



Detecting Stellar Flares in Photometric Data Using Hidden Markov Models

J. Arturo Esquivel^{1,2}, Yunyi Shen³, Vianey Leos-Barajas^{1,2,4}, Gwendolyn Eadie^{1,2,5}, Joshua S. Speagle^{1,2,5}, Radu V Craiu^{1,2}, Amber Medina⁶, and James R. A. Davenport⁷

¹ Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada; a.esquivel@mail.utoronto.ca

² Data Sciences Institute, University of Toronto, Toronto, ON, Canada

³ Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA

⁴ School of the Environment, University of Toronto, Toronto, ON, Canada

⁵ David A. Dunlap Department of Astronomy & Astrophysics, University of Toronto, Toronto, ON, Canada

⁶ Las Cruces, NM, USA

⁷ Department of Astronomy, University of Washington, Box 351580, Seattle, WA 98195, USA

Received 2024 May 16; revised 2024 November 19; accepted 2024 November 20; published 2025 January 22

Abstract

We present a hidden Markov model (HMM) for discovering stellar flares in light-curve data of stars. HMMs provide a framework to model time series data that are nonstationary; they allow for systems to be in different states at different times and consider the probabilities that describe the switching dynamics between states. In the context of the discovery of stellar flares, we exploit the HMM framework by allowing the light curve of a star to be in one of three states at any given time step: *quiet*, *firing*, or *decaying*. This three-state HMM formulation is designed to enable straightforward identification of stellar flares, their duration, and associated uncertainty. This is crucial for estimating the flare's energy, and is useful for studies of stellar flare energy distributions. We combine our HMM with a *celerite* model that accounts for quasiperiodic stellar oscillations. Through an injection recovery experiment, we demonstrate and evaluate the ability of our method to detect and characterize flares in stellar time series. We also show that the proposed HMM flags fainter and lower energy flares more easily than traditional sigma-clipping methods. Lastly, we visually demonstrate that simultaneously conducting *detrending* and flare detection can mitigate biased estimations arising in multistage modeling approaches. Thus, this method paves a new way to calculate stellar flare energy. We conclude with an example application to one star observed by TESS, showing how the HMM compares with sigma clipping when using real data.

Unified Astronomy Thesaurus concepts: [Stellar flares \(1603\)](#); [Bayesian statistics \(1900\)](#); [Astrostatistics tools \(1887\)](#); [Time series analysis \(1916\)](#); [M dwarf stars \(982\)](#); [Stellar activity \(1580\)](#); [Low mass stars \(2050\)](#); [Astrostatistics techniques \(1886\)](#); [Credible region \(1962\)](#)

1. Introduction

Almost all stars in the Universe with convection zones produce *stellar flares*—bursts of energy emitted from the star that are thought to be caused by magnetic reconnection (see, e.g., T. Forbes 1991; J.-F. Donati & J. Landstreet 2009). Properly estimating the energy distribution of flares as a function of a stars' mass, age, and other characteristics is fundamental for understanding (i) the evolution of stellar magnetic fields, (ii) stellar rotation and mass-loss rates, and (iii) the highly energetic radiation environment to which planets orbiting these stars are subjected. Past studies have found relationships between flare energies and decay times, frequency of flares and stellar rotation rates, and flare duration and peak luminosity (e.g., see S. W. Chang et al. 2015; J. R. A. Davenport 2016; T. V. Doorsselaere et al. 2017; A. A. Medina et al. 2020; S. Raetz et al. 2020).

Flares are detected in the time series data of a star's brightness measurements—a sudden, sharp increase in brightness followed by a slower decay usually indicates a stellar flare. However, detecting stellar flares is complicated by the fact that most stars also exhibit small, quasiperiodic oscillations in their brightness over time.

It is common practice to identify flares in time series data using nonparametric models, before assuming anything about

their time series signature or shape (S. W. Chang et al. 2015). Current methods to detect stellar flares in time series data rely on multistage data processing and “sigma clipping”; after the stationary and quasiperiodic part of the time series is modeled and removed (called *detrending*), points lying outside a predefined confidence interval are highlighted as potential flares (e.g., L. M. Walkowicz et al. 2011; R. A. Osten et al. 2012; S. L. Hawley et al. 2014; J. R. A. Davenport 2016; H. Yang et al. 2017; M.N. Günther et al. 2020). A. A. Medina et al. (2020) also note that in the 3σ approach to flare detection and flare energy estimation, the largest source of uncertainty comes from defining the end of the flare. The sigma-clipping approach may also struggle to identify compound flares, although change-point detection does not suffer from this problem (S. W. Chang et al. 2015).

After the flares' locations in the time series are detected, a template or model for flare shape (e.g., J. R. Davenport et al. 2014) is often used to estimate flare parameters. The *detrending* process is often done using a flexible model, such as Gaussian processes (GPs; e.g., *celerite*) (D. Foreman-Mackey et al. 2017). However, *detrending* methods are typically done before the flare detection step, and this preprocessing may absorb lower energy flares in the time series data, and bias the energy profile. Therefore, it would be beneficial to model the trend of the time series and the flares simultaneously.

Other methods, such as change-point detection, have also been explored to identify potential flares (S. W. Chang et al. 2015). While this approach detects the most energetic flares, it



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

can struggle to detect the medium- to low-energetic flares that are part of the flare energy distribution. The *STELLA* software, which uses a convolutional neural network to find flares in TESS data, is an efficient tool for finding flares, but still relies on a probability threshold for flare detection (A. D. Feinstein et al. 2020). Moreover, *STELLA* needs labeled data to be trained on, so its accuracy is dependent on the precision of the pipeline used to label these training data.

In this work, we introduce a new approach to stellar flare detection using hidden Markov models (HMMs). The advantage of using HMMs in the context of stellar flare detection is that they are more likely to detect lower energy flares than traditional sigma-clipping methods. HMMs are flexible time series models that are popular in many domains, including ecology (V. Leos-Barajas et al. 2017; T. Adam et al. 2019; B. T. McClintock et al. 2020), health (J. P. Williams et al. 2020), and sports (M. Ötting et al. 2021). Within astronomy, A. Stanislavsky et al. (2020) used a two-state HMM to model daily solar X-ray flux emissions, aiming to predict future solar flares.

HMMs provide a way to model different states underlying a time series, with a probability associated with the transitions between them (W. Zucchini et al. 2017). This is a very natural scheme to approach the detection of stellar flares—a star may be in a “quiet,” “flare-firing,” or “flare-decaying” state. Thus, if an HMM is fit to the light curve of a star, every point in that time series can be estimated to come from one of these three states. This allows one to discover both the firing and decaying phases of each flare. The decaying state is particularly helpful, as it helps characterize the end of a flare as the star transitions back to the quiet state.

In addition to using an HMM, we simultaneously fit *celerite* to model the quasiperiodic trend in the star’s light curve. We show that this simultaneous fitting of *celerite* and the HMM not only removes the need for iterative fitting when searching for flares, but also improves the *celerite* fit overall. That is, *celerite* does not as easily absorb small flares nor the decaying portion of larger flares. For our entire analysis, we adopt a Bayesian approach.

This paper presents our HMM for the detection of stellar flares, shows its merits, and applies it to a stellar light curve measured from TESS. Our paper is organized as follows: In Section 2, we describe the data that motivated this study. In Section 3, we thoroughly describe our method; we begin with a quick overview of *celerite* (Section 3.1), followed by an introduction to HMMs (Section 3.2); then we proceed to describe our observational model (Section 3.3), our particular HMM (Section 3.4), our model fitting techniques (Section 3.5), how we identify and characterize flares (Section 3.6), and the injection recovery experiments performed (Section 3.7). The results of our injection recovery experiments and the application of our HMM to TESS data are presented in Section 4. We conclude with a discussion and a summary of future research directions in Sections 5 and 6.

2. Data

To test and demonstrate our HMM approach, we use M dwarf TIC 031381302 2 minute cadence data measured by TESS, available on the MAST webpage.⁸ We use this star’s Pre-search Data Conditioning Simple Aperture Photometry (PDCSAP) light curve for both our injected flare tests

Table 1
Notations Used in This Paper

	Meaning
Y_t	Observed brightness of star at time t
f	Quasiperiodic trend of the time series (modeled with <i>celerite</i>)
f_t	<i>Celerite</i> -modeled trend at time t
Z_t	Flaring channel (time series without trend)
μ	Mean flux in quiet state
\mathcal{K}	Kernel for <i>celerite</i>
σ^2	Variance of measurement noise
S_t	State of time series at time t
Q	Quiet state
F	Firing state
D	Decay state
$\log(\lambda)$	Log average flux increase during firing state
$\logit(r)$	Logit of the decay rate
$P_{Q Q}, P_{F Q}$	Transition probabilities from quiet state
$P_{F F}, P_{D F}$	Transition probabilities from firing state
$P_{Q D}, P_{F D}, P_{D D}$	Transition probabilities from decay state

(Section 3.7) and for a case study demonstration detecting real flares (Section 4.2). This star was chosen because it has long portions of the time series undergoing quiescent oscillations, in which simulated flares could be injected to test our method. At the same time, there are parts of the time series for this star that have known flare events—these portions of the light curve are used to demonstrate that our HMM can recover the same flares as other methods.

3. Methods

To identify flares in a stellar brightness time series, we simultaneously model the quasiperiodic changes of the star and the star’s flares. For the former, we use *celerite* (D. Foreman-Mackey et al. 2017), and for the latter, we use the HMM described in this paper. Readers familiar with *celerite* may want to skip ahead to Section 3.2. Those readers familiar with HMMs may want to skip to our particular setup in Section 3.4.

For quick reference, a list of our mathematical notation is shown in Table 1.

3.1. *Celerite* and Detrending

To account for the star’s quasiperiodic changes as well as the mean brightness, we use *celerite*, a physically motivated GP widely used to model the trend of stellar light curves (D. Foreman-Mackey et al. 2017). In principle, one could use any kernel provided in *celerite*, but in this study, we use the rotation kernel, the same one used in A. A. Medina et al. (2020), which consists of a sum of two simple harmonic oscillators.

For the GP, let μ , \mathcal{K} , and f denote the mean function, the kernel parameter, and the (latent) trend, respectively. The value of f at time t is denoted f_t . For a single star, the observed light curve Y_t is modeled as the sum of the trend f_t and a flaring channel Z_t ,

$$f|\mathcal{K}, \mu \sim \text{celerite}(\mu, \mathcal{K})$$

$$Y_t = f_t + Z_t. \quad (1)$$

We use priors for the parameters of μ and \mathcal{K} recommended in D. Foreman-Mackey et al. (2017). The HMM for the flaring channel Z_t is described in Section 3.4.

⁸ doi:10.17909/kk7d-hg51

3.2. An Introduction to HMMs

In its basic form, an HMM is a doubly stochastic process composed of an observable, state-dependent process $\{Z_t\}_{t=1}^T$ and a state process $\{S_t\}_{t=1}^T$. At each point in time, t , the time series is assumed to be in one of N possible states (i.e., the time series follows a state process S_1, S_2, \dots, S_T). The states are taken to be discrete latent variables generated from a first-order Markov chain that evolves over time according to an $N \times N$ transition probability matrix with entries $p_{i,j} = \Pr(S_t = j | S_{t-1} = i)$, for $i, j \in \{1, \dots, N\}$. The state at time $t = 1$ is taken to be generated according to an initial state distribution.

The observations are modeled assuming they are emitted from a set of state-dependent distributions, i.e.,

$$g_n(Z_t) = g(Z_t | S_t = n), \quad \text{for } n \in \{1, \dots, N\}. \quad (2)$$

For example, if the HMM is a three-state model ($N = 3$), then there would be three different distributions g , which describe a different data-generating process conditioned on each state n . In practice, the number of states N and the distributions g are defined using scientific domain knowledge. The parameters of these distributions may be estimated within the model or fixed. The $N \times N$ matrix describing the probability of transition between states is also estimated in practice.

When fitting an HMM to real time series data, one can use the estimated model to obtain the state sequence $\hat{s}_1, \dots, \hat{s}_T$; the most likely, under the assumed model to underlie the observations. This is known as state decoding and can be efficiently carried out through the Viterbi algorithm (A. J. Viterbi 1967; see also G. D. Forney 1973 for a detailed description).

The latter uses the estimated state-dependent distributions \hat{g}_n to compute the probability density $\hat{g}_n(z_t)$ of each observation when a specific state $n \in \{1, \dots, N\}$ is active. The algorithm combines these probability densities with the estimated transition probabilities to recursively determine the most likely sequence leading to each possible state n , at each time $t \in \{1, \dots, T\}$. This is done through recursively computing the quantities

$$\xi_{t,n} = \left(\max_i (\xi_{t-1,i} \hat{p}_{i,n}) \right) \hat{g}_n(z_t)$$

for $t = 2, \dots, T$, initialized with $\xi_{1,n} = \Pr(S_1 = n) \hat{g}_n(z_1)$. i.e., $\xi_{t,n}$ corresponds to the likelihood of the state sequence from time 1 to $t - 1$ most likely to lead to state n being the one active at time t , given the observations between 1 and t . The most likely full state sequence is then determined by recursively maximizing over these likelihoods, starting with

$$\hat{s}_T = \operatorname{argmax}_{n=1, \dots, N} \xi_{T,n},$$

and setting

$$\hat{s}_t = \operatorname{argmax}_{n=1, \dots, N} \xi_{t,n} \hat{p}_{n, \hat{s}_{t+1}},$$

for $t = T - 1, T - 2, \dots, 1$.

Additional references and some foundational papers about HMMs are L. E. Baum et al. (1970), L. R. Rabiner (1989), and W. Zucchini et al. (2017).

Table 2

Interpretation of Transitions between States

		Q	to F	D
from	Q	Remain quiet	Start firing	<i>Forbidden</i>
	F	<i>Forbidden</i>	Increased firing	Start decaying
	D	Return to quiet	Start compound flare	Decaying continues

3.3. Observational Model

In our application, we assume an observational model for the random variable Y_t at a given state S_t :

$$Y_t | S_t = f_t + Z_t | S_t, \quad (3)$$

where f_t is the *celerite*-modeled trend at time t (note that it does not depend on the underlying state), and Z_t is the flaring channel. The latter's distribution depends on whether the star is in a quiet, firing, or decaying state, and is described next.

3.4. A Quiet-firing-decay HMM for Flare Events

We propose a three-state autoregressive HMM (J. D. Hamilton 1990) for modeling flare events in the detrended light curve. Each point in the time series can result from one of three (hidden, unobservable) states: quiet, firing, or decay (denoted as Q , F , and D , respectively). The Q state is used to model the time series when the star is not in any flare event, while the F and D states are used to model the increasing and decreasing phases of a flare.

Recall the probabilities p_{S_{t-1}, S_t} of switching between states at each step in the time series (akin to a Markov Chain); they are conditional probabilities of the form $p_{S_t | S_{t-1}}$, where S_{t-1} is the previous state; for example, $p_{F|Q}$ denotes the probability of transitioning to state F given that the star is in state Q . The interpretation given to each of these transitions is illustrated in Table 2.

The transition from Q to F accounts for the firing rate of flares from the star's quiet state. Once the star is in the flaring state, the transition from F to F accounts for the increasing phase of a flare, while the transition from F to D accounts for the decay of the flare. When the time series is in the decay state, transitioning from D to F corresponds to a compound flare. Note that we forbid the transition from Q to D (i.e., when the star is quiet, it will not suddenly “decay”) and from F to Q (i.e., when the star is flaring, the flare will not spontaneously disappear). All other transitions are allowed, and are each modeled with parameters that account for different physical characteristics.

We model the flaring channel Z_t at time t given the state $S_t \in \{Q, F, D\}$ and the previous step Z_{t-1} as follows:

$$\begin{aligned} Z_t | (S_t = Q, Z_{t-1}, \sigma^2) &\sim \mathcal{N}(0, \sigma^2), \\ Z_t | (S_t = F, Z_{t-1}, \lambda, \sigma^2) &\sim \mathcal{N}(Z_{t-1}, \sigma^2) + \operatorname{Exp}(\lambda), \\ Z_t | (S_t = D, Z_{t-1}, r, \sigma^2) &\sim \mathcal{N}(rZ_{t-1}, \sigma^2). \end{aligned} \quad (4)$$

The distributions above are the g_n distributions mentioned in Section 3.2 and Equation (2). Note that the value of the time series at time t is always dependent on the current state of the star S_t and on the value of the time series in the previous step (Z_{t-1}) when in states F or D . The Q state is modeled as an

Table 3
Prior Setting of the QFD Part of the Model

Parameter	Meaning	Prior Distribution	Hyperparameter Used
μ	Mean flux at quiet	$N(\mu_0, \sigma_0^2)$	$\mu_0 = 0, \sigma_0^2 = 100\sigma^2$
σ^2	Variance of measurement noise	$\text{invGamma}(\alpha, \beta)$	$\alpha = 0.01, \beta = 0.01$
$\log(\lambda)$	Log average flux increasing during firing	$N(\mu_\lambda, \sigma_\lambda^2)$	$\mu_\lambda = 0, \sigma_\lambda^2 = 1e3$
$\text{logit}(r)$	Logit of decay rate	$N(\mu_r, \sigma_r^2)$	$\mu_r = 0, \sigma_r^2 = 1e3$
$p_{Q Q}, p_{F Q}$	Transition probabilities from quiet state	$\text{Dir}(\alpha_Q)$	$\alpha_Q = (1, 0.1)$
$p_{F F}, p_{D F}$	Transition probabilities from firing state	$\text{Dir}(\alpha_F)$	$\alpha_F = (1, 1)$
$p_{Q D}, p_{F D}, p_{D D}$	Transition probabilities from decay state	$\text{Dir}(\alpha_D)$	$\alpha_D = (1, 0.1, 1)$

independent normal random variable with variance σ^2 . When formulating the distributions in Equation (4), we assume that when quiet, the flaring channel will just be measurement noise on top of the quiescent trend; when firing, the flaring channel will be around the previous channel plus an independent flux increase exponentially distributed; and when decaying, the flaring channel will be centered around a scaled value of the previous channel.

Recall that the model for the observed light curve y_t is the combination of the trend and flaring channel; that is, *celerite* and the HMM are fit simultaneously, such that \mathbf{f} is a latent variable (Figure 1). We call this combined model *celeriteQFD*. A summary of the HMM parameters, their prior distributions, and hyperparameter values can be found in Table 3. *invGamma* is the inverse Gamma distribution, and the transition probabilities (e.g., $p_{Q|Q}, p_{F|Q}$, etc.) use Dirichlet prior distributions. This setting is used for all injection recovery and real data examples.

3.5. Computation, Model Fitting, and State Decoding

For computational purposes, we split each time series $\{y_t\}_{t=1}^T$ into smaller chunks of 2000 time steps. We subtract the overall mean from the light curve to center it around zero. Our model was implemented in *Stan* (B. Carpenter et al. 2017), while the C++ code for *celerite* was adopted from the Python library *EXO-PLANET*. The posterior distributions of model parameters and derived quantities are sampled using a dynamic Hamiltonian Monte Carlo algorithm (M. D. Hoffman & A. Gelman 2014). States are decoded using the *Viterbi* algorithm as described in Section 3.2 (see also G. D. Forney 1973). The *Stan* implementation, as well as the injection recovery tests, can be found in the first author's GitHub repository: Esquivel-Arturo/*celeriteQFD*.⁹

To obtain samples from the posterior distributions of the model parameters, we use two Markov Chain Monte Carlo (MCMC) chains and sample 1000 (2000 in total) posterior values, after discarding the first 1000 samples obtained during the warm-up period. For each joint posterior sample of the parameters, we use the *Viterbi* algorithm to uncover the most likely state sequence that could have generated the data. In this manner, we are able to propagate the uncertainty around our parameters to produce 2000 most likely state sequences and capture the variability in state decoding results. Thus, for every point in the time series, we have a “decoding distribution” (see Figure 2 top panel) of the states (Q, F, D). For each point in the time series, we estimate the state of the star to be the one that appears most frequently across the 2000 decodings of that time step.

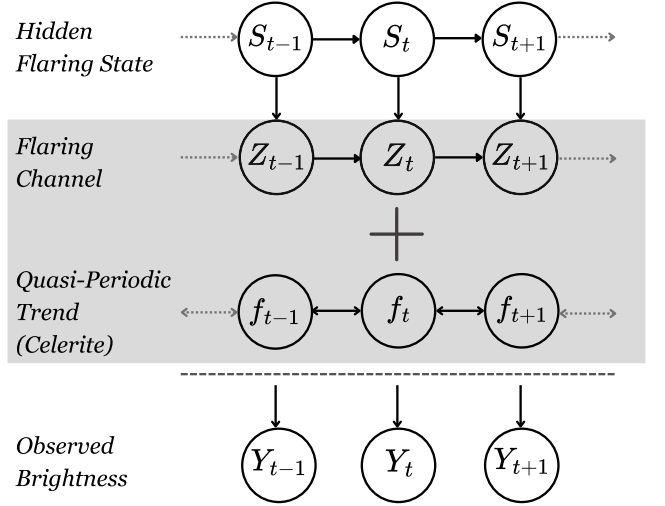


Figure 1. Graphical representation of our complete model for observed brightness decomposed into its various parts. The quasiperiodic trend reflects the average brightness of the star when it is not flaring. The flaring channel represents the extra brightness due to the state of the star. The hidden flaring state represents the (unobserved) state of the star (Q, F , or D).

Each *celeriteQFD* implementation on 2000 time steps took between 1 and 4 hr to run using two cores of an M1 MacBook Pro with 16 GB of memory. When running *celerite* alone and under the same conditions, it usually took between 0.5 and 1.5 hr for the model to run.

3.6. Identifying and Characterizing Flares

To identify flares after fitting *celeriteQFD*, all consecutive points decoded in a non- Q state (i.e., the rise F and fall D of the flare) are used to define the duration of a flare. A flare is considered over once the time series returns to state Q . In other words, the duration of a flare is defined as the time elapsed from when the star enters the flaring state (Q to F) to when the star reenters the quiet state (D to Q). For example, in Figure 2 the flare was estimated to commence with the peak red point and end right before time 1331.30.

This method also allows us to find compound flares (e.g., the time series could be in the decaying phase of a flare, and then start firing again). As the state decoding of a compound flare will have multiple peaks we skip any peaks identified in a flare's duration when searching for the next flare. We can also quantify uncertainty around the *Viterbi* “decoding distributions.” For example, Figure 3 shows two flares detected very close to each other. From the decoding proportions, we can see that about 10% of the *Viterbi* state decodings point to a compound flare instead of two separate flares.

⁹ <https://github.com/Esquivel-Arturo/celeriteQFD>

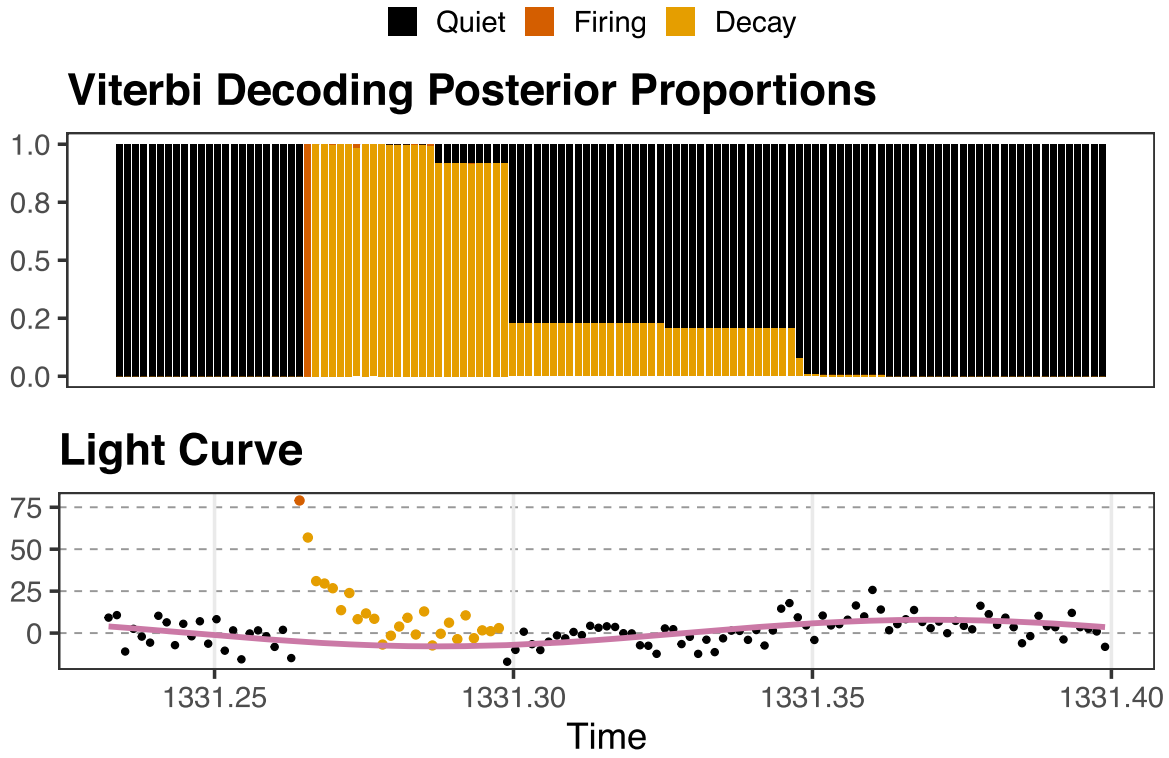


Figure 2. Detected flare example from implementing `celeriteQFD` on the real time series of TIC 031381302. Top panel: state “decoding distributions” per observation across all Viterbi sequences. Second panel: shows the fit of `celeriteQFD`, the estimated trend (purple curve) and assigned state to each point (black points are state Q , red points F , and orange points are D).

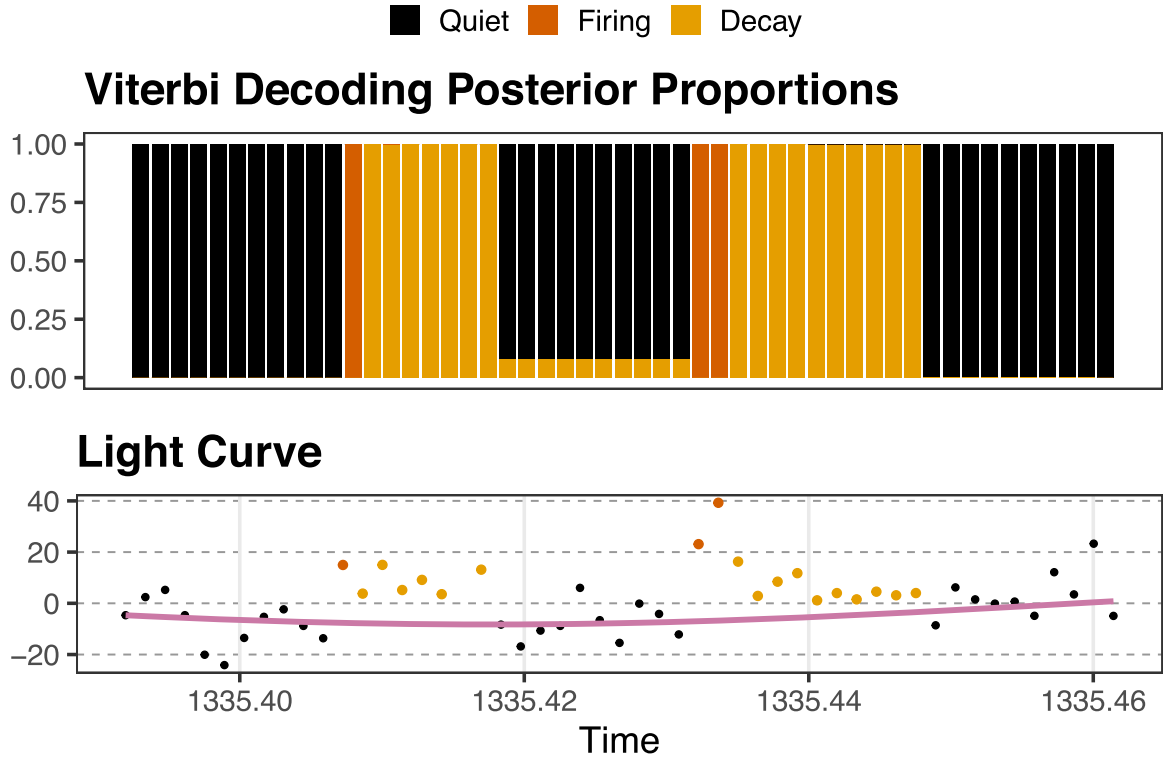


Figure 3. Example of two flares detected from implementing `celeriteQFD` on the real time series of TIC 031381302. Top panel: state “decoding distributions” per observation across all Viterbi sequences. Second panel: shows the fit of `celeriteQFD`, the estimated trend (purple curve) and assigned state to each point (black points are state Q , red points F , and orange points are D).

3.7. Injection Recovery Experiment

To test the ability of our HMM to detect stellar flares, we perform an injection recovery experiment: we inject simulated flares into a real stellar time series, apply our HMM algorithm to detect the simulated flares, and compare our results to the ground truth and those of using a sigma-clipping approach.

We use the mean-centered flux time series data from one star as the base time series for our injection recovery experiment and randomly inject Kepler flares following the procedure outlined in J. R. Davenport et al. (2014). We use TIC 031381302 (from day 1325.292 to day 1327.377 in Barycentric Julian Date (BJD), $n = 1500$). This time segment was chosen to avoid already-discovered natural flares.

We inject five flares, both “small” and “large,” at randomly chosen points in the time series. The timescale of our Kepler-like flares, $t_{1/2}$, is proportional to its peak flux. The peak fluxes of the injections are i.i.d. Pareto (i.e., they follow a truncated power-law distribution):

$$p(x) = \frac{\kappa x_m^\kappa}{x^{\kappa+1}}, \quad (5)$$

where x_m and κ are the distribution parameters.

We use different parameter values ($t_{1/2}$, x_m , κ , δ , ν) to simulate small and large flares. The parameter δ serves as an offset added to the simulated peak fluxes, ensuring that their brightness always exceeds this threshold. Similarly, ν is used as an upper limit; we truncate the Pareto distribution to prevent the simulated values from exceeding the specified maximum ν . We use $(5 \times 10^{-5}, 10, 1, 30, 150)$ for small flares, and $(5 \times 10^{-5}, 50, 1, 0, 300)$ for large flares. These parameters were chosen because they result in significant size differences of the simulated flares while still allowing for a certain degree of variation within each group. Among the small simulated flares, their equivalent durations (EDs; see Section 3.8) vary from 7.28 to 109 s, with a median of 13.4 s. For large flares, their EDs span from 10.9 to 486 s, with a median of 41.3 s.

For each set of parameter values, we separately simulate and inject five flares into the time series, and perform our analysis of flare recovery. We repeat the procedure 100 times for each parameter scheme. Although we do not explicitly study our method's ability to recover compound flares, we do allow the simulated flares to overlap in time and form compound flares in the base time series.

3.8. Flare Detection Evaluation

For each simulated time series, we run `celeriteQFD` and obtain a Bayesian estimate of the state of the star (Q , F , or D) at every time step (see Section 3.5 and Figure 2). Once we have the estimated states for all points in a time series, flares are identified as described in Section 3.6.

To evaluate the accuracy of flare detection using our HMM framework, we compare the number of detected flares to the ground truth (i.e., flares injected). We also compare our results to those obtained with the more commonplace sigma-clipping a - $b\sigma$ rule, where a is the number of consecutive points in the time series that are $b\sigma$ away from the mean flux μ of the detrended time series (i.e., the rule outlined in S. W. Chang et al. 2015 and used in, e.g., A. A. Medina et al. 2020; E. Ilin et al. 2019). In particular, we compare against using a 1- 3σ approach, and with `celerite` for detrending. We use the same priors and kernel as with `celeriteQFD`.

An example of one injection recovery experiment under the small flare scheme is shown in Figure 4. The flux was centered, i.e., it is the raw flux data from which we subtracted the grand mean. It shows the estimated trend and state sequence from our HMM (second panel), along with the ground truth (top panel), and with the result of using the 1- 3σ approach (third panel). We use 1- 3σ because it is more sensitive than the usual 3- 3σ and so it is more likely for it to detect small injections.

For each fitted model (100 per method and flare scheme), we calculate the true positive, false positive, and false negative rates of detection, as well as the sensitivity and the positive predictive value (PPV):

$$\text{sensitivity} = \frac{\text{TFD}}{\text{TFT}}, \quad (6)$$

$$\text{PPV} = \frac{\text{TFD}}{\text{FD}}, \quad (7)$$

where TFD is the number of true flares detected; TFT the total number of true flares; and FD the total number of flares detected. We also assess flare detection in terms of the full duration of the flaring processes, i.e., we compute the per observation sensitivity and PPV:

$$\text{sensitivity} = \frac{\text{TFD}_o}{\text{TFT}_o}, \quad (8)$$

$$\text{PPV} = \frac{\text{TFD}_o}{\text{FD}_o}, \quad (9)$$

where TFD_o is the number of observations part of a true flare correctly identified; TFT_o is the total number of observations that are part of a true flare; and FD_o is the total number of observations identified to be part of a detected flare. Note that sensitivity and PPV (Equations (6) and (7)) should each be 1 if we perfectly identify all true flares. Similarly, the per observation metrics (Equations (8) and (9)) should be 1 if the entire duration of every flare is correctly identified and no point is wrongly identified as part of a flare.

Lastly, we evaluate our method's energy recovery performance on the injected flares. For that, we compare the ED of every injected flare with that of the corresponding recovered flare. The ED is understood as the amount of time (in seconds) that a quiescent star requires to emit as much energy as that emitted during the flaring event (L. M. Walkowicz et al. 2011). Following R. Gershberg (1972) and T. J. Moffett et al. (1974), we determine each ED by numerically integrating the expression:

$$\text{ED}_f = \int \left(\frac{I_f(t)}{I_0(t)} - 1 \right) dt, \quad (10)$$

where I_0 is the flux of the star before the flare injections and I_f the resulting flux within the injected flare f . The ED of recovered flares is determined by the flux of the observations correctly recovered within the duration of the injected flare. We assess the relationship between the energy (as measured by the ED) of a flare and the amount (percentage) recovered by our model.

3.9. More Complicated Flares

As a final exploration, we challenge our HMM with a well-studied, but more complicated flare from the light curve of TIC 129646813, during days 1341 and 1342 (BJD). M. N. Günther et al. (2020) discuss the complicated nature of

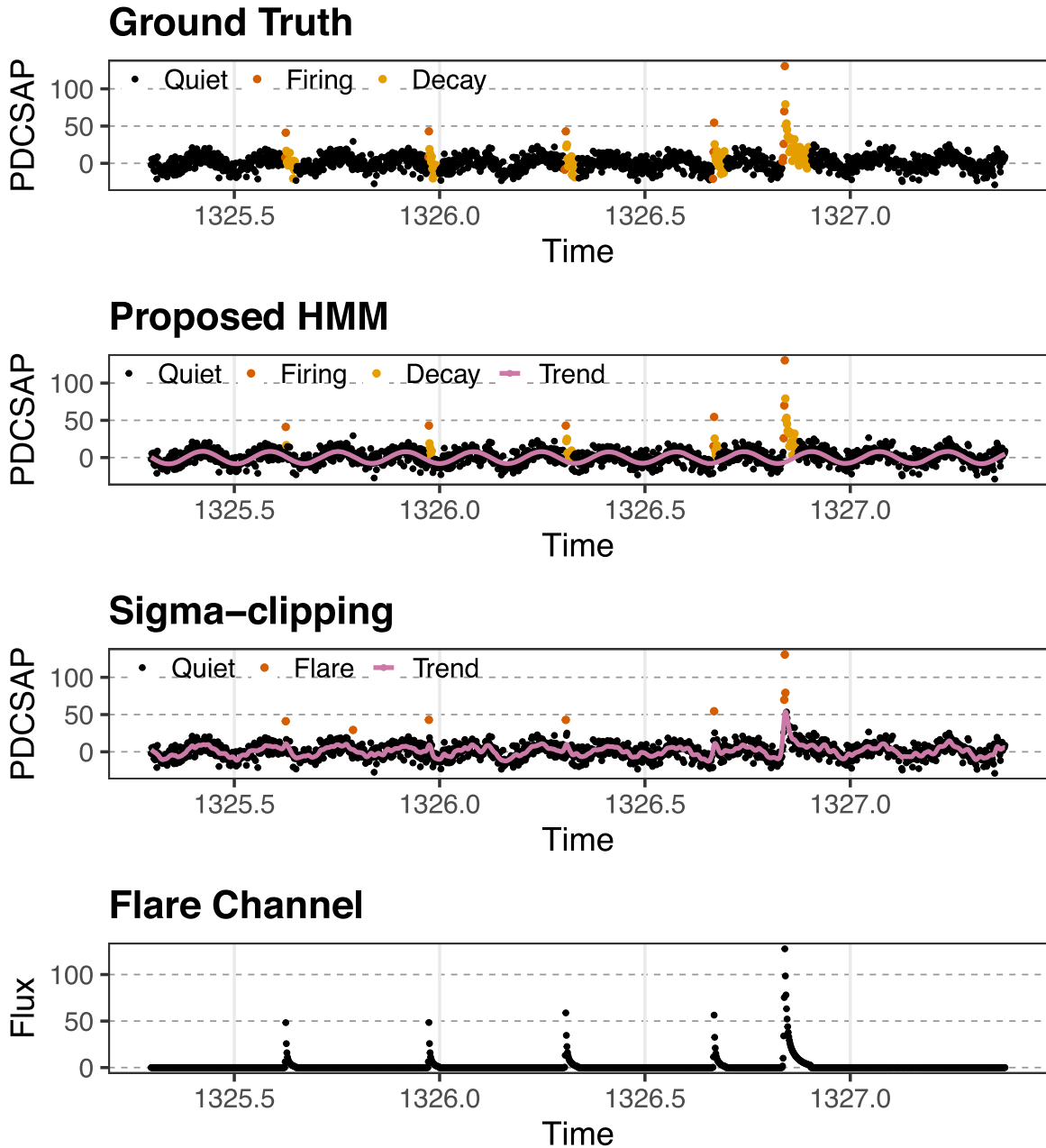


Figure 4. An example of one injection recovery simulation in the small flare scheme. Top panel: ground truth—the real time series for TIC 031381302 with five simulated flares injected (their real states color-coded). Second panel: shows the fit of our proposed algorithm that simultaneously models the trend with *celerite* (purple curve) and assigns states to each point in the time series (black points are state *Q*, red points *F*, and orange points are *D*). Third panel: the sigma-clipping approach that uses *celerite* alone to model the trend (purple curve), with outliers beyond 3σ (red points) used to identify flares. Bottom panel: the flare channel that was injected into the time series.

this flaring outburst, demonstrating it can be difficult to identify the most suitable model to describe it. Their approach consists of a sigma-clipping method which, by employing the Bayes factor (R. E. Kass & A. E. Raftery 1995), sequentially evaluates the improvement in model fit when an additional flare is incorporated into the outburst. This process continues until the inclusion of further flares no longer produces a significant improvement in the model's fit. They mention that a regular sigma-clipping pipeline detects a single flaring peak, while the Bayesian evidence from their approach favors a model with two flares.

Through an exploratory analysis and application of the HMM approach on TIC 129646813, we stretch the limits of our method, compare our results to those shown by

M. N. Günther et al. (2020), and discuss the potential of HMMs in these extraordinary cases (Section 4.3).

4. Results

4.1. Injection Recovery

From the example in Figure 4, it can be seen how, through state *D*, the HMM directly identifies a larger proportion of flaring events than sigma clipping. Moreover, note that *celerite* on its own can absorb part of the flares into the estimated trend of the time series, and so reduce the chance that a flare is detected by sigma-clipping rules.

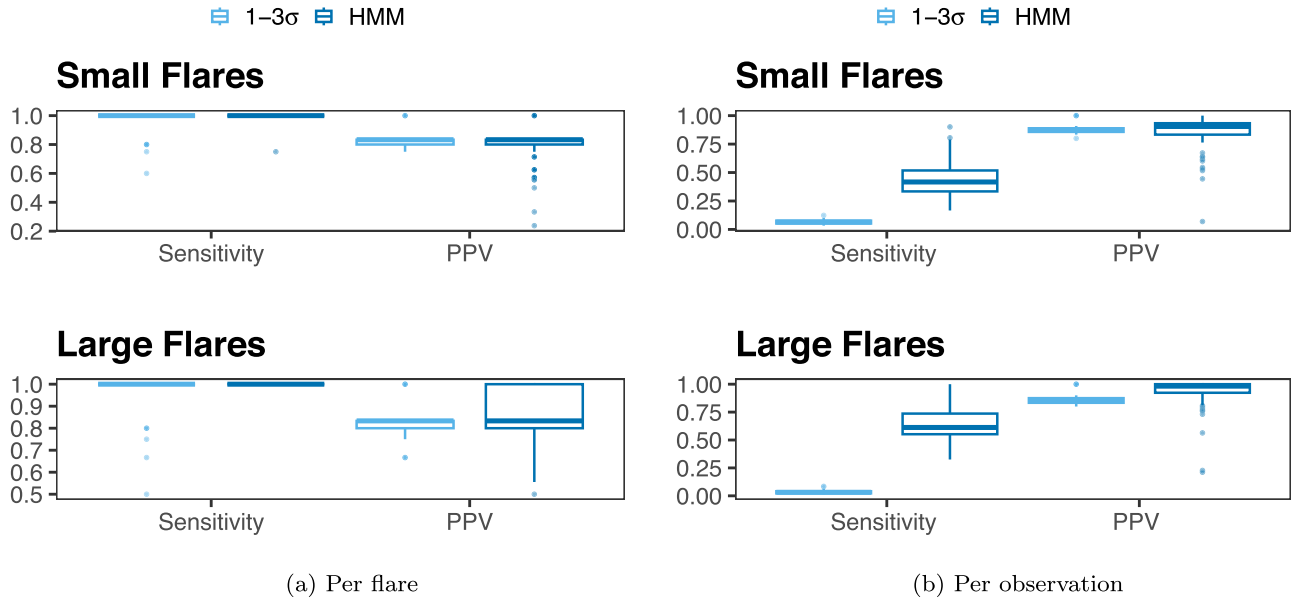


Figure 5. Flare recovery sensitivity and PPV distributions across 100 small and large flare injections, using the $1-3\sigma$ rule and *celerite*QFD (HMM). (a) shows results from detecting flare occurrences (Equations (6) and (7) are used). (b) shows results on a per observation basis (Equations (8) and (9) are used).

The results of the injection recovery experiments are presented in Figure 5. The box plots show a comparison of our method's detection performance with that of $1-3\sigma$ clipping for small and large flares' schemes. They show the sensitivity and PPV distributions across the 100 simulations of each setting, e.g., the first (light-blue) box on the top-left panel corresponds to the distribution of the per flare sensitivity (see Equation (6)) across all 100 sigma-clipping models fitted to the 100 injections of five small flares simulated.

From part (a) of Figure 5, it can be seen that both methods perform almost ideally when it comes to detecting the occurrence of large and small flaring events. Flare detection sensitivity, i.e., the probability of correctly detecting a flare was computed to be 1 in almost all the 100 simulations. They also achieve similar performance in terms of the PPV (see Equation (7)), i.e., the probability of an identified flare indeed being a flare. However, under the large flares scheme, *celerite*QFD often performed slightly better, achieving PPVs of 1 (it was 0.8 almost always with sigma clipping). This indicates that our method is less susceptible to producing false detections.

These results suggest our proposed HMM framework is at least as good as sigma clipping for flaring events detection tasks. But the meaningful difference of our method consists of its ability to directly provide an estimate of the full duration of flaring events. Figure 5(b) contains the distributions of the per observation performance metrics (see Equations (8) and (9)), i.e., the metrics are computed using all points identified as part of a flare by the methods, as compared with all the points that truly belong to injected flares. The plots show that both methods very rarely flagged observations outside of a real flare (PPVs are concentrated very close to 1). Also, one can clearly see the difference between methods when it comes to spotting all light-curve points that are part of a flare; *celerite*QFD consistently identified more than 50% of the observations forming part of a small flare and close to 70% for large flares.

Sigma clipping alone is never used for a full characterization of a flare, which is usually done through further data modeling

steps (see S. W. Chang et al. (2015) for example). Still, this experiment demonstrates the capacity our method has to describe the entire duration of detected flares without the need for extra steps.

Another crucial difference of our method is that it simultaneously carries out detrending and flare detection. By considering the light-curve observations to be a combination of the long-term trend and a flaring channel (see Equation (3)), both components are modeled, accounting for the effects of the other part on the observational process, i.e., the posterior distributions of the *celerite* parameters contain information on the HMM parameters and vice versa.

Figure 6 illustrates the effect of simultaneous detrending and detection. It provides a zoomed-in view of a flare from the example in Figure 4 that was detected by both methods. The first thing to note is that, overall, the *celerite*-estimated trend is considerably less affected by light-curve variability when fitted simultaneously with the HMM. Moreover, in the zoomed flare, note that *celerite*, when used alone, can absorb part of the flares into the trend. Since the data points correspond to an injected flare, we know the increased flux of those observations is not part of the long-term pattern of the light curve and that detrending should ideally ignore it. *celerite*QFD, is not only able to estimate the trend unaffected by the increased flux, it is also capable of identifying many of the observations as part of the decaying phase of a flare. This result is critical, suggesting that better detrending can be achieved, further leading to less biased estimates of the flares' energies.

4.2. Case Study: Photometric Data from the TESS Mission

As a demonstration of real flares' detection, we apply our HMM method to a large portion of the TIC 031381302 light curve. The time series was sliced into pieces of 2000 time steps to more efficiently fit the models and conduct flare detection. Trace plots of the MCMC sub chains of the parameters

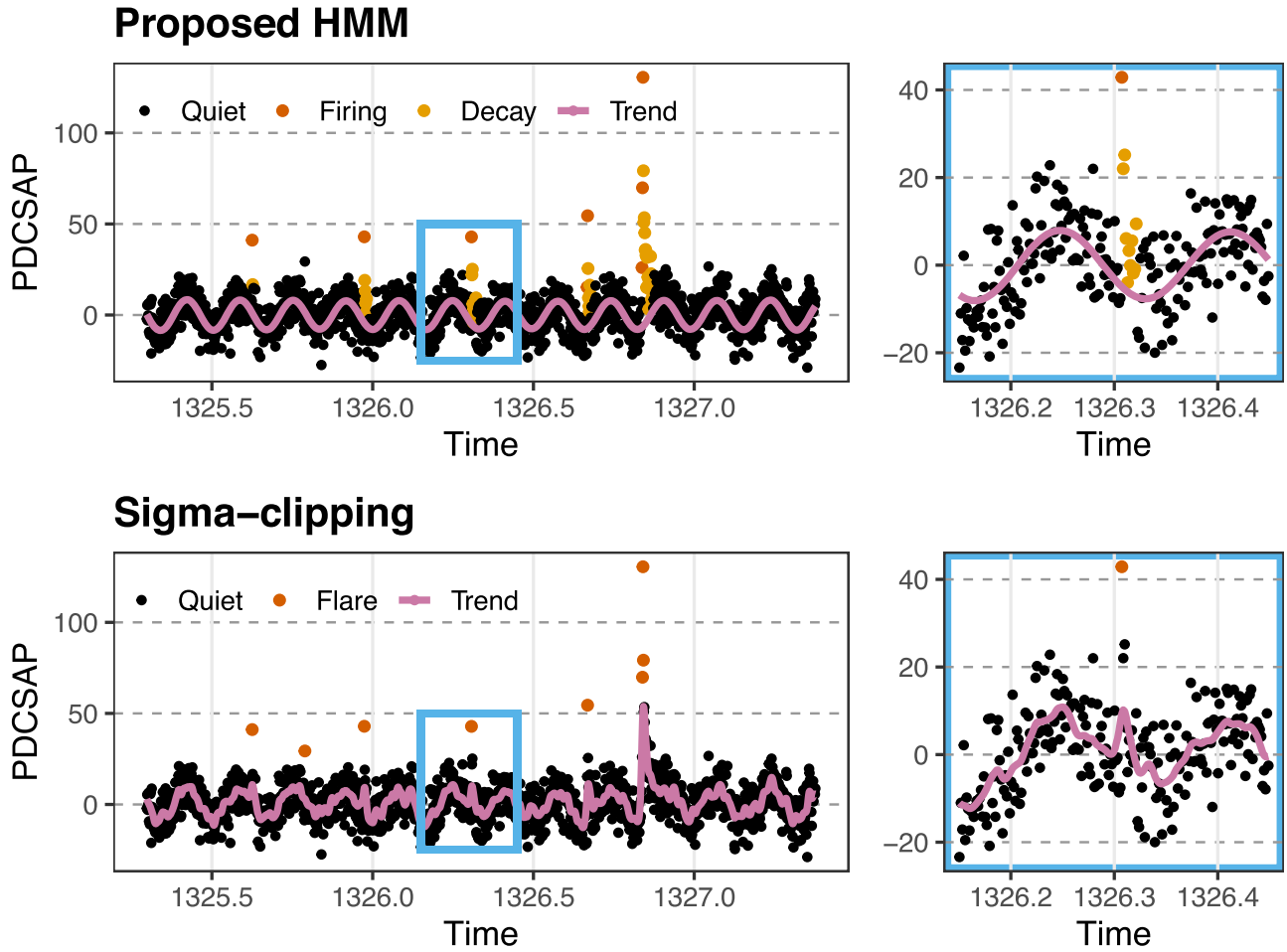


Figure 6. An example comparison of *celeriteQFD* and standard sigma clipping for identifying the injected flares to part of the TIC 031381302 light curve. Top row: *celeriteQFD*, which simultaneously models the trend with *celerite* and assigns states to each point in the time series. Bottom row: the sigma-clipping approach that uses *celerite* alone to model the trend, with outliers beyond 3σ used to identify flares. The right-hand column shows a zoomed-in portion of one of the flares identified by both methods.

sampled were produced and inspected without any indication of lack of convergence.

A large portion (from day 1325.292 to day 1353.177 (BJD)) of TIC 031381302 PDCSAP mean-centered flux, along with the resulting fit and state decoding, are shown in Figure 7. During the period observed, a total of 11 flares were detected, with an average duration of 10.3 observations (approximately 0.01545 days or 1334.88 s). The estimated (using the posterior median) transition probability ($p_{F|Q}$) was 0.00170, with a 95% credible interval of (0.00007, 0.00820). i.e., the estimated probability that this star starts firing at any particular time, given that it was quiet in the previous time step, is 0.17%.

A smaller portion of the TIC 031381302 light curve (from day 1333.627 to day 1336.404 (BJD)) and the resulting fit of both models is shown in Figure 8. The right-hand side shows a zoom-in into a portion containing what both methods identified as a flare event. Note the similarity with many aspects of the flare shown in the right-hand side of Figure 6. The trend modeled using *celerite* only (bottom row) gets distorted, absorbing observations of higher brightness. Given that these points follow two observations of peak brightness, it is rather likely that at least some of them correspond to the decaying phase of a flare. This is precisely the kind of case in which flare energies could be underestimated. Also note that using our approach, the estimated trend remained unaltered in this

window, and the HMM identified multiple observations to be in a decaying state (the same way as in the synthetic case).

By comparing the simulation results, where the ground truth is known, with real data results, it seems *celeriteQFD* can in fact better model the long-term trend of a light curve. Additionally, these results indicate the model is capable of directly identifying light points conforming the decaying phase of a real flare, determining the duration of flaring events. Moreover, through the “decoding distributions” it directly provides a way to quantify the uncertainty about the estimated durations of the flares. For example, in Figure 2, top panel, it can be seen that almost 20% of the Viterbi sequences estimated the decaying phase of the flare detected extends up until time 1331.35. Similarly, it can be seen that around 10% of them identified the flare to end nine time steps earlier than estimated using the majority state. This uncertainty can be easily propagated into the final goal of producing energy distributions, potentially leading to more reliable and comprehensive distributions.

4.3. Case Study: A More Complicated Flare

The results above indicate our method can perform well on classic one-peak flares. However, flares composed of multiple impulsive events sometimes exhibit more complex structures. Sometimes it is not easy to determine the number of peaks

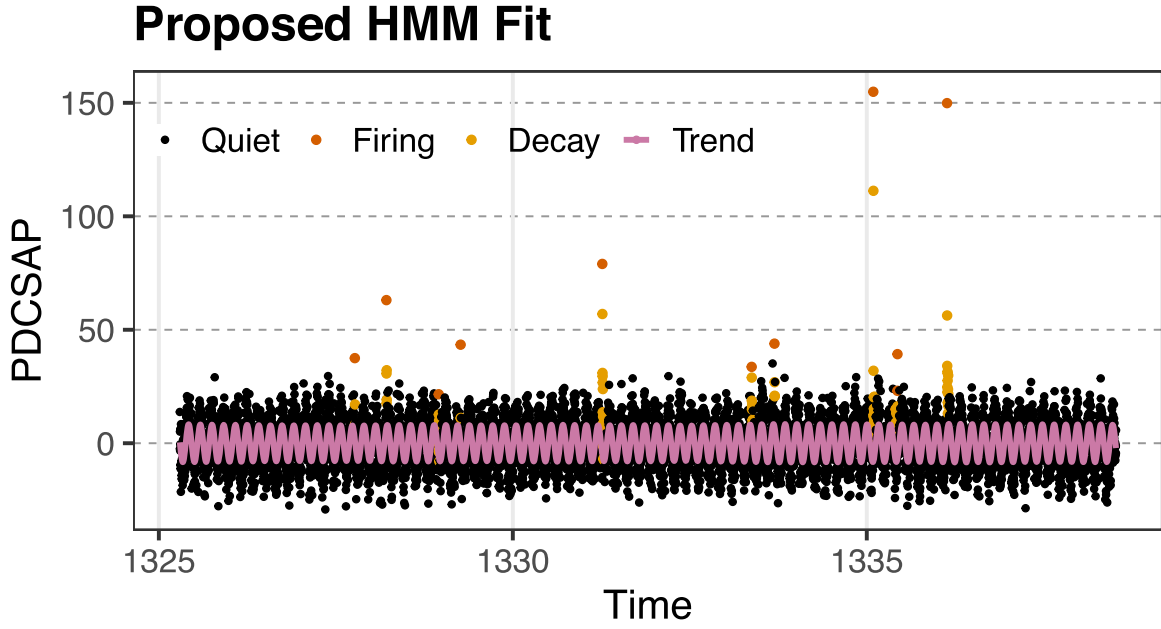


Figure 7. TIC 031381302 mean-centered light curve along with the fit of `celeriteQFD` that simultaneously models the trend using `celerite` and assigns states to each point in the time series using the HMM.

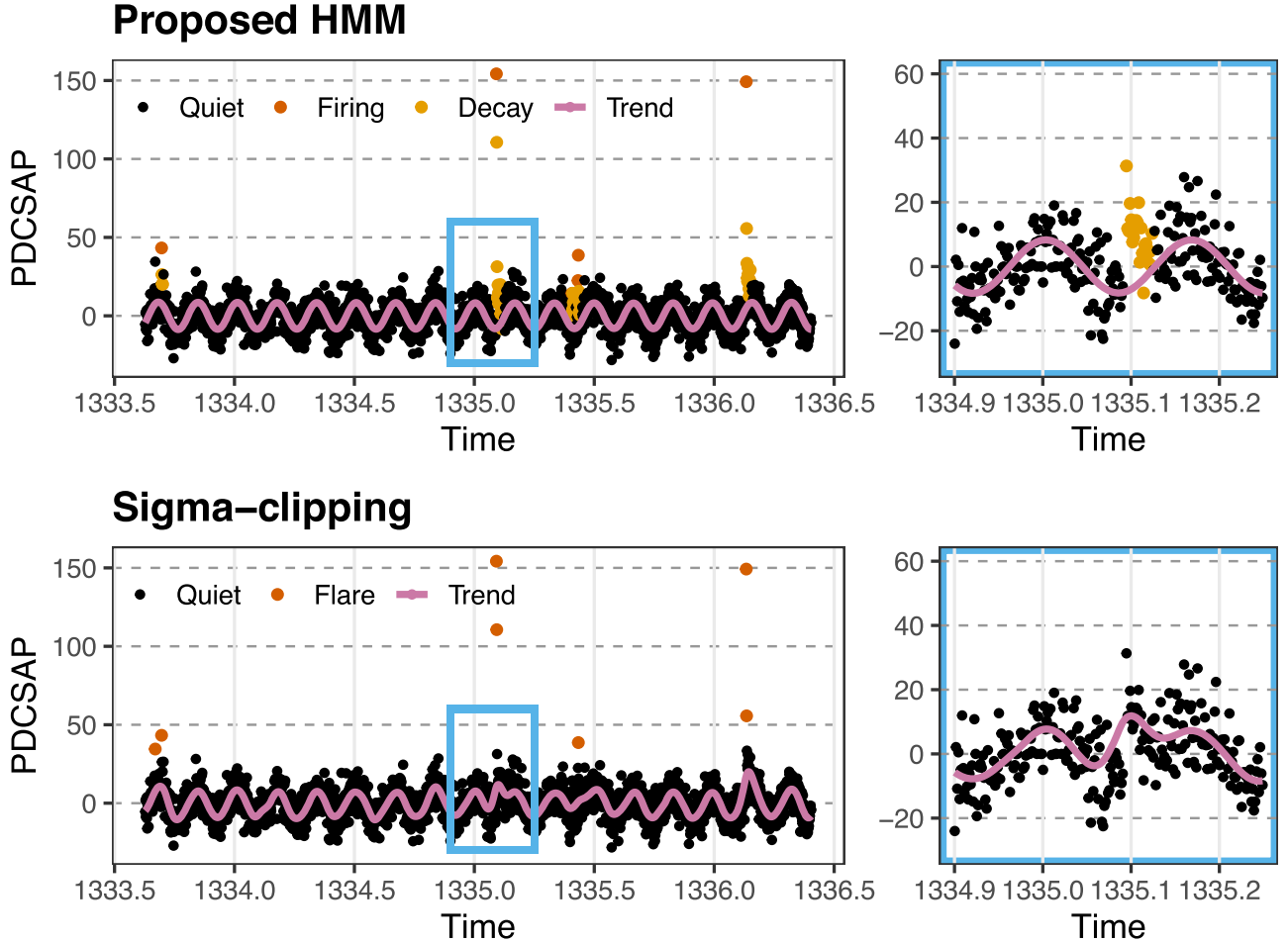


Figure 8. An example comparison of our HMM approach and standard sigma clipping for identifying flares, using real data part of the TIC 031381302 light curve. Top row: `celeriteQFD`, which simultaneously models the trend with `celerite` and assigns states to each point in the time series. Bottom row: the sigma-clipping approach that uses `celerite` alone to model the trend, with outliers beyond 3σ used to identify flares. The right-hand column shows a zoomed-in portion of one of the flares identified by both methods.

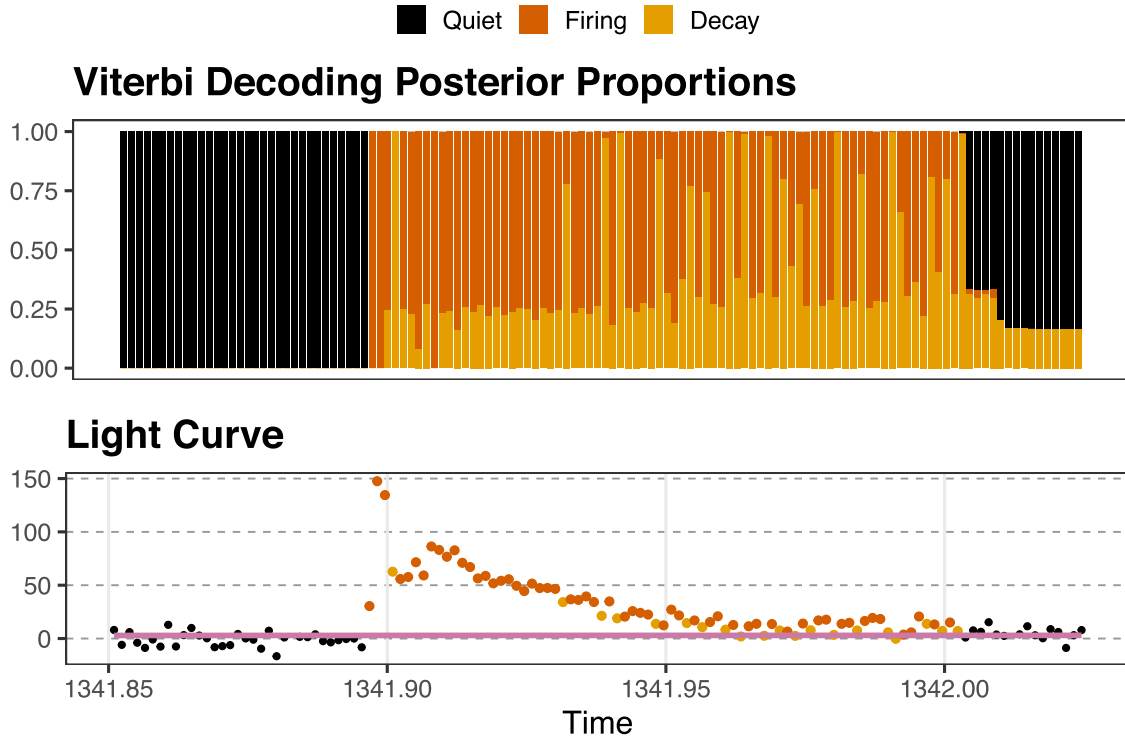


Figure 9. Results obtained through implementing *celeriteQFD* on the real time series of TIC 129646813. Top panel: state “decoding distributions” across all Viterbi sequences per observation. Second panel: shows the fit of *celeriteQFD*, the estimated trend (purple curve) and assigned state to each point (black points are state Q , red points F , and orange points are D).

involved in an outburst. Methodologies that make use of stellar flare templates require the number of flares to be correctly specified to produce the best results.

As a testbed and proof of concept of *celeriteQFD*’s flexibility, and to investigate how it performs in a more complicated setting, we apply our HMM approach to TIC 129646813. As described in Section 3.9, this star exhibits a multi-peak flaring outburst between days 1341 and 1342 (BJD). Figure 9 shows the results of applying *celeriteQFD* to this event.

In the top panel of Figure 9, three points in the time series are unambiguously identified as being in the firing state (the solid red bars, without any orange or black). They correspond to the points of highest brightness within the window of time shown. The first two of these are sequential, and the third is separate. In between these two flaring peaks, there is also an unambiguous decaying state (the solid orange bar). Together, this sequence of observations corresponds to the same first and second flare originally identified by M. N. Günther et al. (2020).

In the bottom panel of Figure 9, each time point is given the color corresponding to the majority state in the top panel. From here, we also can see that *celeriteQFD* identifies both peaks originally detected by M. N. Günther et al. (2020). Thus, a potential advantage of our HMM approach is the detection of multiple peaks in complicated events without the need for an iterative process and/or prior assumptions about the flare count. Moreover, it seems that *celeriteQFD* has enough flexibility to accurately capture the complex structure of this outburst more generally, as described next.

The volatile classification by *celeriteQFD* during the prolonged decay of the flare is particularly noteworthy in the classification across posterior samples (Figure 9, top panel). The proportion of samples assigned to the most frequent state is

lower compared to other observations, where the decoding is more consistent across posterior samples (e.g., all quiet or all firing). Moreover, these states are highly uncertain as the outburst dies down, fluctuating almost periodically between firing and decaying. In M. N. Günther et al. (2020), after fitting their two-flare model, the residuals show what could be an oscillatory pattern. Thus, it is possible that our model’s high uncertainty state fluctuations are the result of true but small oscillations in the data. In other words, there could be an extra layer of complexity to this flaring event that is not being considered. e.g., sometimes some peaks within a flare display quasiperiodic pulsations (QPPs; W. S. Howard & M. A. MacGregor 2022). However, we warn that one should not draw strong conclusions from these exploratory results.

Currently, our method is not meant to describe such complicated mechanisms. The interpretations of states in our particular HMM model are not necessarily applicable to observations such as those in Figure 9. As mentioned above, there could be an extra underlying mechanism other than the star alternating between the firing and decaying phases. As it stands, using our model to capture the behavior of processes as complicated as potential flare QPPs would not be sensible. However, it seems our method could at least hint at the possibility of the existence of such additional processes. As shown with this example, *celeriteQFD* identified a structure incompatible with the one assumed through the state-dependent distributions. It informs about the increased uncertainty that surrounds the observations of the decaying phase of the outburst.

To fully explore cases like this is beyond the scope of this study. However, the flexibility of our method shows potential and it would be relatively easy to expand it to better accommodate these more complicated patterns. For example,

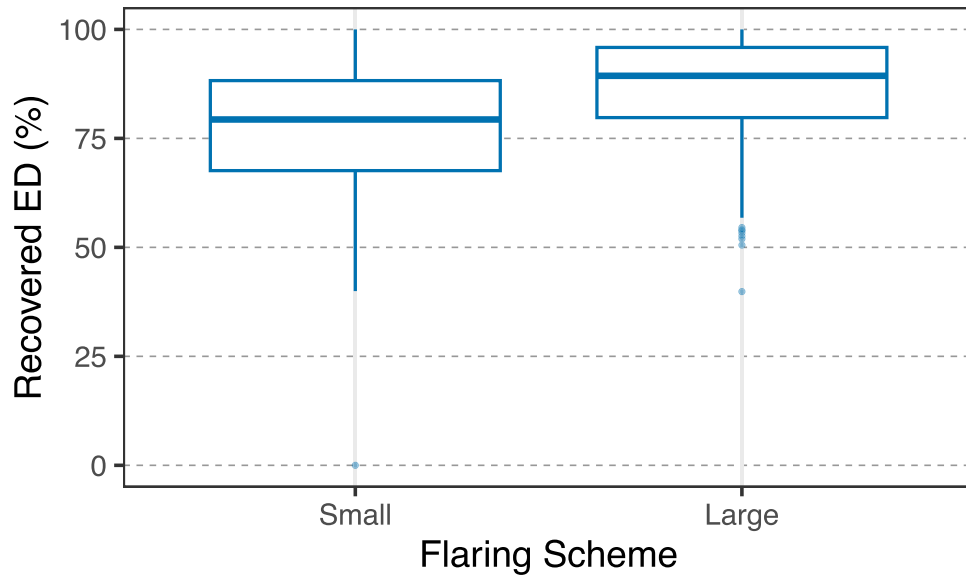


Figure 10. Box plots showing the proportion of the true ED recovered by *celeriteQFD*, across all flares injected in all experiments using the small and large flares schemes.

an additional state could be included with state-dependent distributions that impose a different structure that accounts for possible flare QPPs. This additional state could be considered only during flaring events and could be modeled using simple harmonic oscillators different from the ones used for the *celerite* trend.

4.4. Energy Recovery

To inspect what the detection performance of our method implies in terms of energy estimation, we look at the proportion of ED our model recovered for each injected flare (across the 200 simulation experiments). Figure 10 shows how this proportion is distributed across all the flares from each flaring scheme. The plot shows that *celeriteQFD* energy recovery is generally better on large flares. The recovered ED was at least 75% of the true ED for most large flares, being higher than 90% for more than half of them. In contrast, only in about 50% of the small flare cases, the percentage of recovered ED was found to be 80% or higher. Nevertheless, even for small flares, our model successfully recovered at least 60% of the true ED most of the time.

Figure 11 was produced to further inspect the relationship between the energy of a flare and our method’s capacity to recover it. It shows every flare injected in terms of their recovered ED and the corresponding true value. Points are colored according to the percentage of real ED recovered by *celeriteQFD*. The black line is the identity line, which denotes the perfect recovery threshold. Overall, the points are fairly close to the line. Based on the pattern of the points, and their color, the model’s performance does not seem to be strongly impacted by the energy of the flare. Except for extremely small cases; almost all flares for which the recovered ED was less than 75% had an ED of less than 50 s.

A comparison between Figures 10, 11, and 5(b) shows that the majority of the points missed by the model tend to be of very low energy or luminosity. For instance, in the case of small flares, the per observation sensitivity had a median near 0.50, meaning that in about half of the experiments, 50% or more of the flare-related points were not detected by the model.

However, the median of recovered ED for small flares was nearly 0.80, indicating that the missed points generally accounted for only about 20% of the total ED. Even in the worst performance cases, energy was underestimated by 20%–35%, and this underestimation was primarily associated with extremely low-energy flares. To contextualize, A. A. Medina et al. (2020) identified 1392 flares using TESS observations, of which fewer than 3% had an estimated ED of 50 s or less.

5. Discussion

As shown in this work, an advantage of having the HMM as a flaring model is that we can identify the whole course of a flare via states assigned to each time point. We no longer need to cross-correlate the time series with stellar flare templates and thus remove a step in the analysis process. The HMM approach also allows us to detect compound flares more easily. Moreover, it can be used to give a probabilistic sense of the duration of the flares, making it possible to produce more comprehensive distributions of the energies of the flares.

The inclusion of a state associated with decaying and the capability to simultaneously perform detrending and state decoding constitute a relevant benefit of our proposed method. Through our injection recovery experiment and analysis of a real star, we have shown that *celeriteQFD* can produce better and more stable estimates of the long-term trend of a light curve than *celerite* alone. The agreement between the results obtained for synthetic cases (where the ground truth is known) and real data cases indicates that biased estimation of the trend is indeed an issue that can arise when detrending is done prior to flare detection. It also provides some reassurance that our model is better equipped to handle the problem and prevent such bias. Further, we have shown that our HMM method can detect flares of lower energy that might be missed by other methods, even sensitive methods such as $1\text{--}3\sigma$ clipping. Thus, the HMM approach to flare detection could be well-suited for detecting flares in more “inactive” G-type stars.

It is worth mentioning that our flare recovery experiments are in no way exhaustive of flare morphologies present in stars. While it is true that stellar flare templates are more physically

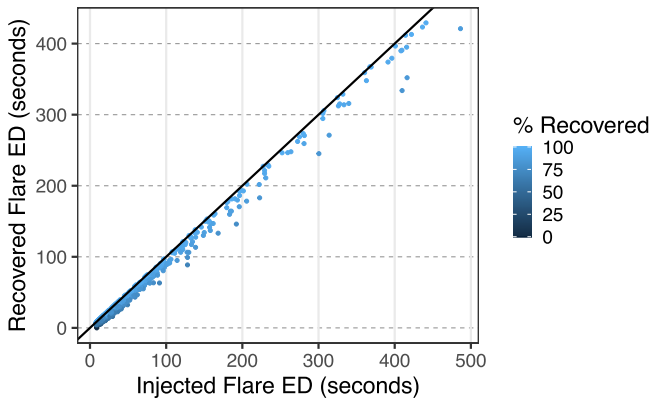


Figure 11. Recovered vs. injected equivalent duration. Each point corresponds to the values of one injected flare. Points are colored based on the proportion of the true ED (x-axis value) accounted for by the recovered ED (y-axis value). The black line indicates the ideal scenario where both values are the same.

motivated than *celerite*QFD, and that *celerite*QFD is not a generative model, HMMs still classify states very well even when the generative process is not specified correctly (S. Ruiz-Suarez et al. 2022). Also, note that we do not make any strong assumptions about the nature of flare events. The few restrictions on the matter are made through the state-dependent distributions, upon which parameters are estimated. This provides flexibility for the model to produce different results in different contexts. We intend to explore this further and conduct detection and energy recovery experiments on different types of simulated flares, assessing our method’s performance and contrasting it with that of other methods.

Considering the ultimate goal of estimating flares’ energies and their distributions, this method paves a new way to calculate stellar flare energy. *celerite*QFD could prove to be generally applicable, versatile, and quantitatively reliable. We plan to continue this work by using our model on other light curves from different stars measured by TESS.

6. Conclusions and Future Work

In this paper, we have introduced an HMM for discovering stellar flares in time series data of M dwarf stars. Our approach has some notable advantages over previous approaches.

First, our method simultaneously fits a *celerite* model for the quiescent state of the star and a three-state flaring model through a hidden Markov process. With this approach, we not only obtain a better estimate of the quiescent state of the star but also eliminate the need for sigma-clipping approaches and iterative fitting of time series data. Moreover, with the combined approach of *celerite* and the HMM, *celerite* does not absorb early or late parts of the flares. Concurrently, the HMM better identifies the whole course of the flare and can also identify compound flares easily.

Second, through our flare injection recovery experiment, we find that our HMM method for flare detection achieves the same or better sensitivity and positive predictive value compared to sigma clipping (Figure 5). Our energy recovery experiment demonstrated that even when the full duration of a flare could not be detected, the missed observations contributed minimally to the total energy. As a result, the estimated ED was rarely as significantly impacted.

Third, our method enables a coherent path for uncertainty quantification. Rather than providing a single most likely flare state sequence for the time series, we obtain a posterior

distribution of most likely flare state sequences by propagating the uncertainty from our parameter estimates. This allows us to capture the variability and uncertainty of the duration of each flare.

While our approach has significant advantages, one potential disadvantage is computation time. Currently, it can take on the order of a couple of hours to run the HMM model on a single star’s time series, as measured by TESS on modest resources (see Section 3.5). We are currently exploring approaches to overcome this challenge, as the advantages of our method seem to outweigh this minor (and surmountable) drawback. Another potential criticism of our approach is that we have not developed the HMM to realistically simulate stellar time series with flares. That is, in this work, we are measuring the capacity of our proposed HMM plus the *celerite* model to identify flares well, but we are not evaluating the model’s generative properties. This is something that could be improved upon and explored in the future.

Overall, this work is a promising initial step and proof of concept in developing a robust flare detection algorithm that does not rely on sigma clipping or iterative approaches. There are many avenues that we plan to explore in future work:

1. We aim to speed up the computation time so that we can apply our HMM approach to a large sample of M dwarf stars observed by TESS and recover stellar flares.
2. In a follow-up paper, we will estimate both the flare energy distribution and flare frequency distribution (FFD) through the posterior distribution of the Viterbi state-decoded sequences. Through this approach, we will be able to propagate the uncertainties in the duration of the flare from the state sequences to the energy of the flare in a coherent way. This, combined with the improved sensitivity and PPV of our method in detecting small and large flares, should produce better estimates and increase our confidence regarding the FFD of M dwarf stars.
3. Ultimately, it could be fruitful to design a hierarchical model that includes the FFD parameters at the population level. In this way, many M dwarf stars could be fit with our HMM approach simultaneously, and both their individual parameters and the population-level parameters of the FFD would be modeled in a coherent way.
4. Extending the model into one with more states is rudimentary. Using additional states to capture extra phenomena, such as QPPs, the presence of transients, or even extremely low-energy observations, could prove useful and lead to better results.
5. After detecting stellar flares in a star, and obtaining the state-decoded sequences that quantify the uncertainty in the duration of each flare for that star, one could still use stellar flare templates to model the flares and obtain parameter estimates of interest. A study comparing this approach with standard approaches in the literature could provide further insight into the benefits (or not) of using HMMs in this framework.
6. The HMM technique presented here could also be further developed for time series data measured across multiple bands, which is now becoming more commonplace (e.g., W. Joseph et al. 2024). This would allow an HMM analysis of stellar flares measured by future data sets (e.g., CubeSat; J. Poyatos et al. 2023).

To our knowledge, this paper is one of the first applications of HMMs to an astronomy problem, and the first to do so for stellar flare detection. This statistical method has promise not only for stellar flare detection but also for other areas in astrophysics with time series data, such as gamma-ray bursts, fast radio bursts, and quasars. Our hope is that this paper is useful as a starting-off point for the astronomical community to use this method in both stellar flare detection and other areas of astronomy.

Author Contributions

J.A.E. and Y. S. contributed equally to this project. Y.S. wrote the original code, performed most of the analysis, and wrote parts of the paper. J.A.E. subsequently modified the code, checked and performed additional analyses, wrote all code and analysis to produce all of the figures (except Figure 1), and contributed substantially to writing Sections 3, 4, and 5.

V.L.B. provided expertise on HMMs and co-supervised J.A.E. and Y.S. V.L.B. also wrote parts of Section 3.2 and made Figure 1. V.L.B. was also a co-I on the DSI grant (see the Acknowledgments) that funded this project.

G.M.E. was the PI on the DSI grant that funded this work. G.M.E. also co-supervised J.A.E. and Y.S. and provided expertise on Bayesian analysis and astrostatistics. G.M.E. edited and contributed to all sections of the paper, and contributed substantially to Section 6.

J.S. co-supervised J.A.E. and Y.S. J.S. contributed to the data handling and interpretation of results and provided comments on the manuscript.

R.V.C. was a co-I on the DSI grant that funded this work. R.V.C. provided expertise on Bayesian analysis and inference, and provided comments on the manuscript.









A.M. co-supervised Y.S., and contributed substantial expertise on M dwarfs. A.M. also helped write parts of Section 1.

J.R.A.D. provided expertise on M dwarfs and stellar flares, provided scientific comments and feedback, and contributed to the editing of the final paper.

Acknowledgments

We thank the referee for the helpful comments that improved this manuscript. Some computations involved in this paper were performed using the compute resources and assistance of the UW-Madison Center for High Throughput Computing (CHTC) in the Department of Computer Sciences. Computing resources in the Department of Statistical Sciences at the University of Toronto were also used, with support from IT specialist Claire Yu. This work was supported by the University of Toronto Data Sciences Institute catalyst fund to G.M.E. (PI), V.L.B., R.C., and A.M. During his work on this project, J.A.E. received funding from the Canadian Statistical Sciences Institute (CANSSI) with support from NSERC.

ORCID iDs

J. Arturo Esquivel  <https://orcid.org/0009-0006-0387-6544>
 Yunyi Shen  <https://orcid.org/0000-0003-2779-6507>
 Vianey Leos-Barajas  <https://orcid.org/0000-0001-8016-773X>
 Gwendolyn Eadie  <https://orcid.org/0000-0003-3734-8177>
 Joshua S. Speagle  <https://orcid.org/0000-0003-2573-9832>
 Radu V Craiu  <https://orcid.org/0000-0002-1348-8063>
 Amber Medina  <https://orcid.org/0000-0001-8726-3134>
 James R. A. Davenport  <https://orcid.org/0000-0002-0637-835X>

References

- Adam, T., Griffiths, C. A., Leos-Barajas, V., et al. 2019, *MEcEv*, **10**, 1536
 Baum, L. E., Petrie, T., Soules, G., & Weiss, N. 1970, *Ann. Math. Stat.*, **41**, 164
 Carpenter, B., Gelman, A., Hoffman, M. D., et al. 2017, *J. Stat. Softw.*, **76**, 1
 Chang, S. W., Byun, Y. I., & Hartman, J. D. 2015, *ApJ*, **814**, 35
 Davenport, J. R., Hawley, S. L., Hebb, L., et al. 2014, *ApJ*, **797**, 122
 Davenport, J. R. A. 2016, *ApJ*, **829**, 23
 Donati, J.-F., & Landstreet, J. 2009, *ARA&A*, **47**, 333
 Doorsselaere, T. V., Shariati, H., & Debosscher, J. 2017, *ApJS*, **232**, 26
 Feinstein, A. D., Montet, B. T., Ansdell, M., et al. 2020, *AJ*, **160**, 219
 Forbes, T. 1991, *GApFD*, **62**, 15
 Foreman-Mackey, D., Agol, E., Ambikasaran, S., & Angus, R. 2017, *AJ*, **154**, 220
 Forney, G. D. 1973, *Proc. IEEE*, **61**, 268
 Gershberg, R. 1972, *ApSSP*, **19**, 75
 Günther, M. N., Zhan, Z., Seager, S., et al. 2020, *AJ*, **159**, 60
 Hamilton, J. D. 1990, *J. Econom.*, **45**, 39
 Hawley, S. L., Davenport, J. R. A., Kowalski, A. F., et al. 2014, *ApJ*, **797**, 121
 Hoffman, M. D., & Gelman, A. 2014, *JMLR*, **15**, 1593
 Howard, W. S., & MacGregor, M. A. 2022, *ApJ*, **926**, 204
 Ilin, E., Schmidt, S. J., Davenport, J. R., & Strassmeier, K. G. 2019, *A&A*, **622**, A133
 Joseph, W., Stelzer, B., Magaudda, E., & Martinez, T. V. 2024, *A&A*, **688**, A49
 Kass, R. E., & Raftery, A. E. 1995, *JASA*, **90**, 773
 Leos-Barajas, V., Gangloff, E. J., Adam, T., et al. 2017, *JABES*, **22**, 232
 McClintock, B. T., Langrock, R., Gimenez, O., et al. 2020, *EcolL*, **23**, 1878
 Medina, A. A., Winters, J. G., Irwin, J. M., & Charbonneau, D. 2020, *ApJ*, **905**, 107
 Moffett, T. J. 1974, *ApJS*, **29**, 1
 Osten, R. A., Kowalski, A., Sahu, K., & Hawley, S. L. 2012, *ApJ*, **754**, 4
 Ötting, M., Langrock, R., & Maruotti, A. 2021, *Adv. Stat. Anal.*, **107**, 9
 Poyatos, J., Fors, O., & Gómez Cama, J. M. 2023, arXiv:2302.12566
 Rabiner, L. R. 1989, *Proc. IEEE*, **77**, 257
 Rietz, S., Stelzer, B., Damasso, M., & Scholz, A. 2020, *A&A*, **637**, A22
 Ruiz-Suarez, S., Leos-Barajas, V., & Morales, J. M. 2022, *JABES*, **27**, 339
 Stanislavsky, A., Nitka, W., Malek, M., Burneck, K., & Janczura, J. 2020, *JASTP*, **208**, 105407
 Viterbi, A. J. 1967, *ITIT*, **13**, 260
 Walkowicz, L. M., Basri, G., Batalha, N., et al. 2011, *AJ*, **141**, 50
 Williams, J. P., Storlie, C. B., Therneau, T. M., Clifford, R. J., Jr., & Hannig, J. 2020, *JASA*, **115**, 16
 Yang, H., Liu, J., Gao, Q., et al. 2017, *ApJ*, **849**, 36
 Zucchini, W., MacDonald, I. L., & Langrock, R. 2017, *Hidden Markov Models for Time Series: an Introduction Using R* (Boca Raton, FL: CRC press),