)	123 (97.6%)	71 (56.3%)
Nontransient/variable	0	3 (2.4%)	55 (43.7%)

Note. Amplitude hyperparameter estimates (σ_{M32}, σ_{SE}) are thresholded at the rule-of-thumb value of 0.67; that is, sources with either hyperparameter larger than this value are considered scientifically interesting. The (η_ν, V_ν) metrics are thresholded according to the minimum values used by A. Rowlinson et al. (2019). The approach using amplitude hyperparameters has a much higher sensitivity of detection (97.6%) compared with the well-established (η_ν, V_ν) method (56.3%).

4.1.3. Comparison with Labeled Data

To validate our claim that (σ_{M32}, σ_{SE}) are useful as variability heuristics, we compared our findings with source classifications made by citizen scientists. In the study by A. Andersson et al. (2023), volunteers were asked to review the light curves and radio images of ThunderKAT sources and label them as stable, extended, transient/variable, artifacts, or uncertain. We used the transient/variable candidates flagged by citizen scientists as ground truth labels.

For this comparison, we again used the naïve thresholds of $\sigma_{M32} > 0.67$ and $\sigma_{SE} > 0.67$ for delineating interesting candidates from noninteresting ones. The third column of Table 3 summarizes this result, where 123 of the 126 sources identified by A. Andersson et al. (2023) sit above this threshold. This results in a classifier with a *true positive rate* or *sensitivity* of 97.6%. Only three sources were not correctly identified as variable or transient.

The left-hand panel of Figure 8 shows this result graphically, where color coding indicates whether citizen scientists voted a particular source as a probable transient/variable or not. In this plot, we observe that all the sources marked as variable/transient have values of either σ_{M32} or σ_{SE} (or both) that sit outside the "stable source" region discussed earlier. The dashed lines indicate the 0.67 thresholds used in this classification, and we can observe three sources in the lower-left quadrant that were missed by this naive thresholding method, but only marginally so.

These results confirm that our approach is at least consistent with the visual judgment of human classifiers and affirm that locating light curves in the (σ_{M32}, σ_{SE}) parameter space is a promising method for identifying variable and transient candidates.

4.1.4. Comparison with η_ν and V_ν

It is instructive to compare the discriminatory power and classification accuracy of using (σ_{M32}, σ_{SE}) with the more commonly used variability measures of η_ν and V_ν (Section 1.1). We first explore the discriminatory power of these metrics by returning to the same four light curves plotted in Figure 7. The right-hand panel of Figure 6 is akin to the left-hand panel except that each light curve is now plotted according to its (η_ν, V_ν) coordinates. Despite being visually very distinct in appearance, these light curves are all situated closely together in (η_ν, V_ν) space and not in the upper-right quadrant where searches for variable sources have traditionally focused. The nominally stable light curve labeled C is somewhat set apart from the other three, but in Table 2, we see that all four have similar η_ν and V_ν values. This contrasts with their location in the left-hand panel of Figure 6, where each source is well separated and points to the (σ_{M32}, σ_{SE}) characterization having stronger discriminatory power.

Figure 12 in Appendix C explores this relationship in the reverse direction by plotting the location in (σ_{M32}, σ_{SE}) space of sources identified as highly variable in (η_ν, V_ν) space. We observe that all six examples, each a known variable XRB

system, lie far away from the region of stable sources described in Section 4.1.1, consistent with their classification based on their (η_ν, V_ν) values.

As was done for σ_{M32} and σ_{SE} in Section 4.1.3, we revisit the citizen scientist results from A. Andersson et al. (2023) and compare them with an approach based on thresholding η_ν and V_ν. For thresholds, we adopt $\eta_\nu > 8.22$ and $V_\nu > 0.179$, as used by A. Rowlinson et al. (2019) in their identification of variable sources. The results are shown in the rightmost column of Table 3, where we see correct identification of 71 of the 126 sources reported by A. Andersson et al. (2023). This amounts to a sensitivity of 56.3%, which is much lower than the 97.6% sensitivity afforded by classifications based on thresholding σ_{M32} and σ_{SE}.

Naturally, we could boost the sensitivity of any classifier by simply lowering the thresholds; however, this comes at the cost of reducing a classifier's *specificity*. A classifier with poor specificity will return many false positives by flagging spurious candidates that are more likely to be nonvariable than variable. The panels in Figure 8 illustrate this trade-off. While large values of η_ν and V_ν in the right-hand panel are consistent with a source being classified as transient (colored points), lowering the thresholds for η_ν and V_ν to identify all these transients correctly would also include most of the other sources *not* flagged by A. Andersson et al. (2023) as likely transients. This is not the case in $(\sigma_{\mathrm{M32}}, \sigma_{\mathrm{SE}})$ space, since transient candidates are largely confined to the upper-right quadrant of the parameter space.

These comparisons show that our GP-based characterization has a parameter space that better segments variable and transient sources than the traditional (η_ν, V_ν) space.

4.2. Accounting for Field-of-view Effects

The statistical characterization of an astronomical source based on its observed properties assumes that these are a reliable and uncontaminated proxy of its true physical properties. However, in all surveys, including ThunderKAT, we expect additional effects related to the observing regime rather than the source itself. Consequently, any fitted hyperparameters will also reflect these contaminants. For example, observing across different fields of view, each with different numbers of pointings, over multiple epochs, has manifested in the distinct binning and clustering of sources by color (field) in the scatter plots in Figure 3.

Observational cadence is an obvious property that distinguishes fields from one another and is a potential explanation for the binning effects described above. In the absence of true astrophysical periodicity, the P kernel term of our GP regression model might be expected to converge onto P patterns such as observational cadence. However, the colored clusters in the posterior medians of T seen in Figure 3(D) are not aligned with the known observation cadences, and inspection of their posteriors showed very little change from their hyperpriors. Refitting the data without the P kernel term also yielded similar posteriors. This suggests that the P kernel has detected neither the survey cadence nor any other P component. Having ruled out cadence, further investigation revealed that the duration of the light curve is the main confounding factor affecting the length-scale hyperparameter estimates (ℓ). This is clearly shown in the observations around GX 339-4, colored yellow in Figures 3 and 4, which is the field with the longest duration light curves and whose hyperparameter estimates are clustered at the largest timescales.

We notice, however, that in Figure 5 there are comparatively fewer binning artifacts in the scatter plot of the amplitude hyperparameters, σ_{M32} and σ_{SE}. This implies that these two measures are less impacted by confounding by the duration of light curves and are more reliable for characterizing sources. However, should data uncertainties be too large, observing campaigns too short, or cadence too sparse relative to the timescale of interest, such light curves will likely be situated in the nonvarying region described in Section 4.1.1.

4.3. Transient Candidates

Following our claim that the $(\sigma_{\mathrm{M32}}, \sigma_{\mathrm{SE}})$ hyperparameter space is more favorable than the (η_ν, V_ν) parameter space for identifying variables and transients, we conducted an elementary search for candidate sources in the ThunderKAT data set using our GP regression approach. Searching the hyperparameter space near known transient sources is a natural place to start, and the most obvious targets are the known XRB systems upon which ThunderKAT fields of view are centered. The light curves and characteristics of the six XRBs used to "seed" our search are described in Appendix C. Rather than a rigorous search for transient sources, we are demonstrating how our proposed metrics might be incorporated as an initial single-pass screening step in a more involved search protocol.

A basic search algorithm is as follows:

1. For a known seed XRB source in the ThunderKAT data set.
2. Select its 30 nearest neighbors in the $(\sigma_{\mathrm{M32}}, \sigma_{\mathrm{SE}})$ hyperparameter space by Euclidean distance, excluding sources from the same field as the original known XRB. The reason for excluding sources from the same field is to ensure that any statistical similarity between sources is not simply due to the similarity in their observing conditions.
3. For each potential candidate, visually inspect its radio image with the highest signal-to-noise ratio. If the source is resolved or otherwise ambiguous, remove it from consideration.
4. For any remaining candidates, visually inspect their light curves and only retain those showing transient behavior.

We expect any candidate sources we identified to have light curves that are statistically, if not astronomically, similar to the target sources. However, as a nonparametric approach, a GP regression model that is not tuned to a particular type of source, say XRBs, may quantify the observed variability patterns of other source types, such as flaring stars, and situate them in the hyperparameter space accordingly. Further work is needed to explore the distribution of different source types within the $(\sigma_{\mathrm{M32}}, \sigma_{\mathrm{SE}})$ hyperparameter space, as well as the impact of using different types of sources to seed searches.

Thirty sources were identified from each of the six known XRBs for an initial list of 180 candidates. After removing sources already identified by citizen scientists as transients in A. Andersson et al. (2023), visual inspection of radio images and light curves was conducted to remove extended sources and observations with spurious data points such as negative flux densities. The complete set of 67 unique candidates is available in a machine-readable format. For illustrative purposes, 10 candidates' light curves and fitted models are plotted in Figure 9 and their variability metrics summarized in Table 4. This sample of 10 was deliberately chosen to showcase light curves from a broad selection of observing fields.
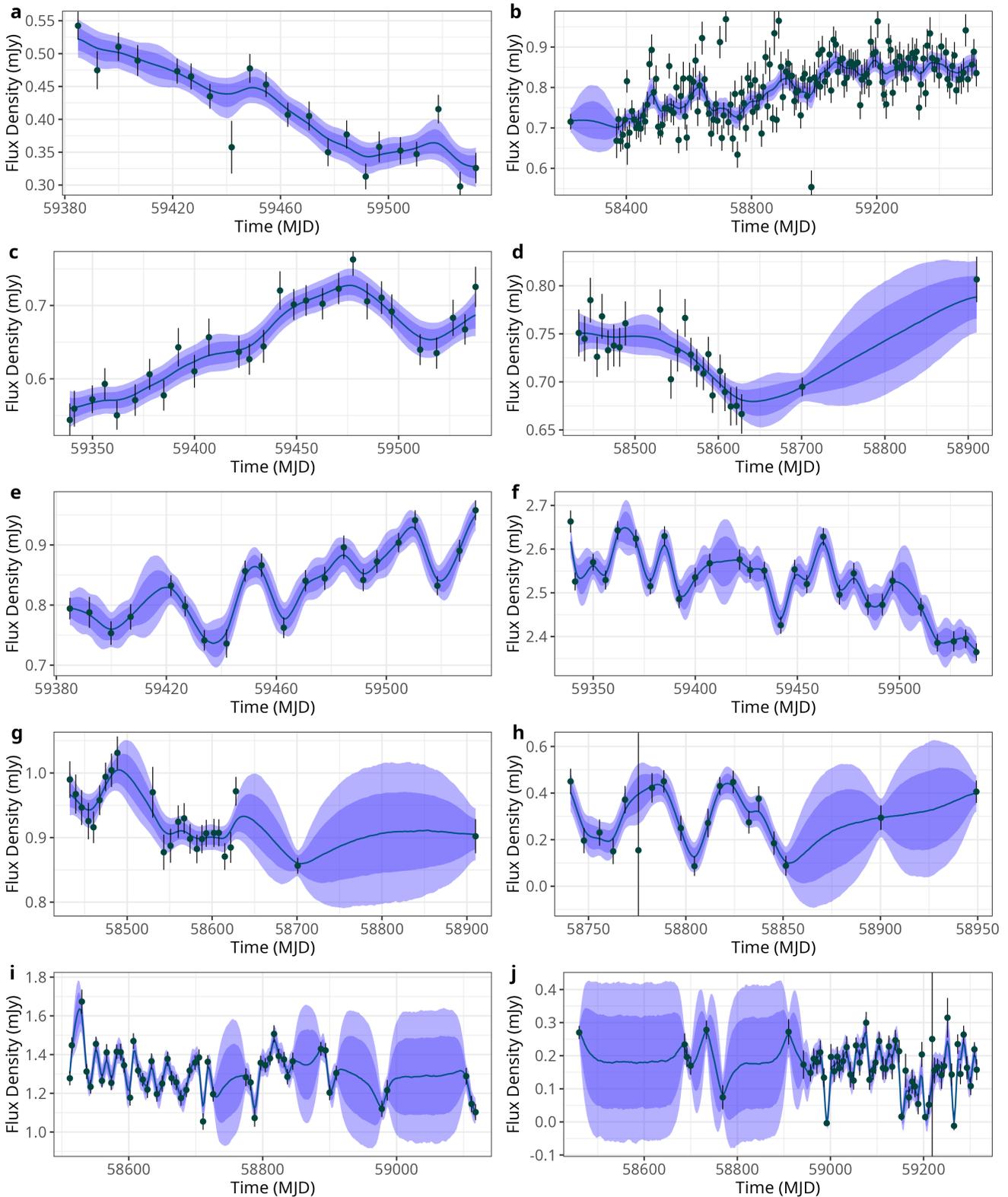
Figure 9. Fitted light-curve models for 10 transient candidates identified using a basic search algorithm. None of these sources were voted as candidate transient or variable sources by citizen scientists in A. Andersson et al. (2023). Location and variability statistics for each source are shown in Table 4. The "seed" light curves to which these candidates were adjacent are shown in Figure 11 in Appendix C.

The 10 candidate light curves shown in Figure 9 are ordered by increasing σ_{M32}, the hyperparameter tuned for detecting short-timescale variability. This is reflected by light curves in the top panels exhibiting long-timescale trends and fewer undulations, compared with the lower panels showing more jagged appearances. Also, while the seed light curves shown in Figure 11 in Appendix C appear burst or flare-like, these 10 candidate sources behave more like persistent variables.
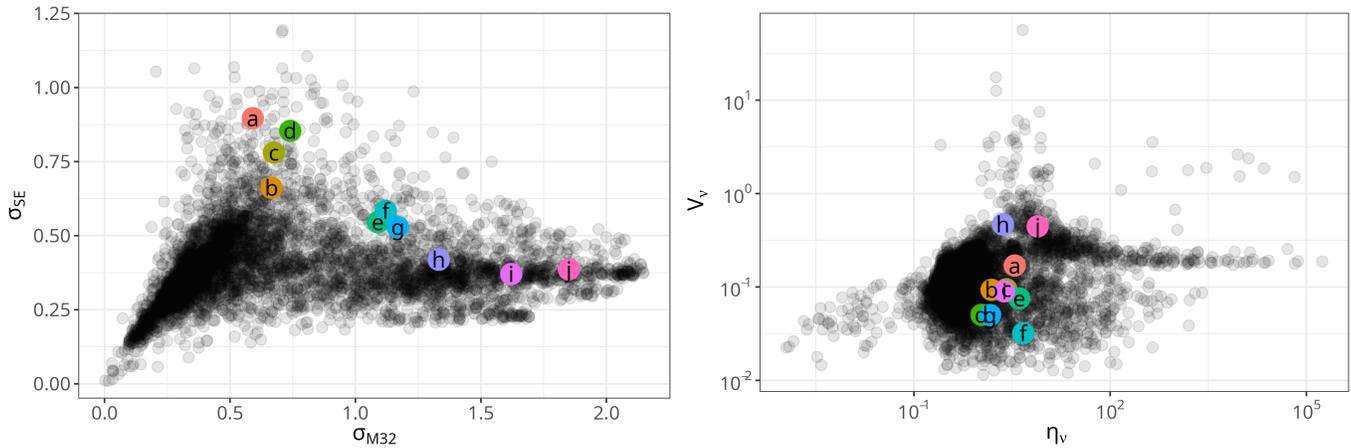
Figure 10. Left: the (σ_{M32}, σ_{SE}) locations of 10 candidates identified by searching the statistical neighborhood of the six known XRB sources labeled in Figure 12. Right: the location in (η_ν, V_ν) of the same sources shown in the left panel. Notice that the locations of these sources in (η_ν, V_ν) are distant from the known XRB sources, as shown in the corresponding panel of Figure 12. Coordinates and variability metrics of each source are summarized in Table 4.

Table 4
Location and Variability Metrics for the 10 Candidate Sources Shown in Figure 9

Figure 9	Seed XRB	Candidate Source							
		Field	R.A. (deg)	Decl. (deg)	Mean Flux (mJy)	η_ν	V_ν	σ_{M32}	σ_{SE}
a	H 1743−322	4U1543	236.9659	−47.6767	0.410	3.48	0.17	0.59	0.90
b	MAXI J1820+070	GX339	256.6843	−48.9579	0.804	1.54	0.09	0.67	0.66
c	H 1743−322	MAXIJ1803	271.1269	−30.4385	0.649	2.58	0.09	0.68	0.78
d	H 1743−322	J1858	284.3228	−8.3813	0.728	1.08	0.05	0.74	0.85
e	GX 339−4	4U1543	237.2070	−48.3865	0.835	4.06	0.08	1.09	0.55
f	GX 339−4	MAXIJ1803	271.2300	−29.6591	2.522	4.68	0.03	1.12	0.58
g	GX 339−4	J1858	284.9503	−8.6815	0.928	1.46	0.05	1.17	0.53
h	MAXI J1803−298	EXO1846	282.3311	−2.5805	0.291	2.29	0.47	1.33	0.42
i	MAXI J1820+070	MAXIJ1348	207.6484	−62.6685	1.306	2.41	0.09	1.62	0.37
j	MAXI J1348−630	GRS1915	289.2072	11.0012	0.169	7.77	0.44	1.85	0.39

Note. Entries are sorted in ascending size of σ_{M32}. None of the (η_ν, V_ν) values are remarkable.

(This table is available in its entirety in machine-readable form in the online article.)

Despite these differences in shape, all of these light curves have median σ_{SE} and σ_{M32} estimates greater than 0.67, identifying them as varying sources. The fact that this model can capture different types of variability reflects the flexibility of GP models and points to our model's generalizability to a diverse range of source types.

Figure 10 compares where these 10 candidate sources sit in two respective parameter spaces. Once again, we observe that the separation of these candidates from both each other and from nominally nonvarying sources is more pronounced in the (σ_{M32}, σ_{SE}) space compared with the (η_ν, V_ν) space. Also, the grouping of light curves within the amplitude hyperparameter space seems more interpretable since sources whose light curves are visually similar, such as sources a–d, are situated away from source j, whose behavior is clearly distinct.

4.4. Proposed Extensions

Several simplifications that we have made require further elaboration. The first is the stationarity of light curves. This strong assumption overlooks situations in which light curves are composites of multiple underlying astrophysical processes, each contributing different components of variability at different phases of activity. This shortcoming could be mitigated by relaxing our modeling choice of using a zero mean function. However, using a nonzero, parametric mean function based on prior knowledge may bias the GP regression model toward particular astrophysical phenomena. One approach is to use a nonparametric function, perhaps even another GP, as the mean function of a more complex hierarchical model.

The second is that we do not postulate a probability model for the observational uncertainty but instead directly use the provided standard errors. Considerable work would be needed to specify the distribution of the observational uncertainties in radio interferometric data. This may entail incorporating work on nonlinear variability found in X-ray light curves (P. Uttley et al. 2005), which may also moderate the effect of outliers such as those seen in Figures 2(G) and 7(D), whose uncharacteristically small standard errors heavily distort the fitted model.

Third, we have used point values (medians) to summarize the posterior distribution of each hyperparameter. A more complete analysis would use posterior distributions.

5. Conclusion

In this work, we have presented an approach using GP regression to characterize the variability of sources based on their light curves. GP models such as ours implicitly handle nonuniform data coverage and cadence, which are properties that are commonplace in time-domain astronomical surveys. We showed that our approach offers greater discriminatory power than the commonly used parameterization of (η_ν, V_ν). Furthermore, it differs from existing applications of GPs in the astronomical literature because it adopts a multiterm kernel function. We specify a composite covariance kernel of SE, M32, and P terms. Then, we use the posterior estimates of their hyperparameters, not the mean posterior predictive curve, to distinguish between variable and nonvariable sources. Nevertheless, the mean curves and their PSDs are useful diagnostics for model assessment.

Our key finding is that the bivariate distribution of the amplitude hyperparameters of the SE and M32 kernels, namely $\sigma_{\rm SE}$ and $\sigma_{\rm M32}$, defines a parameter space where statistically different light curves populate distinctly different regions. Qualitatively, this means that sources exhibiting high variability on long or short timescales (or both) are clustered away from sources whose light curves are nominally nonvarying or stable. This is very useful in screening for transient or variable sources. The astronomer interested in identifying variable or transient source candidates should examine light curves whose $\sigma_{\rm SE}$ or $\sigma_{\rm M32}$ are in the upper quantiles of their survey, since these should prove the most promising. We demonstrated this by identifying sources in ThunderKAT data that were not previously flagged in the literature. We claim this is a generalizable approach that can be used to discover transient and variable candidates, agnostic of source type. Having tested this method on one survey with reasonable results, applying this approach to other large surveys is the subject of upcoming work.

Computational notebooks in R and PYTHON describing our method are available[11] to help facilitate its translation and application to surveys other than ThunderKAT.

Appendix A
Justification of Hyperpriors

Given the anticipated diversity of objects observed in a large survey such as ThunderKAT, it is unlikely that the same kernel hyperparameter values would be suitable across all sources. Yet some physical constraints should be consistent regardless of the underlying astrophysical phenomena observed. Below, we elaborate on how we express such constraints in terms of kernel hyperprior distributions and justify the choice of these hyperpriors (Equations (7)–(9)).

Length scale. We assigned identical hyperpriors to each kernel's length-scale hyperparameter (ℓ). As suggested by the STAN User's Guide (Stan Development Team 2023), we used an inverse gamma (IG) hyperprior since it down-weights very small length scales—scales that are more comparable to noise than signal—while still permitting medium to large length scales. The shape parameter of the IG distribution is fixed at $\alpha = 3$, and the rate parameter β is set to one-half the range of the light-curve duration (Equation (8)).

Recalling that the expected value of an IG distributed random variable is $\beta/(\alpha - 1)$, the above initializes the prior mean of these length-scale hyperpriors to approximately one-quarter of the total period of the light curve. This is a good rule of thumb since we want to observe at least four cycles of repeating patterns before judging any pattern as real, not noise. Additionally, length scales were given a lower bound of the shortest separation in time between data points, Δt (Equation (10)). This is because it is unreasonable to infer autocorrelation structures shorter than the finest resolution in the data. Thus, this lower bound prevented length scales from approaching zero too closely.

In addition to a general lower bound for all length-scale hyperparameters, a further inequality was placed on $\ell_{\rm M32}$ such that it must be smaller than $\ell_{\rm SE}$ (Equation (11)). This mitigates identifiability issues arising from the similar behavior of these kernels when applied at similar scales. Depending on the light curve, either kernel is equally capable of describing the same autocovariance structures. By restricting the SE kernel to take on larger length scales, the kernel is impelled to lock onto longer and smoother trends for which it is better suited. Thus, the M32 kernel tends to capture more jagged patterns found at shorter length scales.

Amplitude. For amplitude hyperparameters (σ), standard half-normal hyperpriors were applied to the standardized flux densities (Equation (7)). Indeed, flux densities were standardized to facilitate applying identical amplitude hyperpriors to all light curves, which otherwise have different flux density ranges. A useful rule of thumb for assessing whether a posterior hyperparameter estimate is significant is comparing it with a nominal threshold derived from its hyperprior distribution. The median of the standard half-normal distribution is ≈ 0.67, which is therefore a useful threshold for gauging the influence of this hyperprior on the resultant posterior—values above 0.67 warrant further investigation.

Period. A uniform hyperprior was chosen for the period (T) hyperparameter since we do not have any general information about the presence or absence of periodicity in ThunderKAT light curves. A uniform hyperprior will avoid biasing the detection of periodicities toward any particular period, and any period in the range is plausible. As with the amplitude hyperpriors, the sparsity of observations was used to inform the bounds of the uniform hyperprior (Equation (9)).

Similar to the length-scale hyperpriors, we wanted to observe at least three or four full P cycles in our data before concluding there was any evidence of periodicity. This heuristic sets an upper bound on the uniform hyperprior to

---

be one-quarter of the total time span of the light curve. For a lower bound, we took the shortest time gap between observations (Δt), or conversely, the highest instantaneous sampling rate, and multiplied by two to get the Nyquist frequency. This prevented the model from fitting periods smaller than twice the smallest time gap, i.e., never finer than the best data resolution.

Appendix B
Marginal GP Model

In our model, the uncertainties in the flux density measurements are assumed to be Gaussian distributed. This means we can "integrate out" the latent function f and reexpress Equations (4) and (5) more simply as a *marginal* GP model:

$$S \sim \mathcal{GP}(\mathbf{0}, \mathbf{K}_{N \times N} + \mathrm{diag}(\hat{\mathbf{e}}^2)), \qquad (\text{B1})$$

where $\mathrm{diag}(\hat{\mathbf{e}}^2)$ is an $N \times N$ matrix whose diagonal values are the error variances of the observed data, $\hat{\mathbf{e}}^2$, and the off-diagonal values are zero. A marginal model is an acceptable simplification because the main object of this study is the characterization of light curves rather than explicitly inferring the "true" flux density function f. This formulation has the added benefit of having an analytical expression for the posterior predictive distribution (C. E. Rasmussen & C. K. I. Williams 2006, p. 16):

$$\mathbf{s}^*|\mathbf{t}^*, \mathbf{s}, \mathbf{t} \sim \mathcal{N}(\mathbf{K}_*^\top \mathbf{\Sigma}^{-1}\mathbf{y}, \mathbf{\Omega} - \mathbf{K}_*^\top \mathbf{\Sigma}^{-1}\mathbf{K}_*), \qquad (\text{B2})$$

where $\mathbf{s}$ is the flux density observed at time $\mathbf{t}$, and $\mathbf{s}^*$ is the predicted flux density at the target time $\mathbf{t}^*$. $\mathbf{\Sigma}$ is the covariance matrix as defined in Equation (6), $\mathbf{K}_*$ is the covariance matrix between times $\mathbf{t}$ and $\mathbf{t}^*$, and $\mathbf{\Omega}$ specifies the covariances between the target times, $\mathbf{t}^*$. With this analytic expression for the posterior predictive distribution, drawing samples is much more efficient and straightforward computationally.

Appendix C
Fitted Models of Known XRBs

This section contains more examples of GP regression model fits to light curves. In particular, the six examples shown in Figure 11 are light curves of sources known to be XRB systems, namely, EXO1846, MAXIJ1348, MAXIJ1820, MAXIJ1803, H1743, and GX339. Indeed, these are the sources at the center of the various fields of view observed in the ThunderKAT survey. Each source exhibits transient behavior, namely large spikes in flux density before resuming some previous level. This makes them suitable cases for testing whether our model can identify transient-like sources.

Table 5 summarizes the variability statistics and estimated GP regression hyperparameters for each of these XRB light curves. In every case, the posterior medians of σ_{M32} are in regions of hyperparameter space that indicate nonnegligible variability. This is clearly shown in the scatter plot of Figure 12, where these sources are marked by colored points.
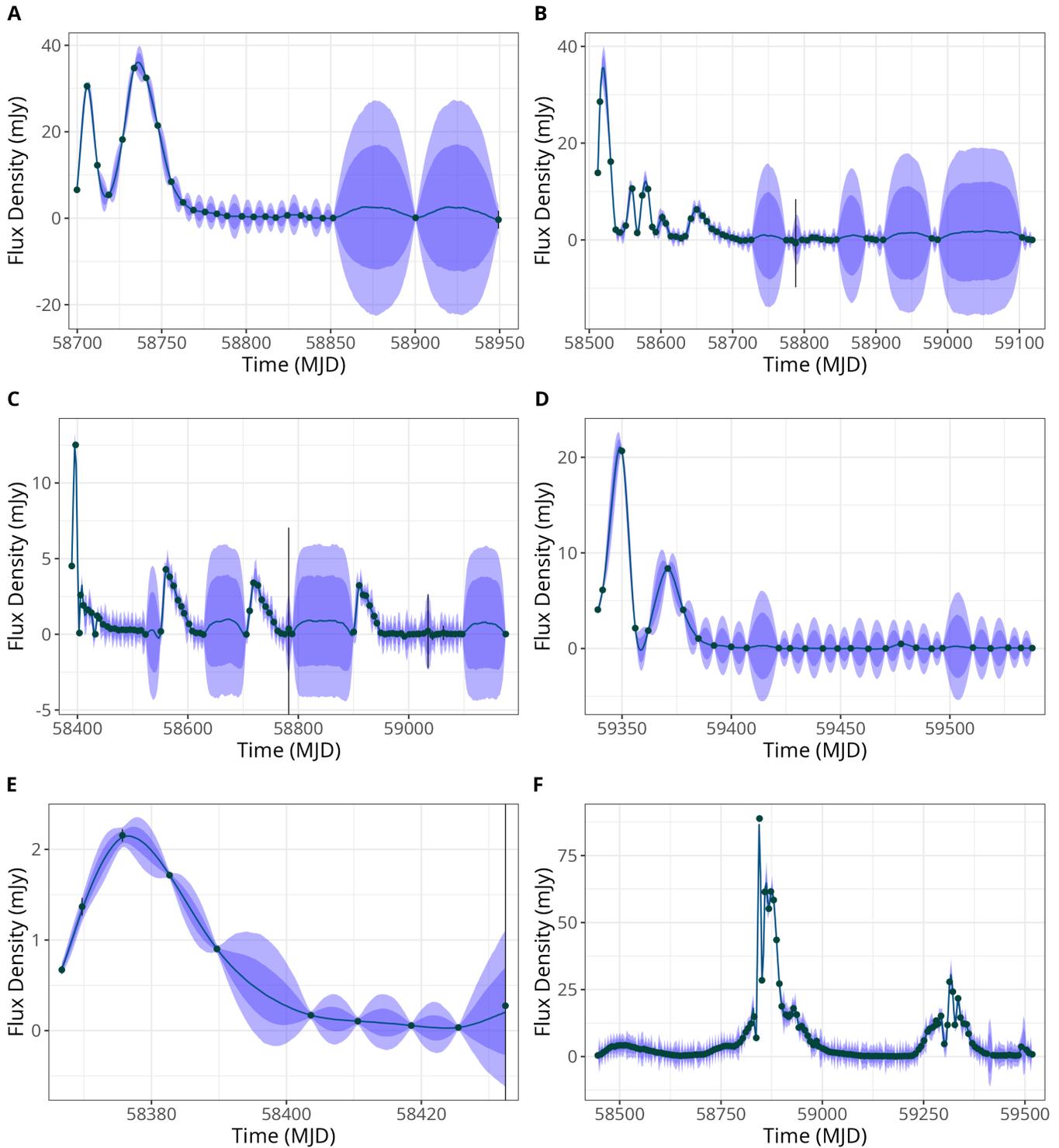
Figure 11. Plots of posterior predictive curves from GP models of light curves of six known X-ray binaries: EXO1846 (A), MAXIJ1348 (B), MAXIJ1820 (C), MAXIJ1803 (D), H1743 (E), and GX339 (F). Coordinates of each source and posterior statistics of each light curve are found in Table 5. The points are the original observations, with black bars indicating standard errors. The lines trace the median of the posterior predictive samples, with shaded regions indicating 68% and 90% quantile intervals, respectively.
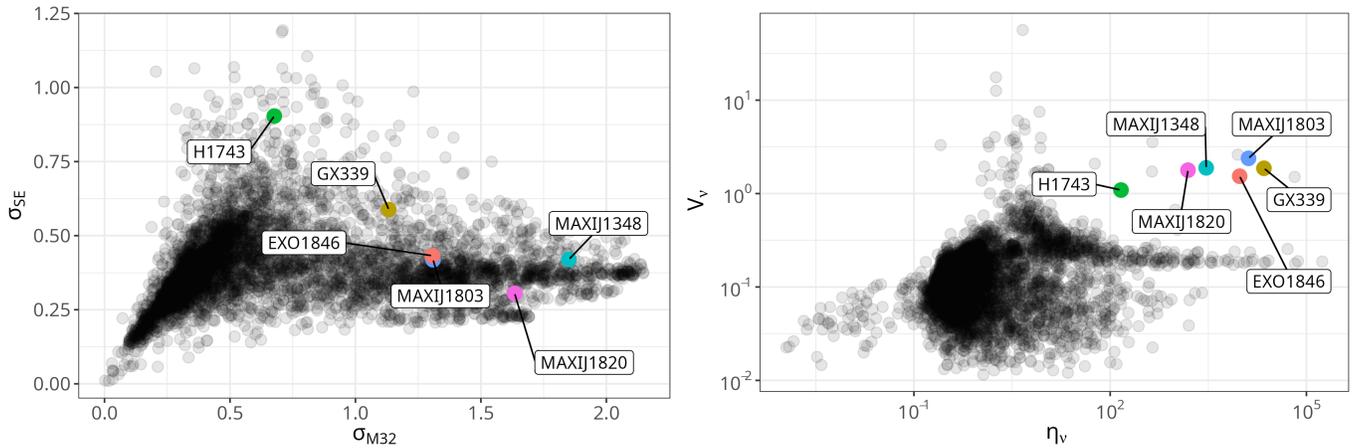
Figure 12. Left: scatter plot of the posterior medians of the amplitude hyperparameter of the SE and M32 kernels. Amplitudes are in standardized units. Right: scatter plot of the (η_ν, V_ν) values for the same light curves. In both panels, six known XRBs are marked by colored points.

Table 5
Fitted Hyperparameter Values for the Example Light Curves Shown in Figure 11, Each of Which Are Known XRB Systems

Figure 11	XRB	R.A. (deg)	Decl. (deg)	η_ν	V_ν	Posterior Median						
						σ_{M32}	ℓ_{M32}	σ_{SE}	ℓ_{SE}	σ_P	ℓ_P	T
A	EXO1846	282.3212	−3.0655	9528.87	1.53	**1.31**	13.89	0.43	48.39	0.48	37.44	86.12
B	MAXIJ1348	207.0532	−63.2747	2936.57	1.88	**1.85**	16.26	0.42	119.09	0.48	90.54	209.01
C	MAXIJ1820	275.0914	7.1853	1546.85	1.78	**1.63**	9.45	0.31	160.82	0.41	145.90	199.04
D	MAXIJ1803	270.7617	−29.8301	13026.88	2.38	**1.31**	9.07	0.42	40.96	0.47	29.52	68.22
E	H1743	266.5652	−32.2340	146.35	1.09	**0.68**	9.36	0.90	13.98	0.49	10.06	23.38
F	GX339	255.7055	−48.7897	22427.60	1.86	**1.13**	10.75	0.59	127.85	0.40	159.48	368.66

Note. Medians of the hyperparameter posteriors are shown beside the frequently reported variability statistics of η_ν and V_ν. Boldface indicates median amplitudes greater than 0.67.

Appendix D
Principal Components Analysis of Hyperparameter Posterior Means

Principal component (PC) analysis of the posterior means of the fitted hyperparameters for every light curve reveals that two PCs are sufficient to explain over 77% of the variability in the data. A third component increases this to over 94% (see the second column of Table 6). The loadings for the length-scale ℓ and period T hyperparameters are negative, which shows that these hyperparameters are negatively correlated with the first PC. Conversely, the amplitude σ hyperparameters are positively correlated with the second PC. This suggests that the first PC, which increases as the length scale and period decrease, captures short-timescale variability in the light curves. Such variability is likely to be noise rather than phenomenological, with the remaining variability explained by the amplitude hyperparameters of the second PC.

The biplot in Figure 13 shows the position of each light curve as projected into the 2D plane of the first two PCs. The arrows are scaled representations of the loadings in Table 6, and the colors indicate the field of observation for each light curve. We observe that the progression of colors (fields) from left to right corresponds strongly to the direction of the first PC, which reflects the confounding of these hyperparameters by field-specific properties, such as the typical duration of those observations. The spread of points within each field toward the top right is consistent with the loadings of the amplitude hyperparameters of the M32 and SE kernels, noting that we will ignore the amplitude hyperparameter of the P kernel since it is strongly correlated with the SE.

From these results, we are confident that examining the two hyperparameters of σ_{M32} and σ_{SE} is sufficient to characterize light curves fitted by our GP regression model.

Table 6
Proportion of Variance Explained by Each PC and the Loadings of Each Hyperparameter against the First Three PCs

PC	Proportion of Variance	Loadings						
		σ_{SE}	σ_{M32}	σ_P	ℓ_{SE}	ℓ_{M32}	ℓ_P	T
1	54.9%	0.241	0	0.297	**−0.496**	**−0.385**	**−0.479**	**−0.479**
2	22.6%	**0.433**	**0.630**	**0.502**	0.136	−0.125	0.254	0.254
3	16.8%	0.559	−0.528	0.325	0.127	0.532	0	0

Note. Nonnegligible values are in boldface. Notice that the first PC is dominated by length-scale hyperparameters and the second PC by amplitude hyperparameters.
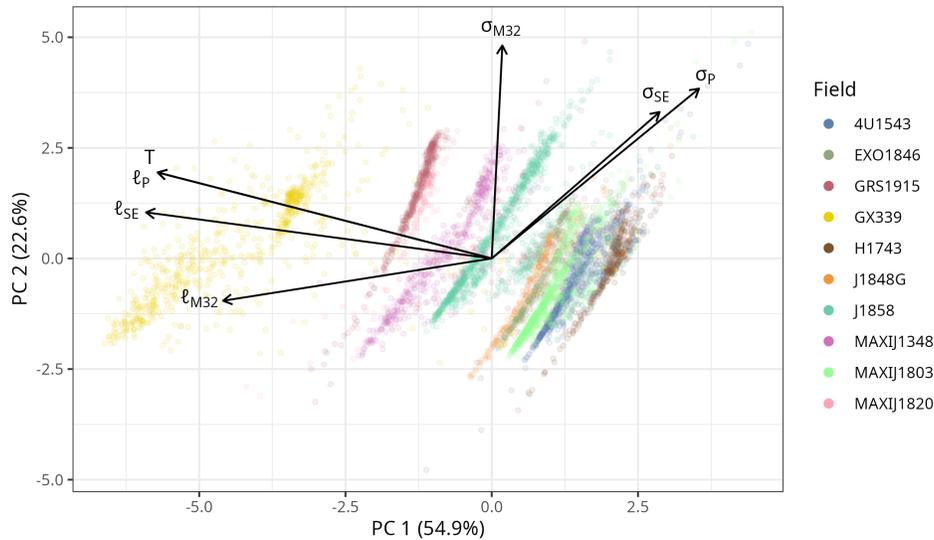
Figure 13. Biplot of the first two PCs of the posterior means of the hyperparameters estimated for each light curve. Hyperparameter values were standardized before applying PC analysis. Colors indicate in which field the light curve was observed, and the arrows indicate the loadings of the original hyperparameters projected onto these first two PCs. The ℓ_P and T hyperparameters are perfectly correlated, so their arrows overlap.

ORCID iDs

Shih Ching Fu (符史青) ● https://orcid.org/0000-0002-9077-6026
Arash Bahramian ● https://orcid.org/0000-0003-2506-6041
Aloke Phatak ● https://orcid.org/0000-0002-0637-7461
James C. A. Miller-Jones ● https://orcid.org/0000-0003-3124-2814
Suman Rakshit ● https://orcid.org/0000-0003-0052-128X
Alexander Andersson ● https://orcid.org/0000-0003-2734-1895
Robert Fender ● https://orcid.org/0000-0002-5654-2744
Patrick A. Woudt ● https://orcid.org/0000-0002-6896-1655

References

Aigrain, S., & Foreman-Mackey, D. 2023, ARA&A, 61, 329
Allen, B., Anderson, W. G., Brady, P. R., Brown, D. A., & Creighton, J. D. E. 2012, PhRvD, 85, 122006
Andersson, A. 2024, AnderssonAstro/BfS-MKT-Analysis, v1.0, Zenodo, doi:10.5281/zenodo.10462688
Andersson, A., Fender, R. P., Lintott, C. J., et al. 2022, MNRAS, 513, 3482
Andersson, A., Lintott, C., Fender, R., et al. 2023, MNRAS, 523, 2219
Andersson, A., Lintott, C., Fender, R., et al. 2025, MNRAS, 538, 1397
Angus, R., Morton, T., Aigrain, S., Foreman-Mackey, D., & Rajpaul, V. 2018, MNRAS, 474, 2094
Antoniadis, J., Arzoumanian, Z., Babak, S., et al. 2022, MNRAS, 510, 4873
Babu, G. J., & Mahabal, A. 2016, International Statistical Review/Revue Internationale de Statistique, 84, 506
Bell, M. E., Murphy, T., Hancock, P. J., et al. 2019, MNRAS, 482, 2484
Bell, M. E., Murphy, T., Kaplan, D. L., et al. 2014, MNRAS, 438, 352
Bortle, A., Fausey, H., Ji, J., et al. 2021, AJ, 161, 230
Covino, S., Landoni, M., Sandrinelli, A., & Treves, A. 2020, ApJ, 895, 122
Covino, S., Tobar, F., & Treves, A. 2022, MNRAS, 513, 2841
Czekala, I., Mandel, K. S., Andrews, S. M., et al. 2017, ApJ, 840, 49
D'Emilio, V., Green, R., & Raymond, V. 2021, MNRAS, 508, 2090
Dewdney, P. E., Hall, P. J., Schilizzi, R. T., & Lazio, T. J. L. W. 2009, IEEEP, 97, 1482
Driessen, L. N., Stappers, B. W., Tremou, E., et al. 2022, MNRAS, 512, 5037
Fender, R., Woudt, P. A., Armstrong, R., et al. 2017, arXiv:1711.04132
Feng, L., Vaulin, R., Hewitt, J. N., et al. 2017, AJ, 153, 98

Fu, S. C. 2024, Computational Notebooks Accompanying "New Metrics for Identifying Variables and Transients in Large Astronomical Surveys", v1.0.0, Zenodo, doi:10.5281/ZENODO.14376242
Gabry, J., Češnovar, R., Johnson, A., & Bronder, S., 2024 cmdstanr: R Interface to "CmdStan", https://mc-stan.org/cmdstanr/
Gelman, A., & Rubin, D. B. 1992, StaSc, 7, 457
Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Natur, 585, 357
Hoffman, M. D., & Gelman, A. 2011, arXiv:1111.4246
Hunter, J. D. 2007, CSE, 9, 90
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Kovacevic, A. B., Popovic, L. C., & Ilic, D. 2020, OAst, 29, 51
Krige, D. G. 1951, PhD thesis, Univ. Witwatersrand
Matheron, G. 2019, Matheron's Theory of Regionalised Variables (Oxford: Oxford Univ. Press)
McLaughlin, S. A. J., Mullaney, J. R., & Littlefair, S. P. 2024, MNRAS, 529, 2877
Mooley, K. P., Hallinan, G., Bourke, S., et al. 2016, ApJ, 818, 105
Moore, C. J., Berry, C. P., Chua, A. J., & Gair, J. R. 2016, PhRvD, 93, 064001
Murphy, T., Kaplan, D. L., Stewart, A. J., et al. 2021, PASA, 38, e054
Pope, B. J. S., Parviainen, H., & Aigrain, S. 2016, MNRAS, 461, 3399
Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning (Cambridge, MA: MIT Press)
R Core Team, 2023 R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing), https://www.R-project.org
Rowlinson, A., Stewart, A. J., Broderick, J. W., et al. 2019, A&C, 27, 111
Song, F.-F., & Mao, J. 2024, ApJ, 966, id.209
Stan Development Team 2023, Stan Reference Manual, https://mc-stan.org/
Stephenson, W. T., Ghosh, S., Nguyen, T. D., et al. 2022, arXiv:2106.06510
Stevance, H. F., & Lee, A. 2023, MNRAS, 518, 5741
Swinbank, J. D., Staley, T. D., Molenaar, G. J., et al. 2015, A&C, 11, 25
The Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167
Uttley, P., McHardy, I. M., & Vaughan, S. 2005, MNRAS, 359, 345
van Haasteren, R., & Vallisneri, M. 2014, PhRvD, 90, 104012
Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. 2021, BayAn, 16, 667
Villar, V. A., Hosseinzadeh, G., Berger, E., et al. 2020, ApJ, 905, 94
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261
Wiecki, T., Salvatier, J., Vieira, R., et al. 2023, pymc-devs/pymc: v5.3.0, Zenodo, doi:10.5281/zenodo.7827191
Wikle, C., Zammit-Mangion, A., & Cressie, N. 2019, Spatio-Temporal Statistics with R (New York: Chapman & Hall/CRC), 9