QUB MSc Astro Analytathon Summer 2021

Classification of astronomical lightcurves

Prof Stephen Smartt, Dr Dave Young, Dr Ken Smith
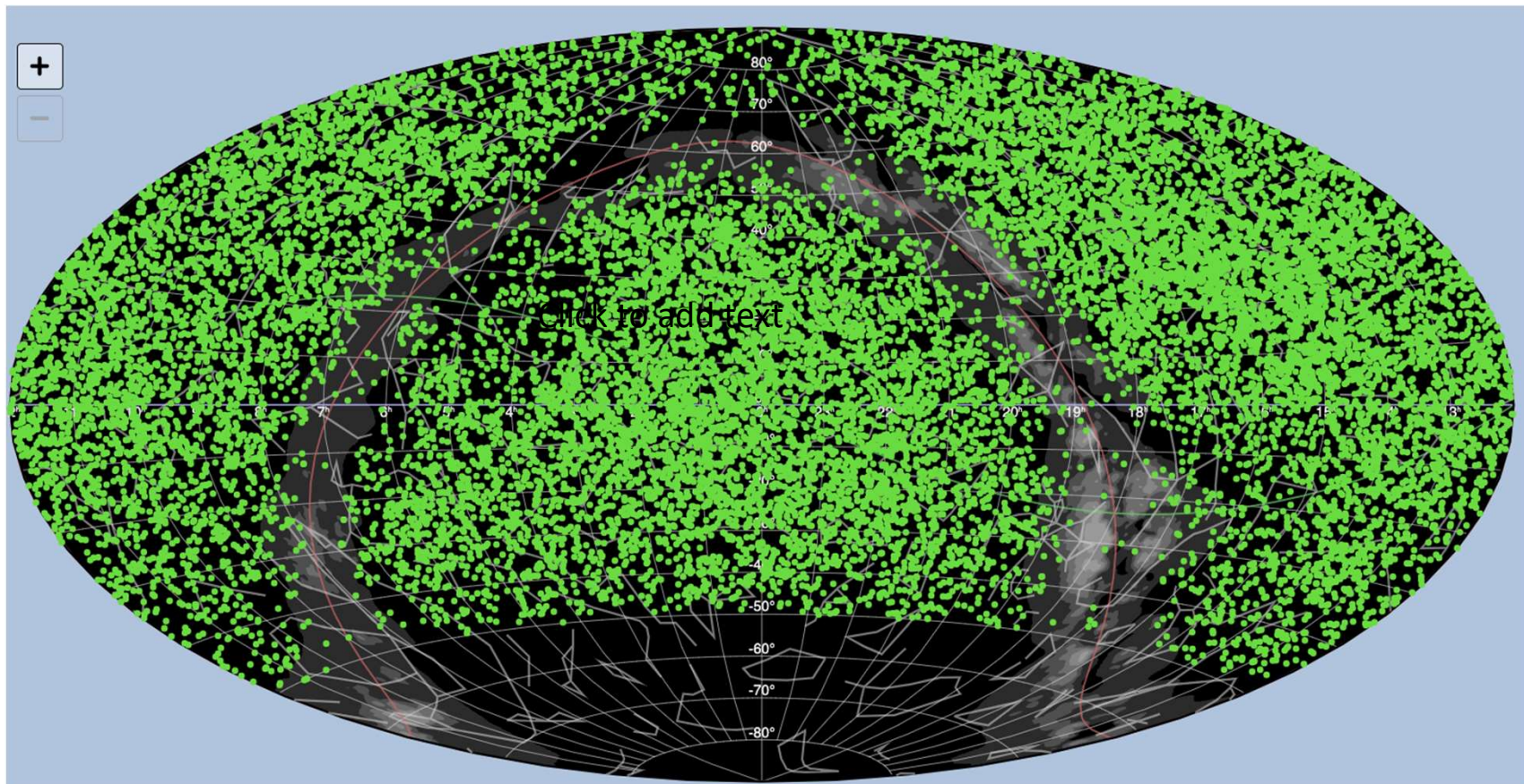
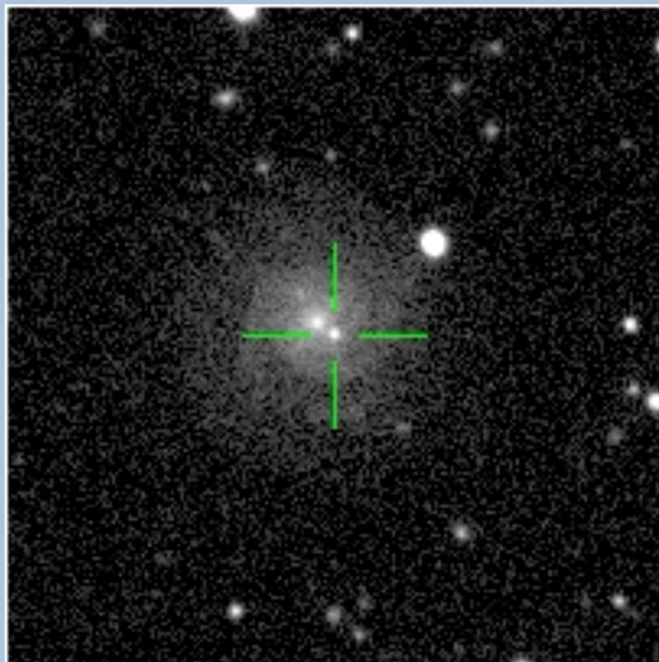# The ATLAS telescopes in Hawaii

O'ahu

IfA, Manoa

PS1 + PS2
Haleakala
Maui

ATLAS1 :
Haleakala
3000m

ATLAS home page :
https://fallingstar.com

Big Island

ATLAS2 :
Mauna Loa
3400m

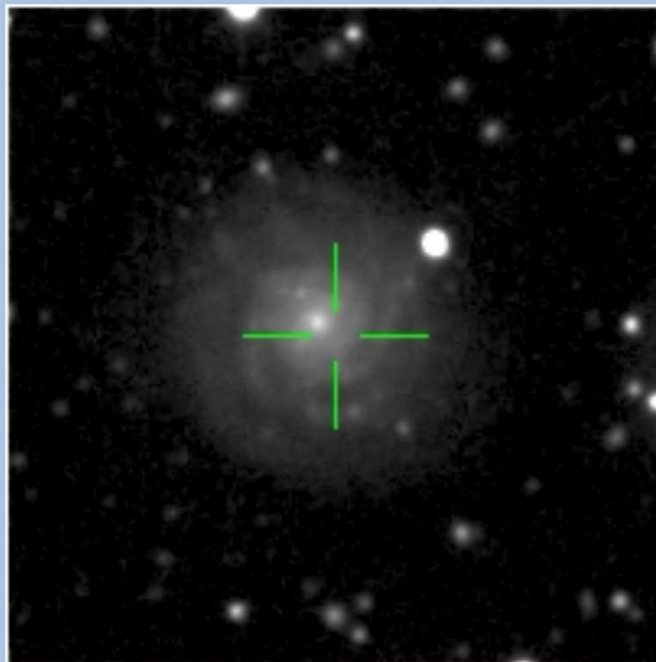# Hawaii - Ireland : perfect longitude match !

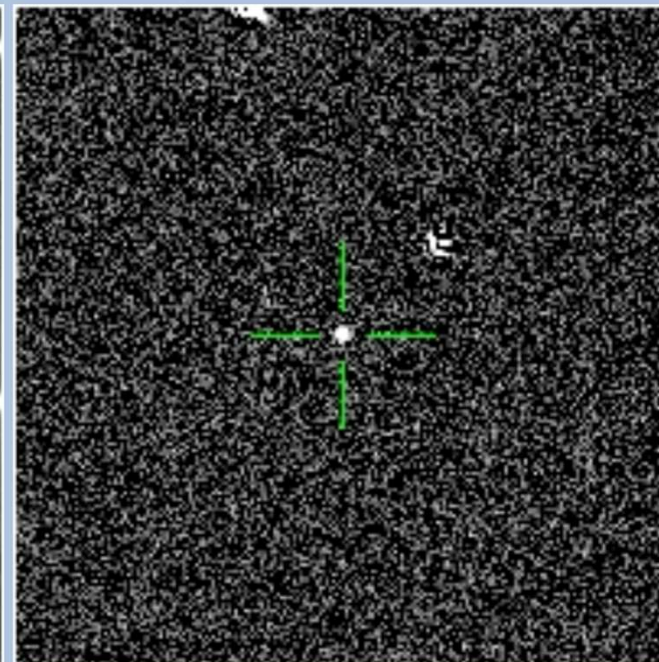# Supernovae and exploding stars – flashes in the sky
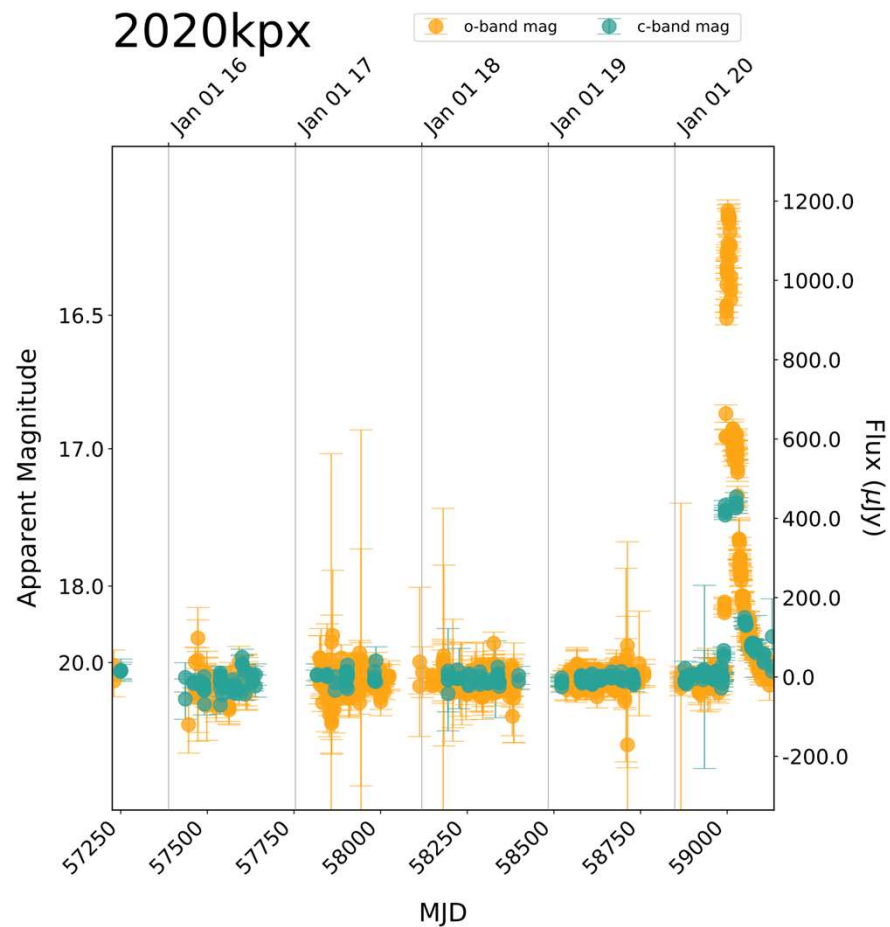
target (c) mjd: 59403.27687   ref (c) mjd: 59403.27687   diff (c) mjd: 59403.27687

Time series data = a lightcurve

X axis : modified Julian date. This is in units of days (decimal)

Y axis : Flux, which is a measurement of power, or energy per second received on earth

The flux unit is microJanskys
1 Jansky = 10^{-23} Watts m^{-2} Hz^{-1}
(where m is a metre on the earth)

1 microJansky = 10^{-6} Jansky

Orange or o-band : light measured through a filter at 690 nm
Cyan or c-band : light measured through a filter at 535 nm

The dataset – overview

https://github.com/thespacedoctor/astro-analytathon

- 645 objects, each with a lightcurve
- Lightcurve contains measurements over about 5 years
- Most points will be at zero flux (within the errors)
- Different shapes, timescales, peak values : star exploded sometime over last 5 years
- Supernovae can be more luminous than 10^{11} Suns, or brighter than a whole galaxy
- They can last somewhere between 1 – 6 months
- ATLAS is effectively a large camera : observes 4 exposures (each 30 seconds) at each position each night
- Within the night – not much variation

Your challenge

We think we understand the physics of most supernovae – many are very similar lightcurve shapes.
We have matched many of these lightcurves to physical models and also to each other
But we bring our own bias to this.

- Treat the dataset purely from a data analytics perspective – no physics !
- Group the objects into clusters or classes by their shape – some will be incredibly similar, implying the same physical mechanism causes the energy release
- How many classes can you find ?
- How many in each class ?
- How many can't be classified comfortably into the main classes ?
- Why can they not ?
- How many are true anomalies – in that they don't fit in any class ?
- How may are too noisy to tell ?

Pay close attention to errors and missing data ! See next slide ….

The dataset - details

```
 1   ###MJD           m      dm    uJy   duJy F err chi/N    RA       Dec        x         y     maj  min   phi  apfit mag5sig Sky    Obs
 2   57248.473434 --18.658  0.737  -125   92 c  0   0.94   0.95507  7.46675  2526.19  5286.52 3.85 3.53 -58.7 -0.429 19.54 21.62 02a57248o0308c
 3   57248.491712 --19.084  1.333   -84  113 c  0   1.16   0.95507  7.46675  2515.93  5290.44 3.86 3.58 -61.4 -0.462 19.52 21.63 02a57248o0331c
 4   57248.524200 --18.044  1.632  -220  359 c  0   0.98   0.95507  7.46675  3977.22  3853.77 3.85 3.69 -55.8 -0.758 19.48 21.59 02a57248o0370c
 5   57248.550040  18.657  6.364   125  796 c  0   1.10   0.95507  7.46675  6607.15  6745.05 3.85 3.62 -88.9 -0.620 18.83 20.85 02a57248o0404c
 6   57313.373363 --19.717  0.411   -47   19 c  0   0.87   0.95507  7.46675  8086.04  5315.77 4.19 3.89  82.8 -0.445 19.39 21.32 02a57313o0175c
 7   57313.396284 --21.068  1.500   -14   20 c  0   0.94   0.95507  7.46675  8078.49  5319.98 4.20 3.90 -89.8 -0.448 19.40 21.34 02a57313o0207c
 8   57313.427114 --19.513  0.286   -57   16 c  0   1.07   0.95507  7.46675  6610.53  6764.57 4.19 3.73 -79.8 -0.445 19.37 21.31 02a57313o0250c
 9   57313.458158 --21.480  2.182    -9   20 c  0   0.79   0.95507  7.46675  6608.58  6766.60 4.26 3.78 -75.1 -0.445 19.35 21.31 02a57313o0294c
10   57318.262930 --18.883  0.452  -102   46 o  0   0.73   0.95507  7.46675  2524.29  5306.71 4.14 3.35 -55.4 -0.475 18.33 19.49 02a57318o0107o
11   57318.284315 --18.383  0.239  -161   38 o  0   1.24   0.95507  7.46675  2522.76  5309.72 4.31 3.34 -58.3 -0.491 18.42 19.63 02a57318o0140o
```

Example content of object lightcurve file

Not all columns are useful for time-series analysis; but here are the definitions for those that are:

Column	Definition
MJD	Modified Julian Date. A datetime-format preferred by astronomers, with units of days.
uJy	The object's flux (brightness) as measured from a single telescope image.
duJy	The error associated with the measured flux above.
F	Colour-filter used during the telescope observation. o orange, c cyan.
chi/N	A *possible* indication of the quality of the flux measurement (unrelated to duJy , lower is better)

Data issues to look out for

- Uncertainty and noise – each flux point comes with an associated error. This is a noisy data set, pay close attention to the error bars
- Season gaps. Sometimes the lightcurve of a transient event is truncated because it either disappears behind the sun, or appears from behind the sun.
- Sudden discontinuities in the lightcurve. This probably implies that the "template" image was changed while the transient event was active (and maybe contains flux from the event).
- Random single point high flux or low flux events. This is caused by satellite trails, moon glints, bright star spikes, clouds, etc.
- A transient event might be very red or very blue, hence very different flux in each filter. You might want to treat the filters separately. Note that historically, the cyan filter is much more sparsely sampled.

Example – discontinuity
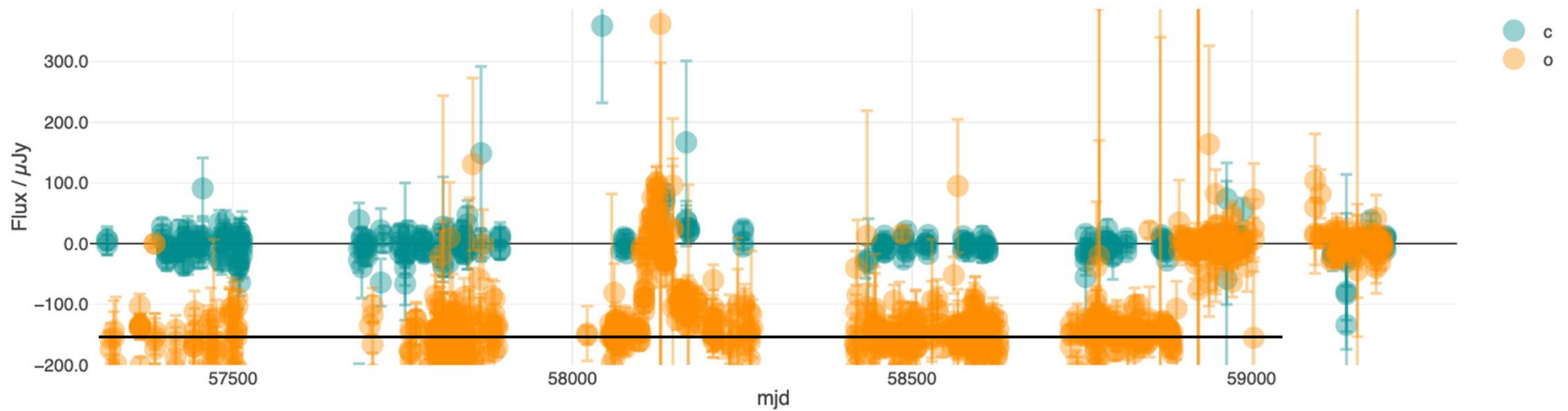
Simple solution : add constant to all points to bring the baseline level to Flux = 0

Hints and tips

- Many of these lightcurves can reasonably well reproduced with simple analytic functions of the form :

  $f = \alpha\ t\ exp\ [-(t-t_0)/\beta]$

  Where t is time in days, t0 is a time constant to shift the start time of the function, β sets the timescale or width of the rise and fall, and α is just an overall scaling factor.

- But not all, some will have rise and long flat phase of 80 – 100 days

- There are more complex functions that can work better – up to you to decide

- All objects are at different distances (a factor 10 difference between the nearest and furthest), which means the peak flux could differ by a factor 100

- You could choose to normalise the peak flux of all of them to 1, for example. But then you lose the information of how intrinsically bright they are

- We give you the estimated distance to each object in a file on github. If you wanted to remove distance, then pay attention to the units of flux. Since a Jansky is measured in Watts m^2 Hz^{-1}, and we define intrinsic luminosity as that emitted by the source :

  $L = 4\pi r^2\ f$

  then to get L in Watts Hz^{-1}, (thus removing the distance dependence) simply multiply f by $4\pi r^2$ where r is in metres

  1 parsec = 3.086 x 10^{16} metres

  1 megaparsec = 3.086 x 10^{22} metres