Tools For Astronomical Big Data

Characterizing the variable sky with CRTS

Matthew J. Graham
Center for Data-Driven Discovery, Caltech
and NOAO

March 10, 2015

The extended CRTS team

S. George Djorgovski, Caltech (PI)

Andrew Drake, Caltech (Co-PI)

Ashish Mahabal, Caltech

Ciro Donalek, Caltech

Daniel Stern, JPL/Caltech

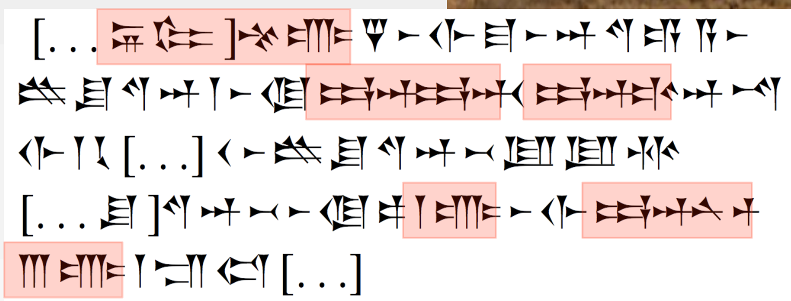
Eilat Glikman, Middlebury

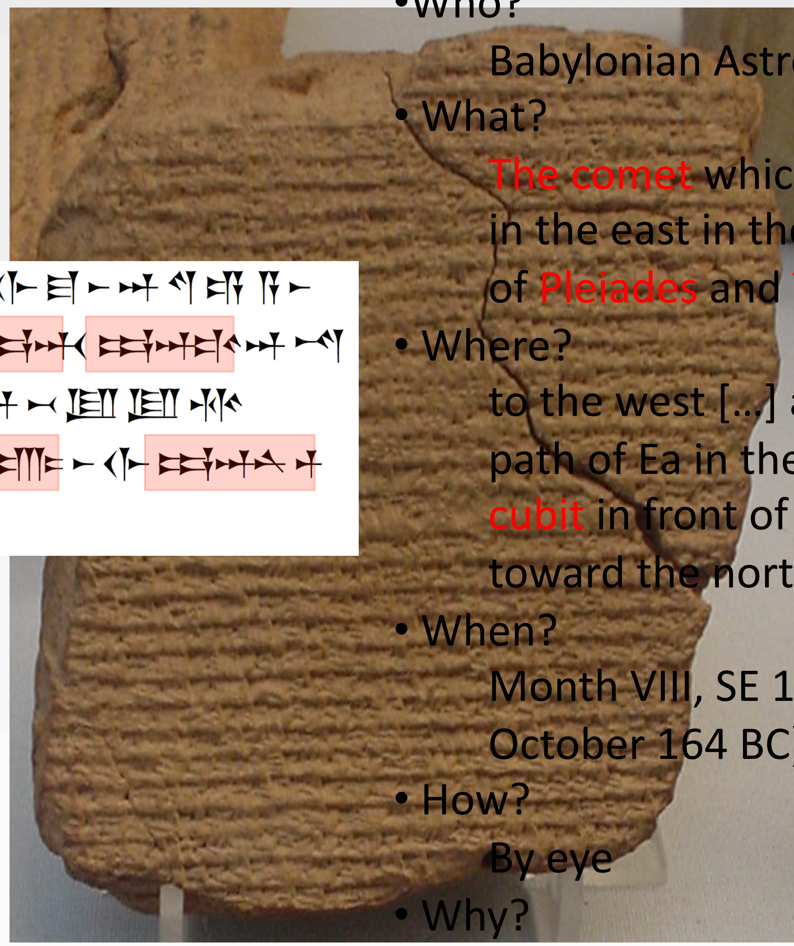
Eric Christensen, UA

Steve Larson, UA

+ students

Time domain astronomy: old style



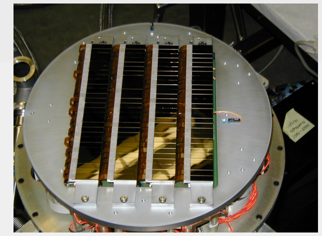
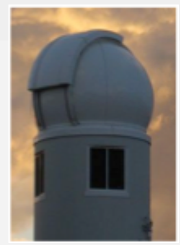


- Who?
  - Babylonian Astronomical Diary
- What?
  - The comet which previously had appeared in the east in the path of Anu in the area of Pleiades and Taurus
- Where?
  - to the west […] and passed along in the path of Ea in the region of Sagittarius, 1 cubit in front of Jupiter, 3 cubits high toward the north […]
- When?
  - Month VIII, SE 148 (lunar month beg. 21 October 164 BC)
- How?
  - By eye
- Why?
  - Celestial divination

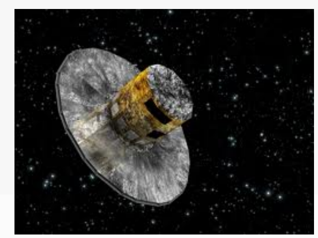
Rebooting the time domain

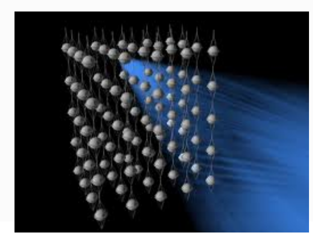
- Palomar-Quest Synoptic Sky Survey
- SDSS (Stripe 82)
- Catalina Real-time Transient Survey
- Palomar Transient Factory
- Pan-STARRs
- SkyMapper
- ASKAP
- ThunderKat (MeerKAT)
- GAIA
- LIGO
- IceCUBE
- LOFAR
- LSST
- SKA

...

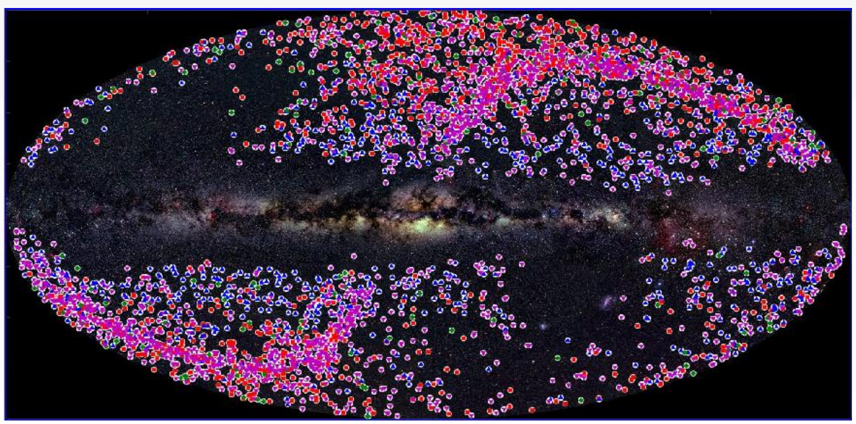


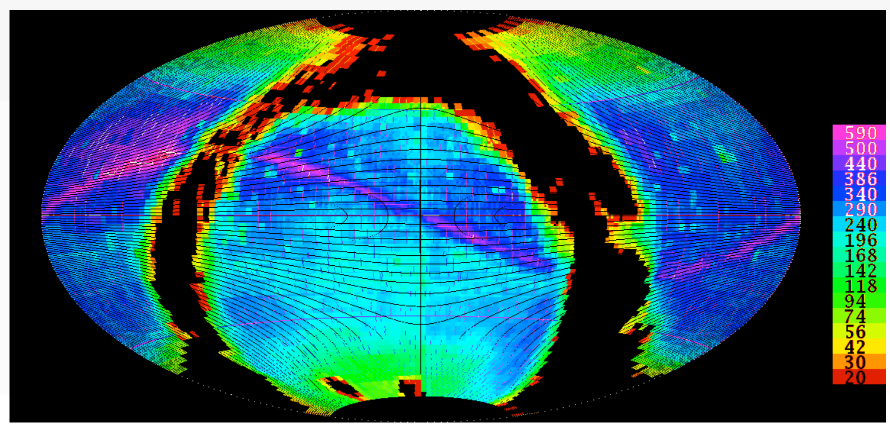


Catalina Real-time Transient Survey (2005-)

- Collaborative survey with Catalina Sky Survey (LPL, UA)
- Unfiltered observations 21 nights/lunation covering up to 2000 deg^2/night
- Covers 33000 sq. deg. (0 < RA < 360, -75 < Dec < 70).
- Calibrated photometry for 500 million objects (> 100 billion data points)
- Depth V = 19 to 21.5
- 100 – 600 observations in most regions (median ~ 250)
- More published SNe and CVs than any other survey (public instantly)
- Open data policy (http://catalinadata.org)
- ~3% LSST



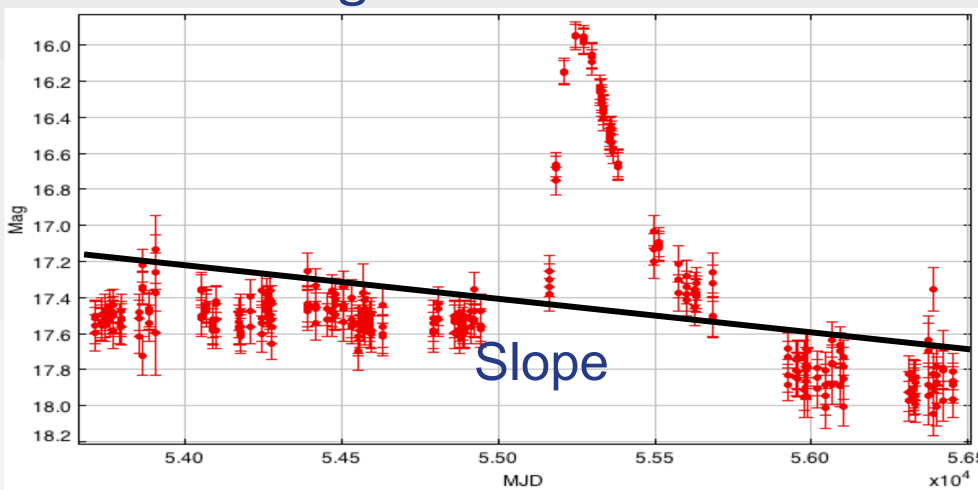


Understanding astronomical time series

- Why is this interesting:

$$\sum_{i=1}^{n} A_i \sin(\omega t + \phi_i)$$

Fourier

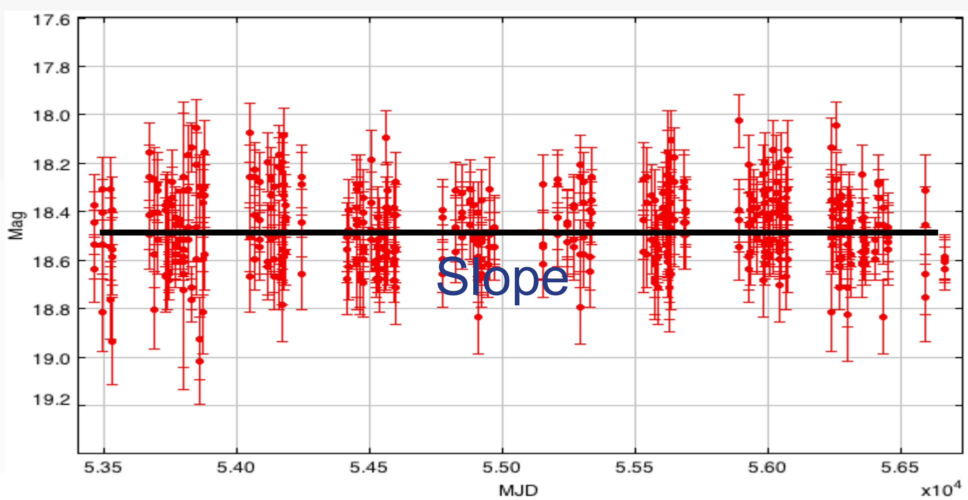


Amplitude

- And this is not:

$$\sum_{i=1}^{n} A_i \sin(\omega t + \phi_i)$$

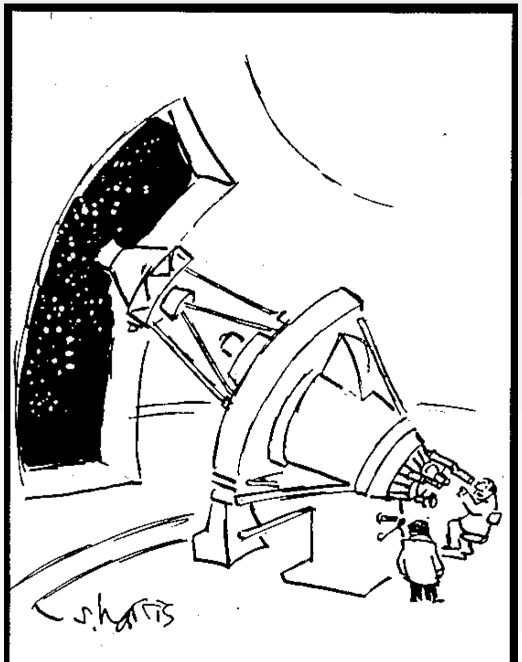
Fourier



Amplitude

Characterizing phenomenology

- Data is irregular, noisy, gappy and heteroscedastic
- Measuring:
  - Morphology (shape): skew, kurtosis
  - Scale: Median absolute deviation, biweight midvar.
  - Variability: Stetson, Abbe, von Neumann
  - Timescale: periodicity, coherence, characteristic
  - Trends: Thiel-Sen
  - Autocorrelation: Durbin-Watson
  - Long-term memory: Hurst exponent
  - Nonlinearity: Teraesvirta
  - Chaos: Lyapunov exponent
  - Models: HMM, CAR, Fourier decomposition, wavelets
- Defines high-dimensional (d > 100) feature space



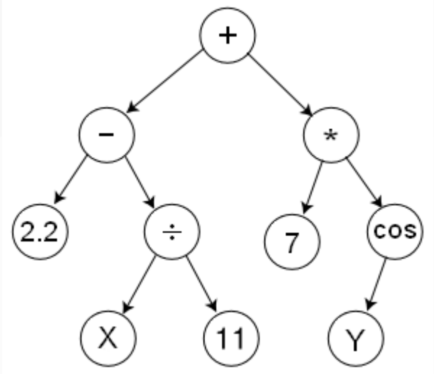
"Actually they all look alike to me."

What are "interesting" features?

- For a given class of object, which features contribute to its classification:

$$class = g(f(x_1, x_2, x_3, \ldots, x_n))$$

  - g(x) is Heaviside step function or logistic function
  - f(x) is an equation for the discriminating hyperplane which defines a class in some high-dimensional feature space
  - x_i are the useful features
- Symbolic regression finds a function, in symbolic form, that fits a sample of data
  - Genetic algorithm
  - Employs an alphabet of mathematical building blocks
  - Produces a small set of final candidate analytical expressions on accuracy-parsimony Pareto front



$$\left(2.2 - \left(\frac{X}{11}\right)\right) + \left(7 * \cos(Y)\right)$$

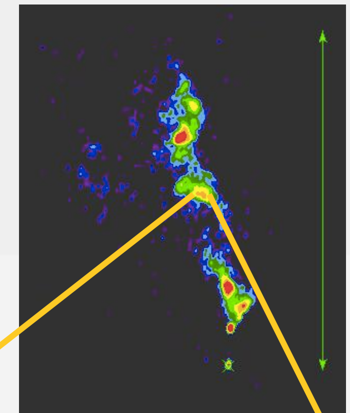
(Graham et al. 2013)

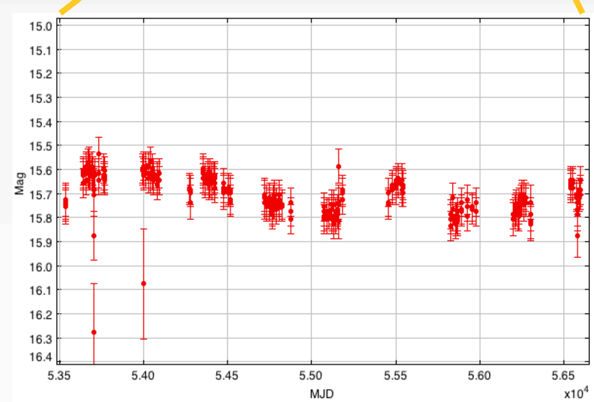
Quasar variability

- First quasar identified 3C 48 – most striking feature was that the optical radiation varied
- Physical origin of photometric variability in optical/UV is unclear
- Many studies based on small sample size or (very) sparse time sampling
- The current best statistical description is damped random walk (CAR(1)):

$$dX(t) = -\frac{1}{\tau} X(t)dt + \sigma\sqrt{dt}\,\varepsilon(t) + b\,dt \quad \tau, \sigma, t > 0$$

characterized by σ^2 and τ





- 335000 spectroscopically-confirmed quasars in CRTS archive

An interesting feature: Slepian wavelet variance

- Wavelets allow localized time and frequency analysis
- A time series can be decomposed by applying a set of wavelet filters

$$W_{j,t} = \sum_{l=0}^{L_j - 1} h_{jl} X_{t-l};\ \mathrm{t}=0, \pm 1, \ldots;\ j = 1, 2, \ldots;\ L \geq 2d$$

- The wavelet variance at a given scale:

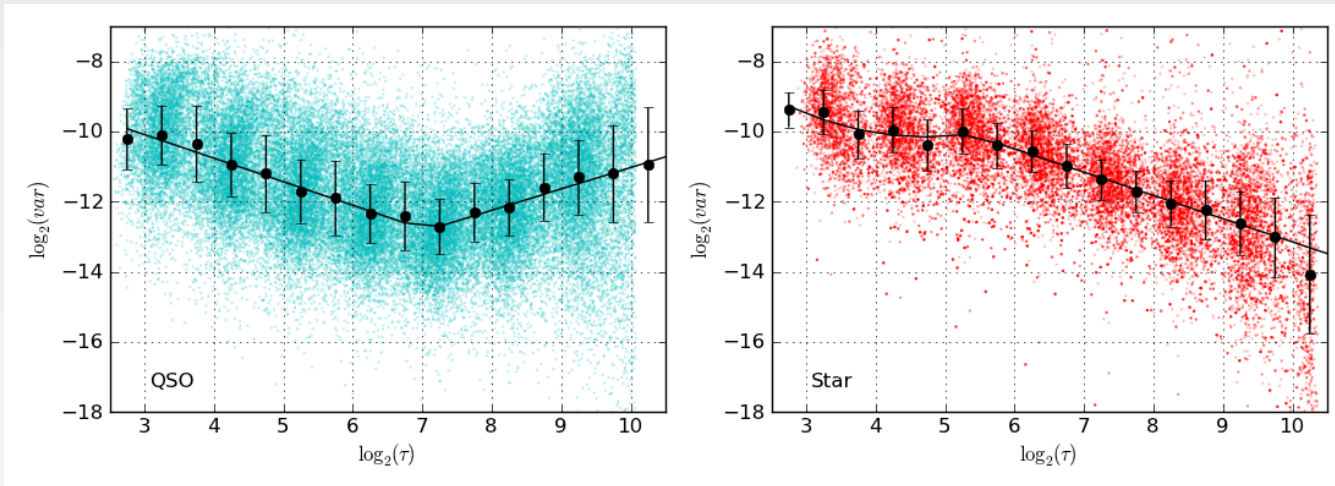
$$\tau_j = 2^{j-1}\overline{\Delta}\ ;\ \ \nu_X^2(\tau_j) = \mathrm{var}(W_{j,t})\ ;\ \ \mathrm{var}(X_t) = \sum_{j=1}^{\infty} \nu_X^2(\tau_j)$$

gives the contribution to the total variance of the time series due to scale τ_j

- Characteristic scales are indicated by peaks or changes of behavior in $\log(\nu^2_X)$ vs. $\log(\tau_j)$
- Slepian wavelets work with irregular and gappy time series

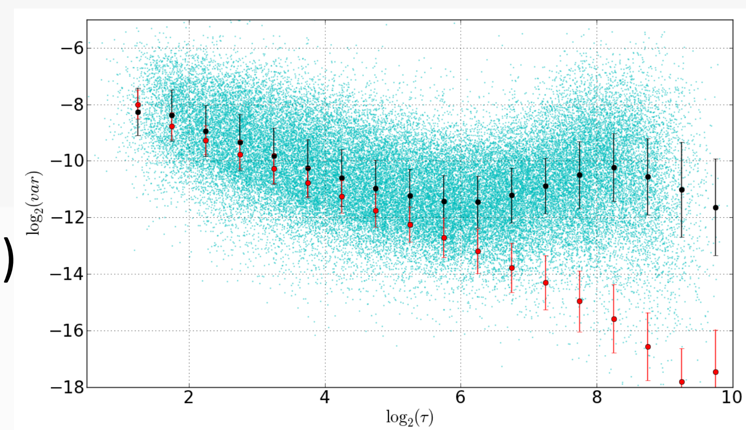
Quasars are different

Quasar   Star

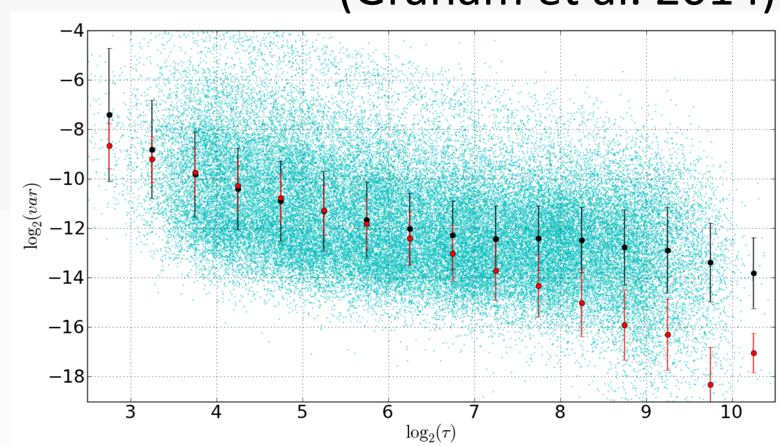


(Graham et al. 2014)

- Real
- CAR(1)

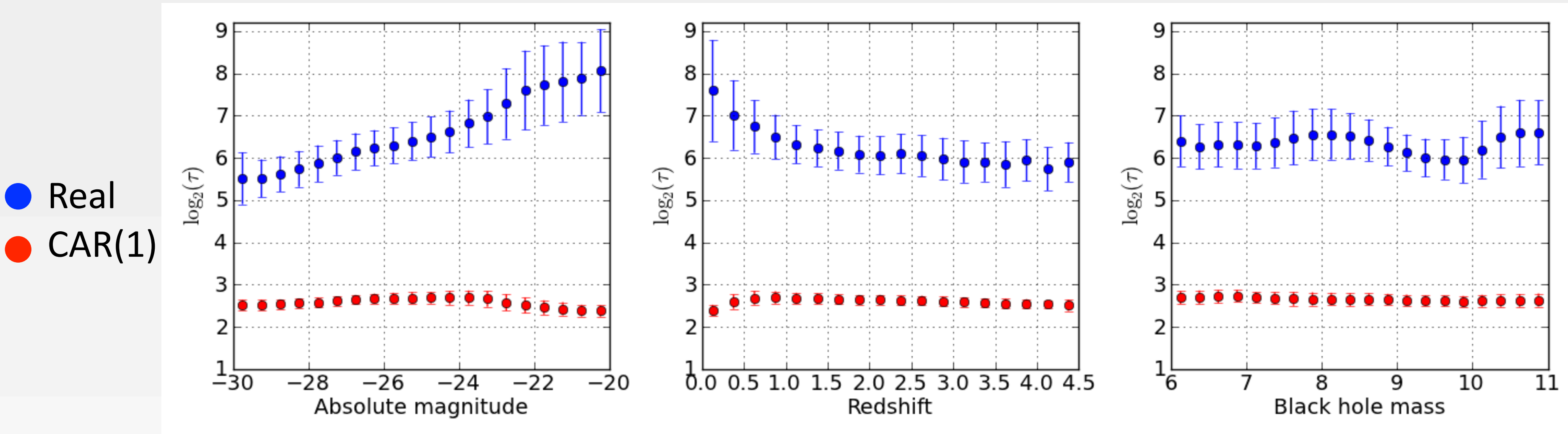


CRTS



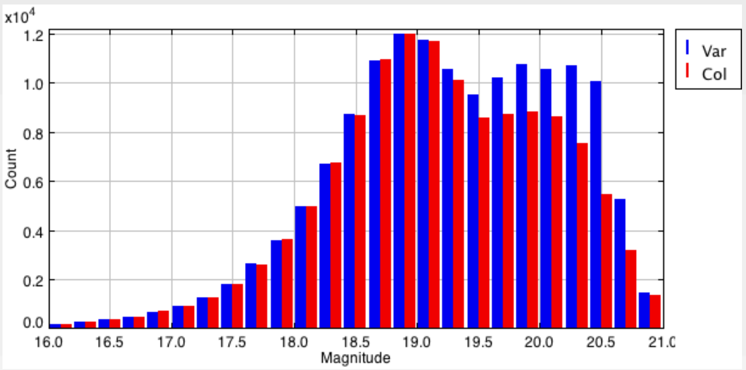
Stripe 82

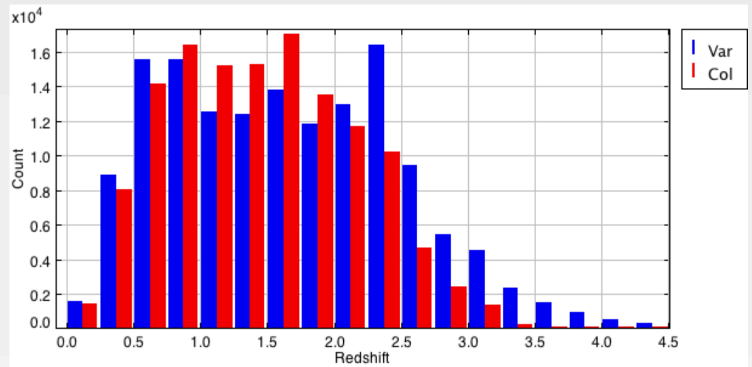
A restframe characteristic timescale



- Related to the viscous crossing time of a noise field driving the perturbations causing photometric variability?

Quasar selection via variability



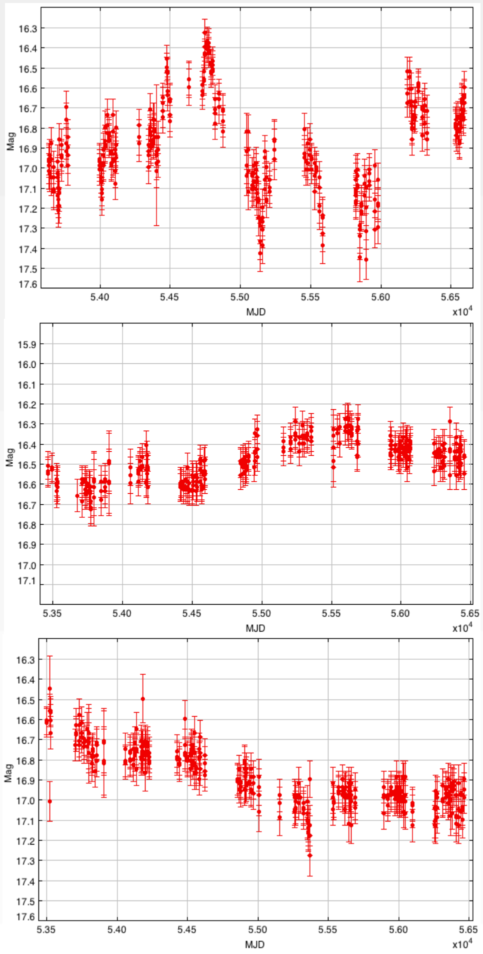


	Completeness	Purity
Structure function	71%	96%
Damped random walk	91%	66%
Slepian wavelet	93%	80%
Ensemble variability (stacking)	97%	95%
Ensemble var. + WISE color	99%	99%

- CRTS Quasar Catalog in preparation: 1M quasars

Subtype selection via variability

- Analyze feature distributions for identified subclasses:
  - Blazar
  - Radio
  - X-ray
  - Type 2
  - LoBAL
  - HiBAL
  - Strong reddening In UV
  - Weak emission lines
  - Strong iron emission



- Variability amplitude:
  - Blazar, radio, type 2, weak lines, strong Fe II
- Characteristic timescale:
  - Blazar, radio, X-ray, type 2
- WISE color (W1 – W2):
  - Radio, X-ray, HiBAL

Beyond CAR(1) = CARMA(1,0)

"Why should the accretion disk around a SMBH behave in a stationary fashion?" – Eric Feigelson

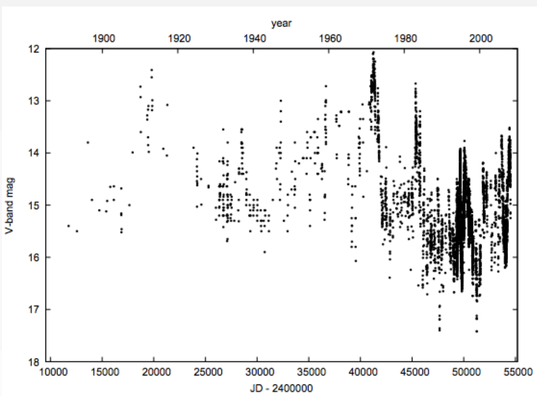
- Hints that CAR(1) not sufficient from Kepler, OGLE, CRTS
- Kelly et al. (2014) describe a formalism for CARMA(p,q) models:

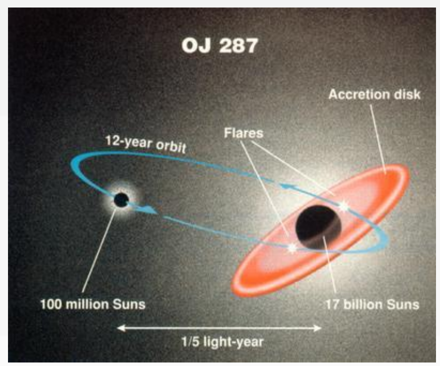
$$\frac{d^p y(t)}{dt^p} + \alpha_{p-1}\frac{d^{p-1} y(t)}{dt^{p-1}} + \ldots + \alpha_0 y(t) = \beta_q \frac{d^q \varepsilon(t)}{dt^q} + \beta_{q-1}\frac{d^{q-1}\varepsilon(t)}{dt^{q-1}} + \ldots + \varepsilon(t)$$

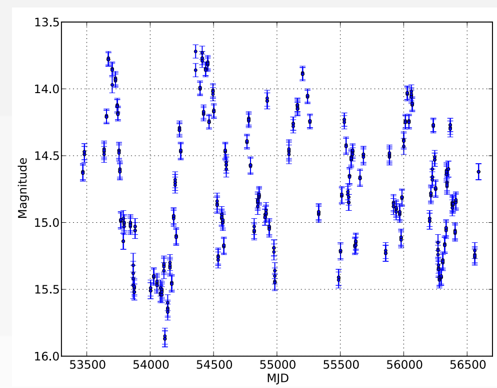
- Initial best fit results from 2634 quasars (using Microsoft Azure allocation) show that lower *p* (autoregressive) are favored and smaller *q* (moving average) for a given *p*
- Higher *p* correlates with variability amplitude (MAD) – complexity?
- No moving average term means no noise term
- (CO)GARCH models (stock market volatility) are too complicated

SMBH binaries

- Expected consequence of galaxy mergers
- Closest (sub-parsec) system are not resolvable except possibly with long baseline radio interferometry
- OJ 287 shows a pair of outburst peaks every 12.2 years for at least the last century



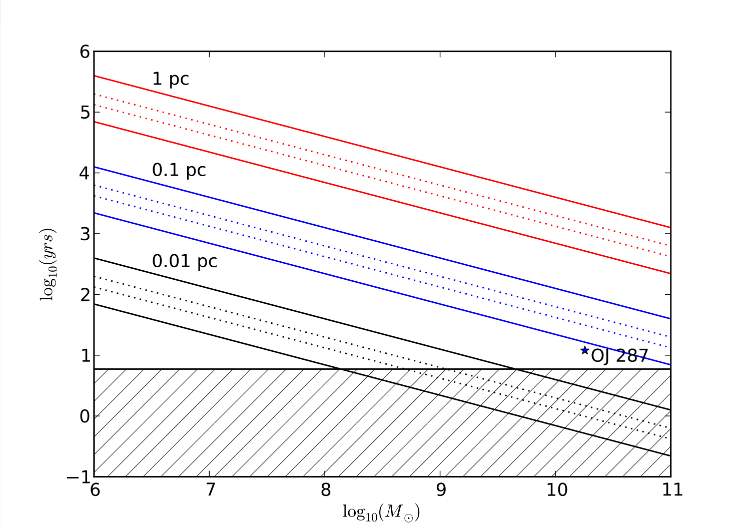




- Systematic searches have so far looked for broad-line velocity offsets in spectra

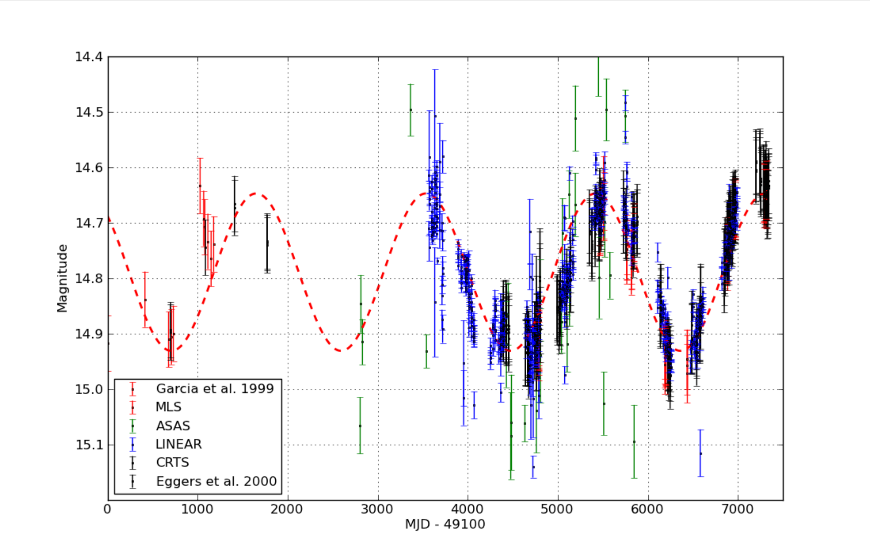
Searching for SMBH binaries

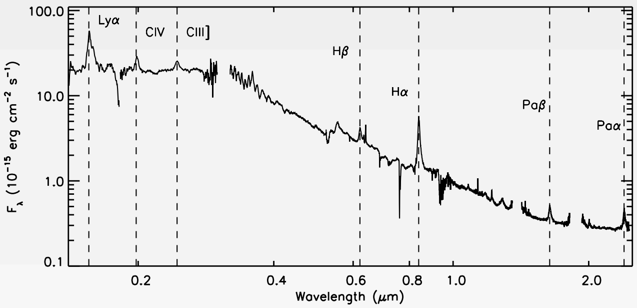
- Joint wavelet and autocorrelation function-based technique to look for strongly periodic behaviour
- Interested in a kinematic (Keplerian) signal
- Assuming period from 20 – 300 weeks, sky coverage of 2π ster., V < 20, 0.5 < z < 4.5, 450 binaries are predicted
- 111 candidates identified out of 340000 quasars
- Simulated data set of objects following a CAR(1) model gives 0 candidates

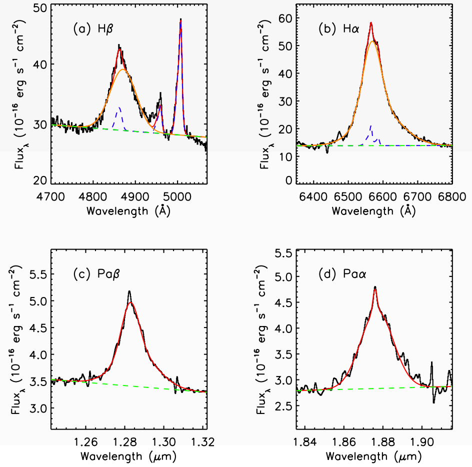


PG 1302-102

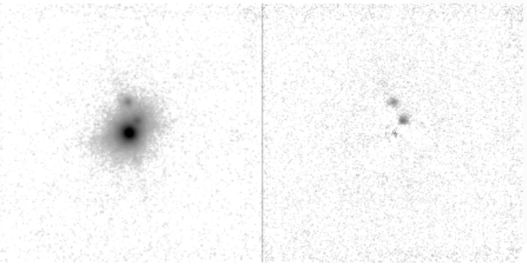
- z = 0.28, core-dominated flat-spectrum radio source
- Luminous elliptical host with nearby companions
- Coincident radio and optical structural features





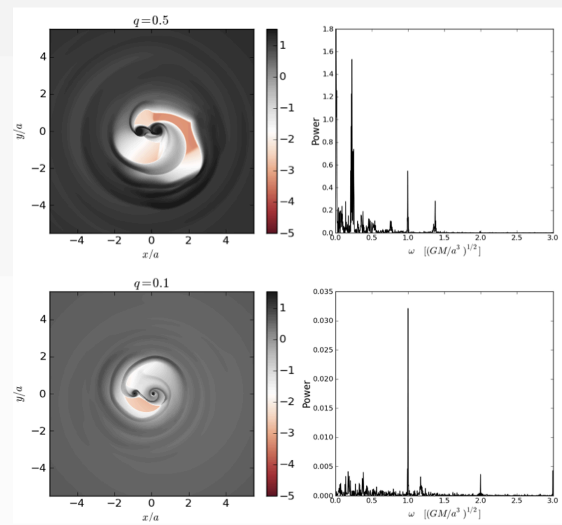


(Graham et al. 2015, Nature, 518, 74)



Possible interpretations

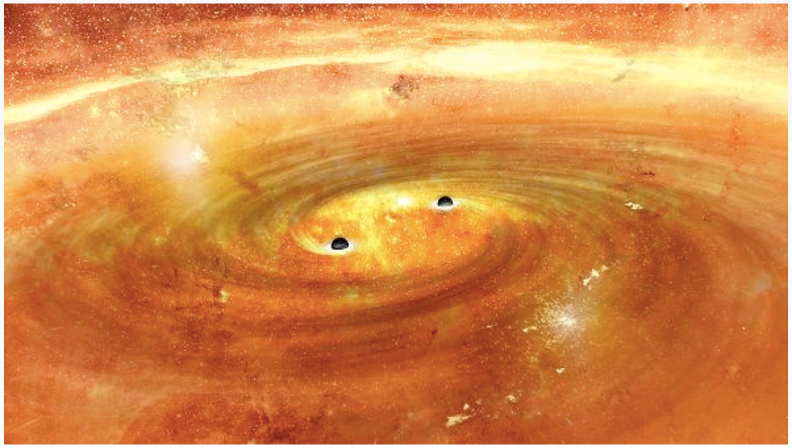
- Superposition of thermal emission from accretion disk and non-thermal contribution from a precessing jet driven by SMBH binary
- Warped accretion disk caused by SMBH binary
- Periodic accretion rates from a SMBH binary can lead to an overdense hump in the inner circumbinary accretion disk
- Hydrodynamical simulations suggest that strongest periodicity associated with cavity in circumbinary disk => true binary period 3-8 times shorter



(D'Orazio et al. 2015)

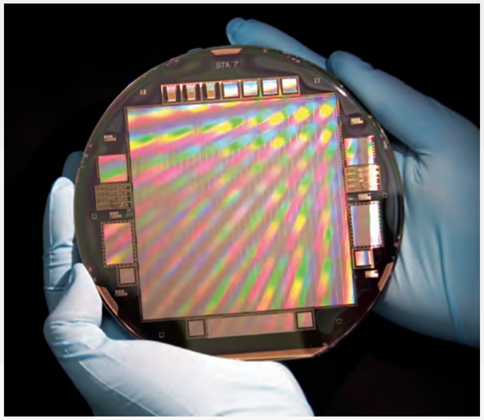
If it is a SMBH binary…

- Restframe period of 1443 days and total mass of $10^{8.5}$ M_{sol} gives an upper limit separation of ~0.01 pc (276 R_S) (~90 R_S?)
- Further monitoring is needed
- Multiple periods with cavity and broad line widths
- Reverberation mapping to distinguish between different scenarios
- Strong candidate for gravitational wave experiments sensitive to nanohertz frequency waves



CRTS-II

- Same telescopes but bigger cameras (already built):
  - MLS 1.5m: 1.2 sq. deg. -> 5 sq. deg.
  - CSS 0.7m: 8.2 sq. deg -> 19 sq. deg.
- 3600 sq. deg. per night
- 1600 sq. deg. at daily cadence
- Offers:
  - 5x transient rate
  - Earlier detection
  - Better time coverage
  - Increased sensitivity to transients and variables:
    - Microlensing, TDEs, orphan GRBs, SN shock breakouts, …
- Funded by NSF until 2017



AURA · LSST · NOAO · VAO · NSF · CONICYT · NLHPC National Laboratory for High Performance Computing Chile · CMM Center for Mathematical Modeling · fcfm · MAS · REUNA Ciencia y Educación en Red

THIRD ANNUAL SCHOOL DATES: 16-23 AUGUST 2015

- Training the next generation of scientists (in fields of astronomy, mathematics, computer science, and others) in the tools and techniques of massive data in Astronomy
- International program: funding for students from Chile and the U.S.
- Target students: senior undergraduate and beginning graduate students

For further information, visit our website:

http://www.aura-o.aura-astronomy.org/winter_school/

Registration still open – late abstract deadline: 15 April 2015

http://hotwireduniverse.org