Time series analysis in astronomy

Simon Vaughan
University of Leicester

What do I hope to achieve?

For astronomers to explore more approaches to time series analysis

For time series experts to start exploring astronomical data

But the scope of this talk is rather limited…

berat: Iuppiter à ſequenti occidua min. 5. hæc ve-
rò à reliqua occidentaliori min. 3. erant omnes eiuſ-

Ori. * O * * Occ.

dem proximè magnitudinis, ſatis conſpicuæ, & in
eadem recta linea exquiſitè ſecundum Zodiaci du-
ctum.

Die decimaſeptima H.1. duæ aderant Stellæ, orien-
talis vna à Ioue diſtans min.3. occidentalis altera diſtas

Ori. * O * Occ.

min. 10. hæc erat aliquanto minor orientali. Sed hora
6. orientalis proximior erat Ioui diſtabat nempè mi e.
ſec.50.occidentalis verò remotior fuit, ſcilicet min. 12.
Fuerunt in vtraque obſeruatione in eadem recta,& am-
bæ ſatis exiguæ,præſertim orientalis in ſecunda obſerua
tione.

Die 18. Ho.1.tres aderant Stellæ, quarum duæ occi-
dentales, orientalis verò vna: diſtabat orientalis à Ioue

Ori. * . O * * Occ.

min.3. Occidentalis proxima m.2.occidentalior reliqua
aberat à media m.8. Omnes fuerunt in eadem recta ad
vnguem,& eiuſdem ferè magnitudinis. At Hora 2.Stel-
læ viciniores paribus à Ioue aberant interſtitijs:occidua
enim aberat ipſa quoque m.3. Sed Hora 6.quarta Stel-
lula viſa eſt inter orientaliorem & Iouem in tali configu
ratione. Orientalior diſtabat à ſequenti m.3.ſequens à

Ioue m.1.ſec.50. Iuppiter ab occidentali ſequenti m. 3.

Ori. * * O * * Occ.

hæc verò ab occidentaliori m.7.erāt ferè æquales,orien
talis tantum Ioui proxima reliquis erat paulo minor,
erantque in eadem recta Eclypticæ parallela.

Die 19.Ho.0.m.40.Stellæ duæ ſolummodo occiduę
à Ioue conſpectæ fuerunt ſatis magnæ, & in eadem re-

Ori. O * * Occ.

cta cum Ioue ad vnguem, ac ſecundum Eclypticæ ductū
diſpoſitæ. Propinquior à Ioue diſtabat m.7. hæc verò
ab occidentaliori m.6.

Die 20. Nubiloſum fuit cœlum.

Die 21. Ho.1.m.30. Stellulæ tres ſatis exiguæ cerne-
bantur in hac conſtitutione. Orientalis aberat à Ioue

Ori. * O * * Occ.

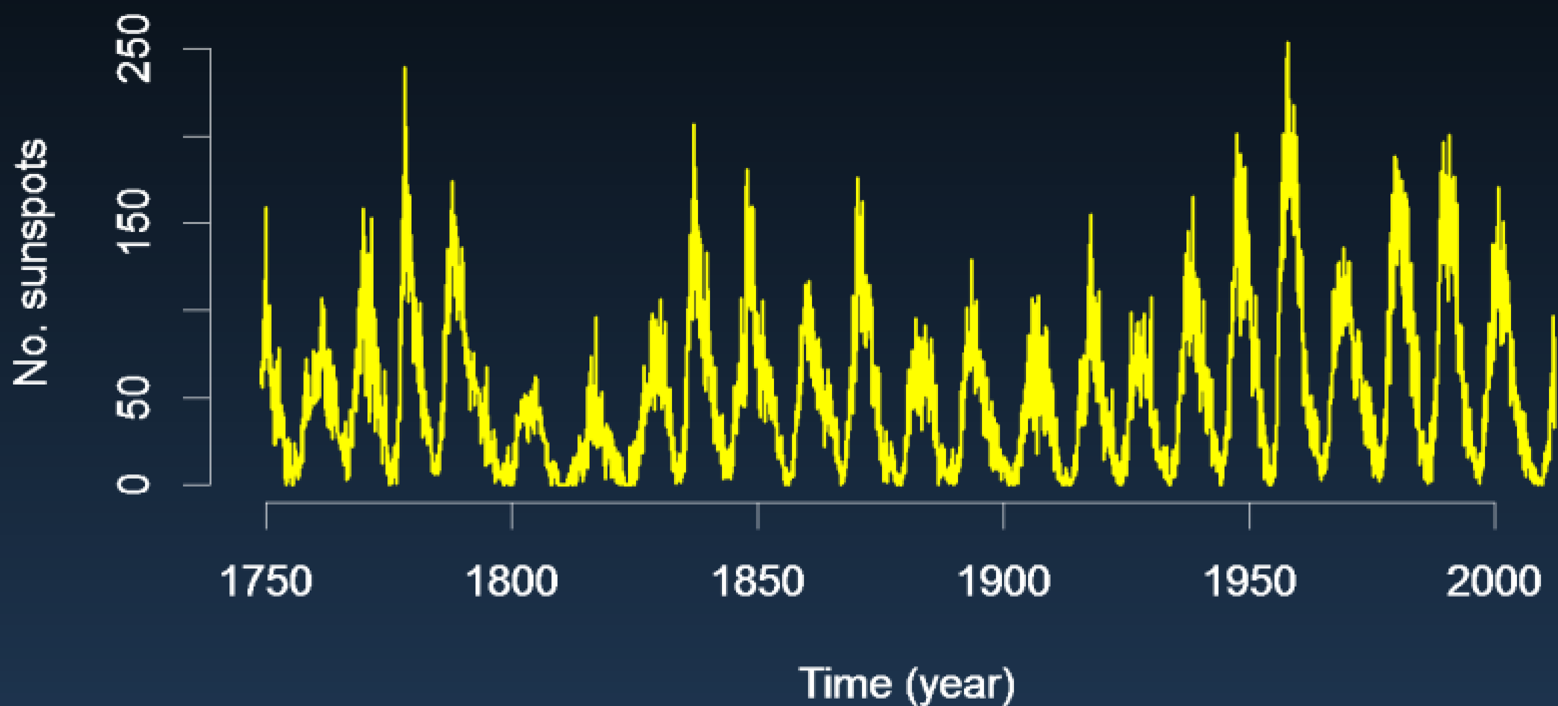
m.2. Iuppiter ab occidentali ſequente.m.3.hæc verò ab
occidentaliori m.7. erant ad vnguē in eadem recta Ecly-
pticæ parallela.

Die 25.Ho.1.m.30.(nam ſuperioribus tribus nocti-
bus cœlū fuit nubibus obductum) tres apparuerūt Stel

Ori. * * O * Occ.

læ.Orientales duæ, quarum diſtantiæ inter ſe, & à Ioue

A classic



No. sunspots
0
50
150
250
1750
1800
1850
1900
1950
2000
Time (year)

Time series processes: periodic

orbits – planets, binary stars, comets, …

rotation/cycles – solar cycle, pulsars, Cepheids, …

identification of new periods →

orbits, rotation, system identification

estimate parameters:

period

amplitude

waveform,

(small) perturbations in period

[see Phil Gregory's talk]

Time series processes: transient

supernovae

stellar “activity”

novae

gamma-ray bursts

detection, identification/classification,

automatic triggers

estimate parameters:

- duration, fluence
- profile shape
- multi-λ comparison

[see RS discussion meeting in April]

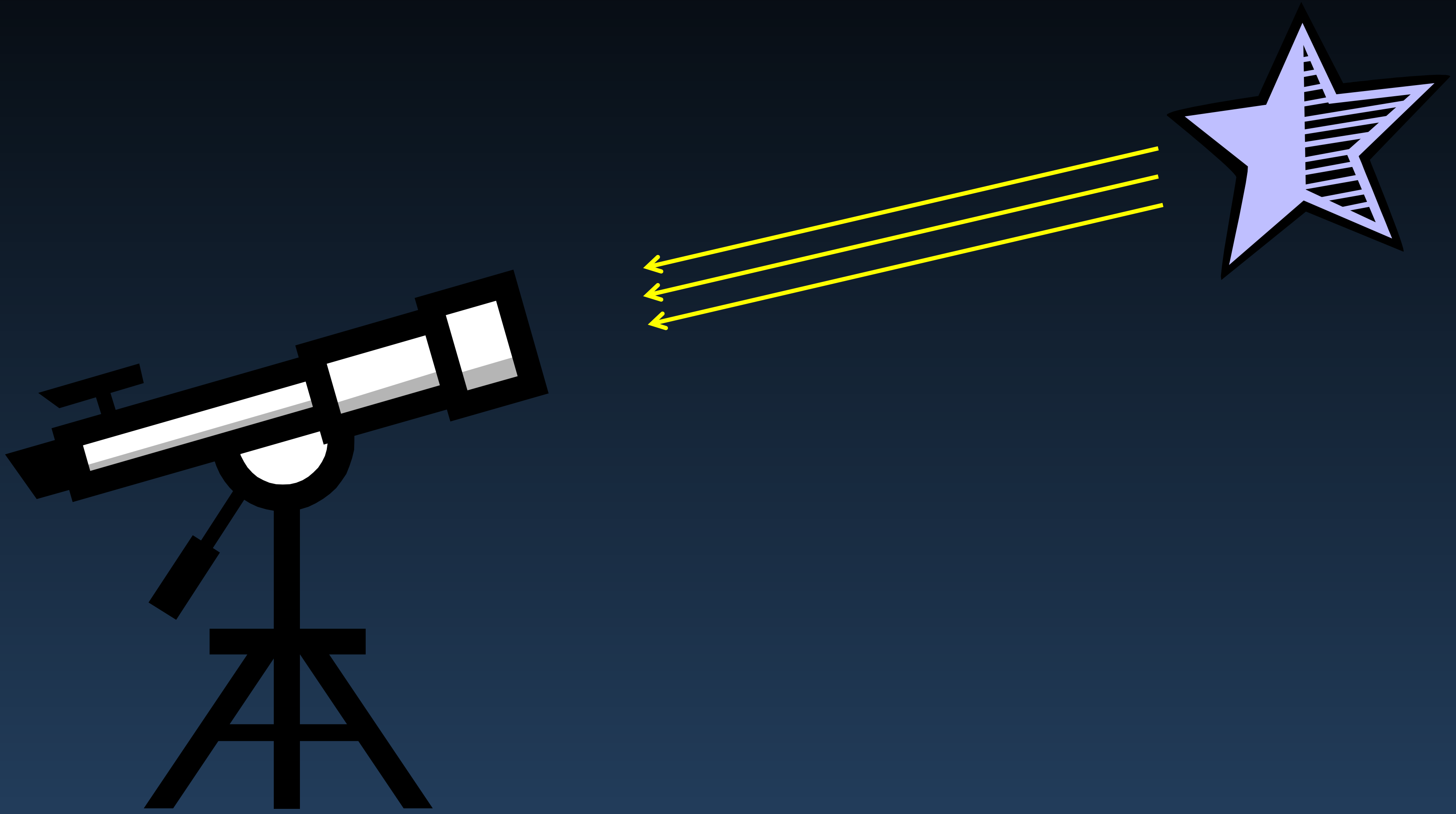
Time series processes: stochastic

accreting systems (neutron stars, black holes), jets

cannot predict (time series) data exactly

statistical comparison between data and model
to infer physics of system

Data in astronomy



Data in astronomy

Astronomy is observational (not experimental)

Limited to light received (but see Neil Cornish's talk later)

images: position on the sky (2 dimensions)

wavelength: "spectrum" (photon energy)

polarization

time "light curves"

combination of the above

Time series in astronomy

Gapped/irregular data

- Diurnal & monthly cycles
- satellite orbital cycles
- telescope allocations

measurement errors & heteroscedastic

- Signal-to-noise ratio

Poisson processes

- Individual (photon) events in X/γ-ray astronomy

Time series – what the books say

trends and seasonal components

autoregressive and moving average processes (ARMA)

forecasting

Kalman filters

state-space models

auto/cross-correlation

spectral methods

…and so on

mostly for even sampling, no (or Normal) “errors”

Astronomers more interested in *modelling* than *forecasting*

stochastically variable objects

interesting: neutron stars, black holes, turbulent inflows/outflows, etc.

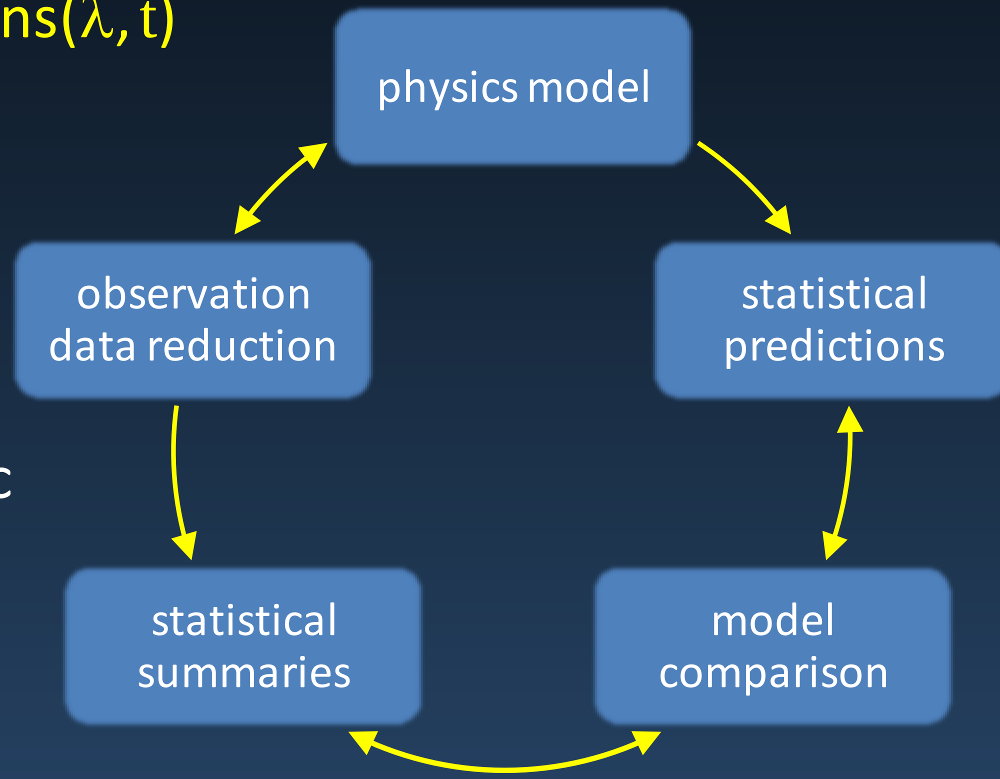
too small to image ($\leq \mu$arcsec)

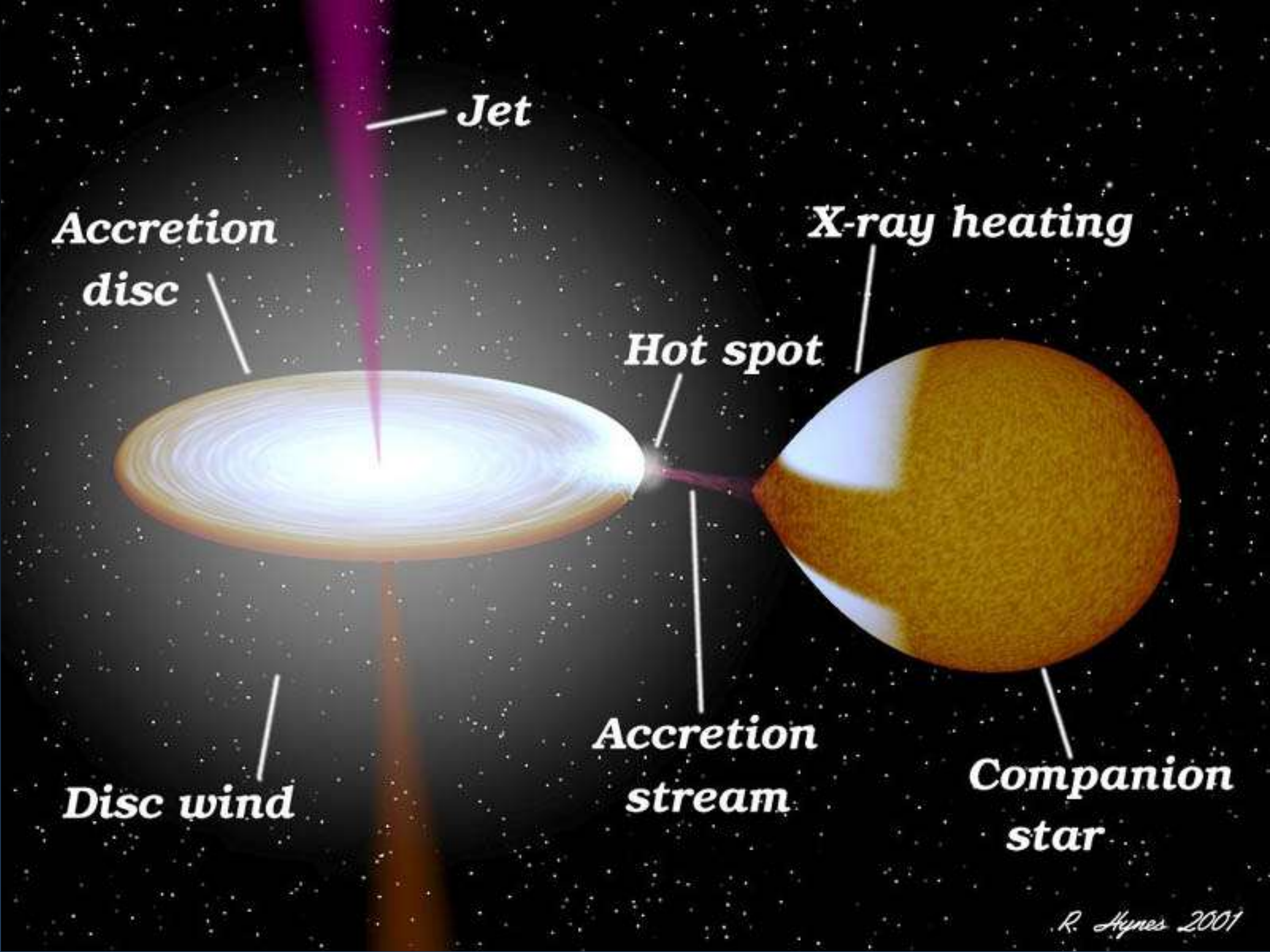
difficult to measure or interpret polarization

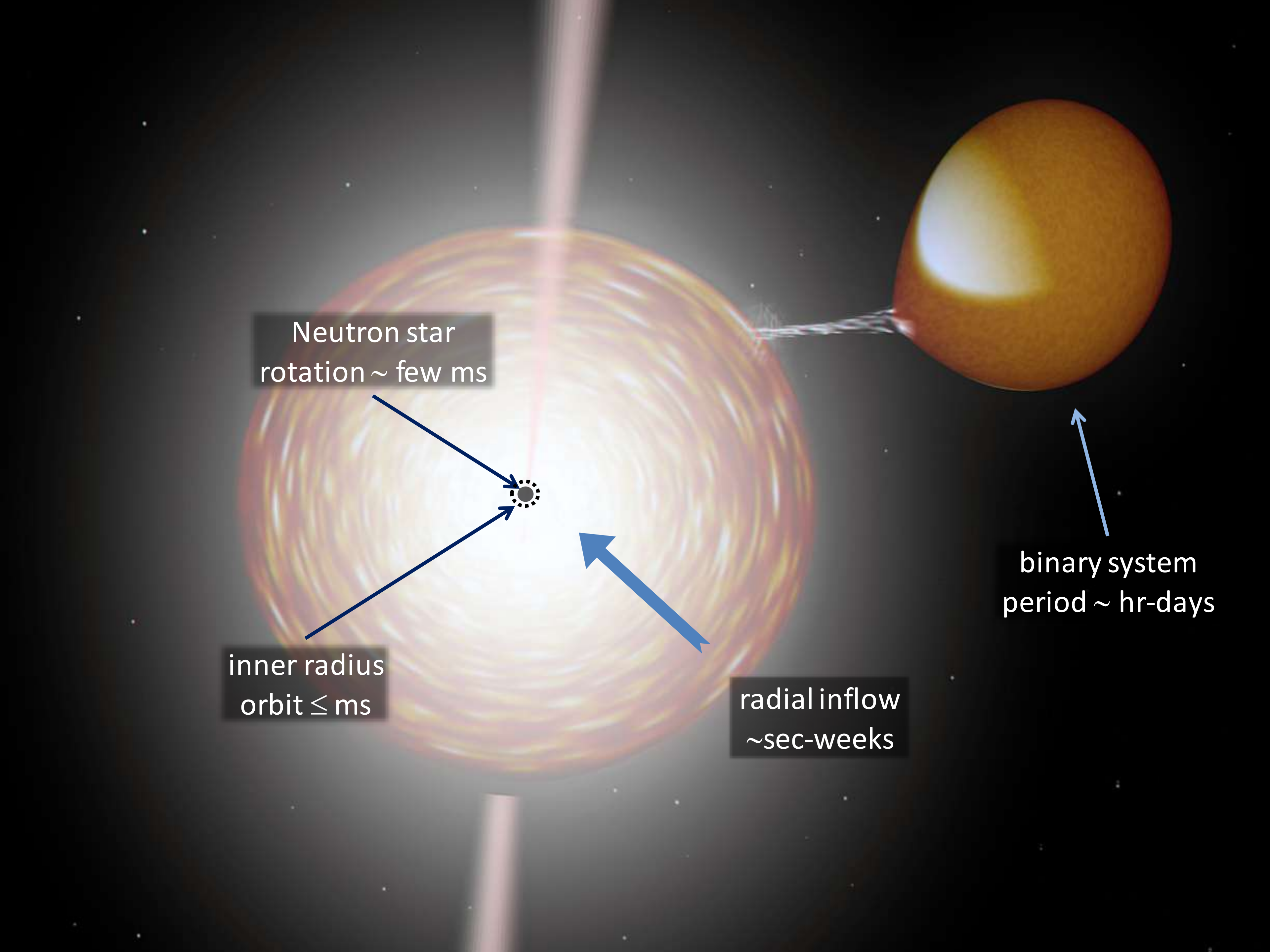
only information from photons(λ, t)

λ-modelling deterministic

t-modelling mixed/stochastic



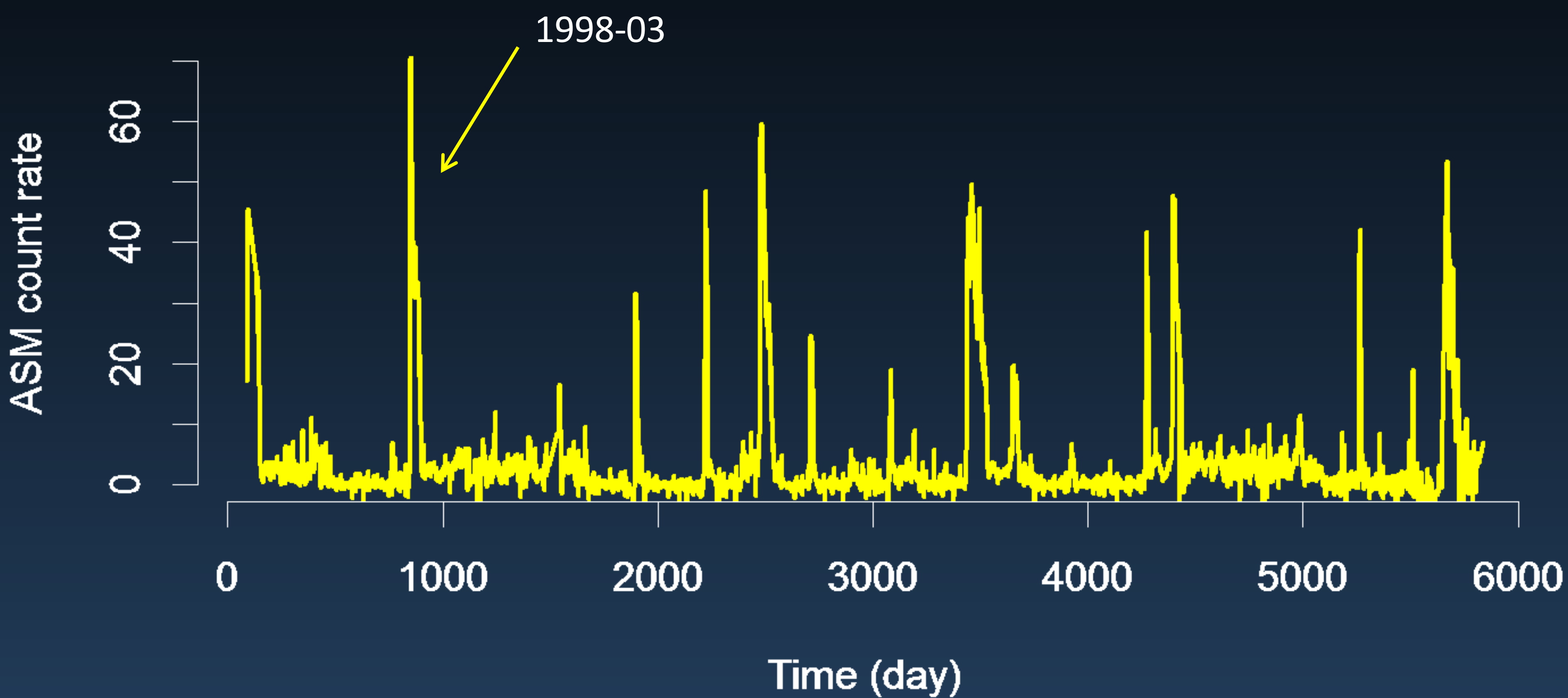

Jet
Accretion disc
X-ray heating
Hot spot
Accretion stream
Companion star
Disc wind
R. Hynes 2001


Neutron star
rotation ~ few ms
inner radius
orbit ≤ ms
radial inflow
~sec-weeks
binary system
period ~ hr-days

data from *RXTE* (1996-2012)

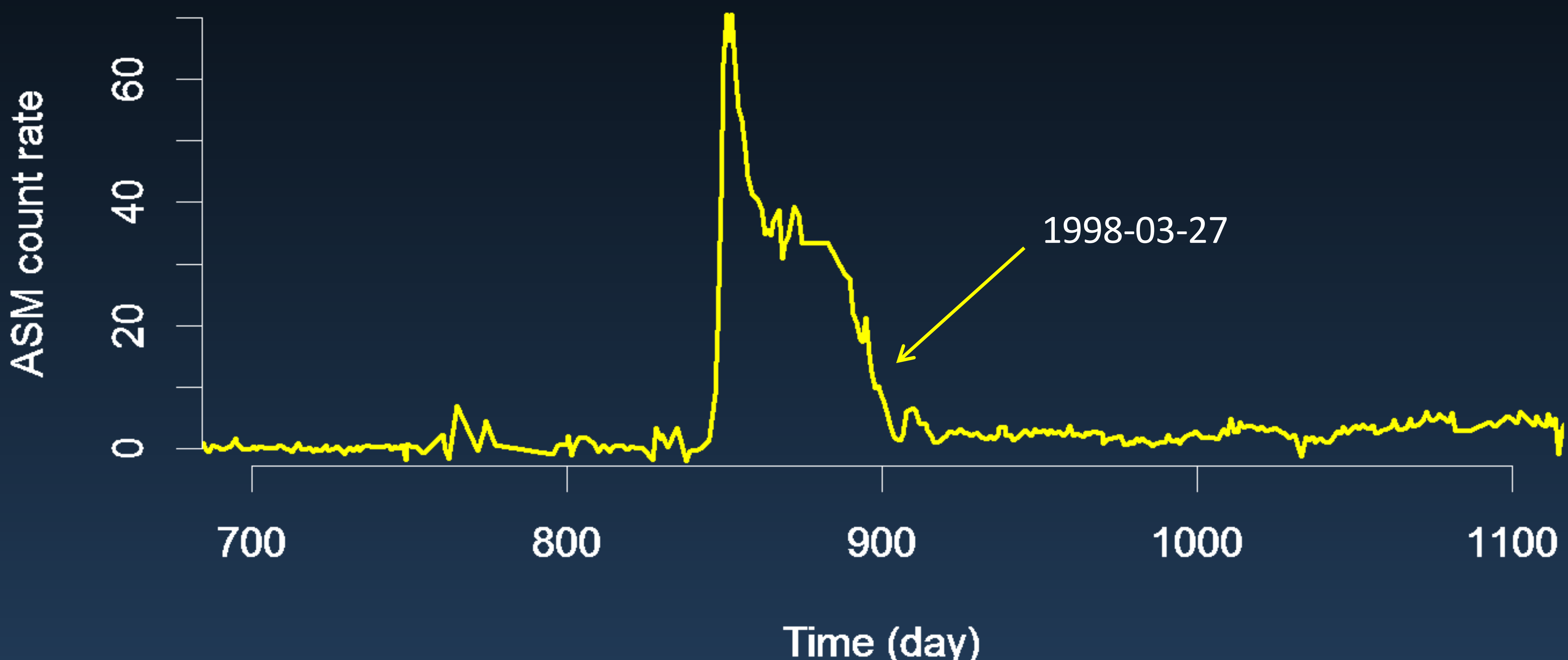
Neutron star X-ray binary 4U1608-52

~daily monitoring with All-Sky Monitor (ASM)



Neutron star X-ray binary 4U1608-52

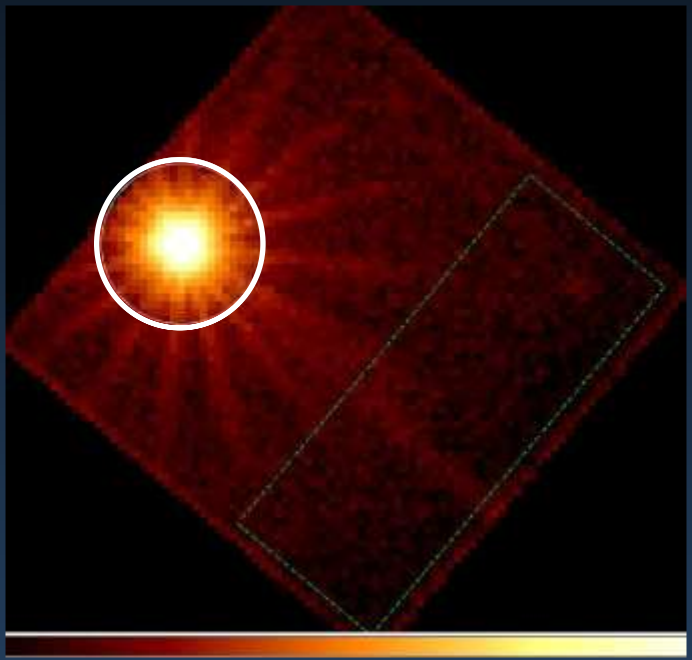
~daily monitoring with All-Sky Monitor (ASM)



(X-ray) event list

Table listing each "event" (mostly genuine X-rays) and their recorded properties e.g. time (detector frame/cycle), channel (event energy), X/Y position, …

time-t.0	chan	X	Y
0.0250	1441	509	517
0.0500	4932	510	491
0.0875	891	488	507
0.1250	5113	494	518
0.1375	6985	510	505
0.1625	2786	488	476
0.2125	2994	504	485
0.2500	6566	504	514
0.2625	2066	460	513
0.2750	1225	492	496
…	…	…	…



(X-ray) data products

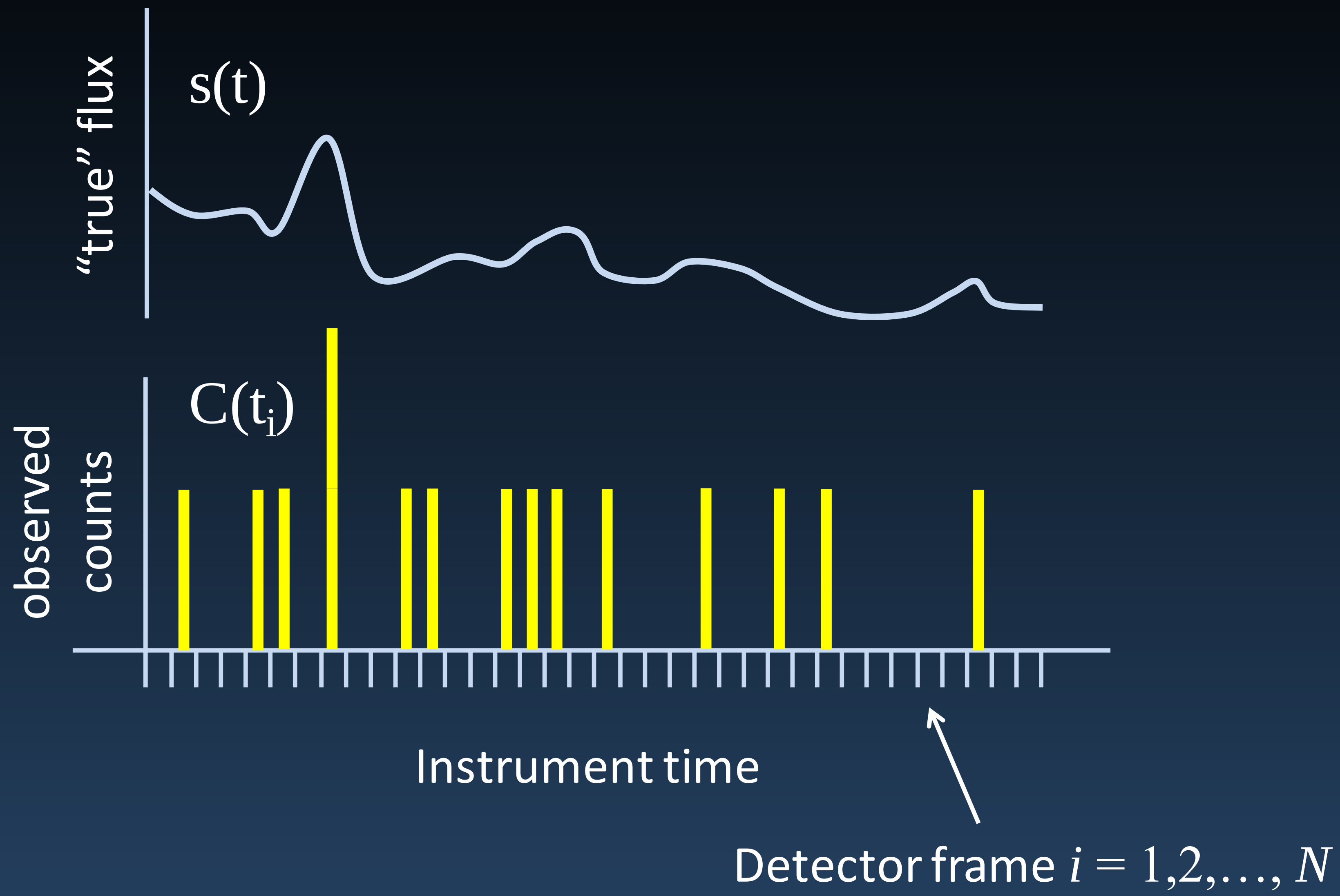
Select using these

time-t.0	chan	X	Y
0.0250	1441	509	517
0.0500	4932	510	491
0.0875	891	488	507
0.1250	5113	494	518
0.1375	6985	510	505
0.1625	2786	488	476
0.2125	2994	504	485
0.2500	6566	504	514
0.2625	2066	460	513
0.2750	1225	492	496
...	...	...	...

Retain these

one, or more, time series

Light curve (time series) product



Light curve (time series) product

observed counts/bin

expected counts
source / background

$$x[i] \sim Pois(s_i) + Pois(b_i)$$

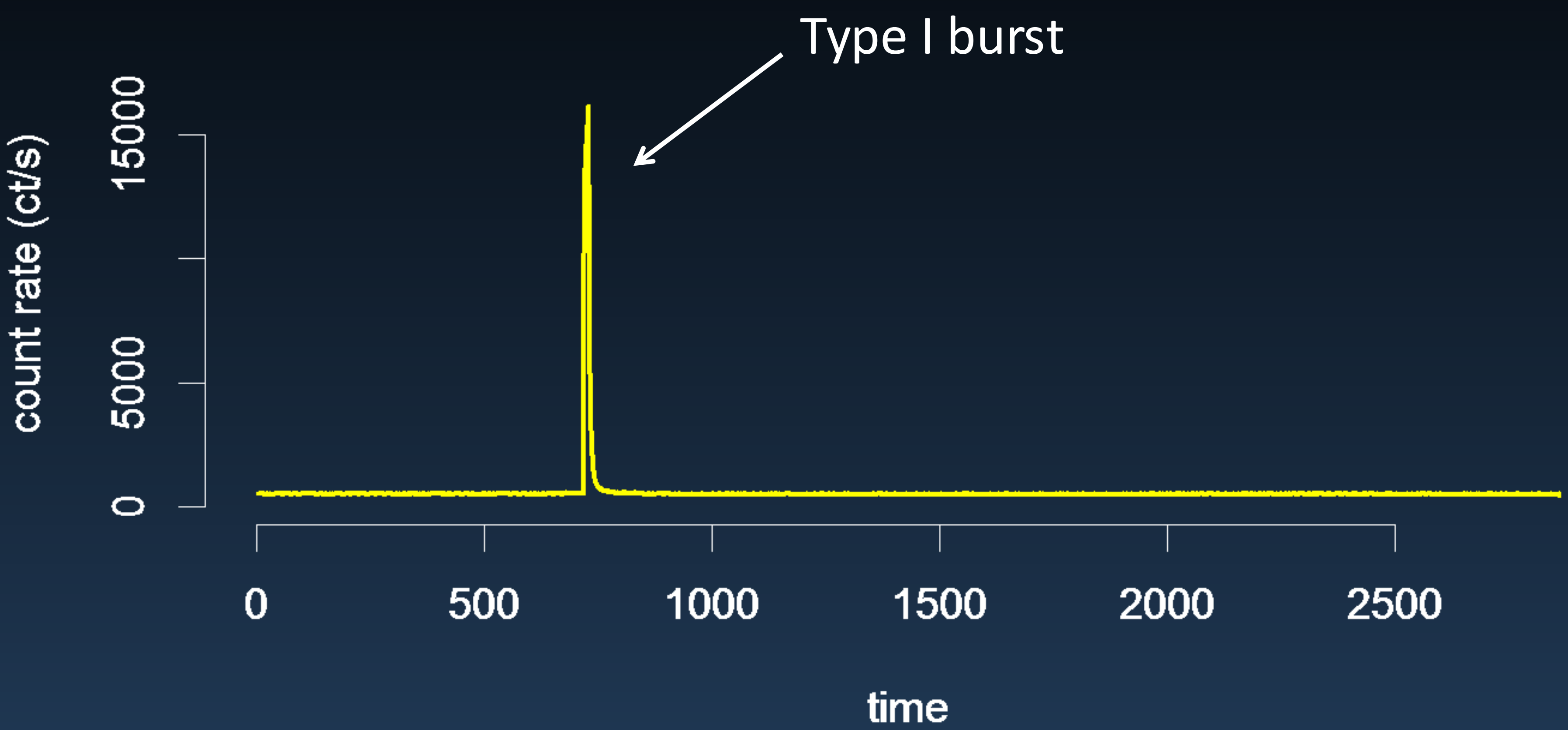
$$s_i = A \int_{t_i}^{t_i + \Delta t} s(t)\,dt$$

sensitivity

true source flux

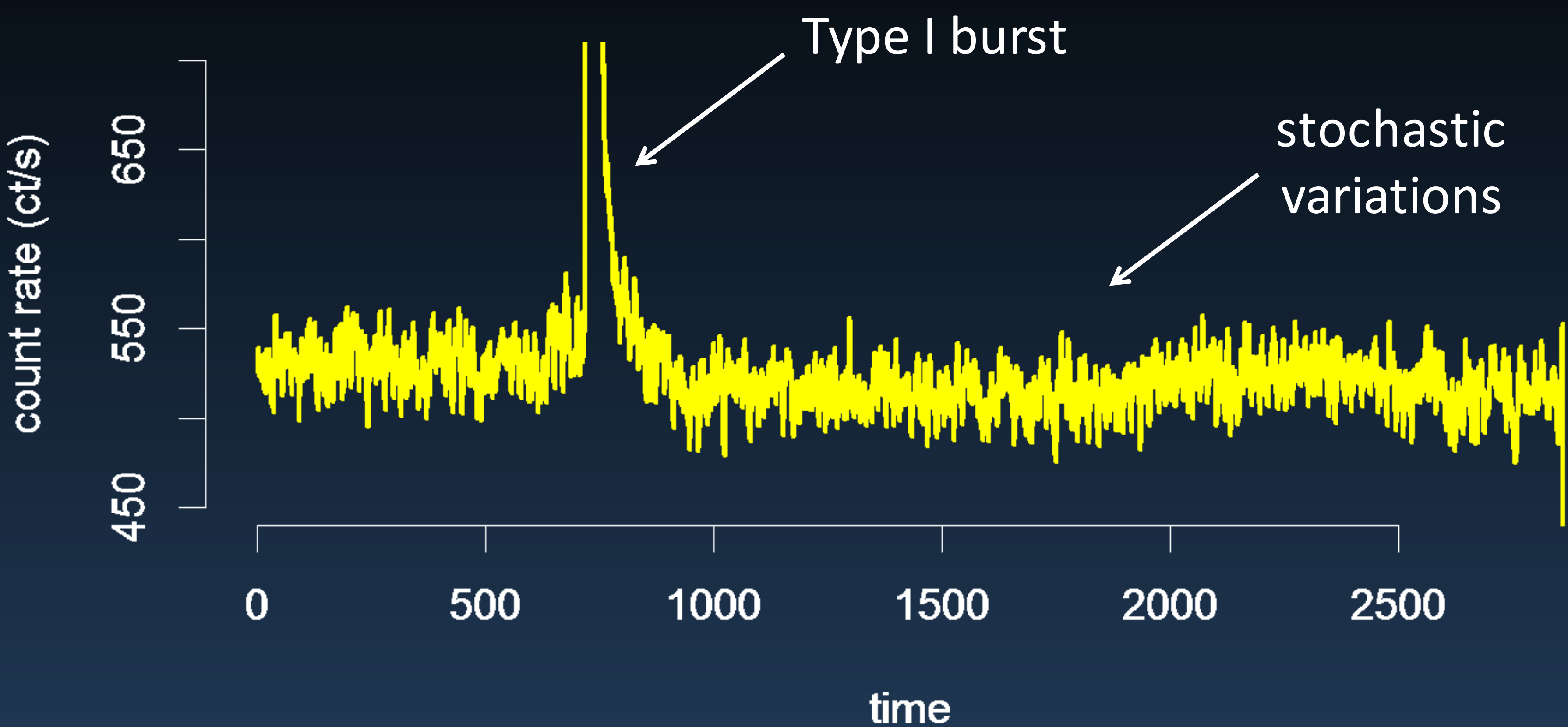
ignoring possible instrumental non-linearities, etc.

X-ray time series of a transient



Data taken with Δt = 125μs resolution, shown binned to Δt = 1s

X-ray time series of a transient



Data taken with Δt = 125μs resolution, shown binned to Δt = 1s

Standard recipe:
Power spectrum analysis

observed = signal + noise
(not quite right!)

$$x = s + n$$

Fourier transforms

$$X = S + N$$

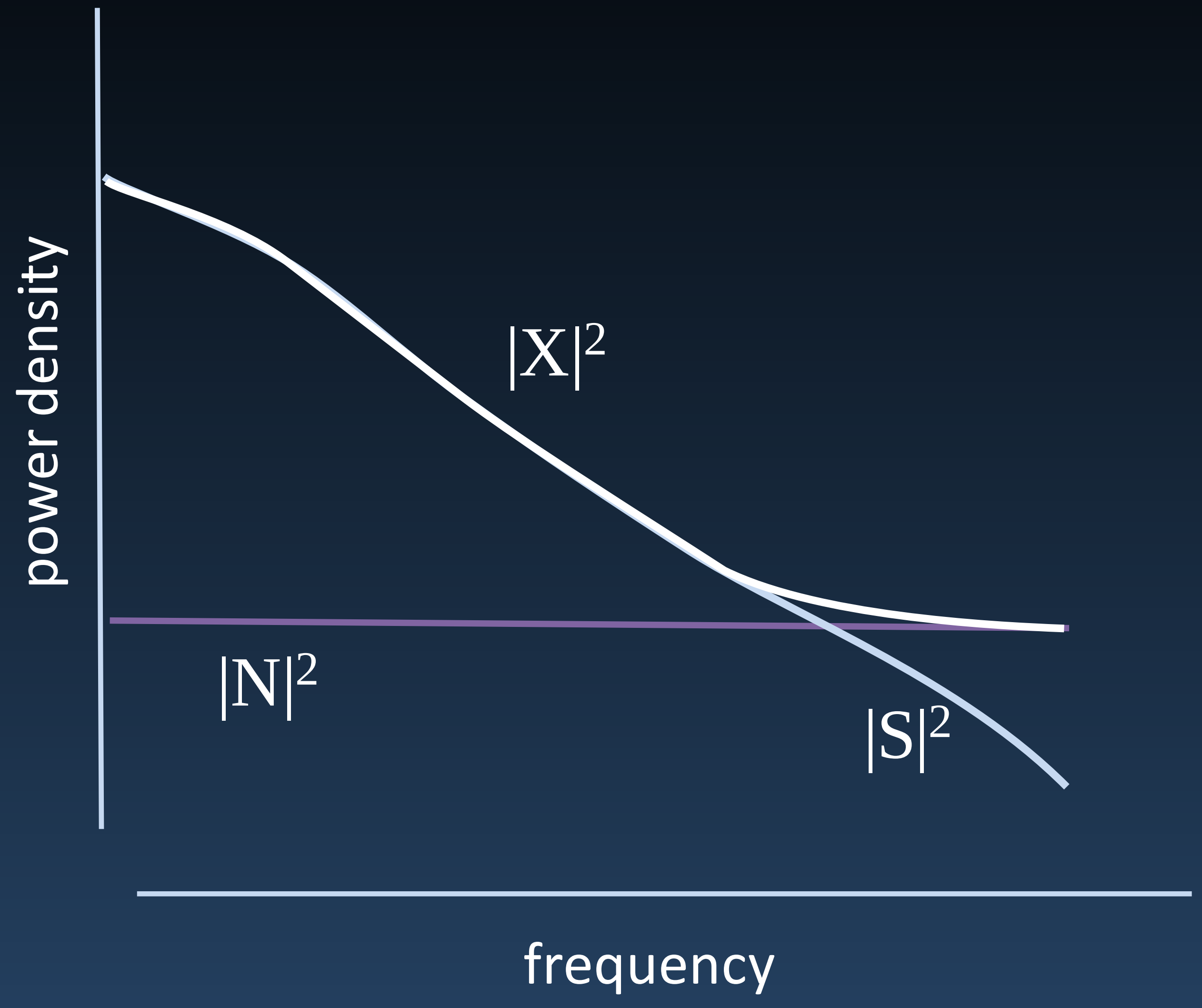
Periodogram

$$|X|^2 = |S|^2 + |N|^2 + \text{cross} - \text{terms}$$

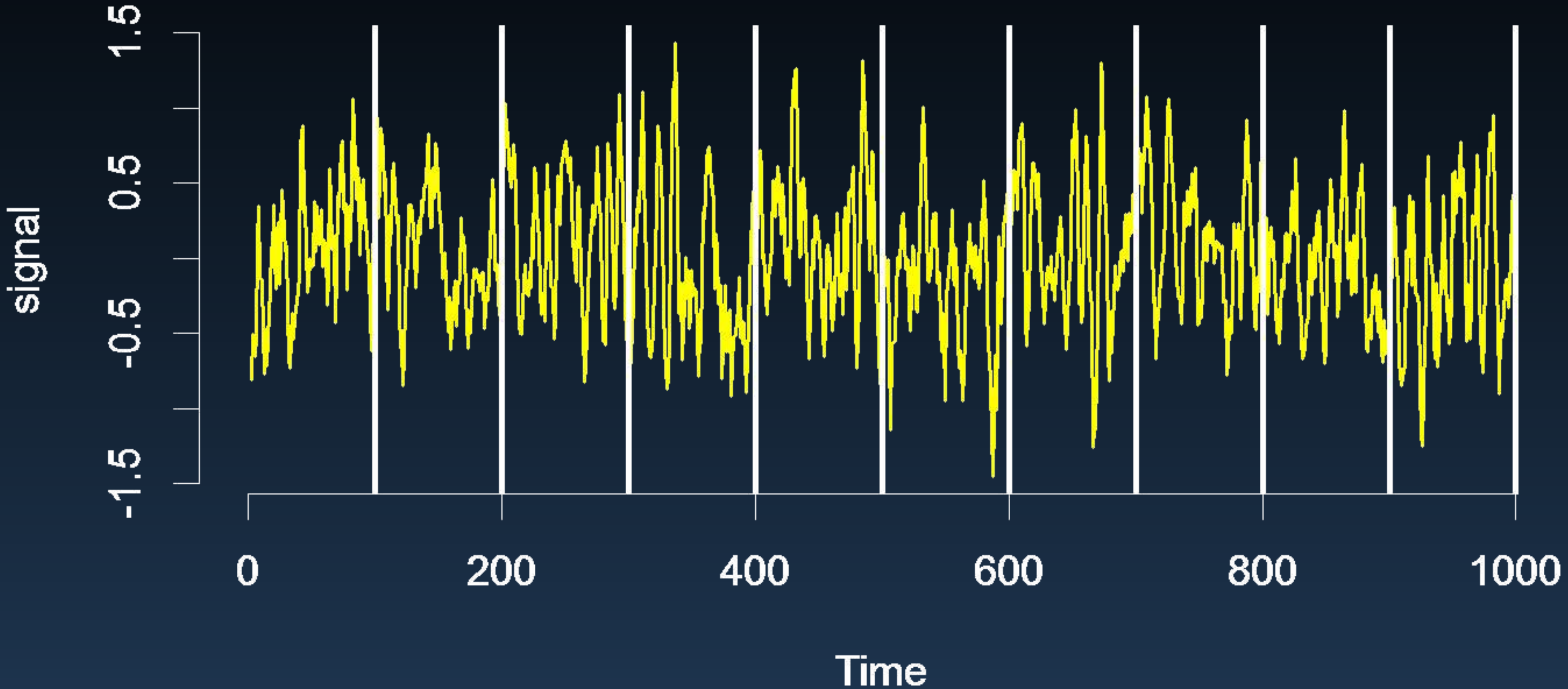
Spectrum

$$P(f) = \left\langle |S|^2 \right\rangle = \left\langle |X|^2 \right\rangle - \left\langle |N|^2 \right\rangle$$

Standard recipe:
Power spectrum analysis



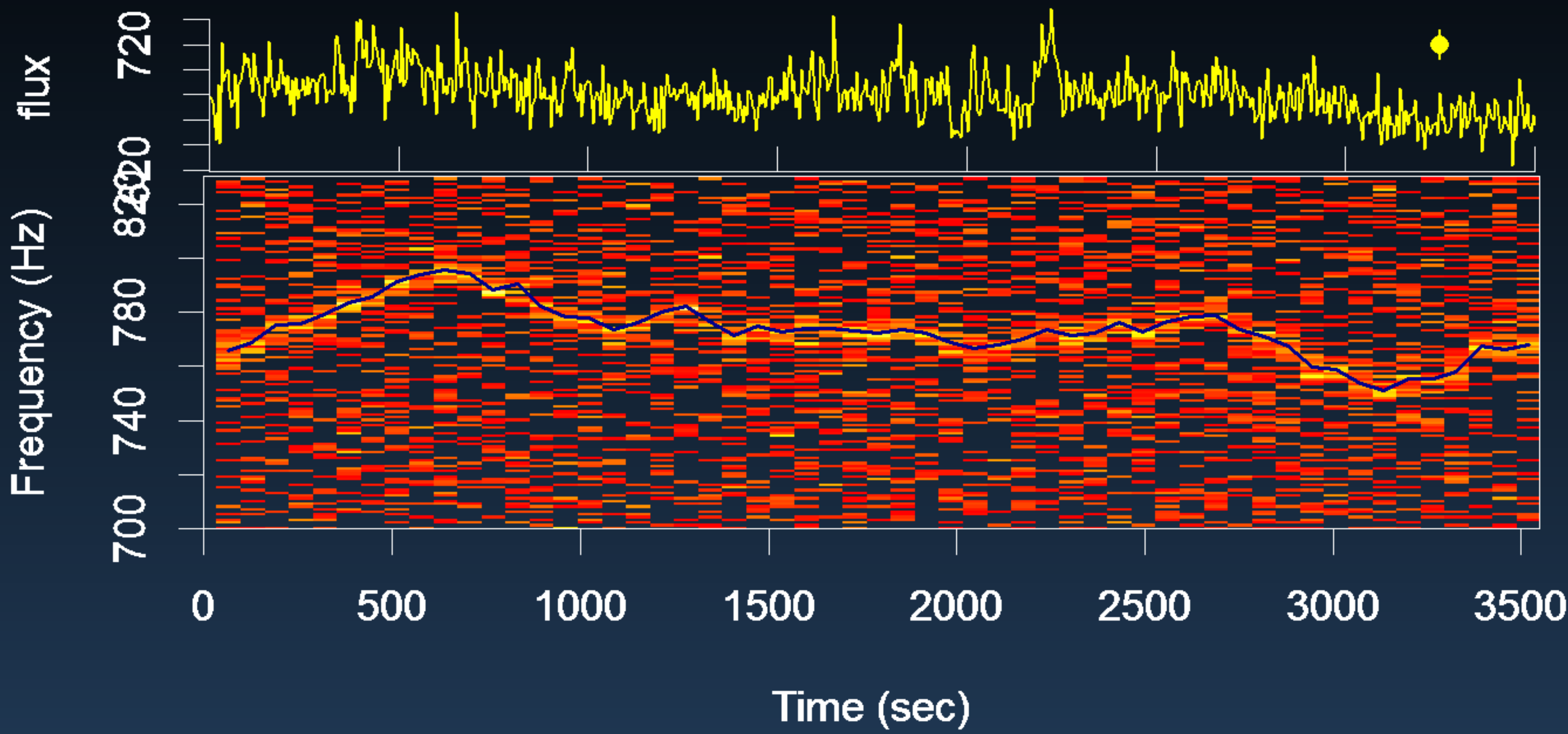
Standard recipe



The most popular spectral estimate (in astronomy, at least) is the averaged periodogram, where periodograms from each of M non-overlapping intervals are averaged.
'Barlett's method' after M. S. Bartlett (1948, *Nature,* 161, 686-687)

stochastic

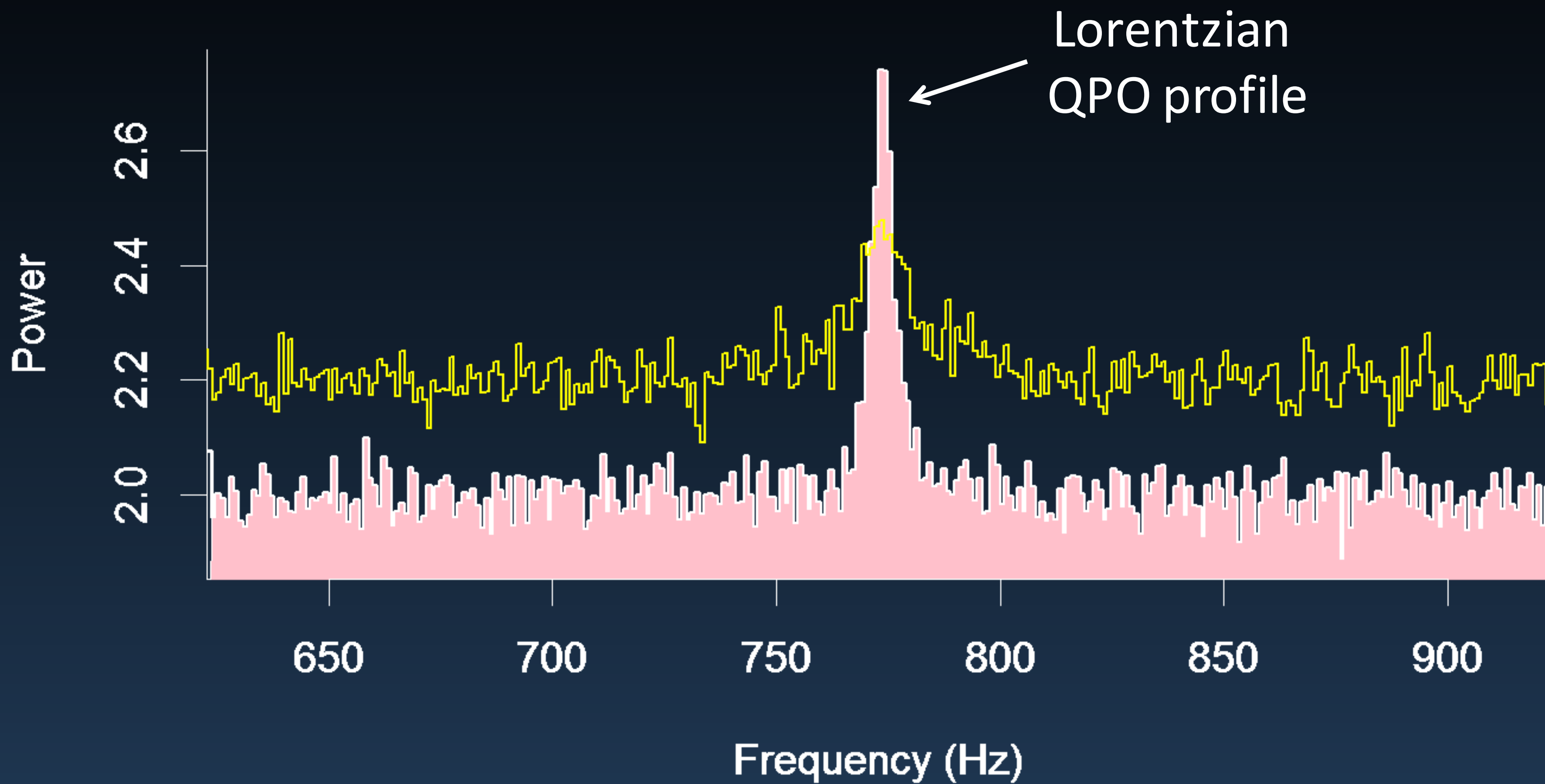
quasi-periodic oscillations (QPOs)



Well-defined peak in power spectrum: $f \sim$ 500-1300 Hz, $\Delta f \sim$ few Hz

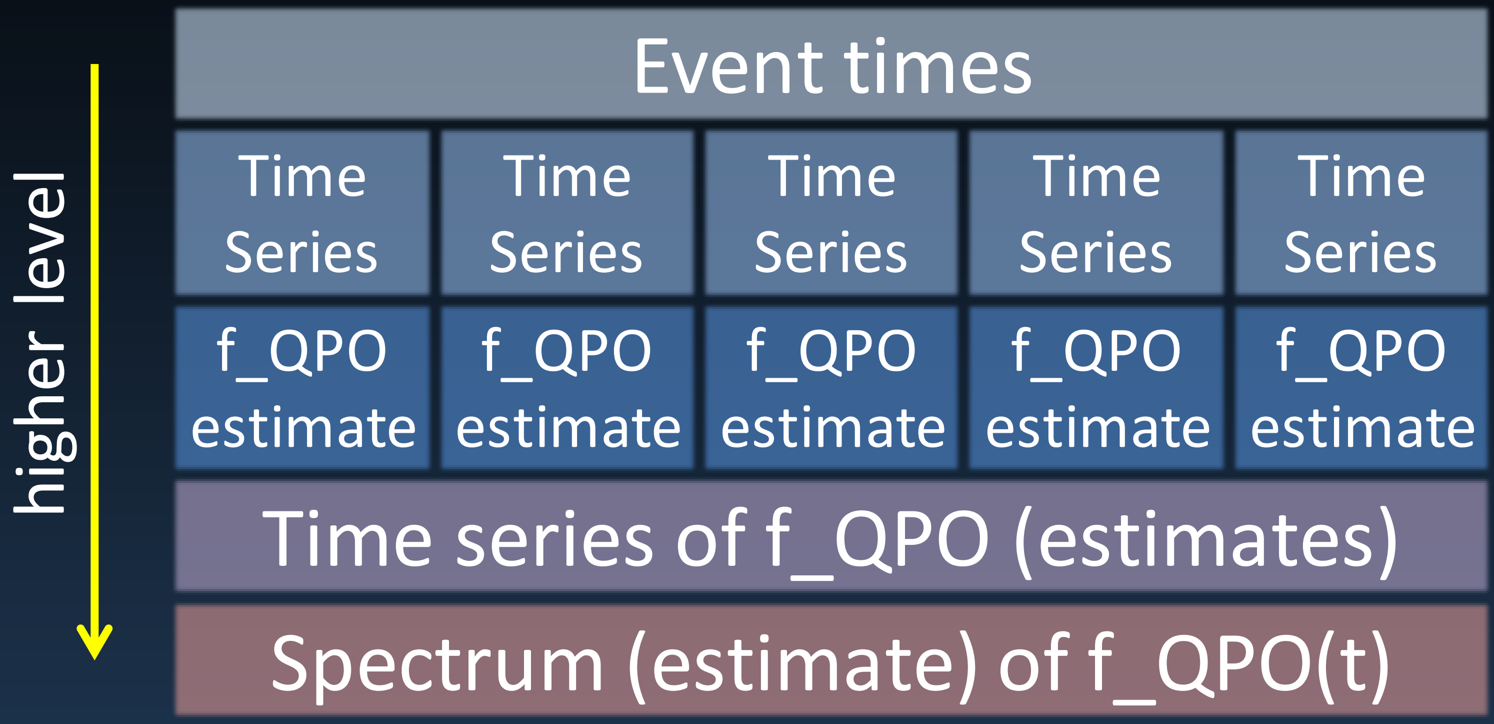
Peak frequency varies on short timescales; strength, width on longer timescales

simple mean vs. "shift and add"



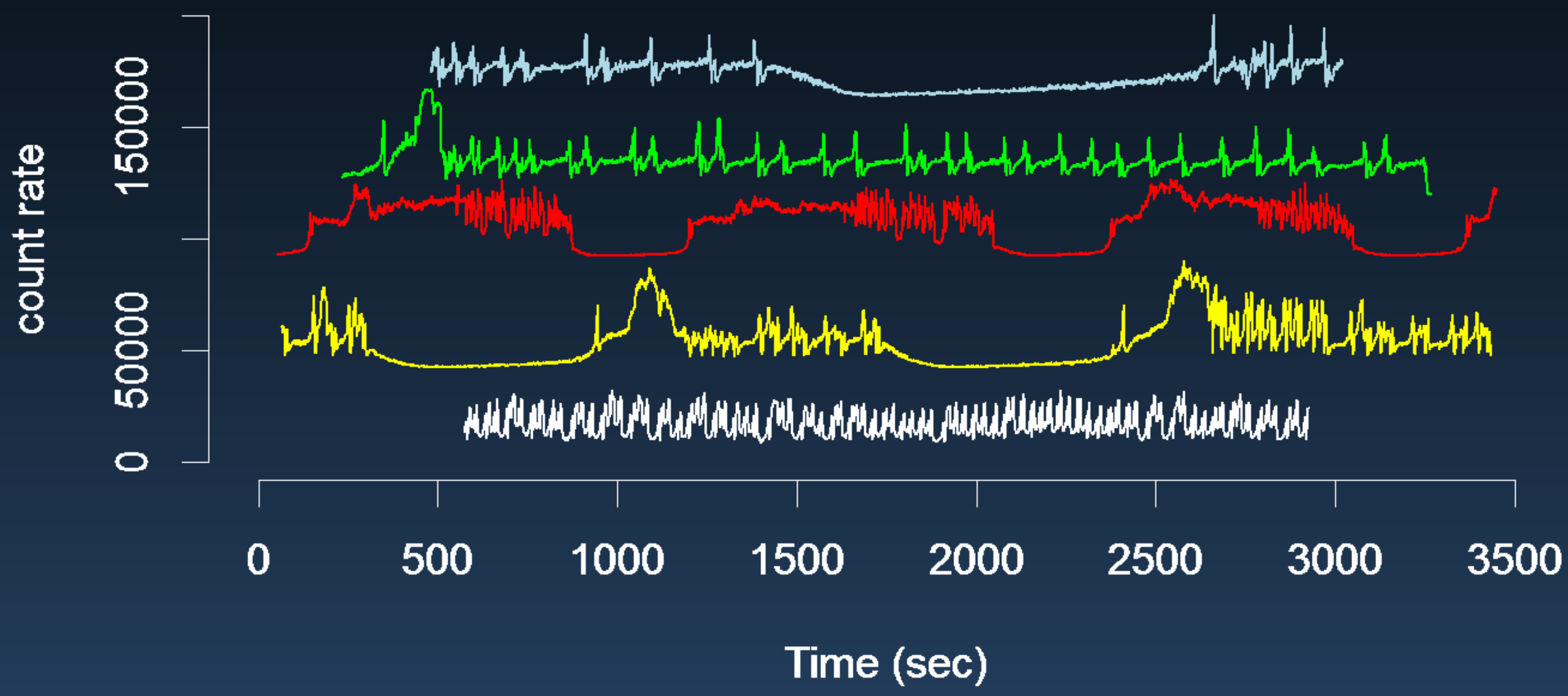
Shift "short time" periodograms to common centroid frequency
(small bias on width, power conserved)
Mendez et al. (1996, ApJ); Barret & Vaughan (2012, ApJ)

An unintentional multi-level model



black hole GRS 1915+105

periodic/stochastic/chaotic/mixed ?



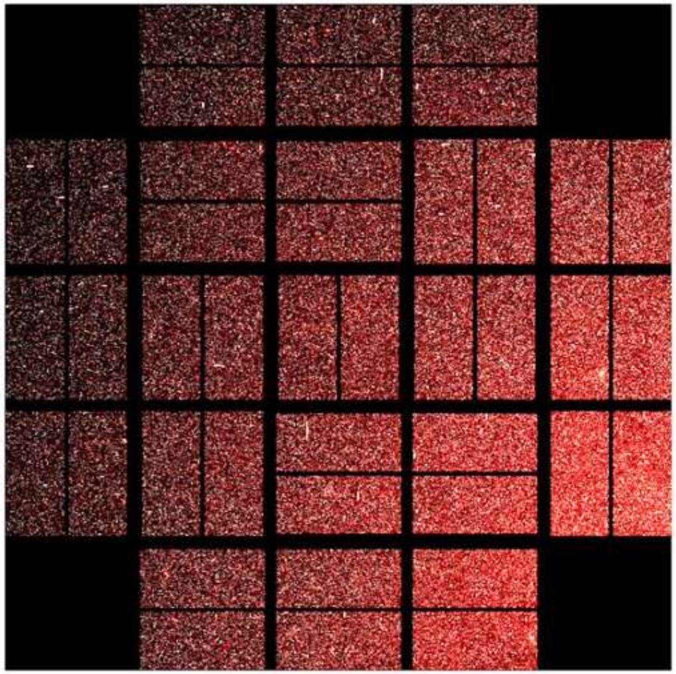
“slow” variations

Kepler mission monitors

10,000s of sources

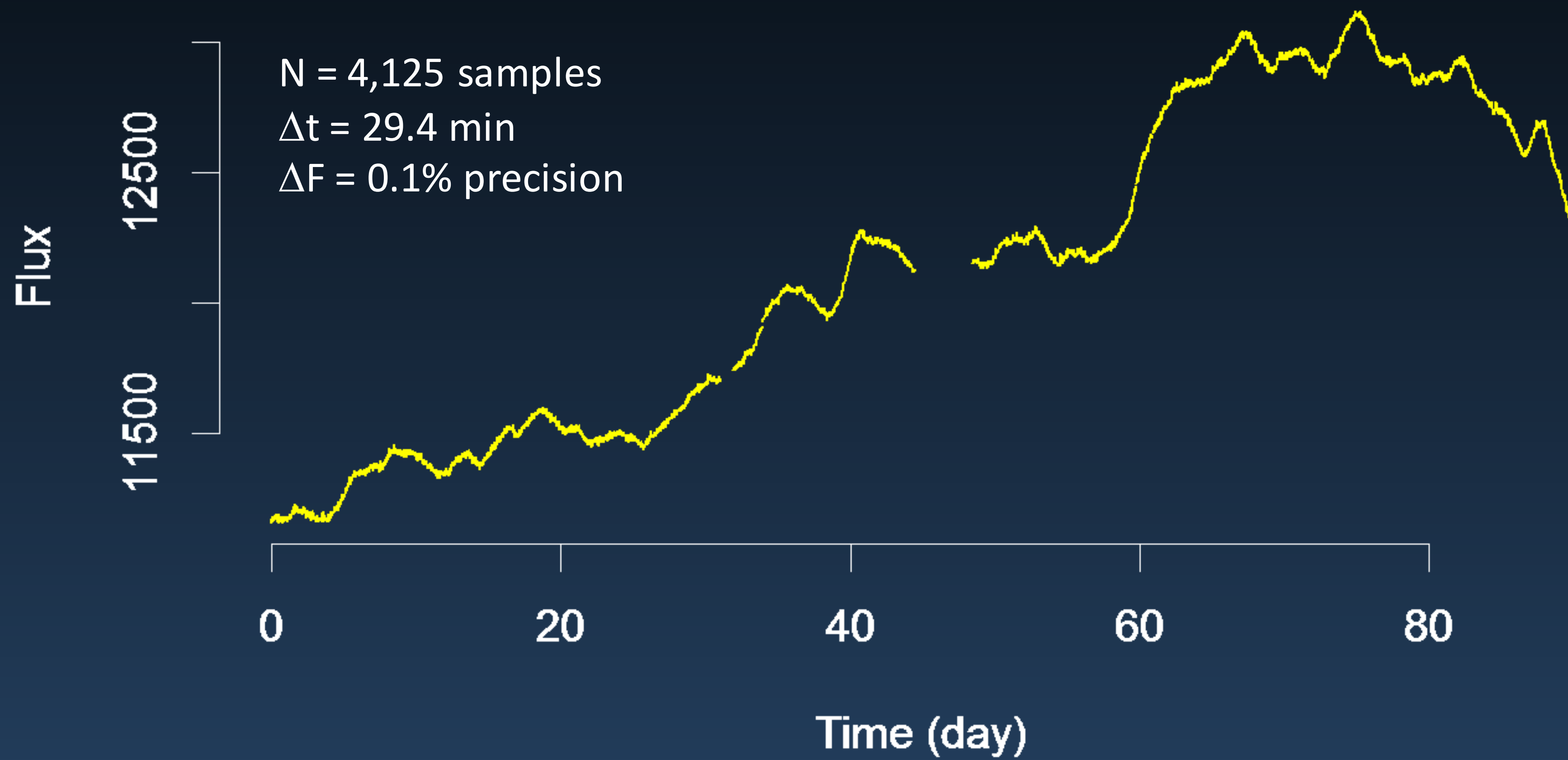
“continuously”

Q1-Q6 now public…



Kepler light curve of Zw 229-15

problem: how to recover spectrum of very “red” processes?
(Deeter & Boynton 1982; Fougere 1985; Mushotzky et al. 2011)



A message from Captain Data:

"More lives have been lost looking at the raw periodogram than by any other action involving time series!"

(J. Tukey 1980; quoted by D. Brillinger, 2002)

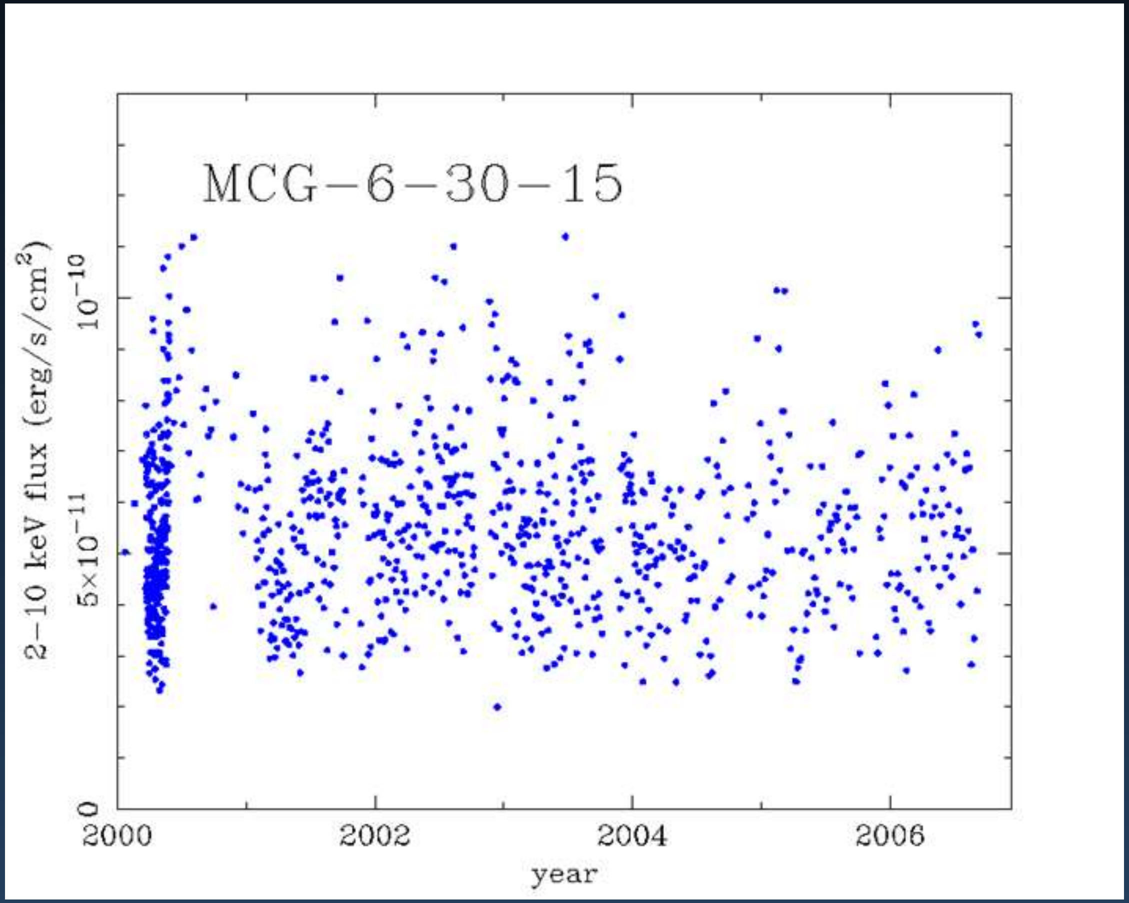
Fejer kernel

$$E[I_j] = \int_{-f(Nyq)}^{+f(Nyq)} F(f_j - f')P(f')\,\mathrm{d}f' = P(f) - bias[f]$$

even when we can "beat down" the intrinsic fluctuations in the periodogram, bias – in the form of ***leakage*** and ***aliasing*** are difficult to overcome

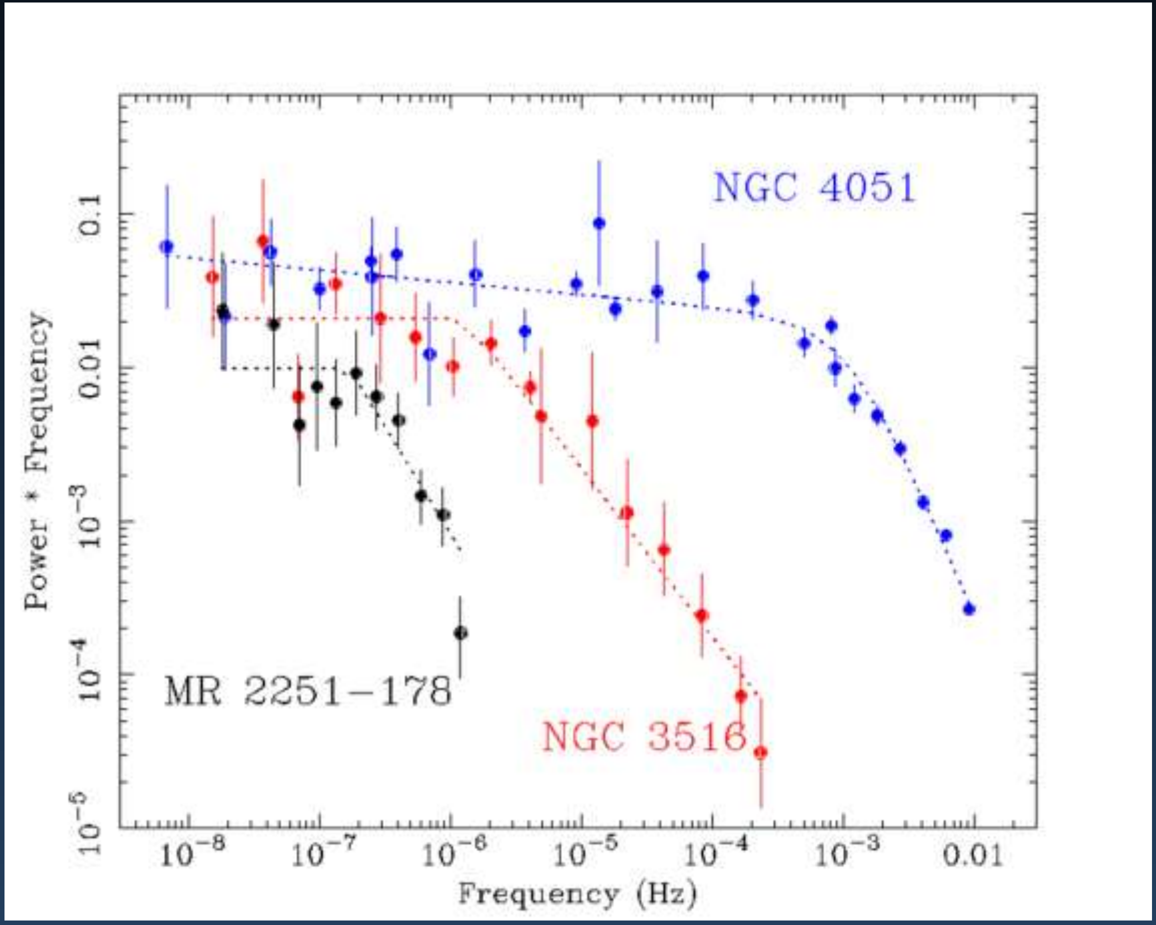
RXTE monitoring of AGN

problem: what about with low efficiency (uneven) sampling?
(Uttley et al. 2002; MNRAS)



RXTE monitoring of AGN

solution: `forward fitting' through sampling window
(Uttley et al. 2002; MNRAS)



Forward fitting

- Extract time series data x(t)
- compute spectrum estimate P_{est}(f)
- define model spectrum P(f; θ)
- compute dist P_{est}(f; θ) by Monte Carlo (incl. sampling)
- compare data & model – does $\langle P_{\text{sim}}(f) \rangle$ look like P_{est}(f)?

issues:

- how to estimate spectrum?
- how to simulate time series data from P(f; θ)?
- how to compare data & model?
- better in time domain?

Simulation

Time domain methods (e.g. ARMA etc.)

Fourier method (Ripley 1987; Davies & Harte 1987; Timmer & König 1995)

- define: `P(f)`
- compute: `E[Re(FT)] = E[Im(FT)] = sqrt(P[f_j]/2)`
- randomise: `FT_j ~ ComplexNorm × sqrt(P[f_j]/2)`
- invert (FFT): `FT` → x_{sim}`(t_i)`

but we are trying to simulate sampling a continuous process

so generate using $\Delta t_{sim} << \Delta t_{data}$ and $T_{sim} >> T_{data}$

then resample (subset).

Even so, very fast using FFT. (Assumes linear, Gaussian process)

Bivariate time series

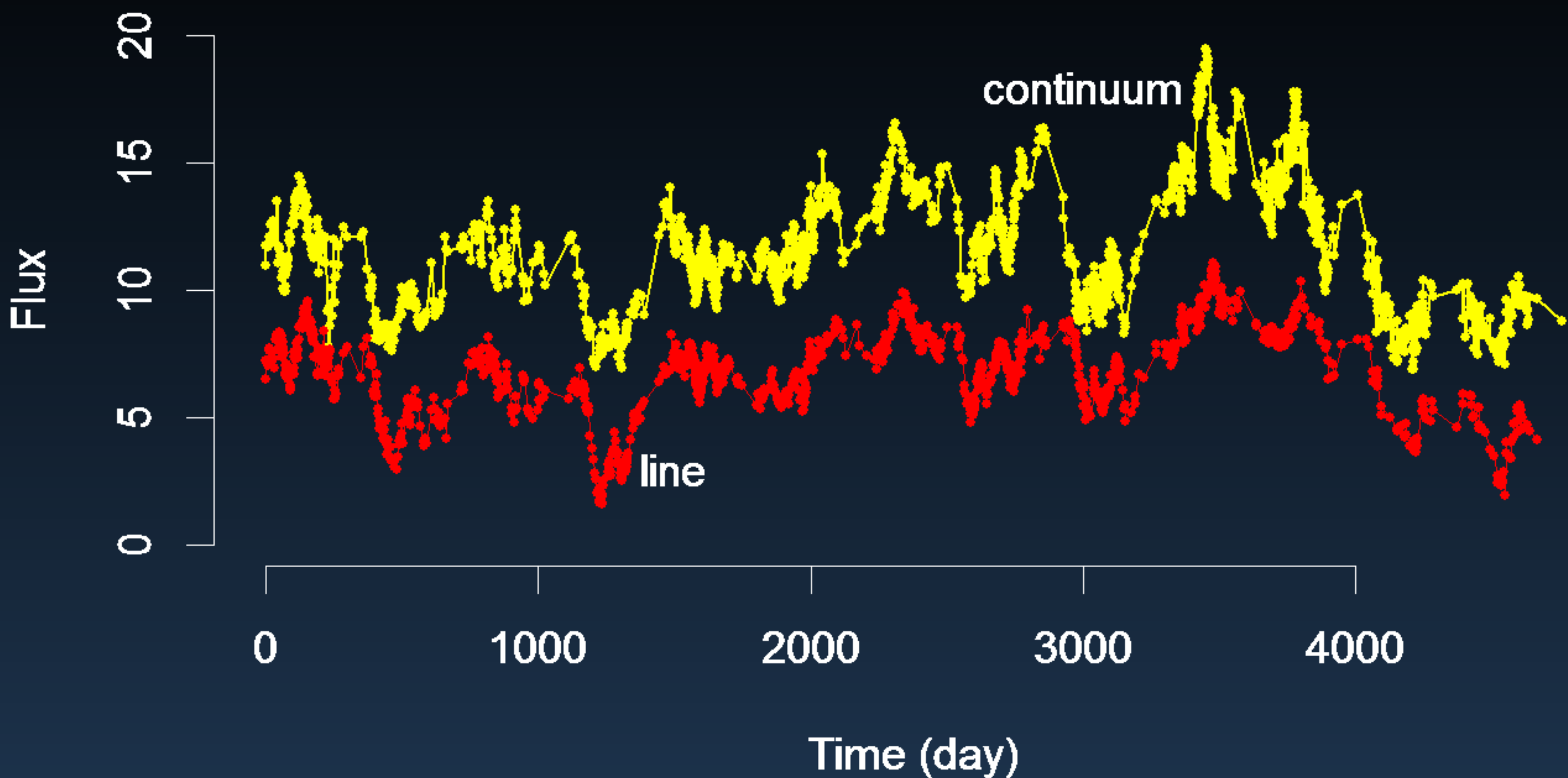
Correlation

+

time delays

potentially powerful
causation and scales

bivariate astronomical time series



$$y(t) = \int_{-\infty}^{+\infty} \psi(t-\tau)x(\tau)\,\mathrm{d}\tau = \psi \otimes x$$

$$Y = \Psi \times X$$

how to recover correlation & delays (more generally the response function) from uneven and/or non-simultaneous time series? when very “red”?

Try for yourself

Sample of astronomical time series (circa 1993)

http://xweb.nrl.navy.mil/timeseries/timeseries.html

Kepler mission archive (continuous optical monitoring)

http://archive.stsci.edu/kepler/

Swift (UK data centre)

http://www.swift.ac.uk/

Plus 1,000s of RXTE time series with <ms resolution!