Using Gaussian Processes to detect AGN flares

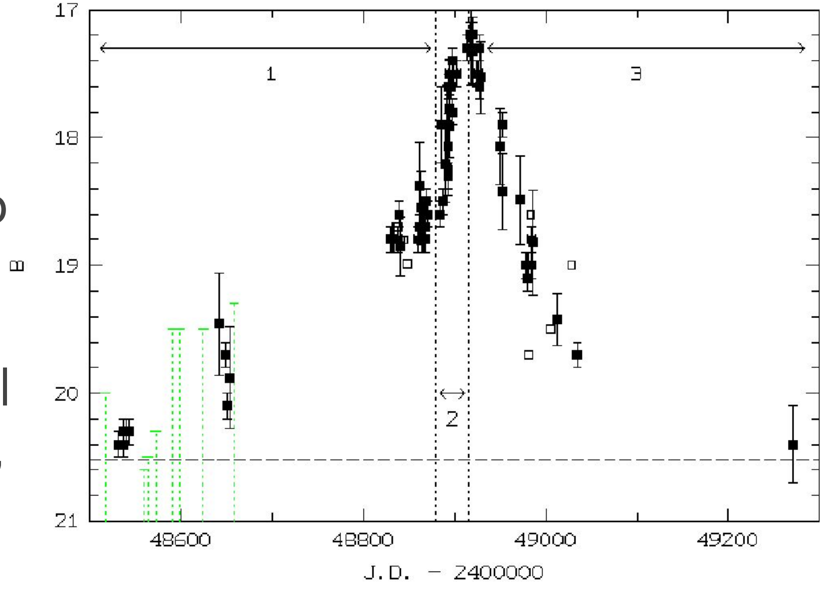
Summer A. J. McLaughlin[1★], James R. Mullaney[1]
[1]Department of Physics and Astronomy, University of Sheffield, Sheffield S3 7RH, UK

Image credit: Artist's impression of a tidal disruption event. NASA/JPL-Caltech

★sajmclaughlin1@sheffield.ac.uk

AGN flares

- Extreme variability of AGN
- Rare: ~ 10^{-4}-10^{-5} yr^{-1} per AGN[1]
- Timescales of several hundreds of days to many years[2]
- Caused by accretion disk instabilities, tidal disruption events, variable accretion rates, supernovae, microlensing… or ?



H. Meusinger *et al* 2010 *A&A* **512** A1

[1] Lawrence, A. *et al* 2016 *MNRAS* **463** 296–331
[2] Graham, J. *et al* 2017 *MNRAS* **470** 4112–4132

What has been done before?

- Lawrence *et al* 2016 - 76 flare candidates in Pan-STARRS1[1]:
    - Define a flare as brightening by > 1.5 mag/decade
- Graham *et al* 2017 - 51 flare candidates from over 900,000 known quasars[2]:
    - De-trend light curves by subtracting Theil-Sen median
    - Then look for contiguous sets of points above zero (mini flare)
    - Define a flare as exceeding the median absolute deviation of these mini flares
    - Ignore flares with a duration < 300 days

[1] Lawrence, A. *et al* 2016 *MNRAS* **463** 296–331
[2] Graham, J. *et al* 2017 *MNRAS* **470** 4112–4132

What has been done before?

University of Sheffield

- Lawrence *et al* 2016 - 76 flare candidates in Pan-STARRS1[1]:
    - Define a flare as brightening by > 1.5 mag/decade
- Graham *et al* 2017 - 51 flare candidates from over 900,000 known quasars[2]:
    - De-trend light curves by subtracting Theil-Sen median
    - Then look for contiguous sets of points above zero (mini flare)
    - Define a flare as exceeding the median absolute deviation of these mini flares
    - Ignore flares with a duration < 300 days
- Amplitude selection
- Timescale selection

[1] Lawrence, A. *et al* 2016 *MNRAS* **463** 296–331
[2] Graham, J. *et al* 2017 *MNRAS* **470** 4112–4132

What has been done before?

- Lawrence *et al* 2016 - 76 flare candidates in Pan-STARRS1[1]:
  - Define a flare as brightening by > 1.5 mag/decade
- Graham *et al* 2017 - 51 flare candidates from over 900,000 known quasars[2]:
  - De-trend light curves by subtracting Theil-Sen median
  - Then look for contiguous sets of points above zero (mini flare)
  - Define a flare as exceeding the median absolute deviation of these mini flares
  - Ignore flares with a duration < 300 days

<u>We want a statistical way to detect flares without having to assume an amplitude or timescale.</u>

- Amplitude selection
- Timescale selection

[1] Lawrence, A. *et al* 2016 *MNRAS* **463** 296–331
[2] Graham, J. *et al* 2017 *MNRAS* **470** 4112–4132

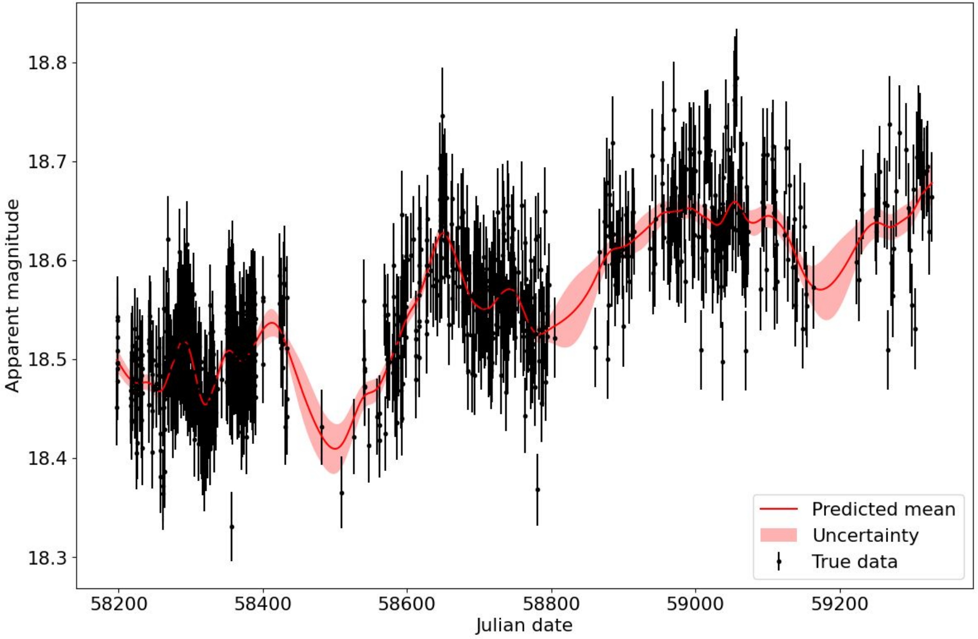
Our problem:

How do you detect a signal in data that is already intrinsically variable?

- ➔ We need a way of quantifying a significant departure from the baseline variability
- ➔ Must be statistically robust and impervious to outliers and poorly sampled data

Gaussian processes

- Non-linear, non-parametric interpolation
- Informed by the sparsity of data
- Uses a kernel (covariance matrix) to describe the best-fit relationship between data points

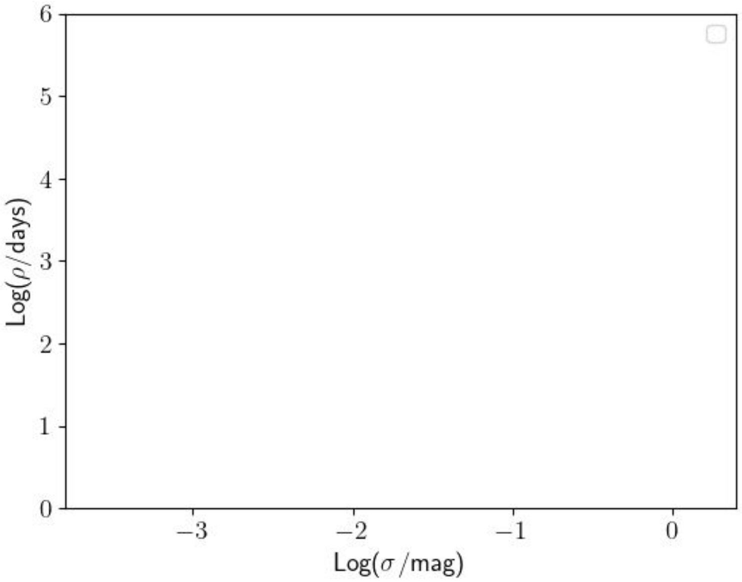


GPs: A feasibility study

University of Sheffield

- A GP is a means of parameterising the covariance of data
- With this in mind, we might expect flares and non-flares to exist in different parts of parameter space
- We simulated 20,000 AGN light curves and injected 10,000 with a flare
- We ran a GP on these light curves with a Matérn 3/2 kernel (right)

$$k(x) = \sigma^2 \left(1 + \frac{\sqrt{3}x}{\rho}\right) \exp\left(-\frac{\sqrt{3}x}{\rho}\right)$$

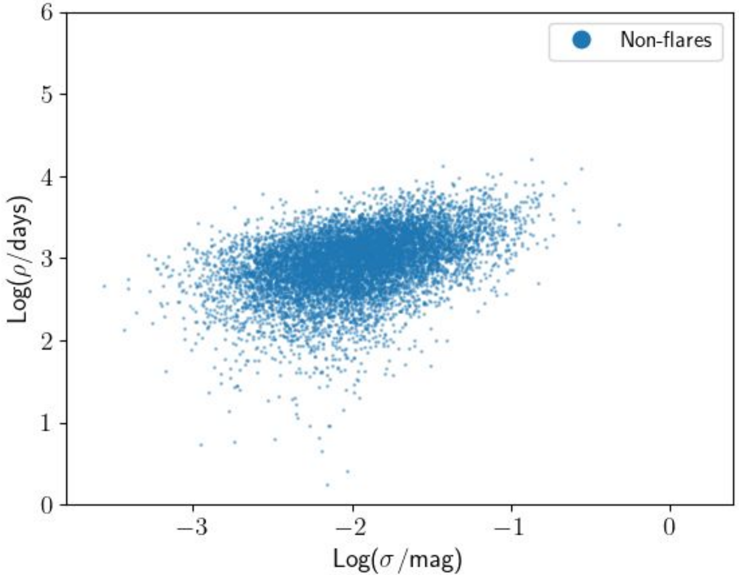


McLaughlin & Mullaney (in prep)

GPs: A feasibility study

- A GP is a means of parameterising the covariance of data
- With this in mind, we might expect flares and non-flares to exist in different parts of parameter space
- We simulated 20,000 AGN light curves and injected 10,000 with a flare
- We ran a GP on these light curves with a Matérn 3/2 kernel (right)

$$\boxed{k(x) = \sigma^2\left(1 + \frac{\sqrt{3}x}{\rho}\right)\exp\left(-\frac{\sqrt{3}x}{\rho}\right)}$$

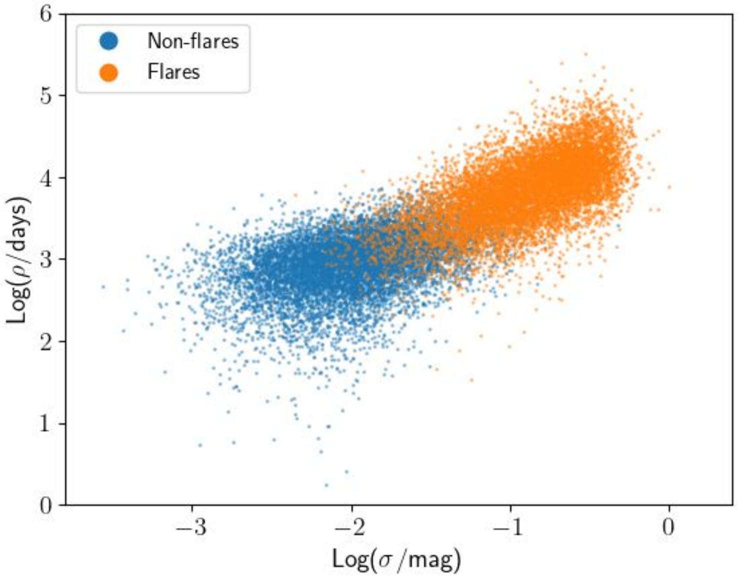


McLaughlin & Mullaney (in prep)

GPs: A feasibility study

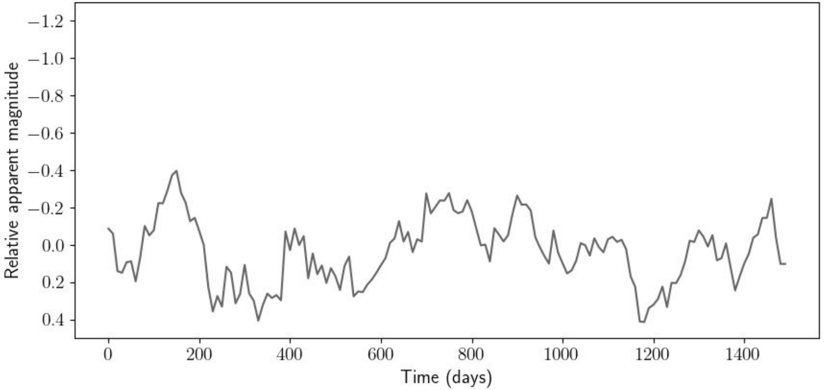
- A GP is a means of parameterising the covariance of data
- With this in mind, we might expect flares and non-flares to exist in different parts of parameter space
- We simulated 20,000 AGN light curves and injected 10,000 with a flare
- We ran a GP on these light curves with a Matérn 3/2 kernel (right)

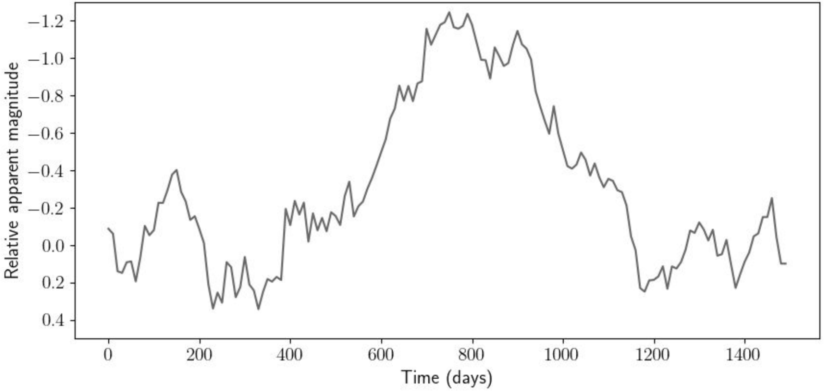
$$k(x) = \sigma^2 \left(1 + \frac{\sqrt{3}x}{\rho}\right) \exp\left(-\frac{\sqrt{3}x}{\rho}\right)$$

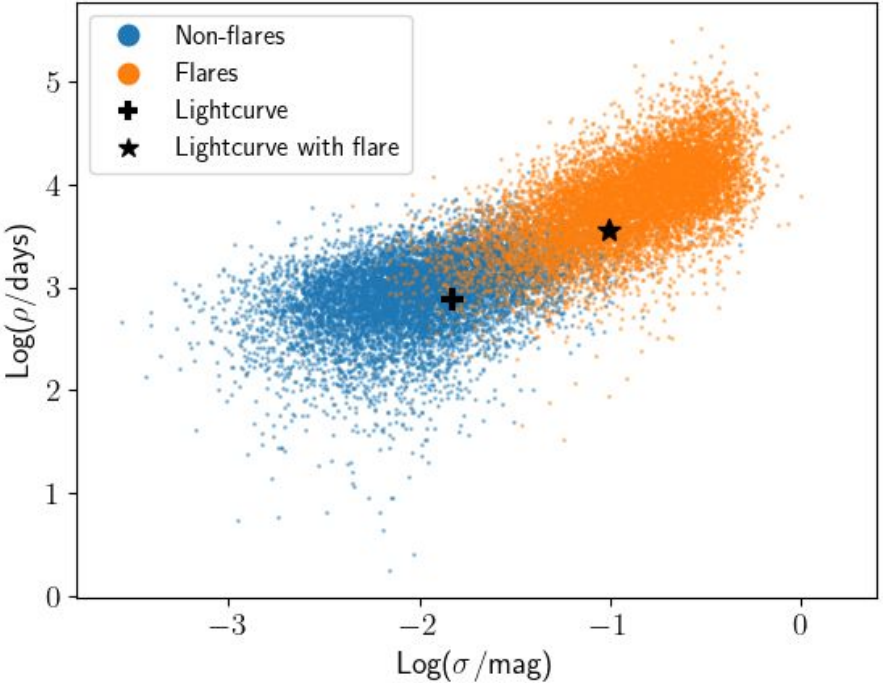


McLaughlin & Mullaney (in prep)

Some examples



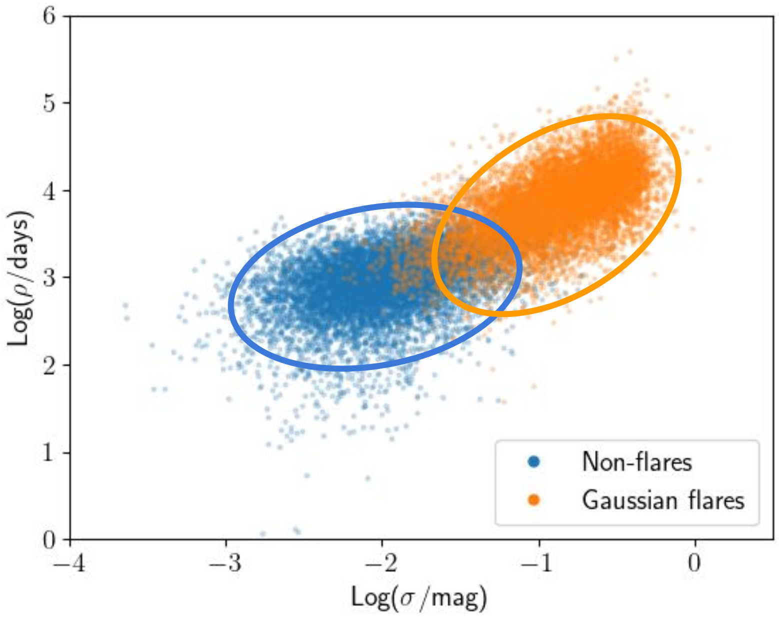




McLaughlin & Mullaney (in prep)

Bayesian hypothesis testing

- Based on a light curve's hyperparameters, what is the probability that it contains a flare?
- Use prior probabilities and likelihoods from the GP
- Define a probability above which a light curve is flagged as a flare



McLaughlin & Mullaney (in prep)

Light curves

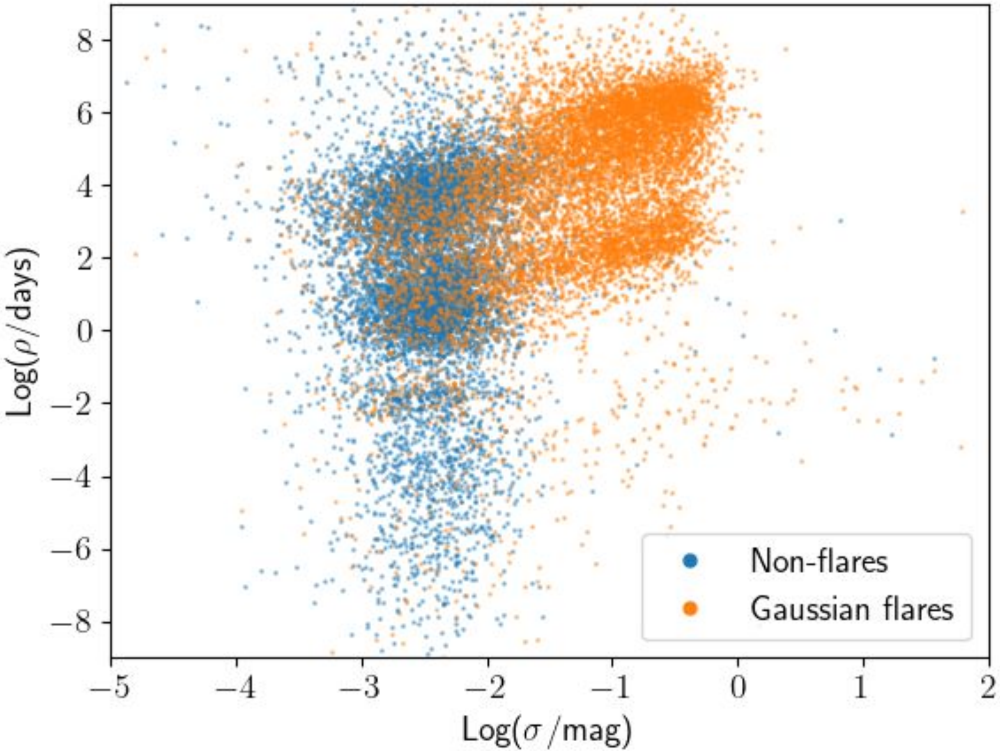
University of Sheffield

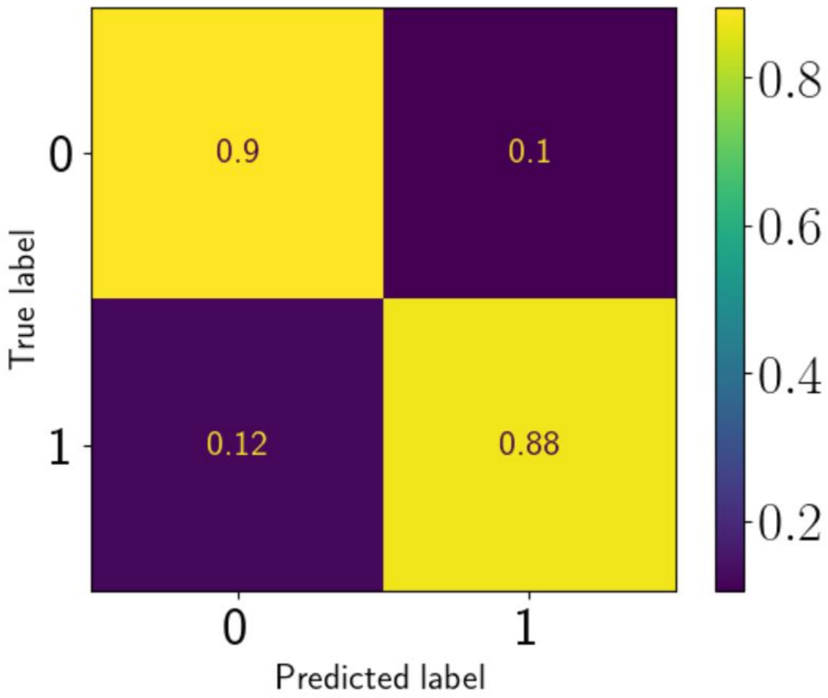
- Light curves sampled every 10 days, half without flares and half with injected flares
- Subsampled light curves
- Added outliers
- Real ZTF light curves, half injected with flares
- Real ZTF light curves - no flares added

Simulated AGN light curves (1D damped random walk)

Real AGN light curves

Provisional results - retrieval rates



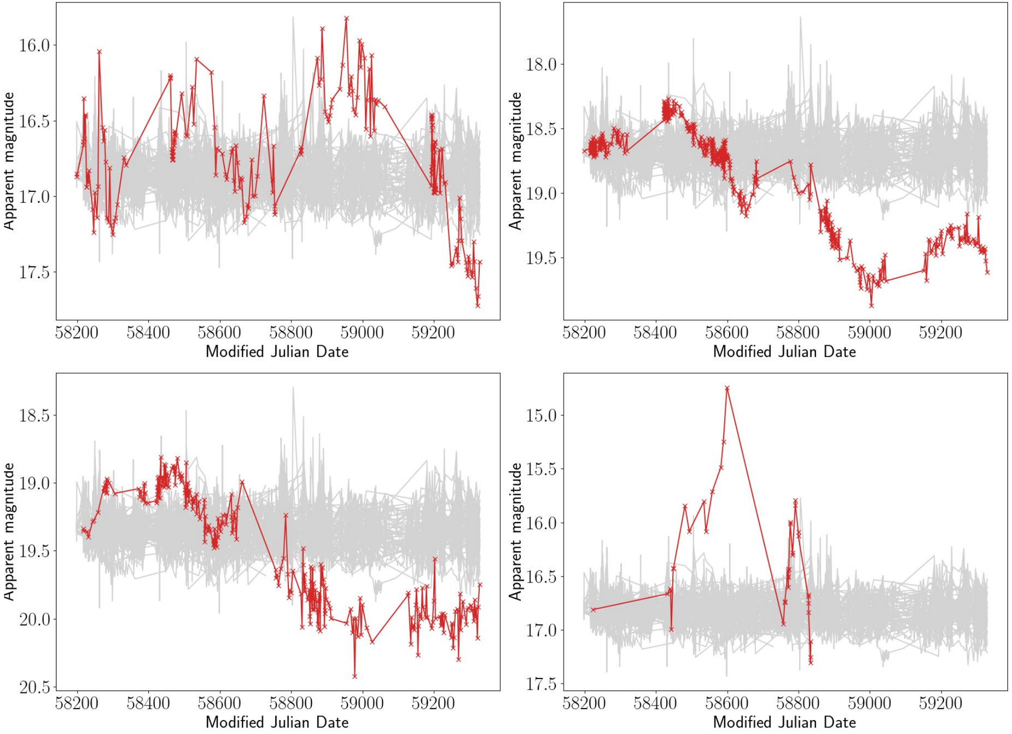


McLaughlin & Mullaney (in prep)

Provisional results - ZTF light curves

- The GP selected 55 out of 9035 spectroscopically selected ZTF light curves of Type 1 AGN as containing flares

- They show extreme variability when compared with a randomly selected sample of 100 ZTF light curves that were not flagged as flares



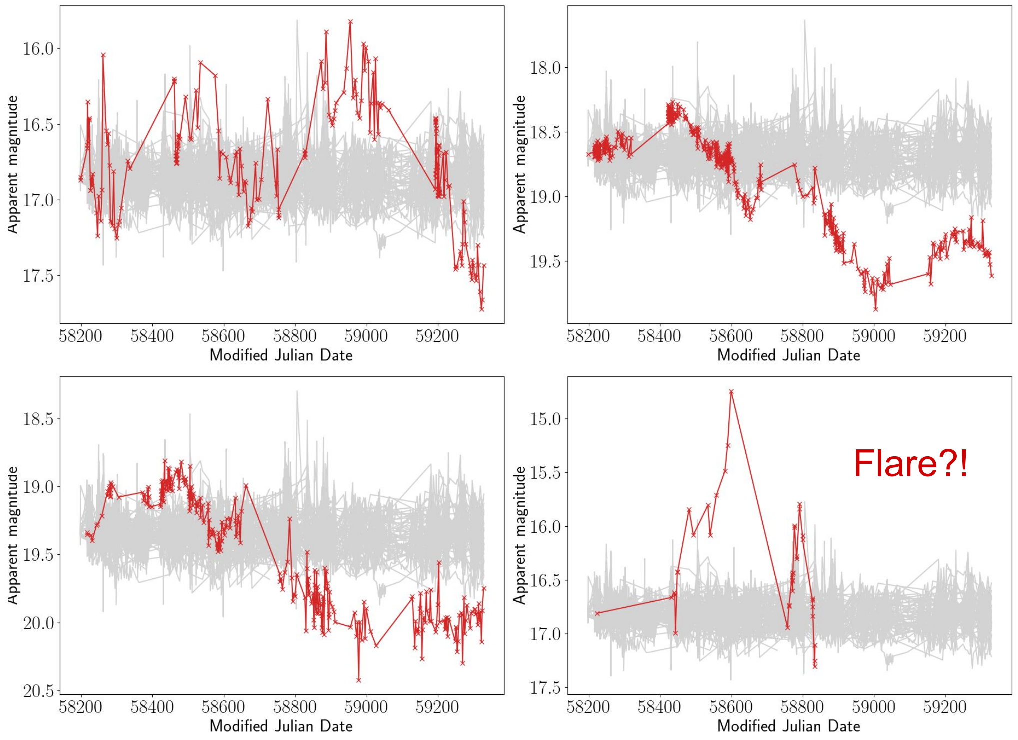
McLaughlin & Mullaney (in prep)

Provisional results - ZTF light curves

University of Sheffield

- The GP selected 55 out of 9035 spectroscopically selected ZTF light curves of Type 1 AGN as containing flares

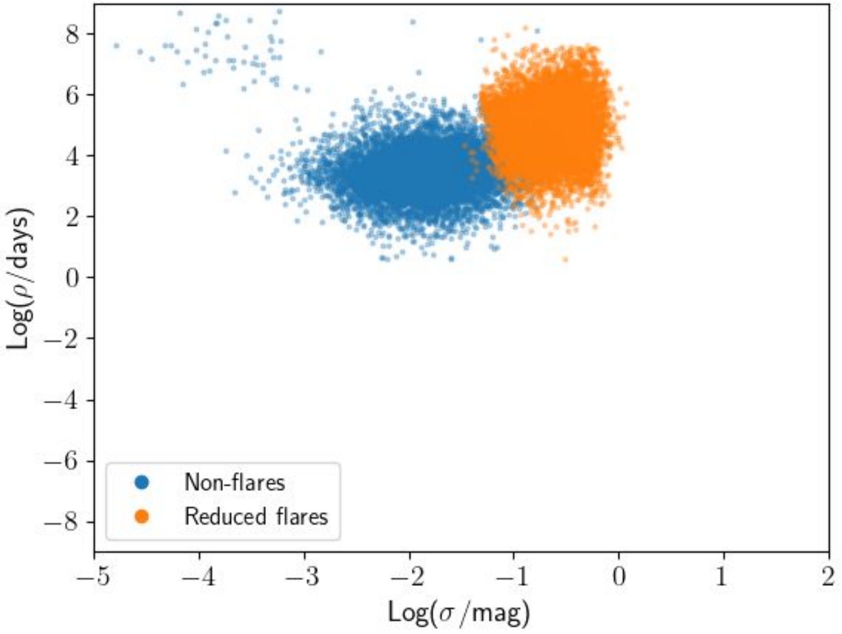
- They show extreme variability when compared with a randomly selected sample of 100 ZTF light curves that were not flagged as flares



McLaughlin & Mullaney (in prep)

What’s next?

- The GP cannot tell you the location of the flare within the light curve
- Deep GPs or regime-switching models are possibilities
- If you can localise the flare, this approach is more effective



McLaughlin & Mullaney (in prep)

Summary

Email: sajmclaughlin1@sheffield.ac.uk

University of Sheffield

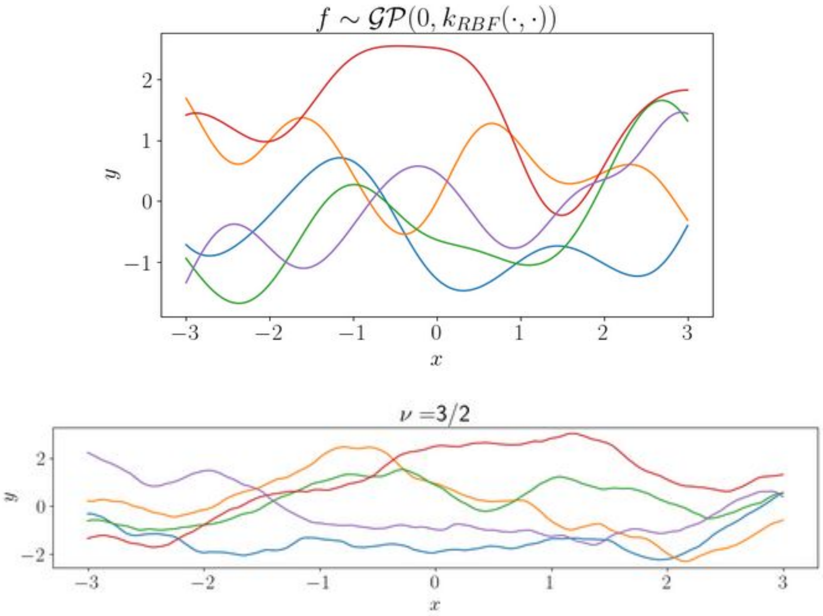
- Ultimate goal is a statistically robust means of identifying flares that is immune to anomalies & sampling

- We have shown that Gaussian Processes can classify flare and non-flare light curves in well-sampled simulated data, and real ZTF data with injected flares

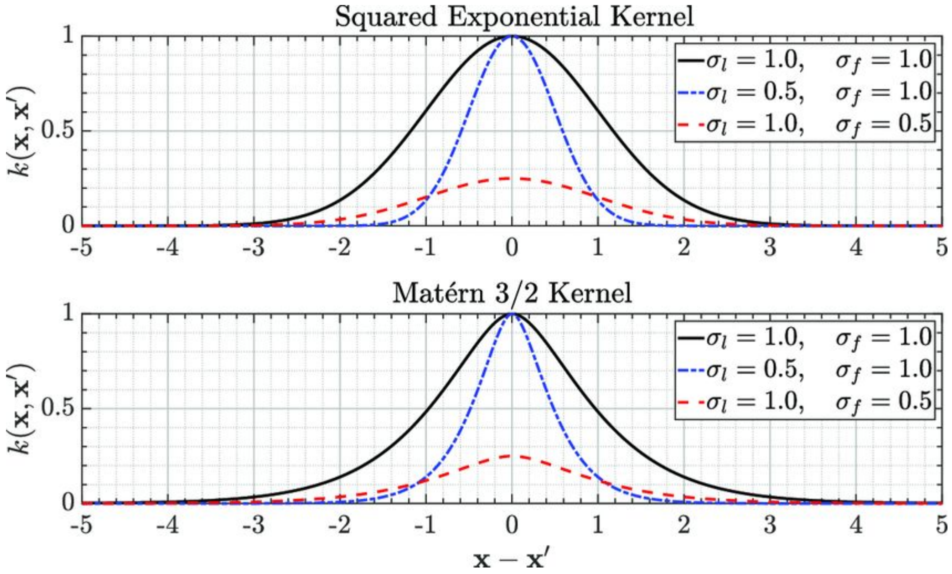
- Regime-switching models or deep GPs may help to localise flares

Extra slides

Gaussian process kernels



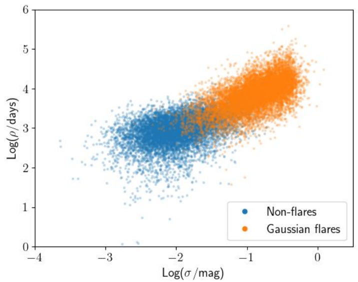
https://andrewcharlesjones.github.io/journal/matern-kernels.html

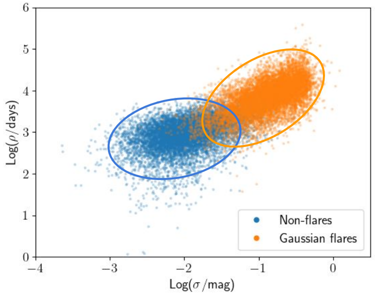


Florian M Heckmeier & Christian Breitsamter 2020 Meas. Sci. Technol. 31 125301

The steps broken down

Run flare and non-flare light curves through a GP and track their hyperparameters



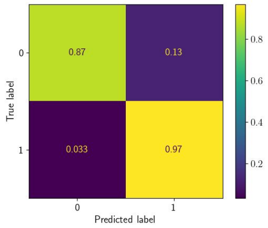


Use the hyperparameter distributions to define priors

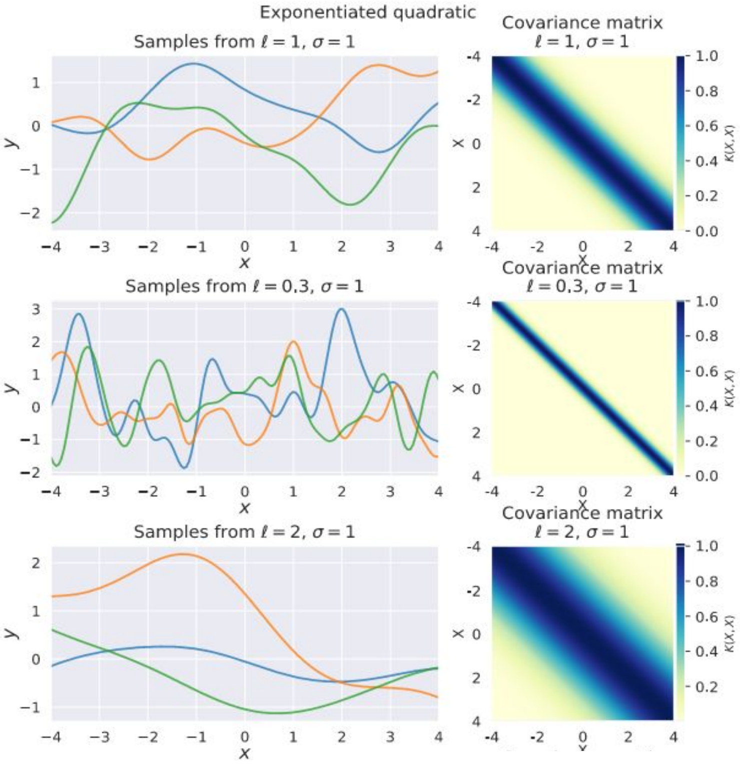
θ = “hyper prior”:
1 (flare), 0 (not flare)

$$P(\sigma, \rho, \theta \mid y) \propto P(y \mid \sigma, \rho) \times P(\sigma, \rho \mid \theta) \times P(\theta)$$

Combine MCMC and Bayesian Hypothesis testing to calculate the probability of a given light curve containing a flare



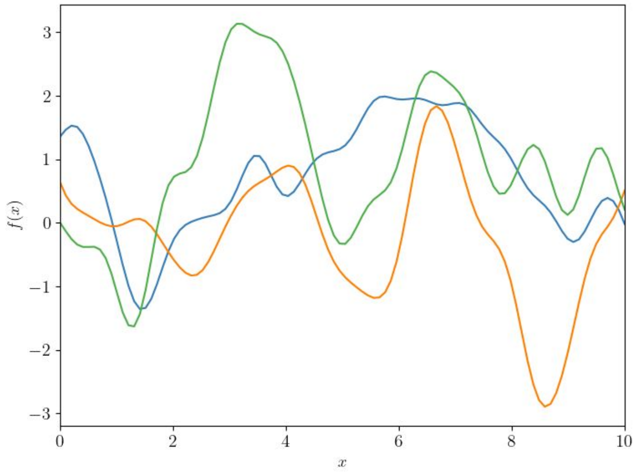
Gaussian process kernels (cont.)



https://peterroelants.github.io/posts/gaussian-process-kernels/

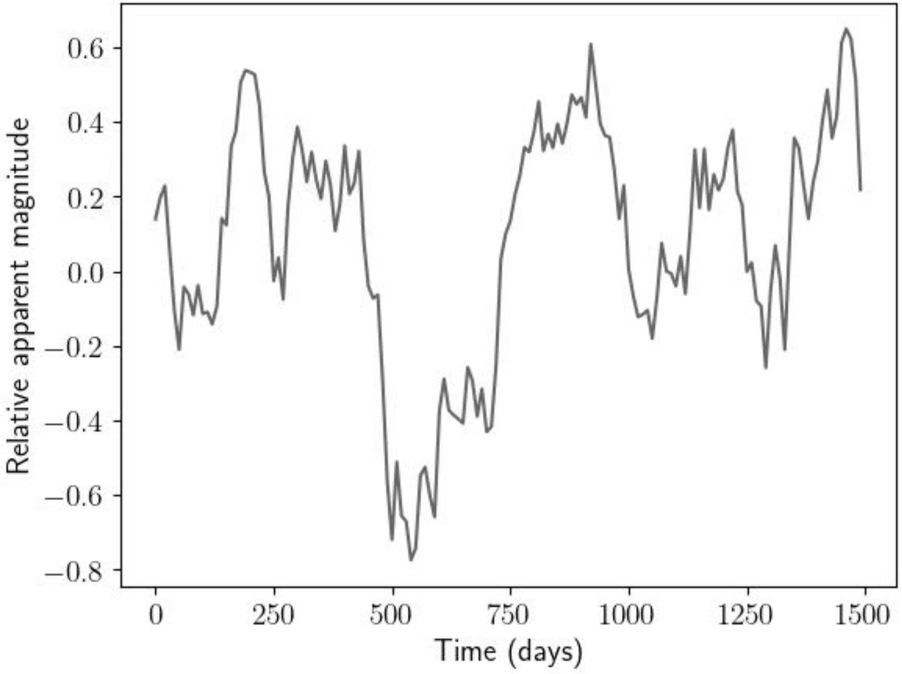
$$k(x) = \sigma^2 \left(1 + \frac{\sqrt{3}x}{\rho}\right) \exp\left(-\frac{\sqrt{3}x}{\rho}\right)$$

Matérn 3/2 kernel

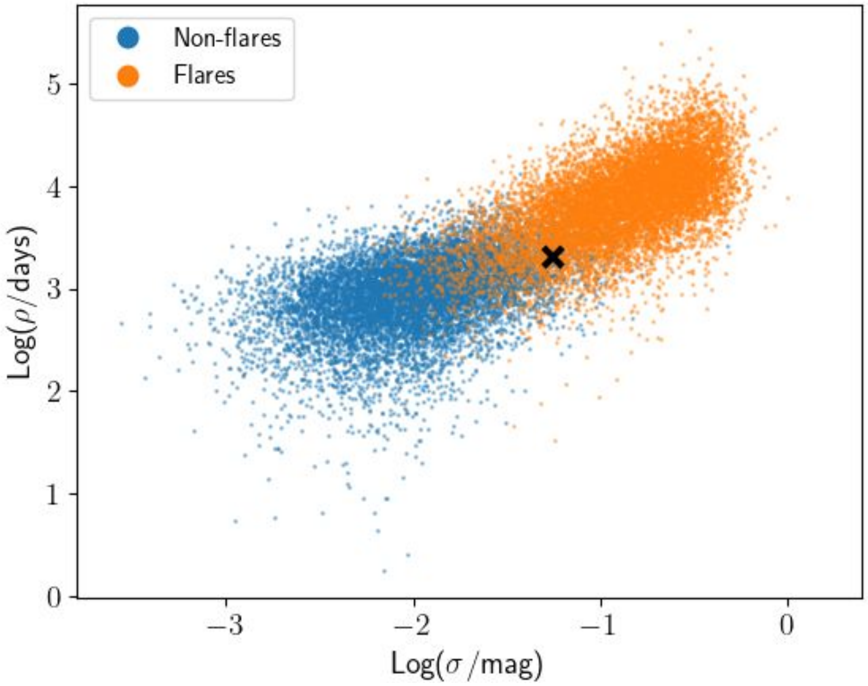


McLaughlin & Mullaney (in prep)

Some examples



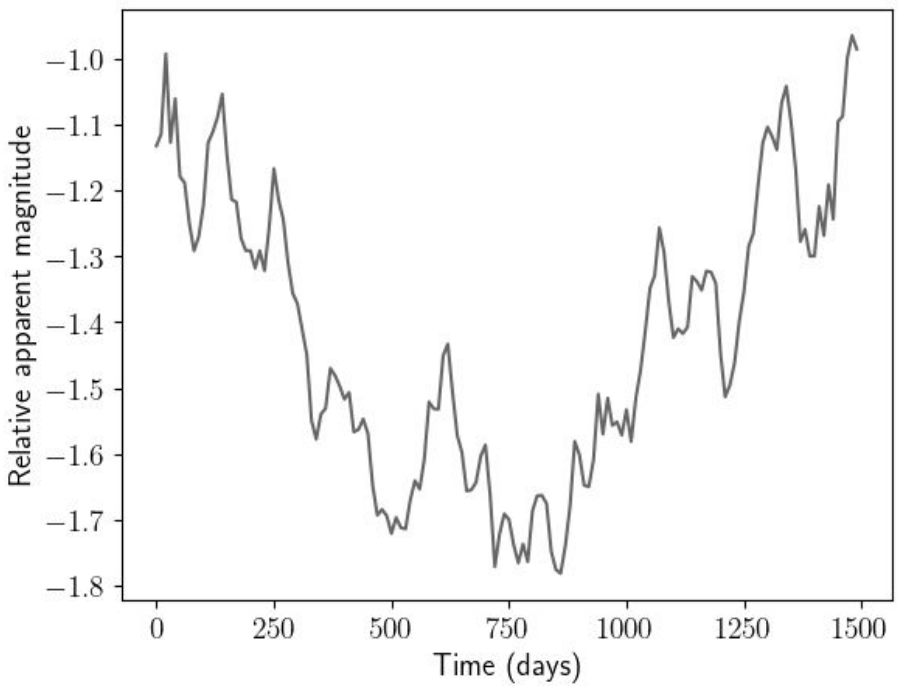
Simulated with a damped random walk



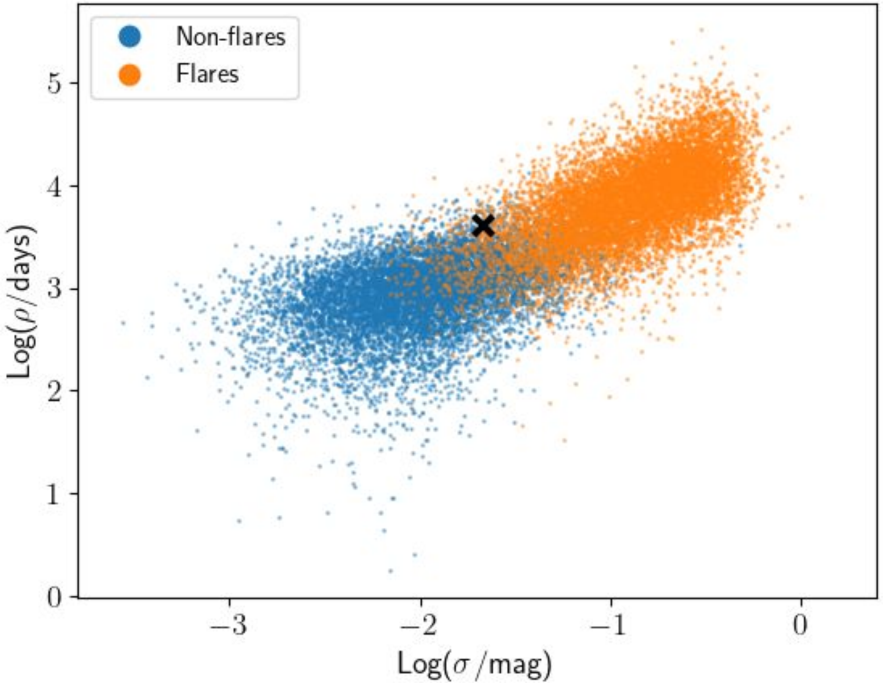
McLaughlin & Mullaney (in prep)

Some examples

University of Sheffield



Simulated with a damped random walk + injected 1.6 mag flare



McLaughlin & Mullaney (in prep)