

GAUSSIAN PROCESS MODELING: A BAYESIAN APPROACH TO DETECTING QUASI-PERIODIC OSCILLATIONS IN SOLAR FLARE DATA

CHRISTOPHER A. ICK,¹ DAVID W. HOGG,² AND DANIELA HUPPENKOTHEN³

¹*New York University
726 Broadway, Ofc 912
New York, NY 10003, USA*

²*New York University
726 Broadway, Ofc 1005
New York, NY 10003, USA*

³*University of Washington
3910 15th Ave NE
Office B356B
Seattle, WA 98195, USA*

ABSTRACT

Despite decades of study at all wavelengths, the exact conditions under which solar flares occur, their energy release mechanisms, and their radiation processes are not well understood. Quasi-periodic pulsations (QPPs) in solar flares have recently emerged as key diagnostics of the plasma conditions and flare triggering processes, but detecting these pulsations reliably has in practice proven difficult. Here, we propose a new reliable and efficient method of modelling solar flare data using Gaussian Processes. This method allows for testing for the presence of a quasi-periodic signal in the presence of a flare envelope and other sources of non-periodic variability. We show the effectiveness of the new approach using both simulated data as well as solar flare observations of M- and X-class flares observed with GOES and the Fermi Gamma-Ray Burst Monitor (GBM).

[ADD RESULTS HERE]

Keywords: keywords go here

1. INTRODUCTION

Solar flares are the most powerful events in the solar system, releasing energies in excess of 10^{32} erg in tens of minutes across the entire electromagnetic spectrum (e.g. ?). They are closely connected to the acceleration of particles into interplanetary space and thus with variations in the solar wind observed as space weather. Solar flares tend to occur in magnetically active regions, and are believed to be caused by a rapid release of energy stored as inductive magnetic fields in the solar corona.

Despite decades of study at all wavelengths, the exact conditions under which solar flares occur, their energy release mechanisms, and their radiation processes are not well understood. Magnetic reconnection is believed to be the primary driver of the energy release, but many open questions about the exact processes occurring remain. The hard X-rays are believed to be caused by bremsstrahlung of accelerated electrons as they precipitate in the solar chromosphere, but both the details of the electron acceleration and precipitation are currently unknown. The detection of quasi-periodic pulsations (QPPs)¹ – roughly coherent oscillations in the light curve on timescales of seconds to minutes – has opened up the possibility of coronal seismology during solar flares. There is particularly strong interest in hard X-ray QPPs, because they are believed to carry information about both solar flare triggering and energy release mechanisms (see e.g. ?). Flares are bright features that can be observed down to fractions of a second, making it possible to resolve the wave transit time across magnetic plasma structures and thus study magnetohydrodynamic (MHD) waves in more detail than possible in any other way.

Understanding these processes in detail is important not only for advancing our understanding of solar MHD processes and plasma physics, but also of interest in the context of stellar flares now observed in abundance with instruments like *Kepler*. Understanding the creation mechanisms of QPPs in solar flares will provide us with key diagnostics for the physical conditions in stars (e.g. ?). The details of the MHD processes and reconnection physics in the sun are highly relevant to other phenomena observable with *Fermi*/GBM where magnetic reconnection likely plays a key role, for example bursts from magnetars, whose triggering mechanisms are currently unknown (?). Magnetar bursts have been likened to solar flares in their statistical properties and morphology before (?) and have also shown QPPs(?), thus it is not inconceivable that, while magnetars are vastly different from the sun, similar processes might occur in terms of the underlying reconnection.

The existence of QPPs in solar flares has been known for decades, but they have only recently emerged as key diagnostics of the plasma conditions and flare triggering processes (e.g. ?, and references therein). A key driving factor is the ability to study active solar regions with a high spatial and spectral resolution in multiple wavelength ranges.

Significant progress has also been made in explaining QPPs with physical models, which generally fall into two classes. In the first, magnetic reconnection proceeds in a “bursty”, somewhat regular fashion, modulating the number of accelerated electrons and thus the intensity of the observed radiation (???). The second class of models directly links the observed oscillations to MHD oscillations or waves in some of the flare parameters, e.g. the rate of energy release, particle acceleration, and particle dynamics (??). In this case, periodicities are prescribed by resonances that act as direct tracer of the magnetic processes in the flare emitting region.

It is likely that QPPs in different flares are generated by different mechanisms, and several mechanisms might occur in the same flare. It has been possible to infer the likely physical mechanisms operating for a very small sample (~ 5) of flares (???). In particular, the detection of multiple independent QPPs in a single flare is highly indicative of the underlying process because various physical models allow only for very specific resonances (??). Multi-wavelength and spatially resolved data has also been crucial for understanding the properties of QPPs and their likely origin.

Despite the successes in identifying the QPP mechanisms in few individual flares, no overall picture of QPP formation has emerged for the majority of flares. Observational data and current data analysis techniques do not permit an unambiguous choice between various models of QPPs in solar flares. To date, none of the proposed models have been confidently excluded, nor have any been identified as the dominant QPP formation mechanism. This is partly due to the lack of high-quality multi-wavelength or spatially resolved data for many flares. However, there is a crucial dearth of large-scale sample studies of QPPs in solar flares, and it has therefore not been possible to relate QPP parameters to other physical properties of the flare. Studies of solar flare variability are often framed as QPP searches, leaving the

¹ The same phenomenon is usually called quasi-periodic oscillation (QPO) in the astronomy community, but we will adhere to the solar physics nomenclature for consistency

information contained in variability patterns largely unexplored. (though some studies of the wavelength-dependence of variability exist, (?)).

The lack of large-scale studies of QPPs and variability is rooted in the technical challenge of QPP identification in the presence of complex structures in the hard X-ray light curve. Until recently, identification was a labour-intensive task, difficult to automate for large samples, and involved the manual de-trending of the light curve to remove the overall rising and falling trends of the *flare envelope*, followed by a visual identification of the residuals to identify QPPs. As recently pointed out by (?) this method is also error-prone, because it does not correctly account for inherent aperiodic variability (also called “red noise”) present in the flare. (?) instead suggested a Fourier-based method taking account of this aperiodic variability based on methodology developed for Active Galactic Nuclei (AGN), while (?) employed a similar approach using wavelets.

Disentangling periodic and quasi-periodic signals from aperiodic variability has been a major problem for this field. The reason lies in the transient nature of these flares and bursts, for which Fourier methods are inherently not well suited, because the assumption that mean and variance of the data do not change over time breaks down in these data sets, as was shown for magnetar bursts in . This especially hurts QPP studies in solar flares because these signals are often observed at low frequencies where the biases introduced by the method are strongest.

The current quandary is that two classes of methods exist, both of which make strong assumptions and neither of which is a good model of the data. On the one hand, detrending techniques assume the solar flare consists of two processes operating on independent time scales, a slowly varying flare envelope, and a faster, possibly periodic process.

On the other hand, Fourier- and wavelet-based models assume a purely stochastic process switching on and off at the start and end of the flare, with a potential QPP on top of this process, but disregard the idea of an underlying envelope entirely. The former model is likely to grossly overestimate the number of QPPs found in solar flares because it often does not account for aperiodic variability. The latter has been shown to underestimate the number of QPPs at lower frequencies (?). Both lead to systematic biases in the results that have not been adequately characterized.

The reliable detection of QPPs in solar flares is a difficult problem, but modern advances in computer science, particularly concerning a flexible class of models called Gaussian Processes (GPs; ?), point to a way forward. Instead of either directly modelling the lightcurve or some transformed version of it, GPs model the *covariance* between data points with a parametric function (also called *covariance function*).

This approach has several clear advantages over alternative methods. (i) It represents the best of both worlds: it is possible to self-consistently include both deterministic components, e.g. a flare envelope, and stochastic components, such as red noise, (?), in the same model. This eliminates common biases in the typical approach of removing the flare envelope by using a running boxcar. Both red noise and QPPs have direct formulations in terms of covariance functions, thus the model can be directly used as a basis to test for QPPs. (ii) The fully probabilistic formulation allows for straightforward model comparisons. It will be possible to test for the presence of various components (e.g. specific formulations of red noise or a flare envelope) directly, allowing for a much more precise model of the flare variability. (iii) Even if no QPP is detected, the parameters of the covariance function serve as a proxy for relevant time scales in the system, allowing for direct interpretation of these time scales e.g. in the context of models for bursty reconnection.

2. DATA

Data for this project was acquired from the Gamma-ray Burst Monitor (GBM) onboard the Fermi Gamma-ray Space Telescope, as well as the GOES-15 satellite’s x-ray imager. Solar flare data from GOES was provided by Andrew Inglis. The data is formatted in a .fits file, with the first two columns corresponding to the flux of the 1-8Å and 0.5-4Å channels respectively.

3. METHODS

3.1. Gaussian Processes

In the regime of 1D continuous data sets, stochastic processes (such as those observed in many solar flare models) (CITATION) are considered Gaussian if the finite set of inputs of our model (in our case, the time indexes of our measurements, t_i) can be correlated with a normally distributed random variable. Furthermore, linear combinations of these random variables, defined as $\tau_{ij} = t_i - t_j$, form a multivariate normal distribution. The distribution of these random variables over our input space can be used to model correlated stochastic processes, such as the quasi-periodic oscillations we aim to observe, using an autocorrelation, or kernel function $k(\tau)$. Combining the autocorrelation

function $k_\alpha(\tau)$ along with some arbitrary mean function $\mu_\theta(t)$, we have our Gaussian process model, where α and θ are the parameters of the kernel and mean functions respectively.

When provided with a training set of data y at input coordinates x , we can define our likelihood function:

$$\ln \mathcal{L}(\theta, \alpha) = \ln p(y|x, \theta, \alpha) = -\frac{1}{2} r_\theta^T K_\alpha^{-1} r_\theta - \ln \det(K_\alpha)^{\frac{1}{2}} - \frac{N}{2} \ln(2\pi) \quad (1)$$

3.2. Modeling Protocol

Once our data is cleanly organized into a time series containing time coordinates of our measurements, x , intensity measurements, y , and error δ_y , in separate arrays, we can begin our modeling algorithm. We model the flare envelope using the flare model:

$$\mu_\theta(t) = A \lambda \exp\left(\frac{-\tau_1}{t - t_s} - \frac{t - t_s}{\tau_2}\right) \quad (2)$$

where t is the time since trigger, t_s is the trigger time, A is the flare envelope amplitude, τ_1 and τ_2 are characteristics of the pulse rise and decay respectively, and $\lambda = \exp(2\frac{\tau_1}{\tau_2})^{1/2}$ (CITATION) The free parameters for our fit are A , τ_1 , and τ_2 , which we will refer to as θ , a parameter vector. We generate our initial guesses for this parameter vector using a simple function that takes the maximum of the flare as A , τ_1 to be $\frac{1}{3}$ of the length of our flare, and τ_2 to similarly be $\frac{2}{3}$ the length of our flare. These initial guesses need only to be approximate, as we'll be optimizing these values after defining all free parameters of our model.

We'll model our phenomena of interest, QPOs, via a kernel function $k(\tau)$. Our modeling protocol is all handled through Celerite a library for 1D Gaussian process regression problems. We model our QPO using a stochastically driven harmonic oscillator, classically defined as:

$$\left(\frac{d^2}{dt^2} + \frac{\omega_0}{Q} \frac{d}{dt} + \omega_0^2\right) y(t) = \epsilon(t) \quad (3)$$

Where ω_0 corresponds to the natural frequency of the undamped oscillator, Q is the quality factor, and $\epsilon(t)$ is a stochastic driving force. There is an additional scaling term S_0 in the PSD of this process that corresponds to the power at natural frequency, $\omega = \omega_0$. Working under the assumption that all quasi-periodic oscillations are triggered by the same mechanism, we set our initial values of these parameters as those we captured from a real flare.

We'll similarly define a rednoise term in the model, defined by kernel function:

$$k(\tau) = a * \exp(-c\tau) \quad (4)$$

Where a and c are similar terms for correlational strength and timescale, that are similarly free, and guessed based on known values from a real solar flare. One convenience of our modeling protocol is that we can combine multiple kernel functions into a composite kernel, k_α , which will automatically concatenate the individual parameter vectors for each kernel into one large parameter vector, which we'll denote as α .

Since we are using a Bayesian probability measurement as our fitting function, we can encode our prior knowledge of our parameters into our models, as prior distributions. For the sake of convenience and simplicity, we applied tophat priors for each of these parameters within reasonable known values of these parameters.²

From here, we have all necessary information to initialize our models: the mean envelope model μ_θ and the kernel function k_α are initialized with initial guesses for their parameters, (θ_0 and α_0), and the prior distribution for each parameter. From here, we can initialize a GP-class (Gaussian Process) object, utilizing the model as the mean function for the process, and the kernel function as a constructor for the kernel matrix.

With these computations ready, we feed the time coordinates (x) as well as intensity error (δ_y) into the GP.compute() function to compute the kernel matrix. With the kernel matrix computed, we can begin the optimization process. For a fit, we want to maximize the likelihood of our fit over parameterspace defined by θ and α .

Where the vector $r_\theta \equiv y - \mu_\theta(x)$. With this function defined, we can define a function that computes the negative log-likelihood ($-\mathcal{L}$) as a function of the joint parameter vector (θ, α). With this, we can use any optimization scheme to

² We may want to think about our priors a bit more before I write out a table for them...

find the optimal values for the parameter vectors; in our case, we called the minimize function from scipy's optimization submodule, utilizing Limited-memory BFGS optimization.

This is generally sufficient to compute the optimal parameters of the solar flare envelope, θ , but for the QPO parameters, α , we can better understand the parameter-space, and to maximize the probability of our fit by using Markov chain Monte Carlo sampling (MCMC). Using the emcee package, we initialize a set of walkers to explore the parameter-space of the model, distributed about the optimal parameters previously computed. Defining a log-probability method as the sum of our likelihood computation as well as our prior at the parameterspace θ acts as our probability distribution for our walkers.

4. DISCUSSION

Let's talk about what happened.