A Bayesian method for detecting stellar flares

M. Pitkin,[1][⋆] D. Williams,[1] L. Fletcher[1] and S. D. T. Grant[1,2]

[1]*SUPA, School of Physics and Astronomy, University of Glasgow, University Avenue, Glasgow G12 8QQ, UK*
[2]*Astrophysics Research Centre, School of Mathematics and Physics, Queen's University Belfast, Belfast BT7 1NN, UK*

ABSTRACT

We present a Bayesian-odds-ratio-based algorithm for detecting stellar flares in light-curve data. We assume flares are described by a model in which there is a rapid rise with a half-Gaussian profile, followed by an exponential decay. Our signal model also contains a polynomial background model required to fit underlying light-curve variations in the data, which could otherwise partially mimic a flare. We characterize the false alarm probability and efficiency of this method under the assumption that any unmodelled noise in the data is Gaussian, and compare it with a simpler thresholding method based on that used in Walkowicz et al. We find our method has a significant increase in detection efficiency for low signal-to-noise ratio (S/N) flares. For a conservative false alarm probability our method can detect 95 per cent of flares with S/N less than 20, as compared to S/N of 25 for the simpler method. We also test how well the assumption of Gaussian noise holds by applying the method to a selection of 'quiet' *Kepler* stars. As an example we have applied our method to a selection of stars in *Kepler* Quarter 1 data. The method finds 687 flaring stars with a total of 1873 flares after vetos have been applied. For these flares we have made preliminary characterizations of their durations and and S/N.

Key words: methods: data analysis – methods: statistical – stars: flare.

1 INTRODUCTION

Solar flares have been intensely studied for decades, using imaging, spectroscopy and light-curve data. The record of soft-X-ray light curves from solar flares, as measured with the Geostationary Orbiting Environmental Satellites (GOES), extends back in a systematic, well-calibrated record to the mid-1970s, and from this, and other (shorter) data sets, many statistical properties of solar flares have been deduced, and a fairly solid understanding of the physical processes in solar flares has emerged. There are other stars, in particular UV Ceti stars, for which large numbers of flares have been observed and statistics gathered, e.g. the observations of Gershberg (1972) and Moffett (1974) used in Lacy, Moffett & Evans (1976), or the studies of Ishida et al. (1991) and Dal & Evren (2012). Flares have also been searched for in M Dwarf stars using data from the *Sloan Digital Sky Survey* (*SDSS*), using both photometric (Kowalski et al. 2009) and spectral data (Hilton et al. 2010). However, the *Kepler* spacecraft, designed to search for exoplanets via the transit method, has provided a great new resource of precision long-duration high-temporal resolution data needed to study a large population of flares from a single star, or a large population of stars. Previous searches for flares in this data set include Walkowicz et al. (2011), Balona

(2012), Maehara et al. (2012) and Shibayama et al. (2013), which have tended to focus on easily recognized large flares.

This paper describes a method for identifying stellar flares of all sizes, based on their expected temporal characteristics, using a Bayesian odds ratio method. Our method assumes a specific flare profile, defined by a Gaussian rise and exponential decay, with a specific range of values, and is therefore most efficient for flares that can be characterized by this model. However, we also demonstrate examples of its performance for other flare shapes. The odds ratio provides a natural quantitative way to discriminate between noise and signal models, which provides additional information over the by-eye judgement relied upon in previous analyses.

1.1 White light flares

Stellar flares are intense, rapid and unpredictable brightenings in the magnetized atmosphere of a star, caused by the release of previously stored magnetic energy (Benz & Güdel 2010). In the case of solar flares, the usual assumption is that energy is converted to the kinetic energy of non-thermal particles which stop collisionally and cause heating in the solar atmosphere and enhanced radiation. Solar flare emission can be detected across the entire electromagnetic spectrum (Fletcher et al. 2011). In the case of the Sun the flare excess spectrum peaks in the optical to near-UV, but optical ('white light') enhancements are hard to identify in light curves simply because the Sun's quiescent spectrum peaks in the optical as well, and even the largest

[⋆] E-mail: matthew.pitkin@glasgow.ac.uk

solar flares produce an observed optical enhancement which is only slightly above the background fluctuation caused by solar p modes (Woods et al. 2004). When they are observed, however, solar white light flares provide the most robust means we have of determining the flare's total radiated energy. Again in the case of the Sun, the optical emission originates from the lower solar chromosphere and/or photosphere, and tends to have a very impulsive shape, often showing a rapid rise and fall consistent with fast heating and equally fast cooling, presumably by radiation in a dense atmosphere (Hudson, Wolfson & Metcalf 2006). On the other hand, solar flares observed at shorter wavelengths, e.g. UV to soft X-rays, have a characteristic fast-rise and exponential decay character, indicative of rapid heating followed by a combination of conductive and radiative cooling in a more tenuous coronal plasma. Solar flare white light emission was first observed in 1859 (Carrington 1859), but we still do not have a settled theoretical explanation for its production. This is partly because, with the difficulty of observing solar white light flares, we do not have the spectroscopic information required to discriminate between different emission mechanisms like enhanced blackbody continuum, or free–bound emission. On the other hand, we do have exquisite imaging observations to guide us in understanding the structure and development of solar white light flares, which are of course inaccessible in the stellar case. In the case of stellar flares we have no such direct spatial information, but spectroscopy of large flares, which dominate the star's radiative output, can be carried out readily (e.g. Kowalski et al. 2010), and evidence for both enhanced (blackbody) heating and free–bound emission can be found. This type of study is critical in understanding flare energetics; for example to produce a blackbody or a free–bound continuum requires heating and ionization in very dense parts of the stellar atmosphere. This in turn has implications for flare energy transport.

White light flares observed on stars can be at levels very much above the star's quiescent level in that wavelength, and mostly show the fast-rise exponential decay pattern (e.g. Moffett & Bopp 1976; Kowalski et al. 2013). Very large stellar flares are readily picked out by eye or by relatively simple thresholding methods, but we know from the case of the Sun that flares occur on all scales of physical size, energy content, and other parameters (Hannah et al. 2011). Identifying the smaller events is very important for understanding magnetic energy release in general, therefore we want to have a robust means to search also for smaller events, not just large ones. Statistical studies of 'superflares' on G-dwarfs in the *Kepler* sample have been carried out by Maehara et al. (2012) and Shibayama et al. (2013) using a method based on identifying signals above some threshold determined from de-trended signals data. They find a superflare occurrence rate $dN/dE \propto E^{-\alpha}$ with $\alpha \sim 2$ for all G-dwarfs, with dN/dE being the rate for flares of total energy E. As with all such efforts, the flare distributions show a turn-over at low energy due to the difficulty of detecting small events, and the location of this turnover is important in many approaches to fitting the spectral index. The value of the spectral index in turn determines whether large or small events contribute most to the overall energization of the stellar corona, with $\alpha \geq 2$ being required for dominance of small events – the so-called 'nanoflare' heating scenario (Hudson 1991). In the case of the Sun, this α value varies between 1.5 and 2.1, depending on the wavelength in which observations are made, the way that background is subtracted, and other parameters. Also in the case of the Sun the flare distributions are produced using radiation signatures which embody only a very small minority of the radiated flare energy – for example the readily observed soft X-ray which is correlated with the total flare energy (identified with the energy content of the fast particles) but with a rather large scatter (Emslie

et al. 2012). It would be much more satisfactory to carry out flare statistics using the energetically significant white light radiation, but this is not possible for the Sun. Therefore, we are motivated to develop methods which will allow this to be attempted for stellar flares, which means we must pay attention to the identification of small events. We must therefore be able to decide in a robust way whether an excursion in the light curve is flare-like or a noise fluctuation. We will do this on the basis of the shape of the light curve, searching for fluctuations consistent with the fast-rise and exponential decay profile observed in larger flares. In principle this method could be adapted to search for different shape profiles.

1.2 *Kepler* light curves

Kepler stellar light curves are known to contain low-frequency variability both from the intrinsic fluctuations of the instrument (see e.g. Jenkins et al. 2010b) and stellar variations (Basri et al. 2010). We are not interested in these variations, but if not dealt with carefully they can influence any flare detection algorithm. The data we have chosen to use is the PDCSAP_FLUX data. This has had the Pre-search Data Conditioning (PDC) module of the *Kepler* analysis pipeline applied, which attempts to remove signatures in the light curves correlated with the spacecraft and detector and also accounts for discontinuities due to pixel sensitivity drop-outs (the method attempts to not remove a true astrophysical signal, although as noted in Basri et al. 2011, some true stellar variability, which we are not interested in, may be removed). In the example analysis we will present we have used data from the *Kepler* data release 21 described in Thompson et al. (2013). For this release there have been several improvements over the original version of the PDC (Jenkins et al. 2010a), to remove unwanted instrumental artefacts. These improvements mean that the PDC data for the majority of stars are processed using the multi-scale maximum a posteriori (msMAP) approach described in Stumpe et al. (2014), which is an extension to the MAP approach described in Stumpe et al. (2012) and Smith et al. (2012). It should be noted that in Walkowicz et al. (2011) the 'raw' data, rather than the PDC data, are used with the detrending of Basri et al. (2011) applied. This means that there is very different data conditioning between that analysis and ours, making direct comparisons unreliable as is reflected in our results (see e.g. Section 5.1).

Various detrending methods have been developed to remove the low-frequency instrumental variations in the light curves (e.g. in addition to the methods in the PDC in Jenkins et al. 2010a; Stumpe et al. 2012, 2014; Smith et al. 2012, or the detrending in Basri et al. 2011, there is the *astrophysically robust correction* of McQuillan, Aigrain & Roberts 2012; Roberts et al. 2013) whilst trying to retain astrophysical variability. These variations often reflect the rotation period of the star and can be due to the presence of spots on the stars' surfaces. We have examined several methods of processing the data to remove, or diminish, the effects of the large sinusoidal variations seen in the light curves, which would otherwise hamper a flare detection algorithm. We wanted a method that we could automatically apply to all light curves without having to tune it for individual stars. We also wanted a method that would not add any major extra artefacts into the data or remove significant power from the flares. Methods such as subtracting a running median, image erosion, whitening the data with an 'average' spectrum, high-pass filtering the data, or removing a running polynomial fit (e.g. with the filtering algorithm of Savitzky & Golay 1964) can all remove the variations, but at the cost of adding artefacts around flares (and other transient signals) and diminishing the signal power. Instead of attempting to remove the variations we have taken the approach

of including them in our model of the data, i.e. fitting the variations *and* the flare model together. This is discussed in detail below.

We will be looking at *Kepler* long cadence data in which there is one photometric data point every 29.42 min. As stated in Walkowicz et al. (2011) this is not ideal for the detection of flares, which generally evolve more quickly than this. However, our method can very easily be applied to short cadence data.

2 DETECTION ALGORITHM

The general outline of our method is to use Bayesian model comparison to create a detection statistic. The detection statistic is formed by calculating the ratio of the probability that a light curve contains a flare-like signal described by a known parameterizable model *and* any background variations (the *signal* model) to the probability that the light curve contains just the background variations *or* other non-flare-like signals and the background variations (the *noise* model). We will refer to this ratio as the odds ratio. The statistic is then characterized by running the algorithm with mock data that contains no signals. The distribution of the statistic yields a threshold value above which we will consider a flare detected for a certain false alarm probability (FAP). Using this threshold and sets of mock data containing simulated signals we can also find the detection probability as a function of signal-to-noise ratio (S/N).

The method of flare detection in Walkowicz et al. (2011) involved smoothing the data with a median filter over a 10 h interval, finding points that crossed a threshold of 4.5 times the data standard deviation and then counting a flare as three contiguous threshold crossings. The values for these three tuneable parameters were found by comparing an automated algorithm to results from a by-eye search on a set of training data. Their method did not make use of the flare signal shape (i.e. information from the below threshold portion of the data was ignored) to try and gain S/N and therefore lacked sensitivity to smaller flares. In our method we attempt to use information on the signal shape by creating a parametrized flare model with which to compare to the data. We also attempt to automatically veto impulsive transients that may be due to instrumental effects by including them in our noise model. However, we note that these models can also veto temporally unresolved flares.

If flare light curves were un-modelled, or far less well modelled than we allow for here, there are other methods based on Bayesian model comparison available. Searle et al. (2008) and Searle, Sutton & Tinto (2009) describe a general method for finding un-modelled bursts based on searching for excess power (as applied to gravitational wave data analysis, but still more widely applicable). Whilst Hambaryan, Neuhäuser & Stelzer (1999) have a Bayesian method to search for flares in X-ray data (based on the Bayesian Blocks method of Scargle 1998) that looks for change points in the statistics of the data.

2.1 The flare model

The simple flare model we use is based on the observed shape of many flares, with a fast rise and exponential decay (e.g. Kowalski et al. 2011). The rise stage is modelled by a half-Gaussian, whilst the decay stage is an exponential fall

$$m(t, \tau_g, \tau_e, T_0) = A_0 \begin{cases} e^{-(t-T_0)^2/(2\tau_g^2)} & \text{if } t \leq T_0, \\ e^{-(t-T_0)/\tau_e} & \text{if } t > T_0, \end{cases} \quad (1)$$

where A_0 is the amplitude at the flare peak time of T_0, τ_g is the standard deviation of the Gaussian rise and τ_e is the exponential
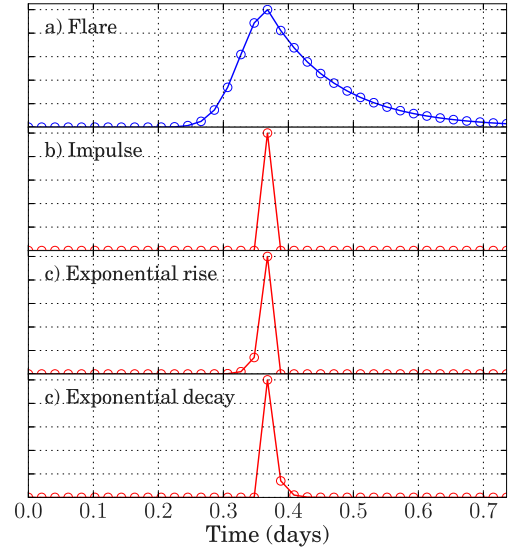

Figure 1. Examples of the signal and noise transient models used in the algorithm. Panel (a) gives the flare signal model and panels (b), (c) and (d) give the noise transient models.

decay time constant. The parametrization also allows estimation of these parameters from the detection algorithm. An example of the flare model is given in Fig. 1. We note that this model does not account for all potential flare shapes, e.g. the 'gradual' flares described in Kowalski et al. (2011), but still has some power to detect flares with different morphologies (see Section 3.5).

2.2 The detection statistic

To create a detection statistic we follow a similar method to that used to search for ring-down gravitational wave signals developed in Clark et al. (2007). In the simplest terms, we want to test the hypothesis that the data, d, contain a flare signal and some background noise (the form of which we discuss below), $\mathcal{F}$, compared to one in which the data just consist of the background noise, $\mathcal{N}$. As we will see below this can be extended to include extra models as required. We can make this comparison by calculating the so-called odds ratio

$$\mathcal{O} = \frac{p(\mathcal{F}|d)}{p(\mathcal{N}|d)} = \frac{p(d|\mathcal{F})}{p(d|\mathcal{N})} \frac{p(\mathcal{F})}{p(\mathcal{N})}, \quad (2)$$

where the first fraction on the right hand side is known as the Bayes factor, and the second part is known as the prior odds, i.e. the prior belief in each hypothesis. From Bayes theorem we can calculate the posterior probability distribution of a set of parameters $\boldsymbol{\theta}$ defining the model in a hypothesis H, given a set of data d, via

$$p(\boldsymbol{\theta}|d, H) = \frac{p(d|\boldsymbol{\theta}, H)p(\boldsymbol{\theta})}{p(d|H)}, \quad (3)$$

where $p(d|\boldsymbol{\theta}, H)$ is the likelihood of the data given H and $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$ is the prior probability on those parameters, and $p(d|H)$ is the probability of the data given the hypothesis (this is a normalization factor often known as the Bayesian evidence). To get to equation (2) the evidence value for each hypothesis must be calculated, which can be performed via marginalization (which is just another term for integration) of the product of the likelihood and prior over the model parameters

$$p(d|H) = \int^{\boldsymbol{\theta}} p(d|\boldsymbol{\theta}, H)p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}. \quad (4)$$

In our detection algorithm we assign the same prior probability to the two (or more) hypotheses and therefore set the prior odds to be one, so the odds ratio is entirely defined by the evidence values (i.e. it is given by the Bayes factor).

For our algorithm we will assume a Gaussian likelihood distribution for the noise in the light-curve data, which means that for a generic model m parametrized by $\boldsymbol{\theta}$ the likelihood of the data given the model parameters is

$$p(d|\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\sum_{j=1}^{n} \frac{[d(t_j) - m_j(\boldsymbol{\theta}, t_j)]^2}{2\sigma^2}\right), \qquad (5)$$

where σ is an estimate of the noise standard deviation (here assumed to be constant over the data) and n is the number of data points. If we assume that any light curve is purely described by Gaussian random noise and flare signals (i.e. for the moment we ignore any other low-frequency variability or correlated noise effects) then for $\mathcal{F}$ the model m is given by equation 1 (so $\boldsymbol{\theta} = \{A_0, \tau_g, \tau_e, T_0\}$) and for $\mathcal{N}$ we have $m = 0$. For our detection statistic we do not care about the actual value of the model parameters other than the flare time, so we can marginalize over a subset of the parameters $\boldsymbol{\theta}' = \{A_0, \tau_g, \tau_e\}$. However, as discussed in Section 4, we can still recover A_0 when performing parameter estimation. Inserting equation (5) into equation (2) as appropriate gives an odds ratio as a function of the flare time of

$$\mathcal{O}(T_0) = \int^{\boldsymbol{\theta}'} \exp\left(\frac{1}{2\sigma^2}\left[\left\{2\sum_{j=1}^{n} m_j(\boldsymbol{\theta}', T_0)d_j\right\}\right.\right.$$
$$\left.\left. - \left\{\sum_{j=1}^{n} m_j(\boldsymbol{\theta}')^2\right\}\right]\right) p(\boldsymbol{\theta}')d\boldsymbol{\theta}', \qquad (6)$$

where the subscript j refers to the data or model value at the time t_j (note that terms involving d^2 and the pre-factors have cancelled out in forming the ratio of likelihoods).

This assumption of a Gaussian likelihood is mainly due to its simplicity, but generally it represents the least informative probability distribution for noise (see e.g. chapter 2 of Bretthorst 1988). Noise in real *Kepler* light curves is not purely Gaussian (see Section 3.3), so some care does need to be taken in estimating noise levels and assessing results when this assumption is used.

Assuming a constant prior probability on the signal amplitude A_0 (which we will discuss more in Section 2.4) the integral in equation (6) is analytic over A_0 between $[0, \infty]$, giving

$$\mathcal{O}(T_0) = \int_{\tau_g^{\min}}^{\tau_g^{\max}} \int_{\tau_e^{\min}}^{\tau_e^{\max}} \exp\left(\frac{D^2}{2\sigma^2 M}\right) \sqrt{\left(\frac{\pi\sigma^2}{2M}\right)}$$
$$\times \left(1 + \mathrm{erf}\left[\frac{D}{\sqrt{2\sigma^2 M}}\right]\right) p(\tau_g, \tau_e)p(A_0)\mathrm{d}\tau_g\mathrm{d}\tau_e, \qquad (7)$$

where $M = \sum_{j=1}^{n} m_j(\boldsymbol{\theta})^2$ (where m is the model from equation 1, but with $A_0 = 1$, i.e. independent of A_0) and $D = \sum_{j=1}^{n} d_j m_j(\boldsymbol{\theta})$. To marginalize over τ_g and τ_e we just perform the integration numerically on a grid over the ranges $[\tau_g^{\min}, \tau_g^{\max}]$ and $[\tau_e^{\min}, \tau_e^{\max}]$ using the trapezium rule. The grid intervals we have used are discussed in Section 2.5. For flares we require that the decay is longer than the rise time-scale ($\tau_e > \tau_g$), but the prior probability distribution for both is otherwise flat, so the distribution we use is

$$p(\tau_g, \tau_e) = \frac{1}{(\tau_g^{\max} - \tau_g^{\min})(\tau_e^{\max} - \tau_e^{\min}) - \frac{1}{2}(\tau_g^{\max} - \tau_g^{\min})^2}. \qquad (8)$$

Note that this prior is correct for the times scale ranges we use in this paper, but the time-scale prior area could be a differently shaped polygon for different ranges.

For each T_0 value at which we calculate $\mathcal{O}(T_0)$ the summation in D requires n operations, so using the time of each light-curve data point as the T_0 values would require n^2 calculations. However, calculating D for each T_0 is just the cross-correlation of the model and the data, which via the convolution theorem can be calculated using Fourier transforms,[1] with of order $n\log_2 n$ operations. This can offer significant speed-up in calculating the odds ratio.

2.3 A variable background

As discussed in Section 1.2 real *Kepler* light curves contain low-frequency variations. We have chosen to incorporate these variations into our signal model by modelling them as a polynomial, giving

$$m(t) = m_f(t) + \sum_{i=0}^{N_p} A_i t^i, \qquad (9)$$

where m_f is our flare model from equation (1), N_p is the number of polynomial terms, and A_i are the polynomial coefficients. If we treat each of the polynomial coefficients and our flare amplitude as independent then we can analytically marginalize over them all (see Appendix A) leaving us still with only the flare time-scales (τ_g and τ_e) to numerically marginalize out to give us a signal odds ratio, $\mathcal{O}_s$. This method requires no detrending of the data, or applications of offsets, as the effects of these types of variation are modelled. With $m(t)$ now being our signal model we want to compare this (i.e. by forming an odds ratio) with the evidence that the data contain *only* a polynomial background, which can be calculated by setting $m_f = 0$ in equation (9). Technically what we are calculating in each case is an odds ratio for our particular model ($\mathcal{O}_s$ for a flare *and* background variability, or $\mathcal{O}_b$, for just background variability) versus Gaussian noise, but if we then form a ratio of these the Gaussian noise case cancels out and we get the odds ratio we require.

Due to the fact that the low-frequency variability can have periods down to less than a day, and individual *Kepler* quarters span many tens of days, it is not practical to try and use very high order polynomials to try and fit all the variability. Instead, the analysis can be performed on a sliding window across the data, thus allowing a relatively low order polynomial to fit out the variability. The flare model has T_0 centred in the window. The sliding window length needs to be chosen such that for the range of flare durations searched for it spans the whole flare whilst also providing enough background on either side of the flare, so that the polynomial does not try to fit out any of the flare power. We discuss the value we have chosen for this analysis in Section 2.5.

As this window slides on to and off of either end of the light curve there will not be as many data points with which to form the odds ratio. To have odds ratios calculated using consistent amounts of data (which is important when assessing a detection threshold) we cut off odds ratio values returned within half the window length of either end of the data (i.e. when there is not full overlap between the window and the data).

---

[1] Cross-correlation of two time series f and g satisfies $f \star g = (\widetilde{\tilde{f}^* \times \tilde{g}})$, where $\tilde{x}$ and $\hat{x}$ are the Fourier transform and inverse Fourier transform of x, respectively.

2.4 Other models

We know that in *Kepler* data some stars contain transit signals from exoplanets (e.g. Batalha et al. 2013) and eclipsing binaries (Matijevič et al. 2012). Due to the above method using a sliding window, and attempting to fit background variations, these transits can occasionally trigger the odds ratio to favour the flare model (this happens as the window starts to slide on to, and off, the transit). Transits or eclipses could be parametrized and included in our background model (the numerator in equation 10), but since transits in *Kepler* data are very well studied we propose just vetoing stars with known transits. In the future it could be that just short stretches of data known to contain a transit are vetoed, or transits/eclipses are added to the noise model.

2.4.1 Short transients

The light curves can also contain impulsing delta-function-like signals, i.e. peaks within a single (or few) time bin(s). These could be caused by short flares that are not temporally resolved into several bins due to the long cadence of the data, but they could also be instrumental in origin. Due to this ambiguity we choose to model any such impulse as part of our noise model. We have three models for such behaviour (see Fig. 1): (i) a transient in a single time bin with a positive or negative amplitude, (ii) a transient with a short exponential decay and a positive amplitude, and (iii) a transient with a short exponential rise and a positive amplitude. This is not an exhaustive list of all potential instrumental artefacts (see e.g. Christiansen et al. 2013, for information on various instrumental effects) and others may be considered in the future. For each of these the unknown amplitude can be analytically marginalized over (along with the variable background polynomial). For model (i) there are no other parameters except the peak time and for each calculation of the sliding window we marginalize the model over all time bins within the window. As well as vetoing real transients this can get rid of detection artefacts around loud flares (S/N of a few tens or above) caused by the effects of having a sliding window. For models (ii) and (iii), as for the flare we fix the time to be the centre of the sliding window, but have to marginalize over the short decay time (see Section 2.5). For these noise models we have odds ratios of $\mathcal{O}_t$, $\mathcal{O}_{e+}$ and $\mathcal{O}_{e-}$, respectively, which gives a final detection statistic odds ratio of

$$\mathcal{O} = \frac{\mathcal{O}_s}{\mathcal{O}_b + \mathcal{O}_t + \mathcal{O}_{e+} + \mathcal{O}_{e-}}. \tag{10}$$

This method biases us against short duration flares, but without some other (instrumental) information that would allow us to veto transient artefacts this will remain a problem for long cadence data.

Note also that this algorithm assumes that the data contain just one flare within the sliding window, whereas in reality it could contain several. This could give a slight bias to the results if there are close flares as it will reduce the noise model evidence.

Examples of the output of this method, which gives a time series (representing the flare peak time) of the natural logarithm of the odds ratio (equation 10), are shown in Fig. 2. Fig. 2(a) shows an example of 'mock data' where Gaussian noise with a mean of zero and standard deviation of unity (in $e^- \, \mathrm{s}^{-1}$) has been generated, a sinusoid with frequency 3.9 d^{-1} has been added along with a flare with parameters $A_0 = 10 \, e^- \, \mathrm{s}^{-1}$, $\tau_g = 1760$ s and $\tau_e = 3768$ s. The algorithm estimated the noise standard deviation (see Section 2.6) to be 0.98 $e^- \, \mathrm{s}^{-1}$. Fig. 2(b) shows the output for *Kepler* Q1 data for the star with *Kepler* ID (KID) 1873543. It can be seen that two flares are obviously found above the threshold set in Sections 3.1 and 3.2.

Conversely, two short duration events seen near the start and end of the light curve show dips in the log odds ratio. This means they are far more consistent with the noise models, in particular the short transient noise models. The noise estimate for this data is 8.0 $e^- \, \mathrm{s}^{-1}$.

Claiming detection of these signals requires some thresholding on these time series', which we discuss in Sections 3.1 and 3.2.

2.5 The parameter space

The ranges of the flare parameters τ_g and τ_e that we have chosen cover flares lasting up to approximately half a day. The Gaussian rise, τ_g, spans from [0, 1.5] h, whilst the exponential decay, τ_e, spans from [0.5, 3] h. These ranges go into calculating the prior given in equation (8). Both these ranges are gridded into 10 evenly spaced points for evaluating and marginalizing the odds ratio. If we define the mis-match as being the fractional power that would be lost by two different flare models (at $[\tau_g, \tau_e]$ and $[\tau_g + \Delta\tau_g, \tau_e + \Delta\tau_e]$, respectively) not completely overlapping as

$$M = \left| 1 - \frac{\sum_{j=1}^n m_j(\tau_g, \tau_e) m_j(\tau_g + \Delta\tau_g, \tau_e + \Delta\tau_e)}{\sum_{j=1}^n m_j(\tau_g, \tau_e)^2} \right| \tag{11}$$

then we can see how well this gridding covers our parameter space. The mis-match can be seen in Fig. 3, where the black box represents our parameter range, and the diagonal line gives $\tau_g = \tau_e$. Given that we require $\tau_g < \tau_e$ only values above the diagonal line are included when calculating our detection threshold. Within our range we find that such a grid spacing gives a maximum mis-match of less than 10 per cent, with a mean mis-match of only 1.5 per cent (i.e. a real flare, that has duration parameters somewhere between our grid points, will be detected with on average 1.5 per cent less power than it really has). This level of mis-match assumes that the flares are fully described by our model, but in reality they are likely to show deviations from this that means that no flare will be perfectly matched to our model. Therefore this mis-match is a best case scenario and more power will be lost for a real search. The grid spacing could be decreased, but with a corresponding linear increase in computational time of the algorithm.

It can also be seen that even for flare values well outside our range we could still detect them without losing too much power.

2.5.1 Amplitude priors

For the flare amplitudes that we are analytically marginalizing over we have assigned a conservative prior range of between 0 and 1 000 000, giving $\log p(A) = \log (10^{-6}) = -13.8$. As we will assess a detection threshold empirically (see Section 3.1) this prior just provides an overall offset in the odds ratio, so is not of great importance, although we include it for completeness. The prior ranges for the polynomial amplitudes will cancel between the denominator and numerator of the odds ratio, so are not required.

2.5.2 Window length

As discussed in Section 2.3 the method uses a sliding window. We have chosen a window 27 h long (i.e. 55 *Kepler* light-curve time bins). This length means that our longest flares will still have almost all their power within the window, whilst being short enough that a fourth order polynomial provides a reasonable background fit for variations down to periods of around two days (see Section 3.1).
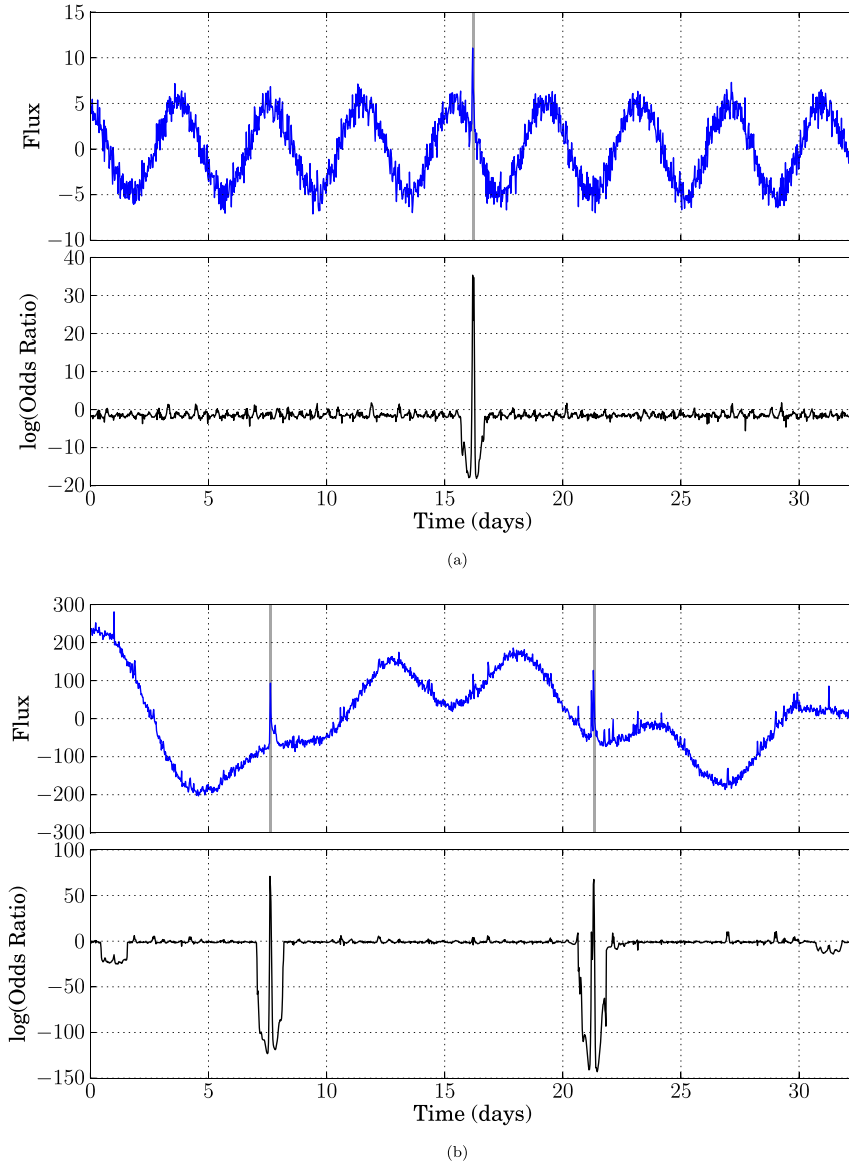
Figure 2. Examples of the output of the algorithm using (a) a simulated light curve containing an injected flare signal, and (b) *Kepler* Q1 data for the star with KID 1873543. Grey vertical bars on the light-curve plots represent times when the log odds ratio is above the threshold value of 16.5 (see Sections 3.1 and 3.2).

2.6 Calculating the noise

The odds ratio calculation (e.g. equation 7) requires an estimate of the noise standard deviation σ of the data. Just using a standard calculation of σ in data containing flares and other variations will lead to overestimates of the noise. We instead use a method to calculate the standard deviation that attempts to veto the effects of flares and the low-frequency background variations.

The method assumes that the underlying data distribution is Gaussian, with outliers (e.g. flares) in the wings of the distribution. However, as we see in Section 3.3, there are further correlations in the noise that make this an approximation when using *Kepler* data. The low-frequency variations in the light curves must initially be filtered out. So, to account for this we first apply a low-pass filter using the Savitzky–Golay method mentioned in Section 1.2 with a window and polynomial order the same as used for our detection algorithm. After this the cumulative probability distribution of the data is calculated. The standard deviation can then be calculated by finding the

values that bound a certain fraction of the probability distribution around the 50 per cent value. For example, one standard deviation (1σ) should enclose $100 \times \mathrm{erf}(1/\sqrt{2}) = 68.3$ per cent of the probability distribution, so one would find the values $x_{\min}$ and $x_{\max}$ that bound the $50 - (68.3/2) = 15.85$ per cent and $50 + (68.3/2) = 85.15$ per cent probabilities, and calculate the standard deviation as $\sigma = (x_{\max} - x_{\min})/2$. For smaller data sets (and provided the outliers do not make up a reasonably large fraction of the data) a more accurate value of the standard deviation can be found by using the value enclosing more of the probability [e.g. the amount enclosing the probability for 2σ, which is 95.4 per cent, and therefore $\sigma = (x_{\max} - x_{\min})/(2 \times 2)$].

This method is quantitatively similar to repeatedly removing outliers (found with e.g. the generalized extreme Studentized deviate test of Rosner 1983) and recalculating the standard deviation, but does not require repeated iterations. Also, unlike simply removing a certain fraction of the data in descending order from the largest
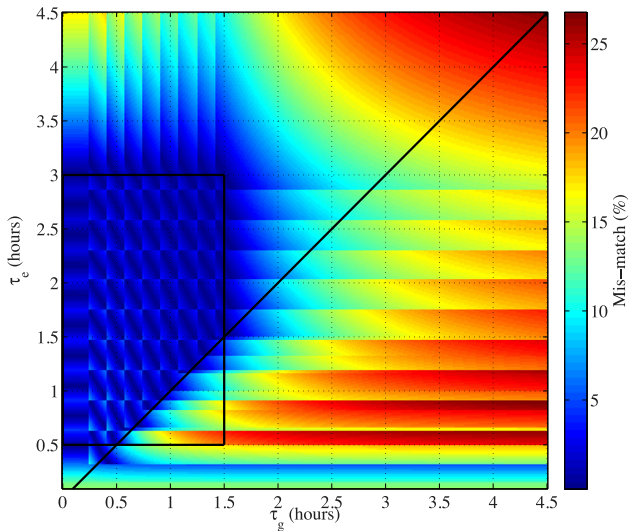
Figure 3. The mis-match across the range of flare parameters given the grid-spacing defined in Section 2.5.

Figure 4. The cumulative distributions of the log odds ratio for simulated light curves containing Gaussian white noise and a sinusoidal variation, and real *Kepler* data for a selection of 'quiet' stars. Thresholds for two false alarm probabilities are shown.

absolute value first, this method is less prone towards underestimating the standard deviation when used on data with few, or no, outliers.

This method can potentially lead to an underestimation of the noise if the distribution is not Gaussian and has broad wings. In *Kepler* data this could be a problem in cases where there is an excess of high-frequency correlations. In practice any such light curves could be vetoed during visual inspection of flare candidate stars. An alternative method, using the power spectrum of the light curve, and assuming the noise is white and Gaussian at high frequencies, is discussed in Appendix B. However, provided flare signals rise and decay very quickly, this method is less prone to overestimating the noise than the method described in Appendix B due to power leakage to high frequencies from loud flares.

3 ALGORITHM CHARACTERIZATION

Here we will detail the characterization of the algorithm. We will first set a threshold value for the natural logarithm of the odds ratio (when calculating the odds ratio it is always easier to work in log-space, and since the logarithm is a monotonic function it makes no difference in practice using this value.) for which we will return a flare detection candidate for a given FAP. Secondly, we will use this threshold to determine the efficiency of the algorithm for a set of simulated flare signals. We will also look at how the algorithm performs on real *Kepler* data to test the validity of our Gaussian noise assumptions. Other tests are performed to see the effects of different flare morphologies on our detection ability.

3.1 Threshold calculation

The odds ratio provides a value that, given a set of data, describes the relative probabilities of competing hypotheses or models. So, provided the prior probabilities for the model parameters and hypotheses are well defined, an odds ratio of greater than or less than one favours the hypothesis on the numerator and denominator, respectively. The number of model parameters and their prior ranges give rise to an Occam factor in the odds ratio calculation (i.e. the model with the smaller parameter space will a priori be favoured due to its simplicity). However, the presence of noise means that
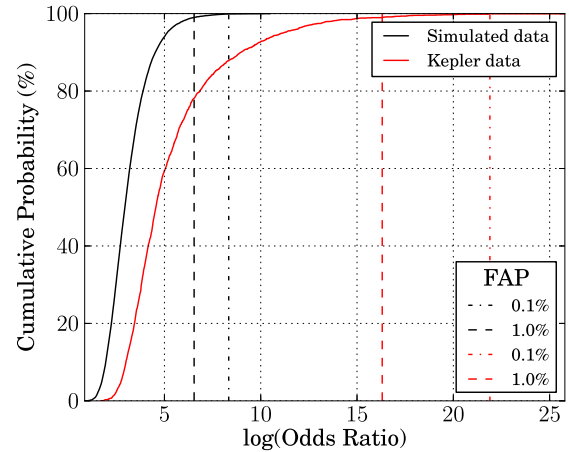
the odds ratio will fluctuate and at low S/N neither hypothesis will be greatly favoured. Jeffreys (1998) (in his appendix) gives a qualitative assessment of how to interpret values of the odds ratio, but the significance can also be assessed empirically.

For a detection algorithm we want to find the distribution of the value of the odds ratio when looking at data that contain no flares. We can then use this to set a threshold at which we expect noise alone to exceed this value with a certain probability, known as the FAP. To try and simulate real data, and incorporate the effect of low-frequency variations, our simulated data sets all contain sinusoids with amplitudes drawn randomly from a uniform distribution between 10 and 100 times the underlying Gaussian noise standard deviation, with frequencies drawn from a uniform distribution between 0.5 d^{-1} and 0.03 d^{-1}, and an initial phase between 0 and 2π rad. We purposely chose this frequency range to be close to the upper end of expected variations (it is in the tail of the period distribution for solar and late-type stars, e.g. fig. 5 of McQuillan, Mazeh & Aigrain 2014), as from our studies we know that these higher frequency variations are more likely to produce outliers in the odds ratio distribution (due to the polynomial background model not fitting the faster changing variations quite so well). This will therefore give us a reasonably conservative threshold in general. Our simulated data set are based on *Kepler* quarter 1 data and are therefore 33.5 d long, with one point every 29.42 min. We have run 20 000 such simulations to derive a threshold. The cumulative probability distribution of the maximum log odds ratio from each simulation can be seen in Fig. 4.

If we choose the FAP to be such that there is a 0.1 per cent chance of noise (Gaussian white noise plus a background sinusoid), giving one false positive per data set (i.e. one false alarm per thousand light curves, or about one per 80.5 yr of observations), it gives a log odds ratio threshold of 8.3. Equivalent thresholds for false alarm probabilities of 0.2 per cent, 0.5 per cent and 1 per cent are 7.9, 7.3 and 6.5, respectively. When calculating the efficiency below we calculate the observed false alarm rate and compare it to these values.

3.2 Detection efficiency

We calculate the efficiency of the algorithm by creating a set of fake signals with varying parameters and S/N in simulated noise
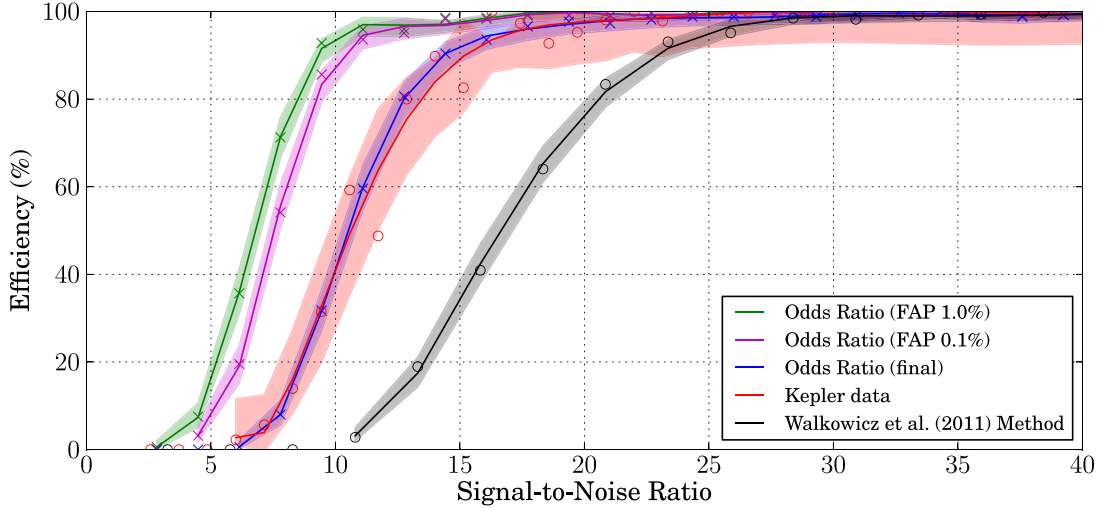
Figure 5. The efficiency of our odds ratio algorithm for various thresholds and the efficiency of our version of the algorithm from Walkowicz et al. (2011). The efficiency of our algorithm is assessed using simulations in which the noise is Gaussian for thresholds at 1 per cent and 0.1 per cent false alarm probabilities, and at a final threshold of 16.5. The efficiency based on adding signals to real *Kepler* data for 'quiet' stars for a threshold of 16.5 is also shown. The shaded error regions are 95 per cent confidence intervals around a best-fitting polynomial, calculated using the Beta distribution described in Cameron (2011).

Table 1. Efficiency of the detection method for different false alarm probabilities.

		S/N for detection efficiency		
		50 per cent	95 per cent	99 per cent
FAP	1.0 per cent	6.6	10.6	25.8
	0.5 per cent	7.0	11.0	26.0
	0.1 per cent	7.4	12.9	26.2
	Final	10.6	19.8	28.2
		σ-threshold method		
4.5σ threshold		16.7	25.1	34.0

(containing sinusoidal variations as above) and determining the fraction that are detected above a given threshold. We define a detected signal as a threshold crossing point (or set of points) that is within two time bins of the known injected central time. Contiguous above-threshold values are counted as a single detection and any two above-threshold segments separated by only one time bin are merged into a single detection. We define the S/N as $\rho = \sqrt{\sum_{j=1}^{n} m_j^2/\sigma^2}$, where m_j is the injected signal calculated from equation (1) and σ is the noise standard deviation calculated as described in Section 2.6. We have performed 10 000 such injections with S/N spanning between 2 and 50, and τ_g and τ_e uniformly drawn from ranges [0, 1.5] h and [0.5, 3] h, respectively, under the condition that $\tau_e \geq \tau_g$ (which is the same range used by the algorithm described in Section 2.5). The detection efficiencies for various FAP thresholds are shown in Fig. 5. For the threshold corresponding to a FAP of 0.1 per cent we find detection efficiencies of 50, 95 and 99 per cent corresponding to S/N of 7.4, 12.9 and 26.2, respectively. The efficiencies for these, and other, false alarm probabilities are summarized in Table 1.

When estimating the efficiency we also count any false positives, which are times that the algorithm exceeded threshold, but for a time that is further than two time bins from the known injected central time. Given a 0.1 per cent false alarm threshold of 8.3 and 10 000 simulations we expect noise to produce 10 false alarms. However, we find a total of 352 false alarms. Investigating these we find that

343 of the false alarms are within half of the length of the algorithm's running window of the injected flare, i.e. they are artefacts related to the injections, which if removed gives a number very close to the expected false alarm rate. Similarly, when calculating efficiencies for thresholds for false alarm rates of 0.2 per cent, 0.5 per cent and 1 per cent, we actually get numbers of false alarms of 397, 490, and 669 which, corrected for artefacts, revert to 12, 28 and 85 out of 10 000, respectively (these suggest that our false alarm probabilities might be slight overestimates).

Given that the algorithm will produce these false positives for loud enough signals it is useful to determine an odds ratio threshold at which they become significant. Using the distribution of log odds ratios for the false positives related to flares we can set a new FAP threshold. As these false alarms are due to flares themselves they will not harm our chances of detection, but will just slightly bias our number of detected signals. Due to this we are less conservative in setting the FAP and allow a 1 per cent value for these artefacts. We find this gives a new threshold of 16.5, for which the efficiency can also be seen in Fig. 5. This gives efficiencies of 50 per cent, 95 per cent and 99 per cent for S/N of 10.6, 19.8 and 28.2, respectively. This is rather conservative (when just assuming the noise is Gaussian) as it is a FAP that really only applies to flaring stars rather than all stars, such that for every 100 flares found one is likely to be a false positive. However, it will also help provide a stronger veto against disturbances in real *Kepler* data as shown below. This could be relaxed in the future, especially if further algorithm development provides stronger artefacts vetoes.

3.3 Characterization using *Kepler* data

To test the robustness of our assumption that the noise in *Kepler* light curves is Gaussian we have run the algorithm on a selection of 2000 'quiet' *Kepler* stars with little variability. To chose this set of 'quiet' stars we have randomly run through all the Q1 *Kepler* long cadence light curves and picked the first 2000 fulfilling the following criteria (after the removal of the best-fitting quadratic from each light curve): (a) the standard deviation of the data before and after removing a running Savitzky–Golay filter (with a fourth 4 order polynomial

and 55 samples in the running window) are within 25 per cent of each other, (b) the maximum and minimum values within the light curve are within 7.5 standard deviations of each other (using the standard deviation without the Savitzky–Golay filtering), and (c) the largest peak in the light curves amplitude spectrum (the square root of the power spectrum) is no more that 7.5 times the median value. These criteria seem to be successful in selecting stars with very little long-duration variability.

For each of these quiet stars we have run our detection algorithm and calculated the maximum log odds ratio. The cumulative distribution of log odds ratios can be seen alongside that for purely Gaussian noise (with simulated sinusoidal variations) in Fig. 4. It can be seen that real *Kepler* data have a broader distribution and are shifted to the right, i.e. *Kepler* data containing no flares more often favour the flare model than data containing Gaussian noise. We see that odds ratio thresholds for 1 per cent and 0.1 per cent false alarm rates per light curve are 16.3 and 21.9, respectively. The 1 per cent false alarm rate threshold is very close to the conservative 16.5 threshold we set in Section 3.2 based on signals themselves producing a 1 per cent false alarm rate.

To check that these 'quiet' stars do not actually contain flares we have visually inspected the light curves for the 17 stars that fall above our previously discussed odds ratio threshold of 16.5. We find that none of these shows signals that conclusively look like flare, but we do see that the noise can have short-term correlations that can mimic flares.

We have also performed an efficiency test equivalent to the one performed using Gaussian noise, but with 2000 simulated injections added to the 'quiet' star light curves, and using an odds ratio threshold of 16.5. This efficiency can be seen in Fig. 5. We find that the efficiency when using real *Kepler* data is consistent with that when using Gaussian noise, i.e. once you have set a threshold you can recover flare equally well from simulated or real data, albeit with a potentially different false alarm rate. From these simulations we found 33 false alarms (or a 1.6 per cent false alarm rate). This is about the number to be expected given that from Fig. 4 we see that 16.5 is roughly the 1 per cent false alarm threshold for *Kepler* data, and is also the 1 per cent false alarm threshold for artefacts associated with the simulated signals.

Based on this for the results presented in Section 5 we will use the assumption of Gaussian noise and a threshold of 16.5, which as shown should lead to a detection false alarm rate of 1 per cent.

In the future there are various options for the analysis to try and account for the non-Gaussianity of *Kepler*, or other, data. The simplest is to keep the assumption of the data consisting of Gaussian noise, but explicitly using a detection threshold that is based on real analysis of 'quiet' *Kepler* stars. Another option is to again keep the assumption of Gaussian noise, but make the method of estimating the noise in the data (see Section 2.6) give a more conservative value for its standard deviation. Finally, a method could be developed to model the correlations in the noise and use this information in the analysis. If the quiet stars truly are representative of the noise in the majority of stars then they could provide a model of the noise in the frequency domain that could be used to 'whiten' all light curves in the analysis. This model could simply be just an average of all the quiet star spectra, or based on a set of principal components (e.g. Shlens 2014) calculated from the quiet star noise that capture the majority of the correlations. We expect the simplicity of the first option (provided quiet, and non-flaring, stars are easy to identify using the criteria we set out above) or second option makes them the most viable options in the short term, although future investigations of the third option may prove illuminating.

3.4 Method comparison

It is useful to compare our method with that used in Walkowicz et al. (2011), which defines a detection as three contiguous positive threshold crossings, where playground data has been used to set a threshold of 4.5 times the data's standard deviation. We will call this the σ-threshold method. In Walkowicz et al. (2011) no empirical FAP is given, but if one assumes that the noise is purely white and Gaussian (which we have shown above is not really the case), then the expected FAP for a single data set of length N would be given by

$$f = 2\frac{N}{\sqrt{\pi}} \sum_{i=3}^{N} \prod_{j=1}^{i} \int_{\sigma_{\mathrm{T}}}^{\infty} e^{-x^2/2} \mathrm{d}x,$$
$$= N \sum_{i=3}^{N} \left(\mathrm{erfc}\left(\sigma_{\mathrm{T}}/\sqrt{2} \right) \right)^i, \tag{12}$$

where σ_{T} is the threshold in number of standard deviations. For $\sigma_{\mathrm{T}} = 4.5$ and $N = 1638$ (the number of points in a *Kepler* Quarter 1 light curve) this gives a FAP of 3.1×10^{-14} per cent, i.e. for any reasonable data set Gaussian noise should never give a false detection. If the threshold is dropped to $\sigma_{\mathrm{T}} = 3$ then the FAP becomes 2×10^{-6} per cent, which is still very low and far below the level we have set with our algorithm. However, dealing with real data may well produce a far higher false alarm rate than that given by these theoretical calculations. This is suggested by the fact that in Walkowicz et al. (2011) the 23 000 stars searched give 5784 stars with candidate flares, of which visual checking confirmed 373 with obvious flares and 565 marginal cases. This is a FAP of 20 per cent. The main thing this again confirms is that *Kepler* noise is far from ideally Gaussian as it contains many noise artefacts, and therefore in reality the noise estimate used was likely to be an underestimate. Our algorithm attempts to account for some of these non-Gaussian noise artefacts by including them in the noise model, and by using the more robust underlying noise estimation method described in Section 2.6.

We have compared the detection efficiency of our method to a version of that used in Walkowicz et al. (2011), by again simulating 10 000 data sets containing flares with S/N between 2 and 100 and time-scale parameters defined in Section 3.2. For this comparison we have not included any sinusoidal variations in the simulated data sets as we find that for short period variations the running median smoothing leaves many artefacts in the data (at the peaks of the variations) that are picked up as false alarms. This may be one of the reasons for the large false alarm rate seen in the real Walkowicz et al. (2011) analysis. As the data are simulated we have not had to apply any *Kepler*-like preprocessing, so do not completely represent the pipeline on Walkowicz et al. (2011). We also estimate the σ in the same way as described in Section 2.6 as we do not know how this was estimated for the original routine (although it is mentioned that extreme outliers are excluded). As in Walkowicz et al. (2011), before applying the detection algorithm we subtract a running median value from the data calculated for a 10 h window. Fig. 5 shows detection efficiencies of 50 per cent, 95 per cent and 99 per cent for S/N of 16.7, 25.1 and 34.0, respectively. It can be seen that our algorithm substantially improves efficiency over this method, albeit at far higher *theoretical* false alarm rates.

We note that many of the light curve figures shown in Walkowicz et al. (2011) seem to have significantly reduced flare heights than seen in the publicly available data (either the raw simple aperture photometry data, or the PDCSAP_FLUX data). An example is their fig. 2(a) (KID 10320656) for which the flare at 5 d is reduced by

an order of magnitude and an obvious flare at 25 d is completely missing. This may well be an aspect of their data pre-processing and the difference between it and the current msMAP processing (Stumpe et al. 2014).

Other methods have been used to find flares in *Kepler* and other data sets. Osten et al. (2012) use a statistic based on the ratio of the relative flux to noise between adjacent bins to search *Hubble Space Telescope* data for flares. Shibayama et al. (2013) detect flares in *Kepler* data by calculating the distribution of brightness variations between consecutive points and selecting only those points with values three times the value of the top 1 per cent of the distribution. Kowalski et al. (2009) searched for flares in *SDSS* data making use of flux changes in two photometric bands. None of these methods makes use of the flare shape as ours does, and are therefore both likely to be less efficient. The second method was specifically designed to look for very large flares, so it was not designed to be efficient at finding small events.

3.4.1 Computational time

Another useful comparison is the computational time of the algorithms. We have coded up both methods in PYTHON,[2] with the core of our method (Appendix A) written in C. Running both algorithms for a single *Kepler* Quarter 1 light curve on an Intel Core Duo 3 GHz machine shows an average time of 6.5 s for our algorithm compared to the far quicker time of 165 μs for the simpler algorithm. Despite the far greater computational cost of our algorithm it is still fast enough that a large number of light curves can be analysed in a reasonable time ($>10\,000$ per day on a single machine). Our algorithm can also make use of parallelization on multi-core machines to speed up the odds ratio calculation.

3.5 Different flare morphologies

Our flare model assumes that flares are well characterized by a fast Gaussian rise and exponential decay shape, with time-scales of $\gtrsim 30$ min. However, this does not describe all flare morphologies or situations in which there are very short flares, or overlapping flares. We have tested our algorithm on a few examples to see whether flares are still detected. We note that this is not an exhaustive test of all potentially flare type, but does point to areas in which our model fails or could be improved.

In the future a more complex model, or set of models, for flares could be developed. If multiple flare models are used then the Bayesian evidence values that are calculated in our method provide a natural way in which to discriminate between these different models. If the algorithm currently identifies flares that do not conform to the expected shape as flares this also has an effect on the parameter estimation (see Section 4), as the recovered parameters will be biased.

3.5.1 Short flares

The noise models we have used (see Section 2.4.1), which count short impulsive events as noise, and the time-scale parameter ranges we use, mean that we are likely to miss flares that evolve on short time-scales. To check the shortest flares that we are likely to miss we have simulated sets of flares with a range of exponential decay

time-scales from 5 min to 1 h at the sample rate of *Kepler* short cadence data (i.e. 30 times more samples than long cadence data). Each of these simulations has been down-sampled to the long cadence rate by averaging consecutive stretches of 30 data points, and then run through our algorithm. Using a log odds ratio detection threshold of 16.5 we find an efficiency of 12 per cent for flares with 5 min decay time-scales, which rises to 100 per cent for flares with 25 min decay time-scales. This is not unexpected, but it is important to acknowledge when presenting results that short flares will be missed.

3.5.2 Superimposed flares

Our algorithm assumes that each analysis segment contains one flare. However, flares may be close together and superimposed over each other. Therefore, it is interesting to see whether superimposed flares would be detected and at what point can we tell the flares apart. We have generated two large flares with equal time-scales and amplitudes, but with a set of peak times that shifts one of the flares with respect to the other between 1 h and 30 h apart. We find that the algorithm detects the two flares as a single flare when they are less than 4 h apart, but after that can distinguish the two flares as separate. An example of this is shown in Fig. 6(a), where the two flares are separated by 4.5 h. If instead we make the second flare five times smaller in amplitude than the first flare we find that only one flare is found until they are separated by 29 h. This is due to the noise models that are meant to deal with artefacts generated by the larger flare are swamping the evidence for the smaller flare when it is nearby. This shows that flares of similar amplitude can be discriminated relatively easily, whilst it gets harder as the amplitude ratio between them increases.

3.5.3 Gaussian signals

One of the most generic signal shapes for a transient outburst would be a Gaussian profile. We have simulated Gaussians of different widths (τ_g) ranging from 30 min to 7 h. For $\tau_g \lesssim 4.5$ h the algorithm detects Gaussians as flares, but does not detect longer Gaussians. The longer Gaussian will not be seen as they start to be fitted out by our polynomial background model. However, for Gaussians with τ_g between approximately 2 and 3 h the algorithm identifies the Gaussian as two or three separate flares as can be seen in Fig. 6(b). This is due to the signal creating extra artefacts in the detection algorithm that are not being vetoed by our noise models.

3.5.4 Flat-topped flares

A final flare shape that we consider is based on the Gaussian rise and exponential decay model, but with a flat peak that varies in width. We have generated such flares with fixed rise and decay times, but with the flat peak lasting between 15 min and 12 h. All these signals are detected, but often identified as two or three separate flares, as in Fig. 6(c). It can be seen that the rise and fall appear to be identified as separate flares.

4 PARAMETER ESTIMATION

Our analysis method also naturally lends itself to being used for estimation of the flare parameters. However, we note that this will generally only provide reliable parameter estimates if the flare can be well characterized by our model. In calculating the odds ratio we marginalized over the flare amplitude, and rise and decay

---

[2] The code is freely available as release v1.0.0 of the BAYESFLARE package http://github.com/BayesFlare/bayesflare/releases
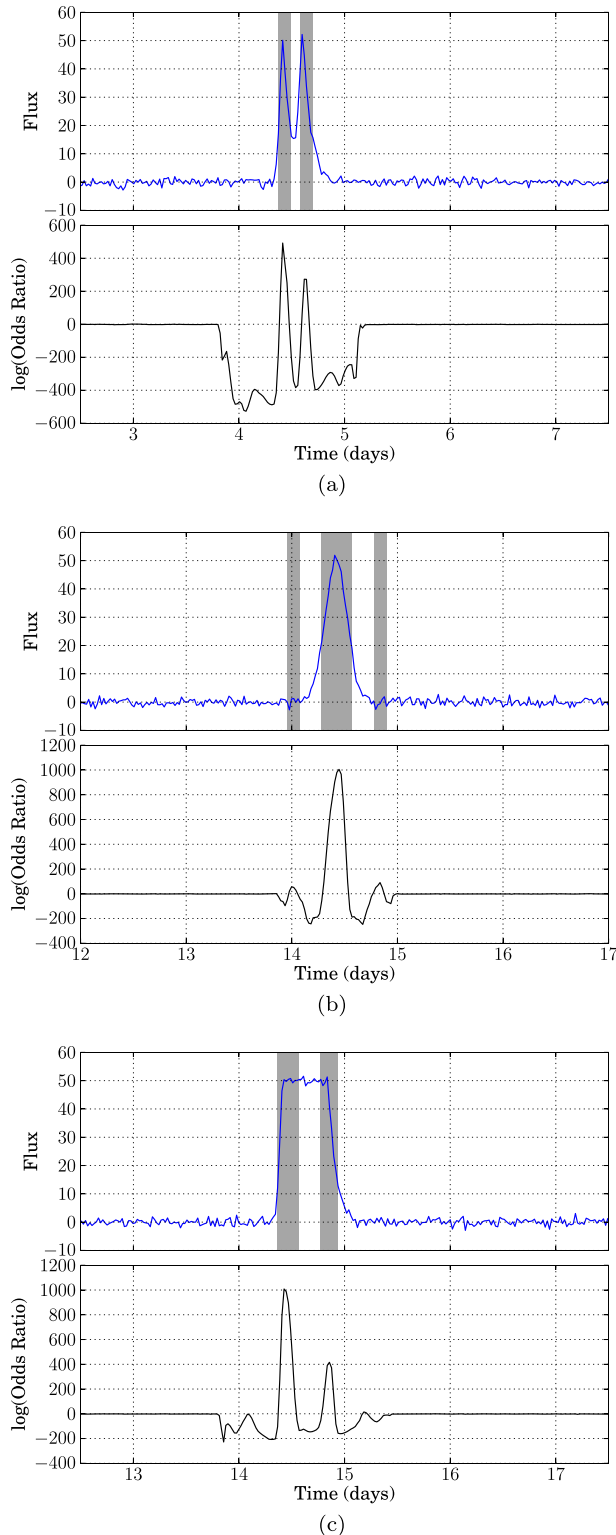
Figure 6. Examples of the algorithm run with different flare morphologies: (a) a superposition of two nearby flares, (b) a Gaussian flare profile, and (c) a flare with a flat top.

time-scales. However, for parameter estimation we want to get the probability distributions of these parameters. For the amplitude, which was analytically marginalized over, this requires a slight change in the algorithm as described in Appendix A2, as we still

want to marginalize over the unknown background variation. With this change we can then just evaluate the signal posterior probability, $p(A_0, \tau_g, \tau_e, T_0|d)$ (removing the integral from equation (6) gives an equation which is directly proportional to the posterior probability), over a grid in A_0, τ_g, τ_e and T_0 around the peak time corresponding to any recovered flare candidate. For parameter estimation we relax the prior stating that $\tau_e > \tau_g$ and just set a flat prior over the whole $\tau_e - \tau_g$ area (the black rectangle in Fig. 3). Once the posterior probability for the whole parameter volume has been calculated the posterior probability distributions for individual parameters are obtained by marginalizing over the other parameters, e.g.

$$p(A_0|d) = \iiint p(A_0, \tau_g, \tau_e, T_0|d)\mathrm{d}\tau_g\mathrm{d}\tau_g\mathrm{d}T_0. \tag{13}$$

These integrals are performed numerically with the trapezium rule. From these distributions the most probable parameter values and credible intervals can be found, or from the posterior volume the joint maximum a posteriori probability (i.e. the global maximum) for all parameters can be found.

When analysing our flare candidates we have used a grid spanning $0 \leq \tau_g \leq 2\,\mathrm{h}$, $0 \leq \tau_e \leq 5\,\mathrm{h}$, an amplitude from 0 to twice the maximum dynamic range in the data (for the 55 bins surrounding the flare), and a flare peak time (T_0) window of an hour either side of the recovered peak log odds ratio time.

An example of this for a simulated flare can be seen in Fig. 7, where the true flare parameters were $A_0 = 80\ e^-\mathrm{s}^{-1}$, $\tau_g = 0.49\,\mathrm{h}$ and $\tau_e = 1.05\,\mathrm{h}$. This simulation was added to real *Kepler* data for the star with KID 893676. The noise for this light curve was estimated to be $6.0\ e^-\mathrm{s}^{-1}$, giving a simulated S/N of 19.1. The maximum a posteriori parameters that were recovered were $A_0 = 83\ e^-\mathrm{s}^{-1}$, $\tau_g = 0.49\,\mathrm{h}$ and $\tau_e = 1.36\,\mathrm{h}$, giving a recovered S/N of 21.3 consistent with the expected value. In this case as the flare time is precisely known the T_0 value was held fixed at this known value. The recovered probability distribution on τ_e, whilst still consistent with the known value, does appear skewed towards high values. This is partially due to larger values of τ_e being strongly correlated with the polynomial background coefficients. These correlations spread the marginalized τ_e distribution towards higher values (see also the discussion in Section 5.1). These correlations could be greatly reduced by using longer data windows (greater than the 55 time bins used here), so that the flare and the background are more easily separated.

Another example of parameter estimation, in this case for a real flare, is shown in Fig. 8. This flare was found on star KID 8376893. In this case the peak time of the flare has also been estimated using a grid in times over a 2 h interval around the time of the maximum odds ratio. The maximum a posteriori recovered values are $A_0 = 529\ e^-\mathrm{s}^{-1}$, $\tau_g = 0\,\mathrm{h}$ and $\tau_e = 0.72\,\mathrm{h}$, which with the estimated noise of $9.3\ e^-\mathrm{s}^{-1}$ gives a recovered S/N of 56.2. The flat posterior probability distribution for τ_g is due to the fact that given the size of the data time-steps small values of τ_g produce indistinguishable models. The best-fitting flare model is overlaid on the real data (after the removal of the median offset) in Fig. 9.

5 ANALYSIS RESULTS

As an initial test of the algorithm we have applied it to *Kepler* Quarter 1 (Q1) long-cadence data as gathered from the public data release on the *Mikulski Archive for Space Telescopes* (*MAST*).[3] We

---

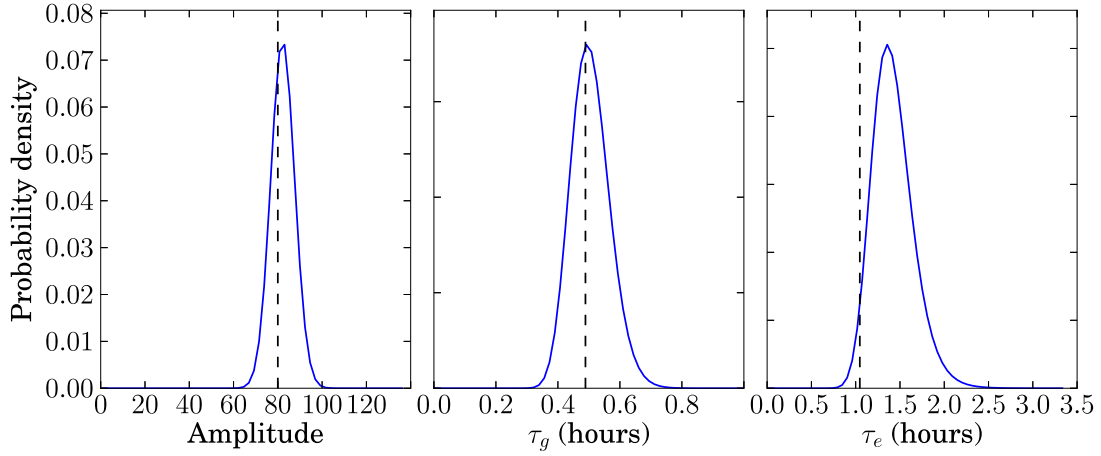[3] http://archive.stsci.edu/kepler/

Figure 7. The parameter posterior probability distributions for a simulated signal added to data from star KID 893676. The true values of the simulated signal are given as vertical dashed lines.
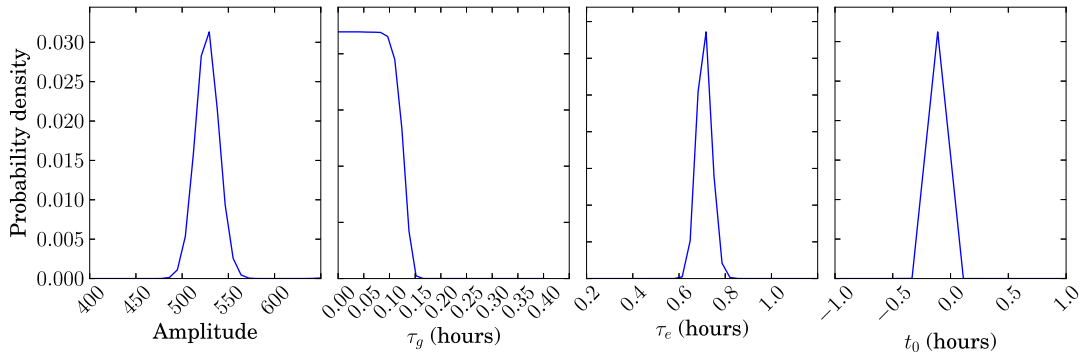

Figure 8. The parameter posterior probability distributions for a real detected signal in data from star KID 8376893.
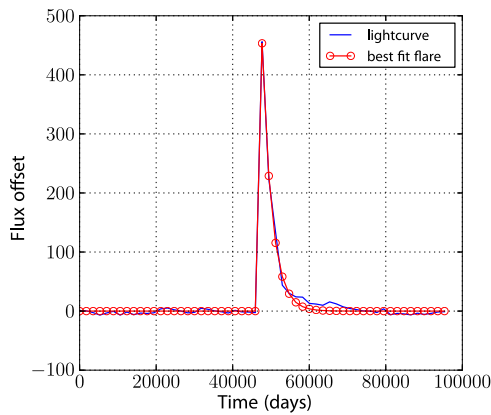

Figure 9. The best-fitting recovered flare model overlaid on the real data for star KID 8376893.

use the same selection criteria as used in Walkowicz et al. (2011) to pick stars (log $g \geq 4.2$ and $T_{\mathrm{eff}} \leq 5150$), which returned 23 301 stars. We veto stars that: have known transits, or are *Kepler* objects of interest; are in eclipsing binaries[4] (Prša et al. 2011; Matijevič et al. 2012) [note that Prša et al. (2011) contains 33 eclipsing binaries that are not flagged in the condition flags returned by the MAST search];

have periods of less than two days (where we take periods from tables 1 and 2 of McQuillan et al. 2014 and periods and secondary periods from Reinhold, Reiners & Basri 2013); are possible red giants; or, exhibit any potential artefacts.[5] This reduced the number of stars to 21 746.

When performing the analysis all light curves were checked for data gaps of more than a single time bin, but none was found (otherwise we would have vetoed them). For single time bin gaps in any of the light curves (i.e. NaNs in the light curve .fits files) we linearly interpolated across the gaps. This is required as the algorithm makes use of discrete Fourier transforms, which required contiguous evenly spaced data.

For all the flares detected by our algorithm we have run the parameter estimation routine (see Section 4) to estimate the most probable set of parameters. We have combined the estimates of τ_{g} and τ_{e} to give a characteristic flare duration, which we define as the time-span that encloses 95 per cent of the best-fitting flare power, centred around the point for which half the flare's power is on either side. We can also use these estimates to derive the recovered S/N of the signals (assuming that they are well described by our flare model).

---

[4] See e.g. the table at http://keplerebs.villanova.edu/

[5] In practice we veto any star for which the MAST/*Kepler* condition flag is defined (e.g. it is not 'None') http://archive.stsci.edu/kepler/condition_flag.html

Table 2. Statistics of detected flares.

Kepler ID	No. of flares	T_{eff} (K)	$\log(g)$ (cm s^{-2})	Median duration (h)	Median S/N	Median amp. (e^- s^{-1})	σ (e^- s^{-1})
1569863	4	3809	4.45	3.14	11.01	55.79	5.9
1570924	3	4923	4.55	2.99	64.85	3774.86	59.5
1722506	2	4270	4.43	12.20	68.62	963.49	9.8
1873543	2	3702	4.52	4.33	25.76	176.47	8.0
2013754	5	4280	4.57	4.55	12.06	79.42	7.8
2140782	1	4946	4.62	2.35	12.87	77.39	5.1
2437317	1	4250	4.70	7.49	575.65	2574.92	8.6
2441562	1	3806	4.46	4.49	37.02	149.61	5.0
2442866	3	4344	4.49	9.44	14.56	62.76	8.6
2570846	1	3865	4.61	2.62	24.98	152.08	5.3

This is an abridged version of the table. A full version is available with the online version of the article.

5.1 Results

On the above data set the algorithm returned 898 flaring stars with a total of 2856 flares. However, despite applying the above selection criterion we have viewed all light curves returning flare candidates by eye to check for other anomalies. In doing this we have vetoed a further 160 stars, many showing strong periodic variations (some stars had high-frequency variations that led to tens of flare signals being produced), or non-white high-frequency noise, or other oddities (e.g. data discontinuities/offsets, nova-like light curves, eclipses). We note that for some stars with detected flares the dynamic range of the underlying light curve variations was so large that visual identification of flares was very difficult (the algorithm may do better than the eye in this case), so in general these stars were vetoed. Other areas where the algorithm more easily identified flares than visual inspection were cases when flare lay on steep parts of any underlying variations.

From the parameter estimation we found that for some of the flare candidates the parameter estimates peaked at the upper edge of the flare time-scale ranges, i.e. we are not including the best-fitting value within the range of our parameter estimation grid. For many flares it could be due to correlations between our polynomial background variation model and the flare time-scales,[6] but for others it may be that the candidates just do not match the flare model very well. However, with these estimates we can further veto flare candidates where this happens [i.e. veto flares for which $p(\tau_e)_{\mathrm{max}} = 5$ h and/or $p(\tau_g)_{\mathrm{max}} = 2$ h]. This will bias us against long-duration flares.

From this analysis, and after the above vetoing, we found 687 stars with a total of 1873 flares (cf. 373 star with 2358 flares in Walkowicz et al. 2011). Prior to the time-scale parameter estimate veto the numbers we found were 738 stars and 2357 flares. We find flares on 305 of the 373 stars identified by Walkowicz et al. (2011) (of the 68 not found in our search 41 of those were missed through our veto criteria) and on these star find on average 70 per cent of the number of flares they found. This means that we have confidently recovered flares on 382 additional stars than previously found. As we are mainly interested in the performance of our algorithm we have just quoted numbers of flares returned by the algorithm, and we

therefore do not include flares that could be identified by subsequent visual inspection of the light curves. Inspection would reveal more flares on the identified flaring stars.

We have visually inspected the light curves and log odds ratio time series' for the 27 unvetoed stars where we found no candidates, but Walkowicz et al. (2011) did. We have also run these through our implementation of the standard deviation threshold detection algorithm (see Section 3.4). Of these 27 only four return a flare candidate using that algorithm, which suggests that the differently processed data sets between our analysis and theirs, and the different noise estimation method, can make a significant difference to the results. For one of these stars the flare is missed as our algorithm removes odds ratios for the beginning and end of the data, and a flare fell into this period. For the other stars there appear to be either: candidate flares just below the log odds ratio threshold; times where potential flare-like signals are vetoed by a strong consistency with the background model (i.e. the events are too short and look more like a short transient, or are inconsistent with our flare model in some other way); or, in some cases no obvious flares were present at all.

Median parameters for the flare candidates after the parameter veto are given in Table 2. The distributions for all flares in both S/N and flare duration are given in Fig. 10. The mean S/N is 36 and the standard deviation of the distribution is 50, with a median value of 20, peak at 14, and range between 6 and 808. The S/N distribution shows the expected fall-off to low values given our efficiency curve in Fig. 5, whilst the fall-off after the peak is related to the true distribution of flares. The duration distribution has a mean of 4.6 h and standard deviation of 3.1 h, a median duration of 3.8 h, a peak at 2.9 h and a range between 0.9 and 19.2 h. It is interesting that there is an anti-correlation between the S/N and the duration. Further simulations are needed to see if this is an effect of the algorithm or a feature of the population. However, it is hard to see why longer high S/N signals would be missed, so it is more likely related to the population of flares. In the future it would be worthwhile to also estimate the relative flare energies using, for example, the method given in Walkowicz et al. (2011) or Shibayama et al. (2013).

We do find a few cases where an obvious loud flare is missed. An example is KID 8176468 where five flares are found, but not the loudest looking event (at around 25.5 d into the data set). Looking at this case we see that this flare is strongly consistent with the background model, because even though it is loud it actually has a decay time that is shorter than our model range. This is consistent with our knowledge that we will be biased against short flares, but in this particular case, where the flare is obvious and has a long tail,

---

[6] With our current analysis window length around a flare of 27 h, flares with $\tau_e \gtrsim 5$ h do not decay to zero by the end of the window. This causes large correlations with the polynomial background that heavily bias the τ_e estimate towards large values. Longer window lengths could be used, but at the expense of a fourth order polynomial not properly fitting the background variations for stars with higher frequency periodicities. Further studies are needed to investigate this more thoroughly.
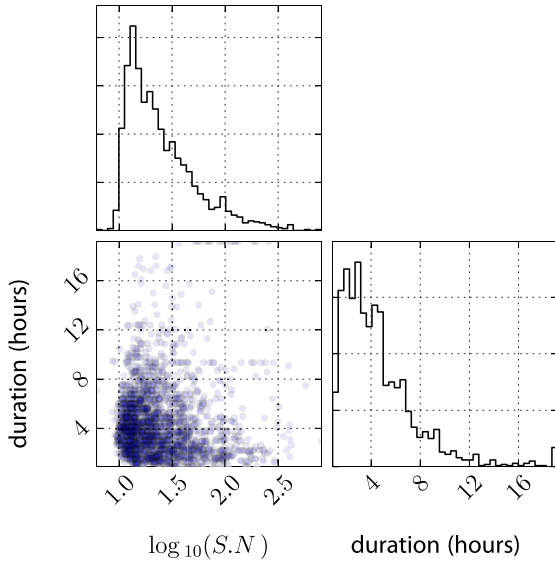
Figure 10. The distributions of the recovered S/N and durations of flares found in the analysis. The figure was produced using an edited version of the python TRIANGLE module from https://github.com/dfm/triangle.py.

it may suggest that the flare model we use may need augmenting to allow a steeper initial decay.

Visual inspection of the flaring light curves identified some interesting candidates. A particularly noteworthy example was KID 9450669, which appeared to undergo an intense period of flaring activity before dropping back to a more normal rate. The light curve and odds ratio time series can be seen in Fig. 11. Inspection of light

curves from all later *Kepler* Quarters does not show such an intense period of activity again.

6 CONCLUSIONS

This paper has been primarily focused on the development and characterization of an algorithm for detecting flares. The algorithm works under the general assumption that flares have a characteristic shape that we model as a Gaussian rise and exponential decay. This assumption seems reasonable from observations of flares, but a more generic model (or indeed a more specific model if there is good theoretical motivation for a particular flare shape) could in the future prove more appropriate. Along with the flare model we have assumed that the data contain unknown low-frequency variations, which we model as a fourth order polynomial. The algorithm computes an odds ratio comparing the probability of the data containing a flare model plus the background polynomial variations (the signal model) to the probability of the data just consisting of the background variations *or* the background variations plus very short duration transients within the analysis window (the background model). This method allows further signal or background models to be added in the future based on updated knowledge of flare morphologies, or better understanding of background artefacts. Another fundamental assumption is that the noise in the data can be modelled with a Gaussian likelihood, which we have seen in Section 3.3 is not wholly reliable, but can be robust provided a conservative threshold is set.

We have characterized this method using simulated data sets consisting of Gaussian white noise plus a low-frequency sinusoidal background. If requiring a FAP of 0.1 per cent (based on data that are the length of *Kepler* Quarter 1 long cadence data) from noise alone this provides a detection threshold on the log odds ratio of 8.3. When this threshold is used for simulated data sets with fake
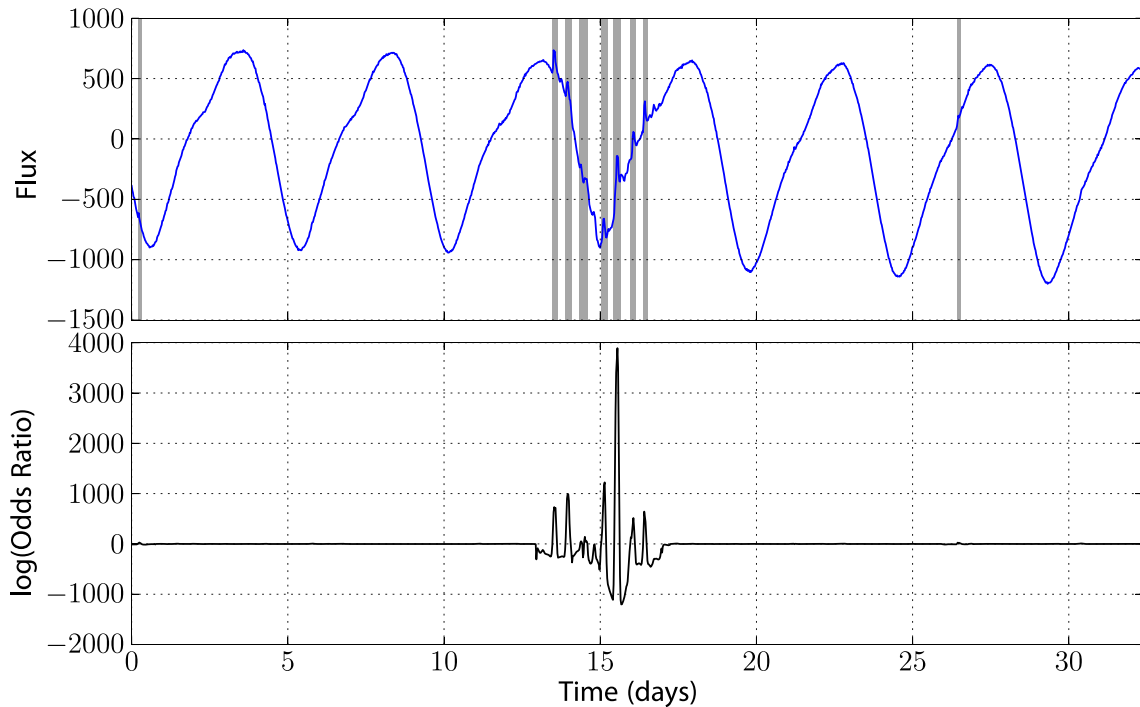

Figure 11. The light curve and log odds ratio time series for KID 9450669 showing an intense period of flaring activity. Shaded areas in the upper plot represent times when the log odds ratio exceeds our threshold of 16.5.

signals of varying parameters and S/N it returns a 95 per cent detection efficiency for signals with S/N $\gtrsim$ 13. However, we find that when adding signals they can occasionally produce spurious extra false detections at times when the analysis window is sliding on to the flare. To reduce these false alarms to occur for only 1 per cent of simulated flares requires increasing the detection threshold to 16.5. This corresponds to a 95 per cent detection efficiency for signals with S/N $\gtrsim$ 20. Studies using real *Kepler* data for 'quiet' stars, in which the noise deviates from the Gaussian assumption, show that this threshold of 16.5 again gives an approximately 1 per cent false alarm rate.

Applying our method to a selection of stars within the ranges $\log g \geq 4.2$ and $T_{\text{eff}} \leq 5150$ in *Kepler* Q1 data (see Section 5 for other veto criteria) we find 687 stars exhibiting flares (cf. 373 stars in the analysis of Walkowicz et al. 2011). We have not aimed at attempting a detailed statistical analysis of distributions of flare properties, but have shown basic distributions of flare duration and S/N.

Future modifications to the algorithm could be to include extra models for the background variations that would be a better fit to the low frequency variations than the polynomial currently used. This could be done by adding a low-frequency sinusoid model with unknown amplitude, frequency and phase to the current polynomial fit. The unknown amplitude and phase could be analytically marginalized over by splitting $m = A \sin(2\pi f t + \phi)$ into $m = B \sin(2\pi f t) + C \cos(2\pi f t)$ and using the algorithm described in Appendix A. However, we would need to numerically search of this frequency range, which would slow down the method. Such a method could allow longer windows to be used and potentially reduce the additional artefacts caused by flares themselves, meaning that non-flare-like transients (such as transits and eclipses) would not trigger a flare detection. This could allow the threshold to be significantly lowered allowing smaller flares to be recovered. As discussed in Section 3.3 further modifications to the noise model to account for short period correlations seen in real *Kepler* data could also be taken into account. Another type of artefact that triggered the algorithm and was observed during visual inspection of *Kepler* data was large step offsets between stretches of data. This type of artefact could easily be modelled in the noise model by including a step function of unknown amplitude (to be analytically marginalized over). The algorithm currently relies on data being contiguous over a *Kepler* Quarter. However, future quarters contain some large gaps. These can easily be dealt with by making the algorithm treat data separated by large gaps as independent analyses.

We plan to run the algorithm on further *Kepler* Quarters to build up a complete picture of flaring activity at statistics on these stars. This would include running on short-cadence data. In further studies it would also be interesting to compare any times highlighted by our method as strongly favouring the noise model with *Kepler* data quality flags. We also expect our method to be applicable to other flare searches, in particular using the large sets of multi-wavelength solar observations.

ACKNOWLEDGEMENTS

REFERENCES

Balona L. A., 2012, MNRAS, 423, 3420
Basri G. et al., 2010, ApJ, 713, L155
Basri G. et al., 2011, AJ, 141, 20
Batalha N. M. et al., 2013, ApJS, 204, 24
Benz A. O., Güdel M., 2010, ARA&A, 48, 241
Bretthorst G. L., 1988, Bayesian Spectrum Analysis and Parameter Estimation. No. 48 in Lecture Notes in Statistics. Springer-Verlag, Berlin, Heidelberg
Cameron E., 2011, PASA, 28, 128
Carrington R. C., 1859, MNRAS, 20, 13
Christiansen J. L. et al., 2013, Technical Report KSCI-19040-00, Kepler Data Characteristics Handbook, NASA Ames Research Center, Moffett Field, CA
Clark J., Heng I. S., Pitkin M., Woan G., 2007, Phys. Rev. D, 76, 043003
Dal H. A., Evren S., 2012, New Astron., 17, 399
Emslie A. G. et al., 2012, ApJ, 759, 71
Fletcher L. et al., 2011, Space Sci. Rev., 159, 19
Gershberg R. E., 1972, Ap&SS, 19, 75
Hambaryan V., Neuhäuser R., Stelzer B., 1999, A&A, 345, 121
Hannah I. G., Hudson H. S., Battaglia M., Christe S., Kašparová J., Krucker S., Kundu M. R., Veronig A., 2011, Space Sci. Rev., 159, 263
Hilton E. J., West A. A., Hawley S. L., Kowalski A. F., 2010, AJ, 140, 1402
Hudson H. S., 1991, Sol. Phys., 133, 357
Hudson H. S., Wolfson C. J., Metcalf T. R., 2006, Sol. Phys., 234, 79
Ishida K., Ichimura K., Shimizu Y., Mahasenaputra 1991, Ap&SS, 182, 227
Jeffreys H., 1998, Theory of Probability, 3 edn. Oxford University Press, Oxford
Jenkins J. M. et al., 2010a, ApJ, 713, L87
Jenkins J. M. et al., 2010b, ApJ, 713, L120
Kowalski A. F., Hawley S. L., Hilton E. J., Becker A. C., West A. A., Bochanski J. J., Sesar B., 2009, AJ, 138, 633
Kowalski A. F., Hawley S. L., Holtzman J. A., Wisniewski J. P., Hilton E. J., 2010, ApJ, 714, L98
Kowalski A. F., Mathioudakis M., Hawley S. L., Hilton E. J., Dhillon V. S., Marsh T. R., Copperwheat C. M., 2011, in Johns-Krull C., Browning M. K., West A. A., eds, ASP Conf. Ser. Vol. 448, 16th Cambridge Workshop on Cool Stars, Stellar Systems, and the Sun. Astron. Soc. Pac., San Francisco, p. 1157
Kowalski A. F., Hawley S. L., Wisniewski J. P., Osten R. A., Hilton E. J., Holtzman J. A., Schmidt S. J., Davenport J. R. A., 2013, ApJS, 207, 15
Lacy C. H., Moffett T. J., Evans D. S., 1976, ApJS, 30, 85
McQuillan A., Aigrain S., Roberts S., 2012, A&A, 539, A137
McQuillan A., Mazeh T., Aigrain S., 2014, ApJS, 211, 24
Maehara H. et al., 2012, Nature, 485, 478
Matijevič G., Prša A., Orosz J. A., Welsh W. F., Bloemen S., Barclay T., 2012, AJ, 143, 123
Moffett T. J., 1974, ApJS, 29, 1
Moffett T. J., Bopp B. W., 1976, ApJS, 31, 61
Osten R. A., Kowalski A., Sahu K., Hawley S. L., 2012, ApJ, 754, 4
Prša A. et al., 2011, AJ, 141, 83
Reinhold T., Reiners A., Basri G., 2013, A&A, 560, A4
Roberts S., McQuillan A., Reece S., Aigrain S., 2013, MNRAS, 435, 3639
Rosner B., 1983, Technometrics, 25, 165
Savitzky A., Golay M. J. E., 1964, Analytical Chem., 36, 1627
Scargle J. D., 1998, ApJ, 504, 405
Searle A. C., Sutton P. J., Tinto M., Woan G., 2008, Classical Quantum Gravity, 25, 114038
Searle A. C., Sutton P. J., Tinto M., 2009, Classical Quantum Gravity, 26, 155017
Shibayama T. et al., 2013, ApJS, 209, 5
Shlens J., 2014, preprint (arXiv:1404.1100)
Smith J. C. et al., 2012, PASP, 124, 1000
Stumpe M. C. et al., 2012, PASP, 124, 985
Stumpe M. C., Smith J. C., Catanzarite J. H., Van Cleve J. E., Jenkins J. M., Twicken J. D., Girouard F. R., 2014, PASP, 126, 100

Thompson S. E. et al., 2013, Technical Report KSCI-19061-001, Kepler Data Release 21 Notes
Walkowicz L. M. et al., 2011, AJ, 141, 50
Woods T. N. et al., 2004, Geophys. Res. Lett., 31, 10802

APPENDIX A: MARGINALIZING OVER A POLYNOMIAL BACKGROUND

In Section 2.3 we introduced a polynomial background variation. Rather than having to integrate numerically over each polynomial coefficient it is possible to analytically integrate each of them. We define a generic model as $m = Af + g$, where $f(x)$ is an arbitrary function with an amplitude that can be factored out as A, and g contains any other components of the model. If we substitute this into the odds ratio defined in equation (6), and place the integral over A between $[-\infty, \infty]$, we can rearrange it to give

$$\mathcal{O} = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(A^2 X + AY + Z)\right) p(A)\mathrm{d}A, \qquad (A1)$$

where $X = \sum_j f_j^2$, $Y = 2(\sum_j f_j g_j - \sum_j d_j f_j)$ and $Z = \sum_j g_j^2 - 2\sum_j d_j g_j$. Provided that the prior on A, $p(A)$ is constant, or varies very little, over the range in which the likelihood ratio is significant (e.g. is very wide Gaussian), this integral is solved to give

$$\mathcal{O} = p(A)\sqrt{\frac{2\pi\sigma^2}{X}} \exp\left(-\frac{(Z - Y^2/4X)}{2\sigma^2}\right). \qquad (A2)$$

For the case where the amplitude A must be purely positive (e.g. a flare) then equation (A1) can be integrated between $[0, \infty]$ to give

$$\mathcal{O} = p(A)\sqrt{\frac{\pi\sigma^2}{2X}} \exp\left(-\frac{(Z - Y^2/4X)}{2\sigma^2}\right) \mathrm{erfc}\left(\frac{Y}{2\sqrt{2\sigma^2 X}}\right). \quad (A3)$$

If we consider a model containing a third order polynomial and a flare signal

$$m = A_0 m_{\mathrm{f}}(t, \tau_{\mathrm{g}}, \tau_{\mathrm{e}}, T_0) + A + Bt + Ct^2 + Dt^3 \qquad (A4)$$

the odds ratio in equation (6) would instead be given by

$$\mathcal{O}(T_0) = \int_{\tau_{\mathrm{g}}} \int_{\tau_{\mathrm{e}}} \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-\frac{\sum(m^2 - 2dm)}{2\sigma^2}}$$
$$\times p(\tau_{\mathrm{g}}, \tau_{\mathrm{e}}, A_0, A, B, C, D)\mathrm{d}\tau_{\mathrm{g}}\mathrm{d}\tau_{\mathrm{g}}\mathrm{d}A_0\mathrm{d}A\mathrm{d}B\mathrm{d}C\mathrm{d}D, \quad (A5)$$

which can be analytically reduced by four applications of equation (A2) followed by a final application of equation (A3). Note that the integral over $[0, \infty]$ can only be used for one parameter and must be performed last. If explicitly performing the integral for the five amplitudes being marginalized over above (e.g. using a symbolic mathematics program) the algebraic expression is far too large to be easily transcribed into an analysis code. However, we have developed an algorithm that can perform the integral for an arbitrary number of model component amplitudes without the need to explicitly write out the whole expression.[7] This is given in pseudo-code below.

If performing all amplitude marginalizations between $[-\infty, \infty]$ an alternative approach to simplify the integral would be to create an orthogonal set of model components (via diagonalization of a model component matrix), which would mean that all cross terms between the *orthogonal* models were zero (see e.g. Bretthorst 1988).

[7] A version of this algorithm is provided in the AMPLITUDE-MARGINALISER function library found at http://github.com/mattpitkin/amplitude-marginaliser

A1 The marginalization algorithm

Given N model components for which an amplitude can be factored out we have a generic model

$$m = \sum_i^N A_i f_i \qquad (A6)$$

where A_i are the amplitude components and f_i are the component functions. We can create a matrix, C, containing the sums of the products of each of the component functions with each other and with the data d, such that

$$\mathtt{C} = \begin{pmatrix} -2\sum_j d_j f_{1j} & 2\sum_j f_{1j} f_{2j} & 2\sum_j f_{1j} f_{3j} & \cdots \\ 0 & -2\sum_j d_j f_{2j} & 2\sum_j f_{2j} f_{3j} & \cdots \\ 0 & 0 & -2\sum_j d_j f_{3j} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

where the sums are over all the model and data points. We can also create a vector containing the squared model terms

$$\mathtt{S} = \left\{ \sum_j f_{1j}^2, \sum_j f_{2j}^2, \sum_j f_{3j}^2, \cdots \right\}.$$

Algorithm 1 can then be applied to produce the X, Y and Z values given in equations (A2) and (A3) and calculate the log of the odds ratio marginalized over the amplitudes of the model components.

Algorithm 1 Calculate $\log \mathcal{O}$ marginalized over model component amplitudes (this assumes C-style array indexing starting from 0).

```
 1: Z = 0
 2: M = N − 1
 3: log 𝒪 = 0
 4: for i = 0 to M − 1 do
 5:     for j = i to M do
 6:         for k = i to M do
 7:             if j is equal to i then
 8:                 if k > j then
 9:                     S[k] = S[k] − C[j][k]²/4S[i]
10:                 else if k is equal to j then
11:                     Z = Z − C[j][k]²/4S[i]
12:                 end if
13:             else
14:                 if k is equal to i then
15:                     C[j][j] = C[j][j] − C[i][j] × C[i][k]/2S[i]
16:                 else if k > j then
17:                     C[j][k] = C[j][k] − C[i][j] × C[i][k]/2S[i]
18:                 end if
19:             end if
20:         end for
21:     end for
22: end for
23: X = S[M]
24: Y = C[M][M]
25: for i = 0 to M do
26:     log 𝒪 = log 𝒪 − log (√S[i])
27: end for
28: log 𝒪 = log 𝒪 + N log σ − (Z − Y²/4X)/2σ²
29: if integrating the final amplitude from [0, ∞] then
30:     log 𝒪 = log 𝒪 + M log (√2π) + log (√π/2)
```

2284 *M. Pitkin et al.*

31: $\log \mathcal{O} = \log \mathcal{O} + \log \left(\text{erfc} \left(Y / \left(2\sqrt{2\sigma^2 X} \right) \right) \right)$
32: **else** {integrating the final amplitude from $-\infty$ to ∞}
33: $\quad \log \mathcal{O} = \log \mathcal{O} + N \log \left(\sqrt{2\pi} \right)$
34: **end if**
35: add $\log$ (amplitudepriors) to odds ratio

A2 Marginalization without the final amplitude

In the above marginalizations all the component amplitudes have been marginalized. However, if, for example, performing parameter estimation one might be interested in the actual posterior probability distribution of one of those amplitudes, or there might be some subset of components that do not have a factorizable amplitude, which therefore cannot be marginalized over with this algorithm. These require a slight modification of Algorithm 1 as given in Algorithm 2, where essentially the final amplitude integral is not performed. In our case when performing parameter estimation we would be interested in the actual posterior probability distribution of the flare amplitude, but would still want the background polynomial coefficients marginalized over.

Algorithm 2 Calculate $\log \mathcal{O}$ marginalized over model component amplitudes except for a final model component, where the sizes of C and S are the same as for Algorithm 1 (again this assumes C-style array indexing and that N refers to the total number of model components including the one being left out of the marginalization).

1: $Z = 0$
2: $Y = 0$
3: $M = N - 1$
4: $P = N - 2$
5: $\log \mathcal{O} = 0$
6: **for** $i = 0$ **to** P **do**
7: $\quad$ **for** $j = i$ **to** M **do**
8: $\quad\quad$ **for** $k = i$ **to** M **do**
9: $\quad\quad\quad$ **if** $j \neq i + 1$ **and** $j \neq P$ **and** $k \neq i + 1$ **and** $k \neq P$ **then**
10: $\quad\quad\quad\quad Z = Z - \mathtt{C}[i][j] \times \mathtt{C}[i][k]/4\mathtt{S}[i]$
11: $\quad\quad\quad$ **end if**
12: $\quad\quad\quad$ **if** j is equal to i **then**
13: $\quad\quad\quad\quad$ **if** $k > j$ **and** $k < M$ **then**
14: $\quad\quad\quad\quad\quad \mathtt{S}[k] = \mathtt{S}[k] - \mathtt{C}[j][k]^2/4\mathtt{S}[i]$
15: $\quad\quad\quad\quad$ **end if**
16: $\quad\quad\quad$ **else**
17: $\quad\quad\quad\quad$ **if** k is equal to i **and** $j < M$ **then**
18: $\quad\quad\quad\quad\quad \mathtt{C}[j][j] = \mathtt{C}[j][j] - \mathtt{C}[i][j] \times \mathtt{C}[i][k]/2\mathtt{S}[i]$
19: $\quad\quad\quad\quad$ **else if** $k > j$ **then**
20: $\quad\quad\quad\quad\quad \mathtt{C}[j][k] = \mathtt{C}[j][k] - \mathtt{C}[i][j] \times \mathtt{C}[i][k]/2\mathtt{S}[i]$
21: $\quad\quad\quad\quad$ **end if**
22: $\quad\quad\quad$ **end if**
23: $\quad\quad$ **end for**
24: $\quad$ **end for**
25: **end for**
26: $X = \mathtt{S}[P]$
27: **for** $i = P$ **to** M **do**
28: $\quad Y = Y + \mathtt{C}[P][i]$
29: **end for**
30: $Z = Z + \mathtt{S}[M] + \mathtt{C}[M][M]$
31: **for** $i = 0$ **to** P **do**
32: $\quad \log \mathcal{O} = \log \mathcal{O} - \log \left(\sqrt{\mathtt{S}[i]} \right)$
33: **end for**
34: $\log \mathcal{O} = \log \mathcal{O} + M \log \left(\sqrt{2\pi\sigma^2} \right) - (Z - Y^2/4X)/2\sigma^2$
35: add $\log$ (amplitudepriors) to odds ratio

APPENDIX B: AN ALTERNATIVE METHOD FOR NOISE ESTIMATION

Here we describe an alternative to the noise estimation method discussed in Section 2.6. Flares have most of their spectral power at low frequencies, and also the background variations are at low frequencies (compared to the sample time), so assuming that the underlying noise is white and Gaussian the high-frequency end of the spectrum can be used to estimate the noise standard deviation. For spectra of light curves with very large amplitude variations a large amount of the power from the variations can leak into the high-frequency part of the spectrum and lead to an overestimation for the noise, so the data must first be detrended using e.g. the Savitzky–Golay filtering algorithm. We then take the mean of the final half of the one-sided power spectrum of the filtered data $\bar{S}_n$ (although different fractions of the spectrum could be used) and get the time series standard deviation via $\sigma = \sqrt{\bar{S}_n/(2\Delta t)}$, where Δt is the time-step between points (in seconds).

For flare signals with S/N of a few 10s their power can start leaking into the part of the spectrum we use for noise estimation. This would mean that for data containing loud flares there would be a bias against finding quieter flares due to noise overestimation.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

(http://mnras.oxfordjournals.org/lookup/suppl/doi:10.1093/mnras/stu1889/-/DC1).

Please note: Oxford University Press are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

This paper has been typeset from a TeX/LaTeX file prepared by the author.