

## 4. Real data

Simon Vaughan \*

July 8, 2016

In this section we discuss an example of a real astronomical dataset.

### The observations

During 2016 the *Swift* satellite made repeated observations of the active galaxy NGC 6814 as part of a Leicester-led project. We obtained data from  $n = 43$  short observations, which provided fluxes (a measure of brightness) in X-ray and ultraviolet (UV) wavebands. Here we will look at both of these.

The data file `2016_uvw1_rate.txt` is an ASCII (plain text) file containing 4 columns of numbers, and a first row with the column names (a *header* row). The columns give

1. `t`: the observation time in Modified Julian Day (MJD)
2. `dt`: the length of the observation (in days)
3. `y`: the flux (brightness) in counts second<sup>-1</sup> units
4. `dy`: the error on the flux in counts second<sup>-1</sup> units

We can use R to load the data file and make a simple plot.

```
# set the working directory
setwd("C:/Users/sav2/Documents/R/gin/vignettes")

# load the file into the 'dat' object
dat <- read.table("2016_uvw1_rate.txt", header = TRUE)

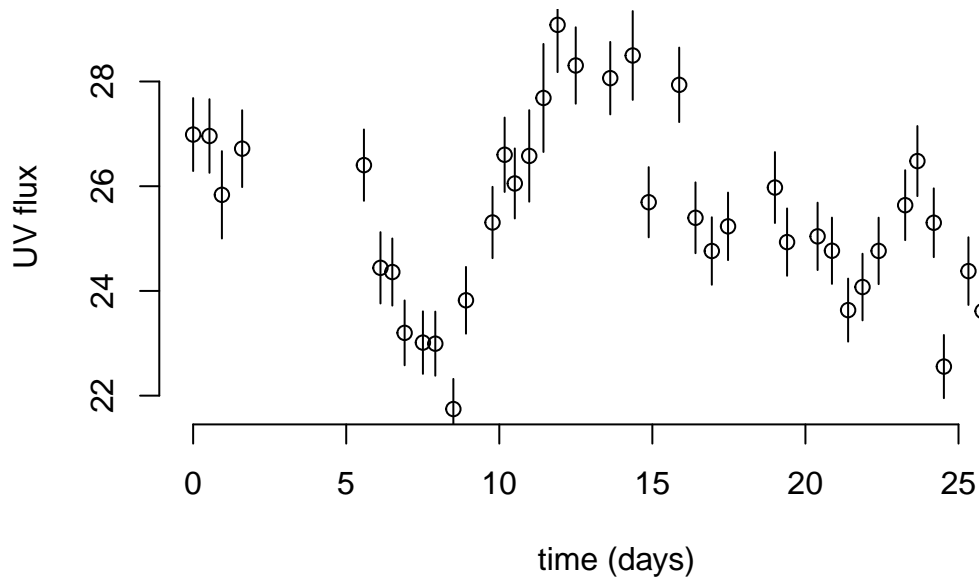
# subtract the start time
dat$t <- dat$t - min(dat$t)

# plot the data
plot(dat$t, dat$y, bty = "n", xlab = "time (days)", ylab = "UV flux")

# add the error bars to the plot
segments(dat$t, dat$y - dat$dy, dat$t, dat$y + dat$dy)
```

---

\*Email: sav2@le.ac.uk



The times are not evenly spaced, and the errors  $dy$  are different for each observations. We should be able to account for both of these properties of the data quite naturally using a Gaussian Process model.

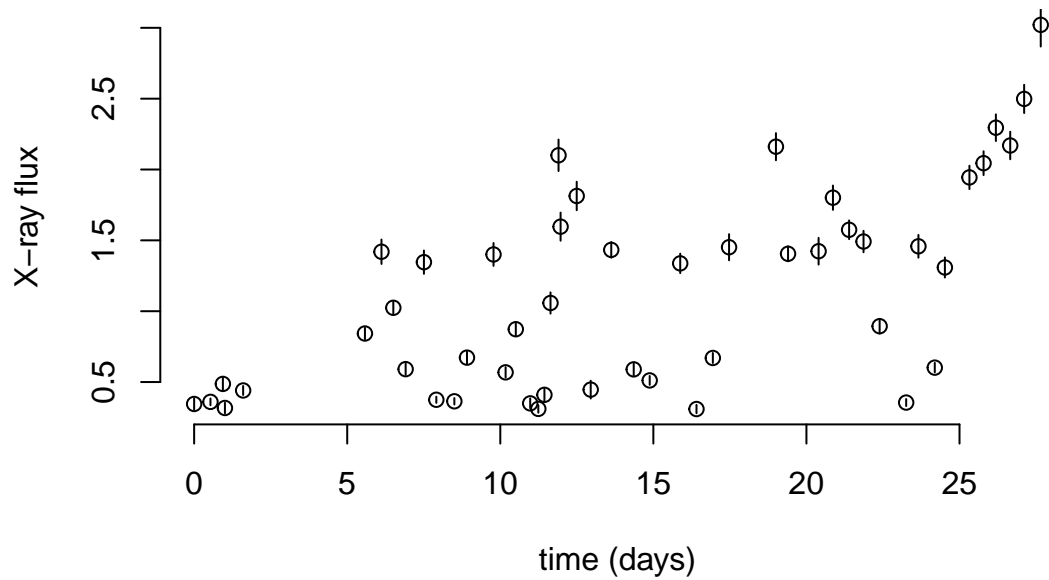
And we can do the same with the X-ray data from the file `2016_xrt_rate.txt` which is in exactly the same format.

```
# load the file into the 'dat' object
dat <- read.table("2016_xrt_rate.txt", header = TRUE)

# subtract the start time
dat$t <- dat$t - min(dat$t)

# plot the data
plot(dat$t, dat$y, bty = "n", xlab = "time (days)", ylab = "X-ray flux")

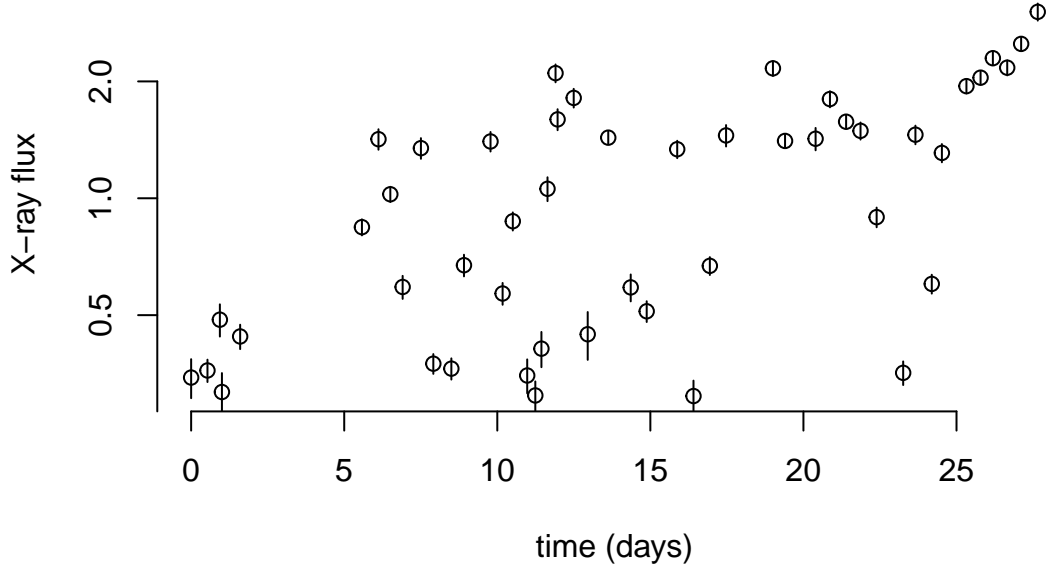
# add the error bars to the plot
segments(dat$t, dat$y - dat$dy, dat$t, dat$y + dat$dy)
```



Notice how much more variable the X-rays are, so we make a logarithmic transformation to the axis and re-plot.

```
# plot the data
plot(dat$t, dat$y, bty = "n", xlab = "time (days)", ylab = "X-ray flux", log = "y")

# add the error bars to the plot
segments(dat$t, dat$y - dat$dy, dat$t, dat$y + dat$dy)
```



Clearly the processes that generates the X-ray and UV variations are quite different; the X-ray variations are much stronger (larger amplitude) and more rapid (less smooth). We would like to be able to infer the properties of the random processes at work here. If we assume these are stationary GPs then the only important characteristic is the auto-covariance function  $ACV(\tau)$ .

If we think the X-rays are produced by a lognormal process (i.e.  $\ln \mathbf{y} \sim N(\mu, \Sigma)$ ), we can treat this by modelling the logarithm of the fluxes (i.e. we apply a simple log transformation to the data, and the corresponding transformation to the errors, prior to fitting with a GP model).

### The errors

A further point is that sometimes we cannot trust the error bars. In the case of the *Swift* data here, the errors are usually quite reliable indicators of the reproducibility of the data. But data taken from other telescopes can be less reliable. To account for this we can allow an additional parameter  $\nu$  that is a multiplicative factor applied to the errors. I.e. the covariance matrix of the measurement ‘noise’ is:

$$\begin{aligned} N &= \nu \sigma^2 I_n, \\ N_{ij} &= \nu \sigma_i^2 \delta_{ij}. \end{aligned} \tag{1}$$

And we should find  $\nu \approx 1$  when the error bars are reliable. The covariance matrix for the GP model is then

$$\Sigma = K + N. \tag{2}$$