Big Data Meet Big Black Holes: Quasars in the Time Domain

S. G. Djorgovski & M. J. Graham

Center for Data-Driven Discovery and Astronomy Dept., Caltech

With: A. Drake, A. Mahabal (Caltech), D. Stern (JPL),

E. Glikman (Middlebury), and many collaborators world-wide

Lecture 4

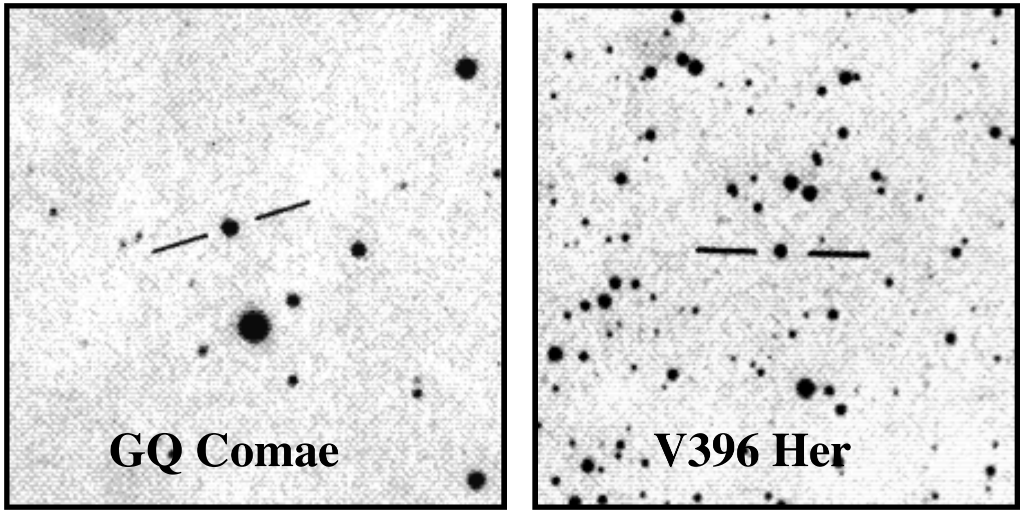
XXX Canary Islands Winter School

November 2018

Caltech

CENTER FOR DATA-DRIVEN DISCOVERY

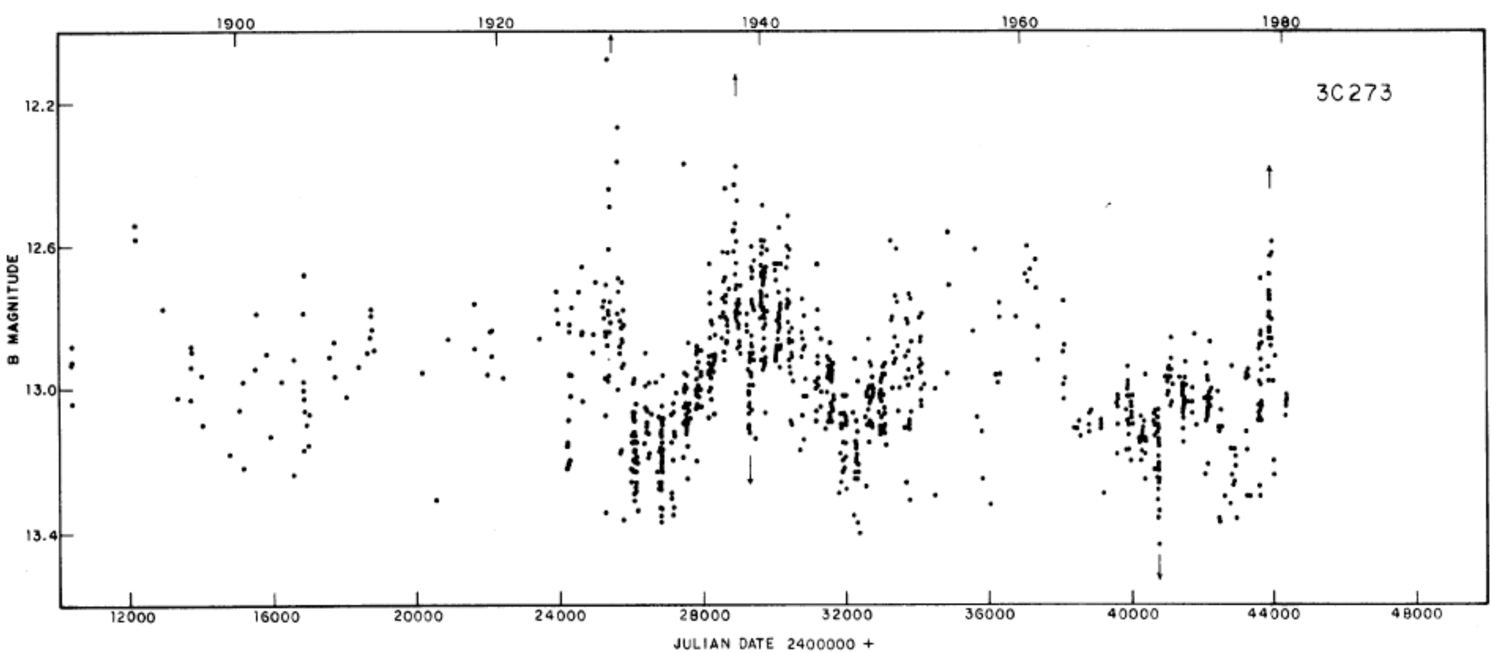
How Quasars Were *Not* Discovered



Noted as variable sources early on, but … ***misclassified as variable stars (e.g., BL Lacertae)***

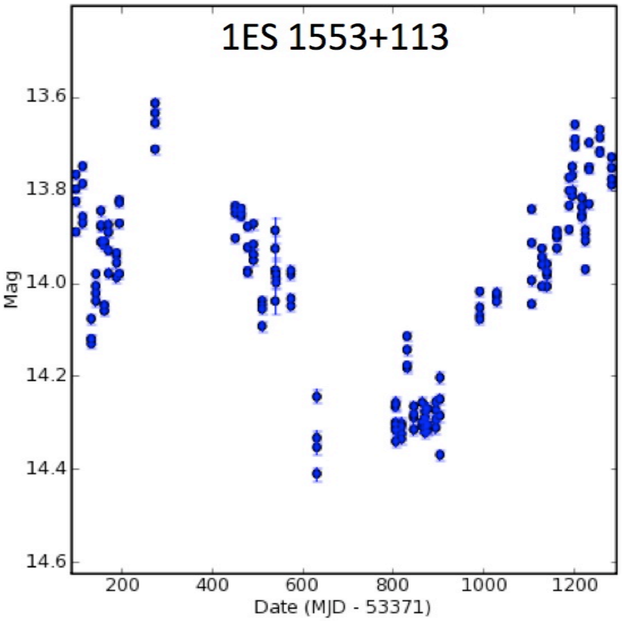
(C. Hoffmeister 1929)

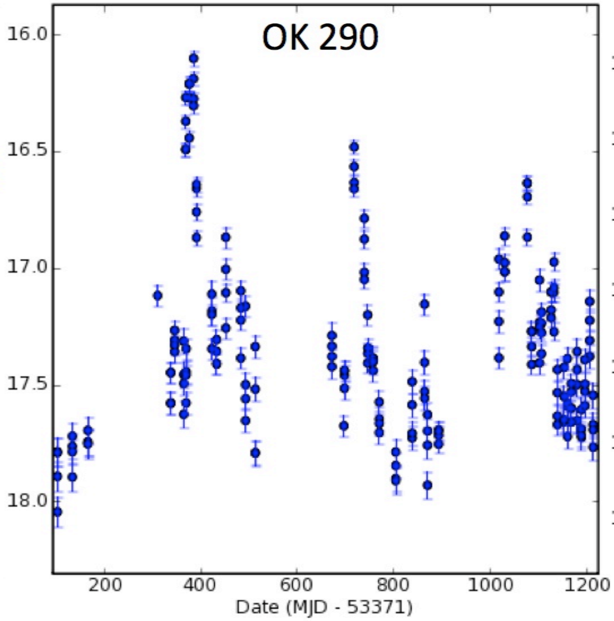
Historical (archival) lightcurve of 3C273, starting from the 1880's …

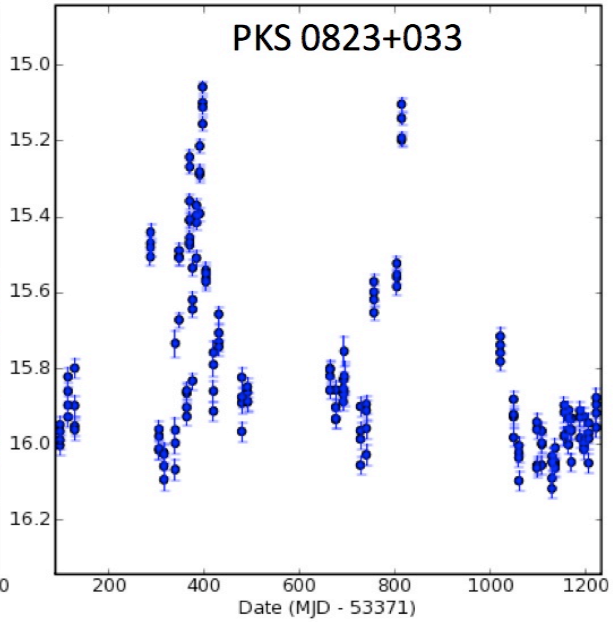


A Systematic Approach to Quasar Variability

- CRTS light curve archive is still unique in terms of the spatiotemporal coverage (80% of the sky, 13 years)
  - It contains ~350,000 spectroscopically confirmed quasars and > 1 million of quasar candidates
- This offers an unprecedented opportunity to study a variability of quasars in a systematic manner







Statistical Descriptors of Quasar Variability

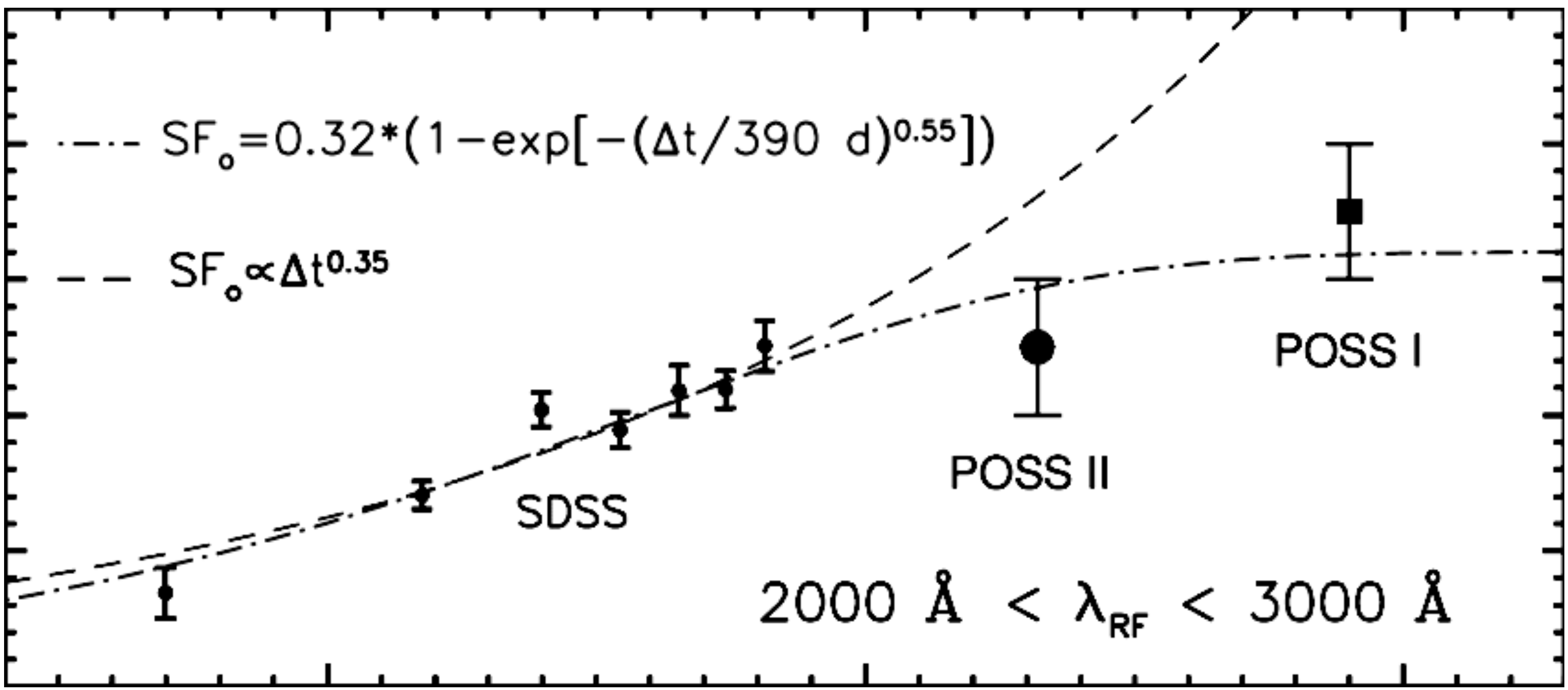
A traditional approach:

Structure Function:

$$S(\tau) = \left(\frac{1}{N(\tau)} \sum_{i<j} [m(i) - m(j)]^2 \right)^{\frac{1}{2}}$$

Variability amplitude as a function of the time lag $\tau = t_j - t_i$

Drawback: very little information



Sesar et al. 2007

Statistical Descriptors of Quasar Variability

A modern approach: model the process as a **Damped Random Walk (DRW)**, or specifically the **Gaussian first-order continuous autoregressive process (CAR(1)**), aka the Ornstein-Uhlenbeck process, defined by the equation:

$$\mathrm{d}X(t) = -\frac{1}{\tau} X(t)\,\mathrm{d}t + \sigma\sqrt{\mathrm{d}t}\,\epsilon(t) + b\mathrm{d}t, \quad \tau, \sigma, t > 0,$$

where *X(t)* is the flux, τ is the relaxation time, σ is the variability on time scales $t < \tau$, $b\tau$ is the mean amplitude, and *ε(t)* is a white noise process

- Characterized by the variability amplitude and time scale
- Basis for the stochastic models of quasar variability
- Not a perfect descriptor: some deviations noted

CAR(1) Process

Its autocorrelation function is: $ACF(t') = e^{-t'/\tau}$

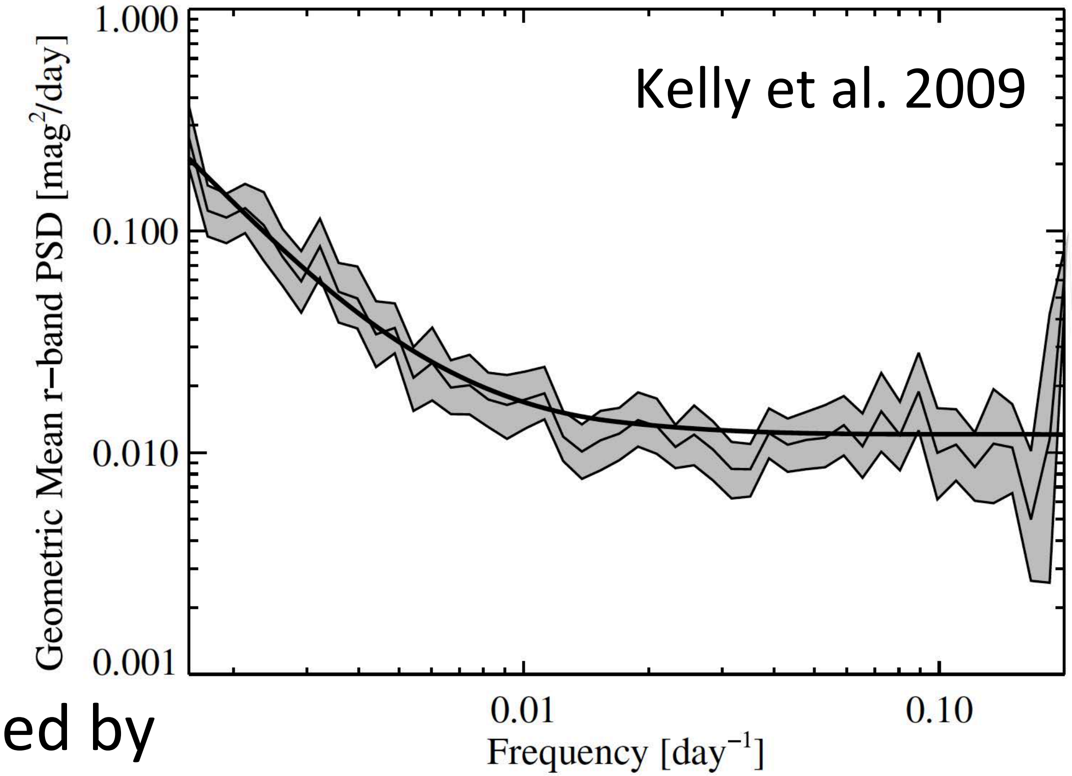
And the corresponding power spectrum:

$$P_X(f) = \frac{2\sigma^2\tau^2}{1 + (2\pi\tau f)^2}$$

τ = the relaxation time

σ = variability for $t < \tau$

$b\tau$ = the mean amplitude



It can be further generalized by allowing for a non-stationarity (moving average):

CAR(1) = CARMA(1,0) = CARIMA(1,0,0) = CARFIMA(1,0,0)

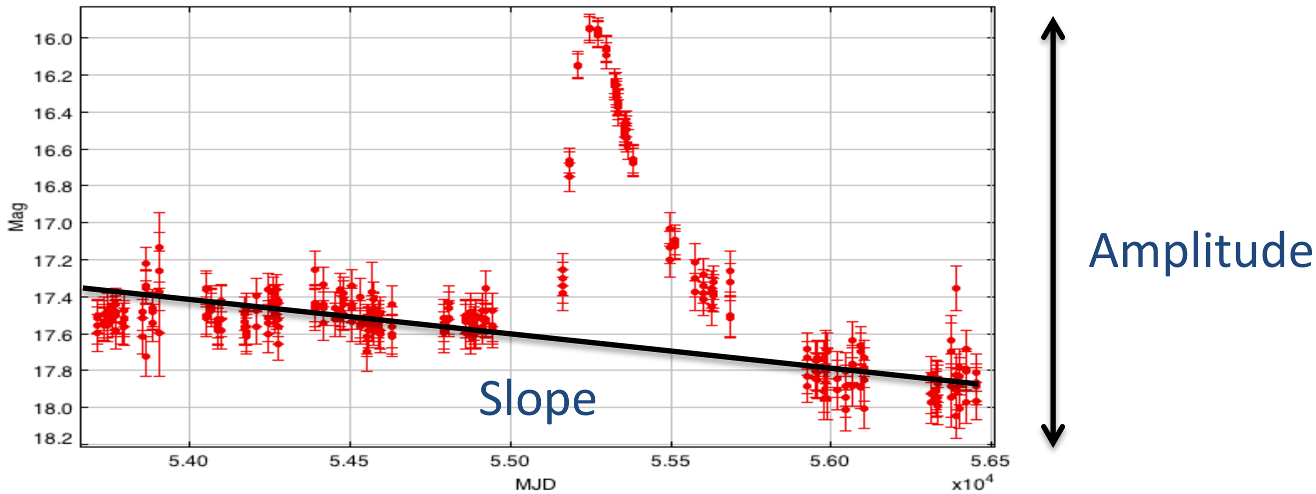
Correct modeling of the stochastic variability is ***essential*** for the analysis of the observed properties of quasars.

Variability Feature Space

- Generate homogeneous representation of time series
- Most Richards et al. (2011) features carry little information
- 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 (representative) feature space

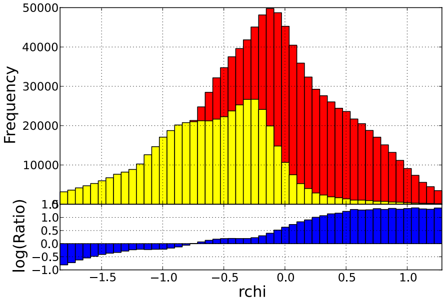
Parameter Spaces of Quasar Variability

Some are simple, but most are not

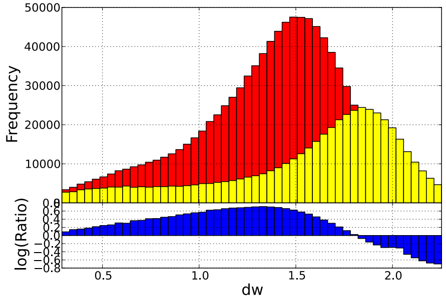


We compare them with the distributions for stars in the same magnitude range, and use machine learning tools to separate them in a multidimensional parameter space

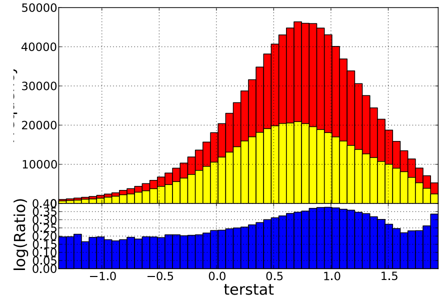
Red = quasars, Yellow = Stars, Blue = Ratio



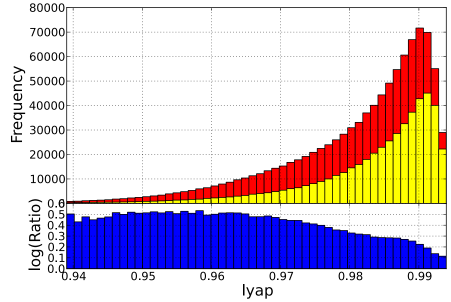
Variability



Autocorrelation

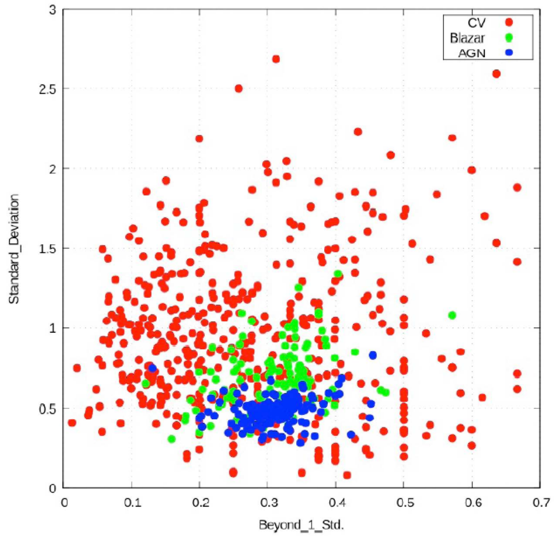


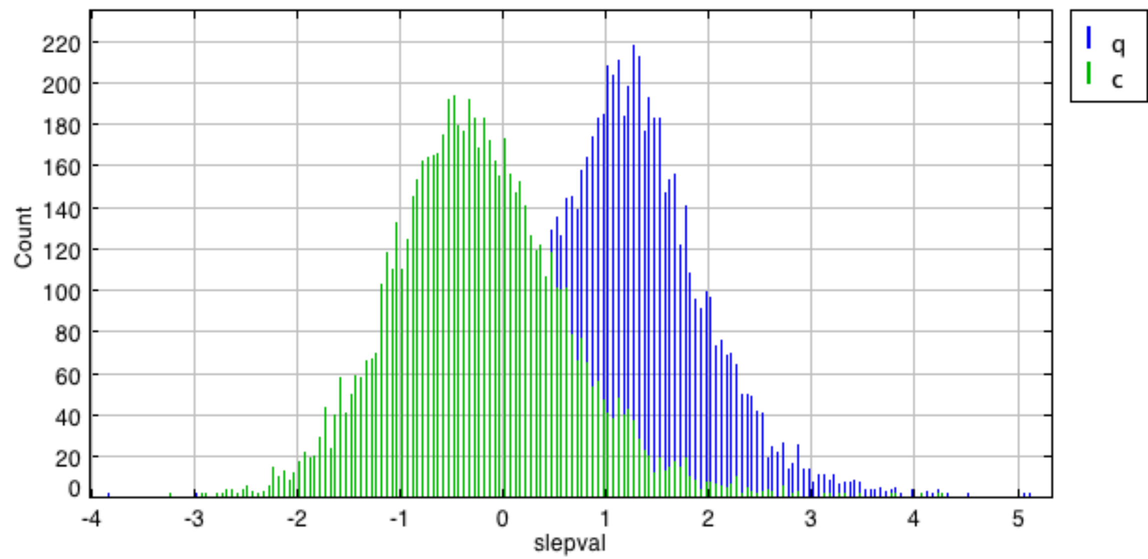
Nonlinearity

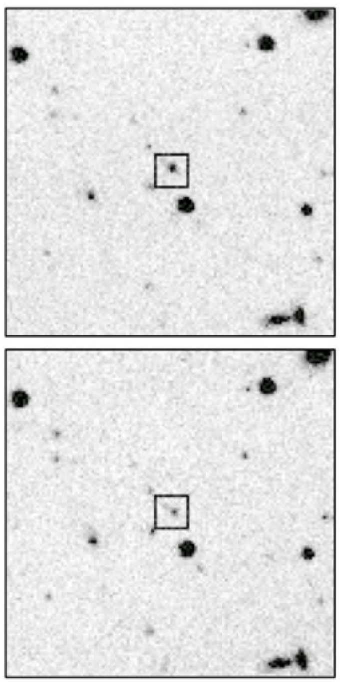


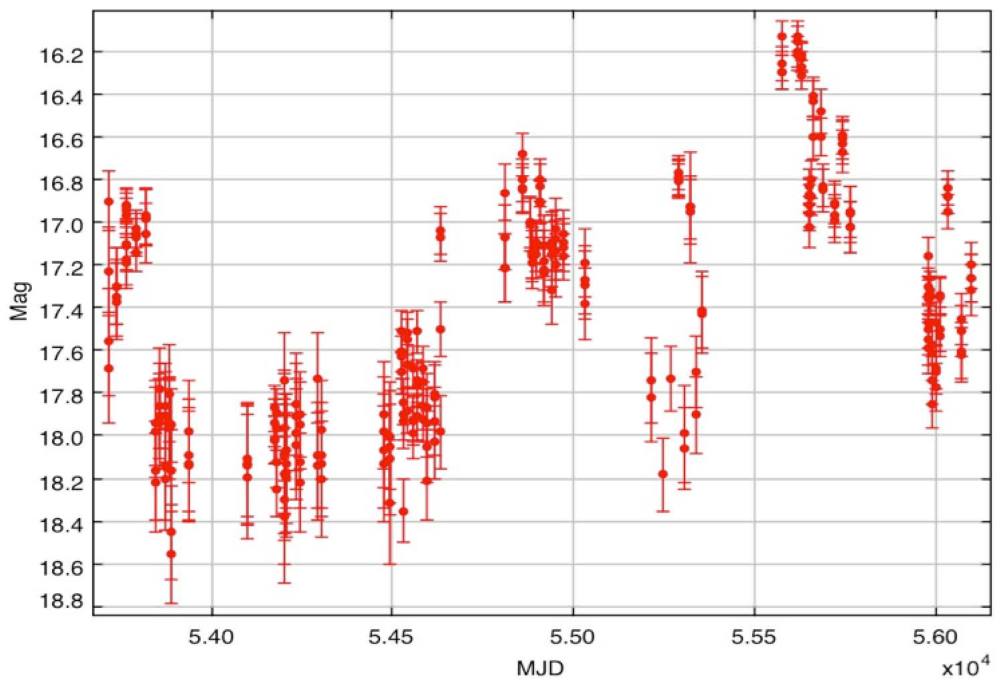
Chaos

Variability-Based Selection of Quasars



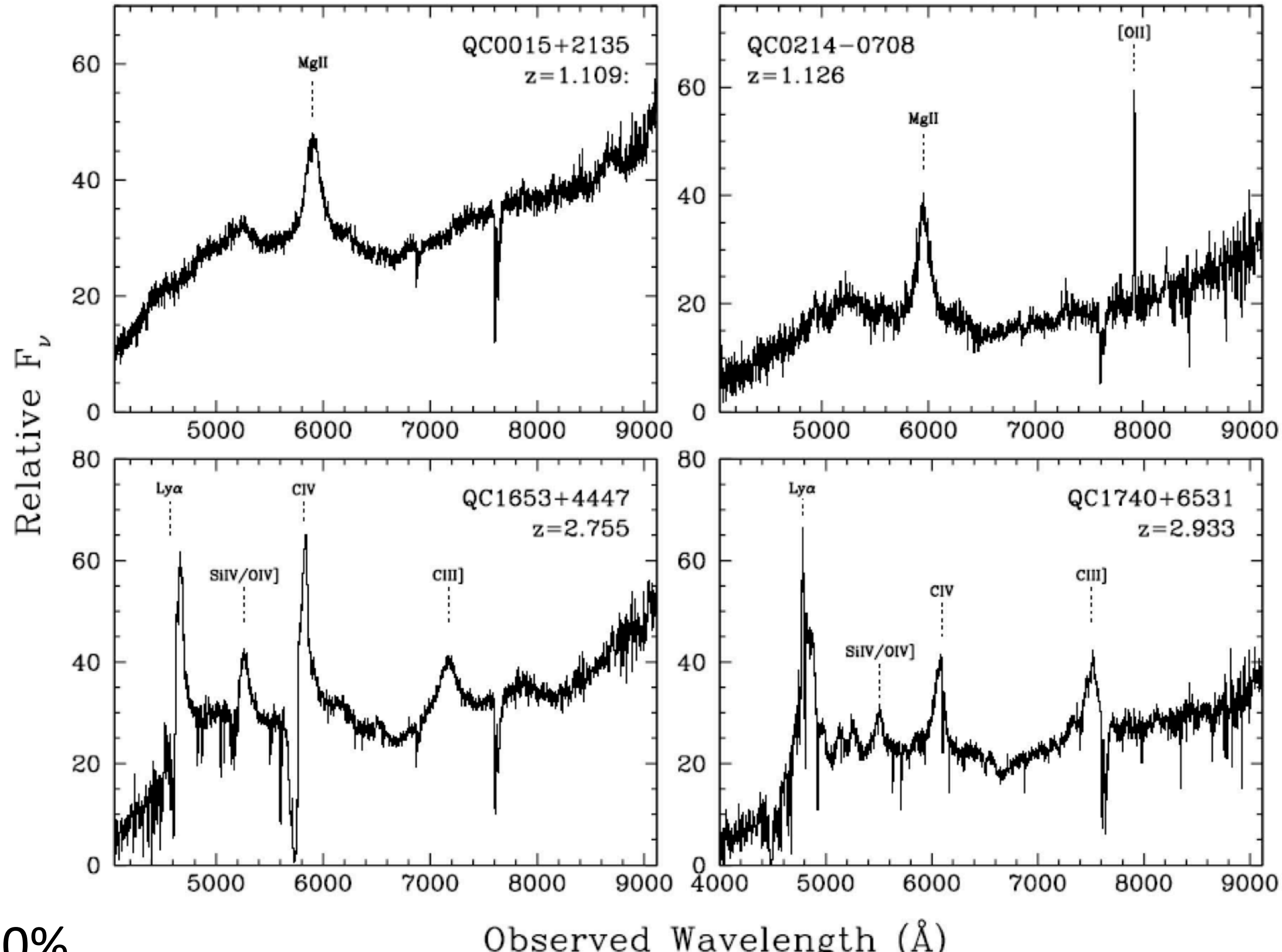






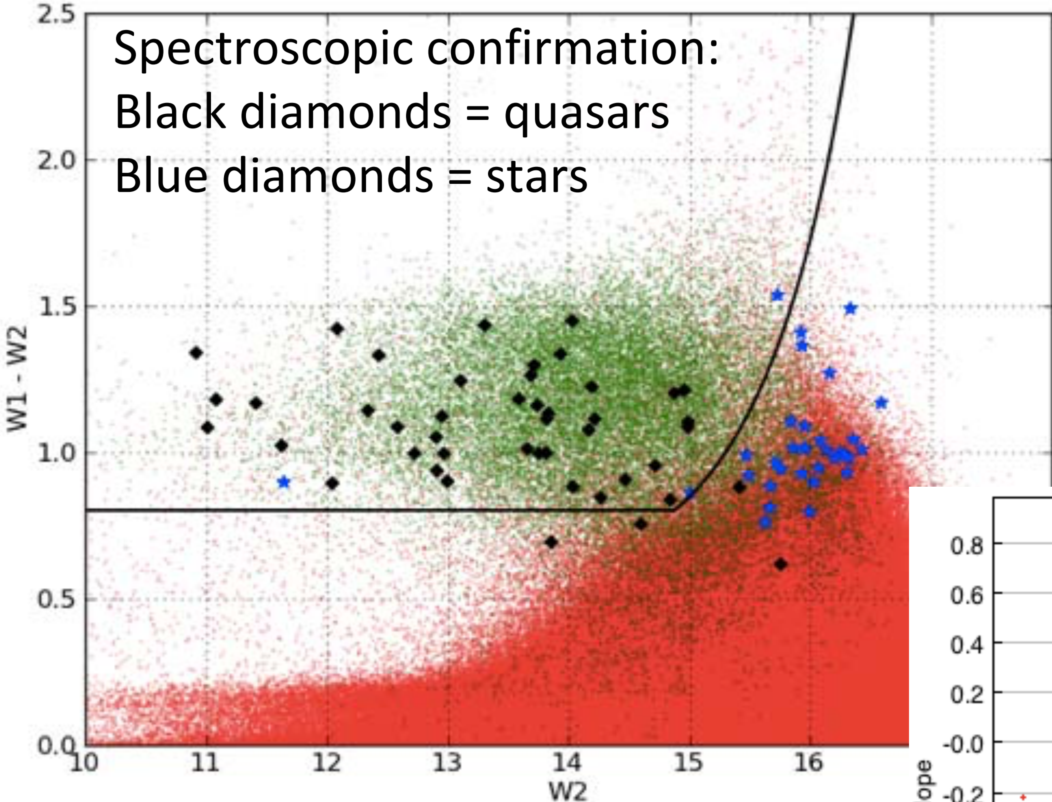
Using the light curve variability parameter space to select quasar candidates

Spectroscopic Confirmation of Variability-Selected QSO Candidates



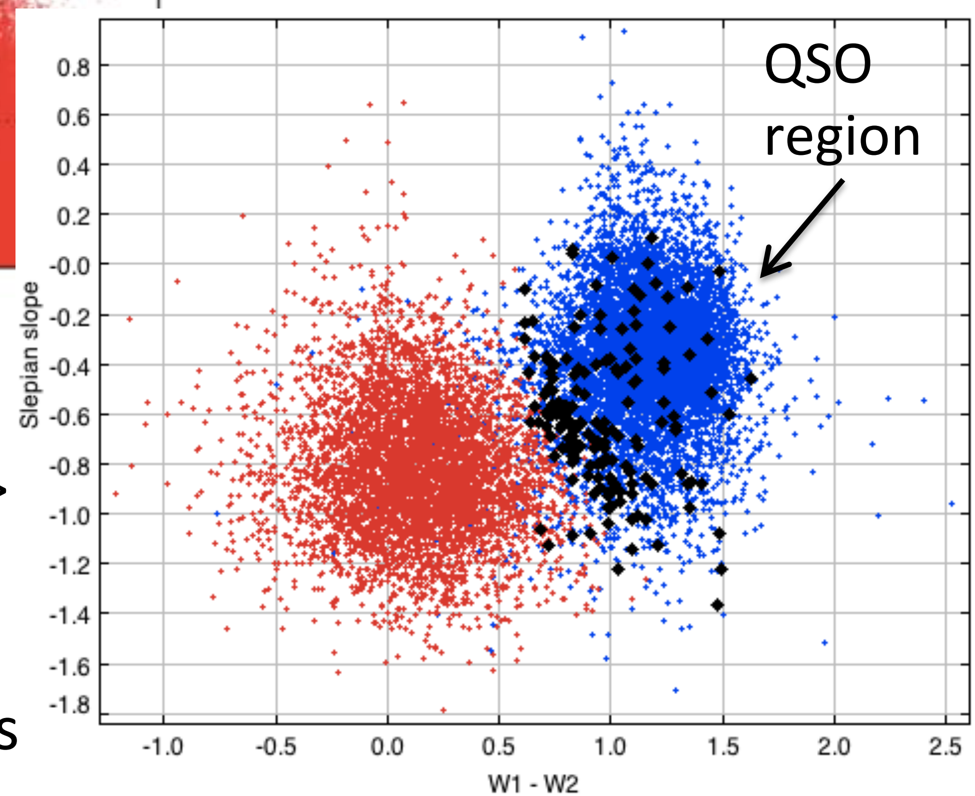
Initial success rate > 80%

Combining Variability and WISE Colors

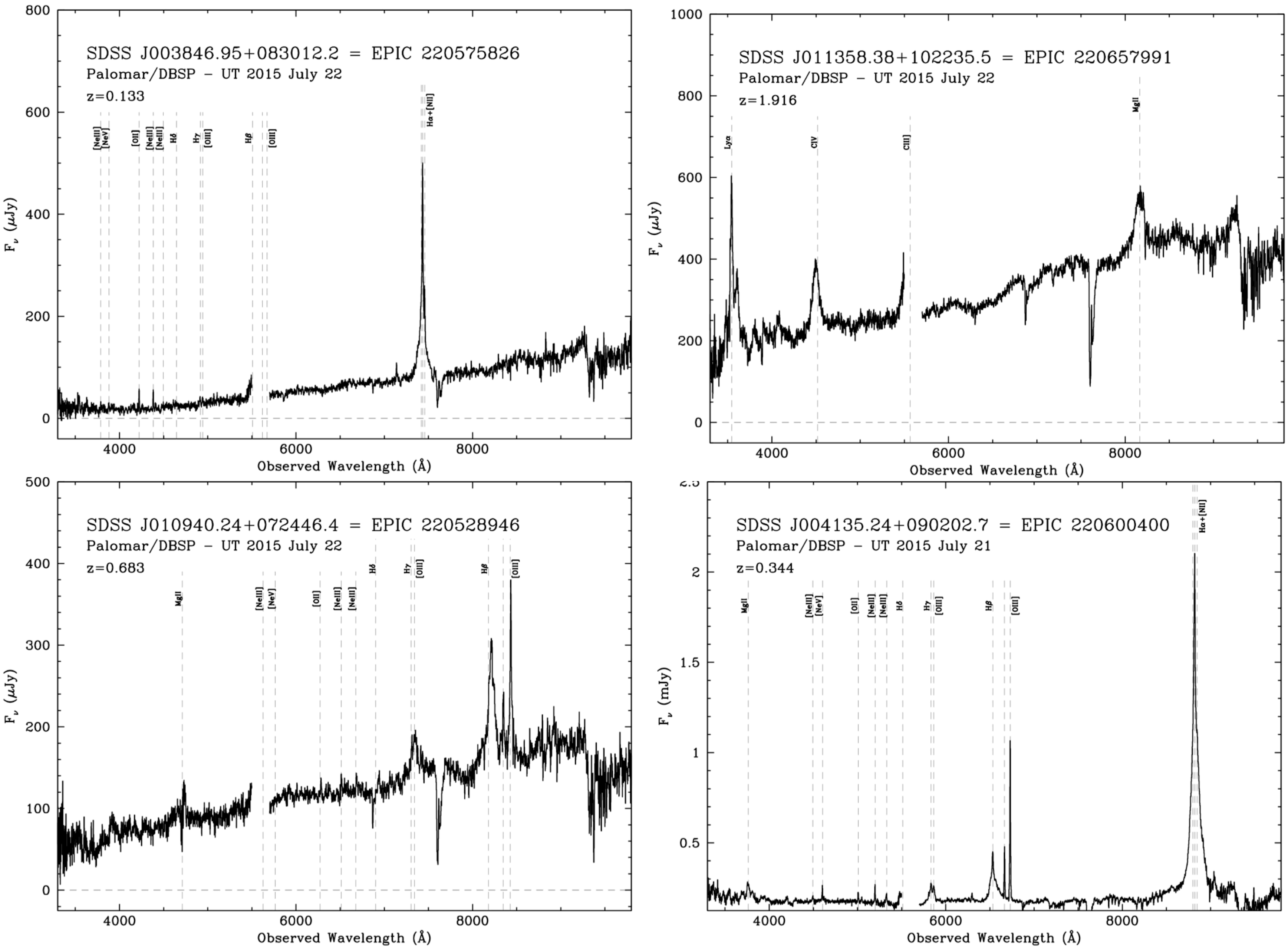


<= WISE color plot showing objects classified as stars (red) and quasars (green) by the ensemble classifier

A combined parameter space => of variability (Slepian slope, Y axis) and WISE colors. Black diamonds are confirmed quasars

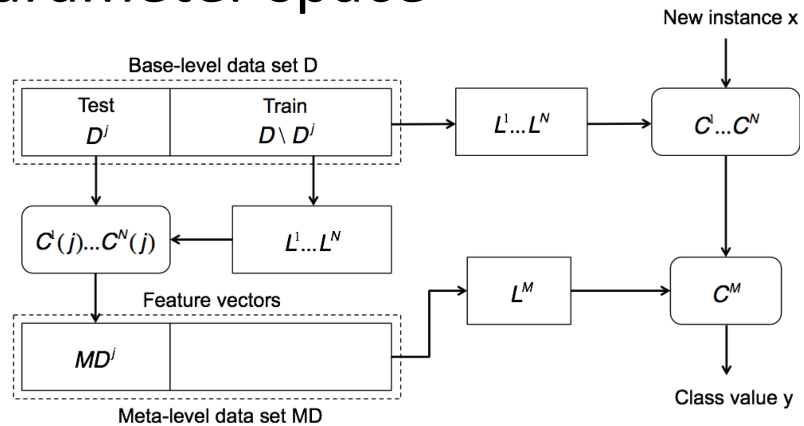


Initial results from the Kepler field: a 100% success rate!

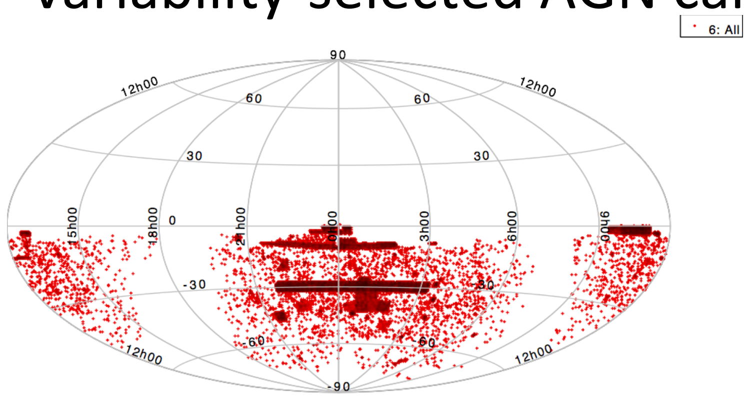


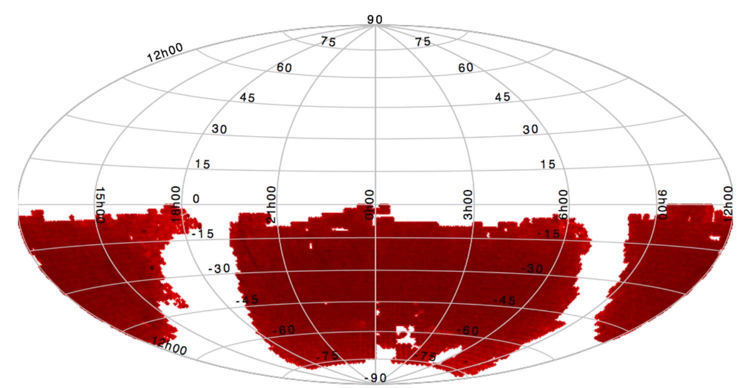
Southern Sky Quasar Catalog

- Data set of 1.5+ million QSOs/QSO candidates, selected in the WISE colors and variability parameter space
  - Stacked framework for ensemble classification



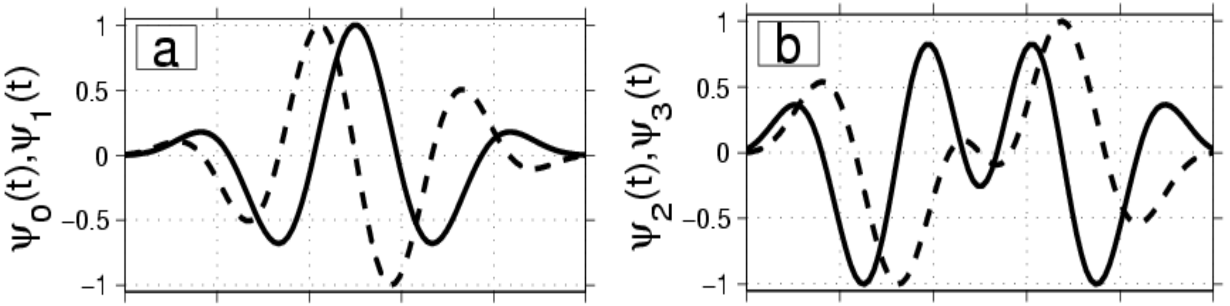
- CRTS Southern Quasar Catalog
  - Within the SSS footprint, there are 25,828 spectroscopically-confirmed AGN
  - The first pass SSS quasar catalog has 454,763 color and variability-selected AGN candidates to V ~ 19.5





Wavelet Decomposition of Light Curves

- 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:

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

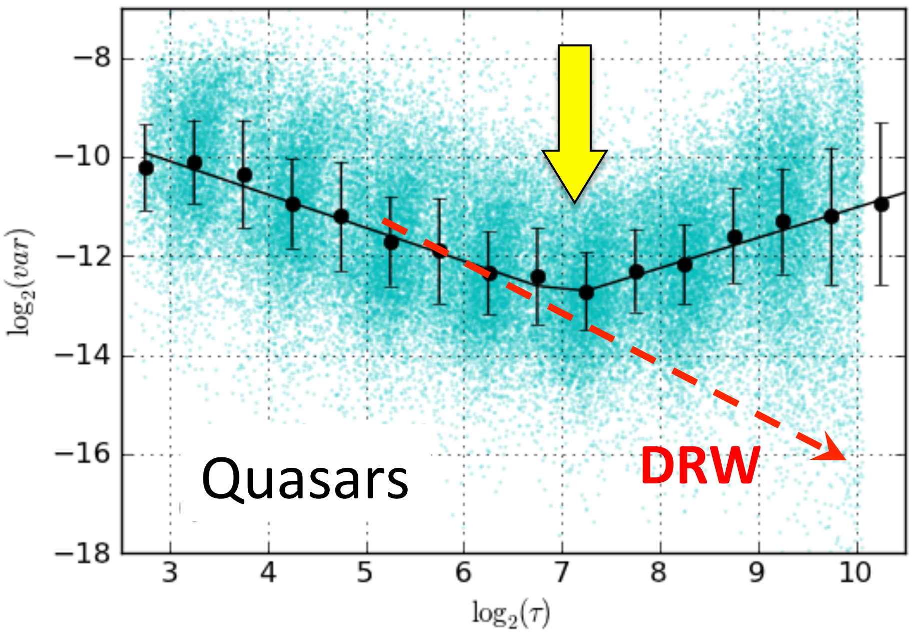
is the total variance contribution due to scale τ_j

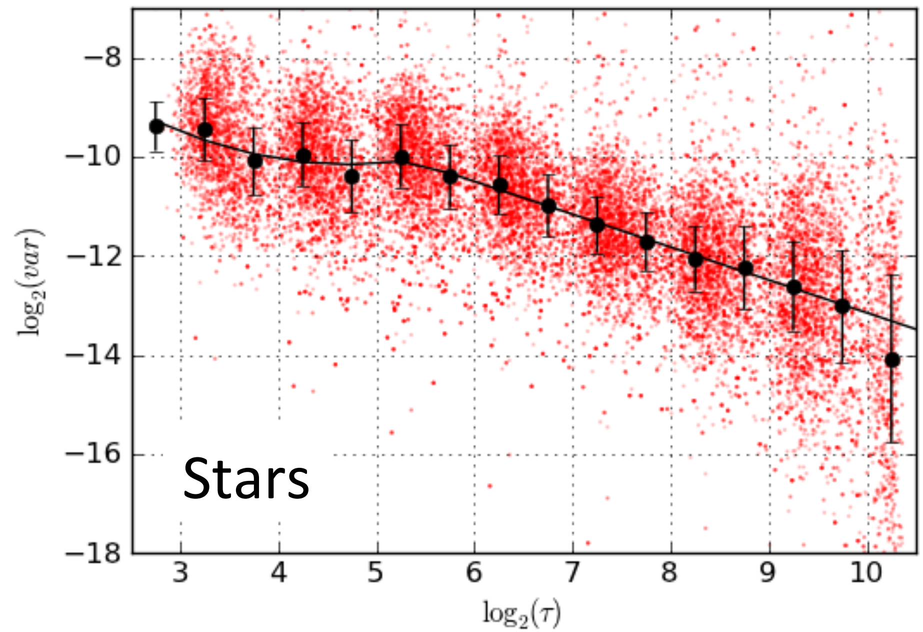
- Characteristic scales are indicated by peaks or changes of behavior in $\log(\nu^2_X)$ vs. $\log(\tau_j)$

- **Slepian wavelets** work with ***irregular and gappy time series***

A Characteristic Time Scale for a Stochastic Variability, from the Wavelet Analysis

(Graham et al. 2014)

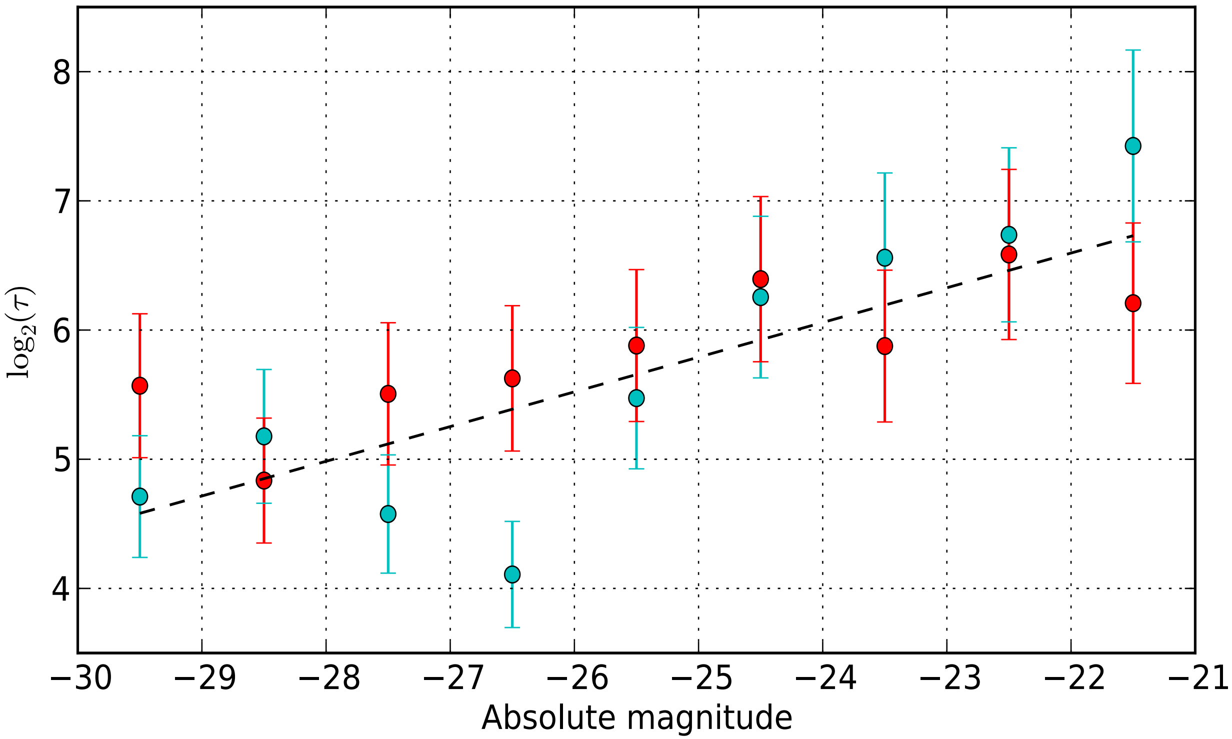




- Quasars deviate from the pure, correlated noise CAR(1) process, that was established by numerous studies
- There is a characteristic time scale, ~ 54 day in the restframe
- Its physical origin is not yet established

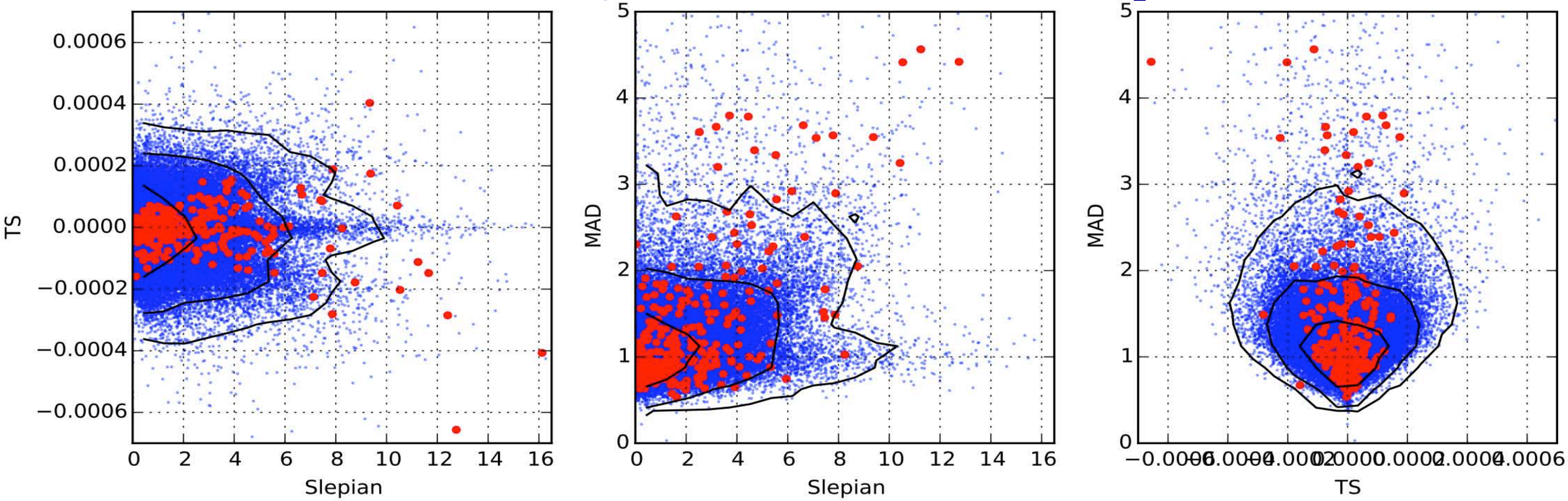
Evidence for a Characteristic Time Scale

- First solid evidence for a characteristic time scale (~ 50 days) associated with the quasar variability in the visible
  - Previous indications in the X-ray
- Possible probe of the accretion disk physics
  - Diffusion time scale in the outer regions of the accretion flow?
- Anticorrelated with the luminosity, and possibly with other physical parameters (work in progress)



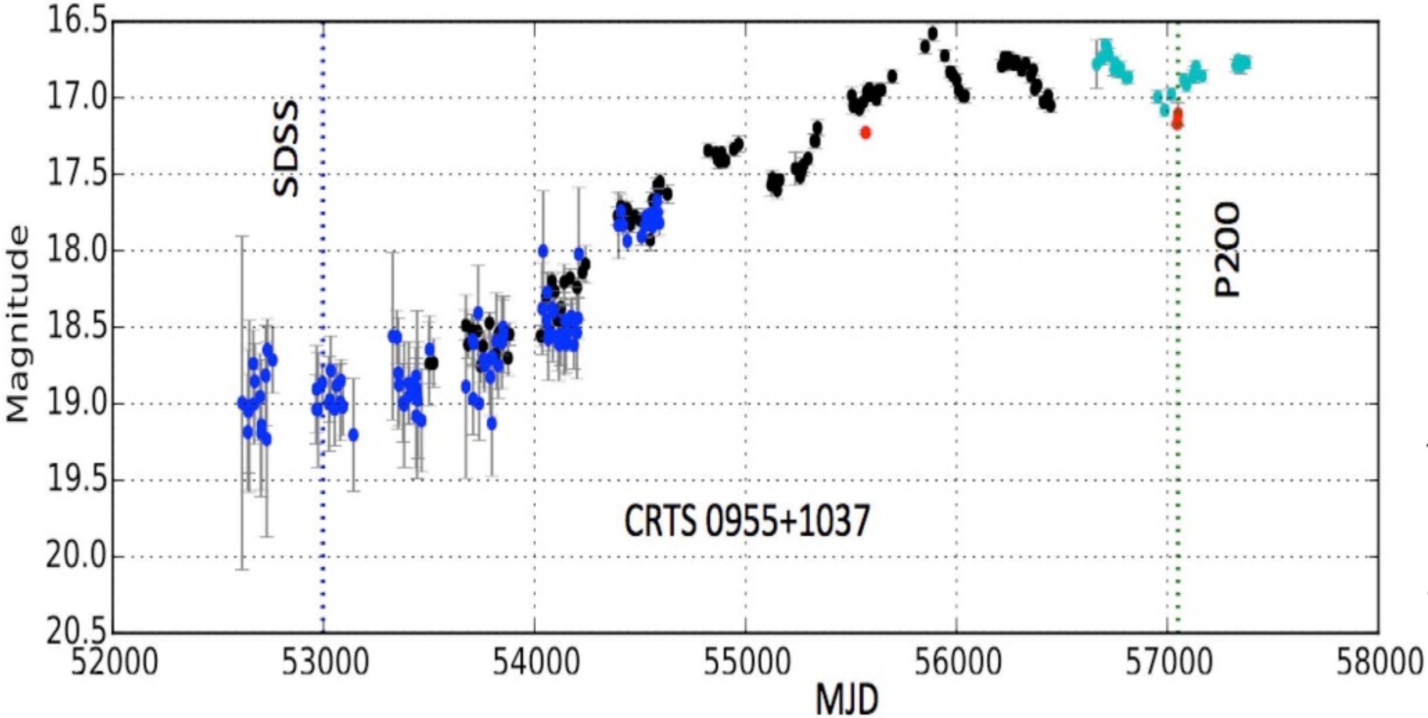
(Graham et al. 2014)

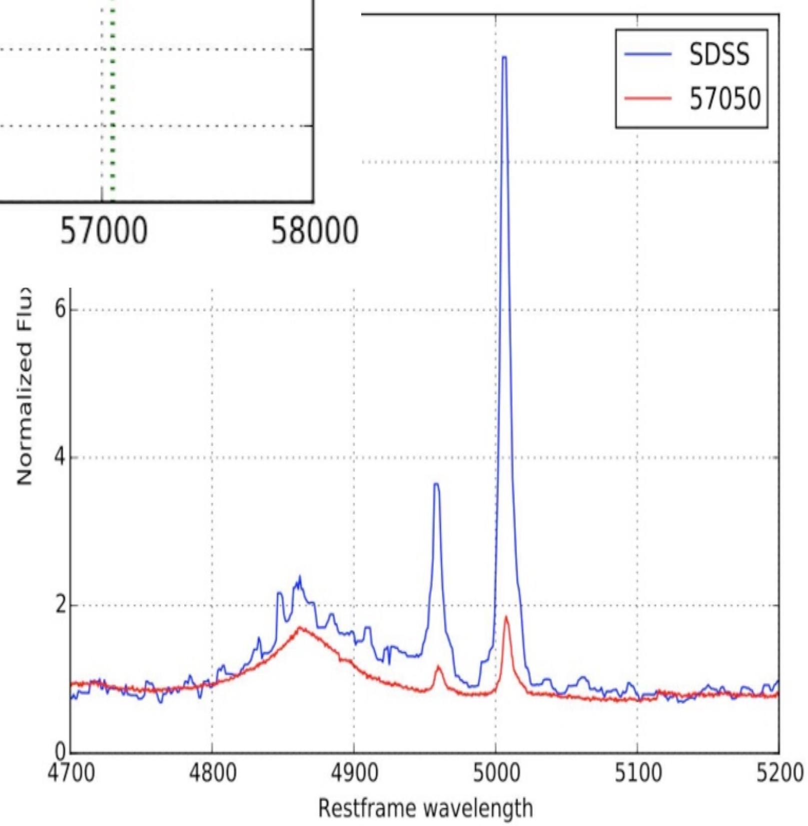
Looking for Outliers in the QSO Variability Parameter Space



- Identify quasars with anomalous/unusual variability patterns as outliers in the variability parameter space
- Spectroscopic follow-up + archival spectra, to look for a **correlated photometric and spectroscopic variability**
- Quasars with large, gradual changes in flux contain at least three different types of interesting objects

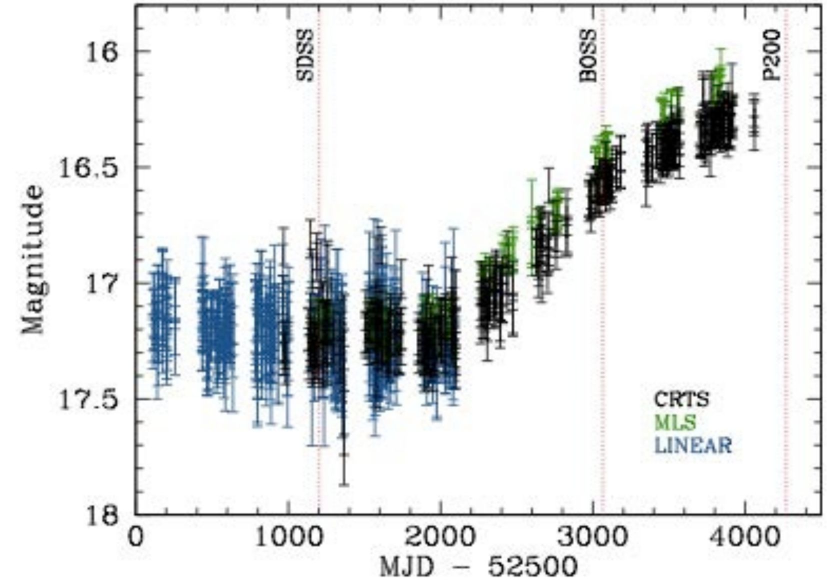
Some Are Changing Look (Type) Quasars

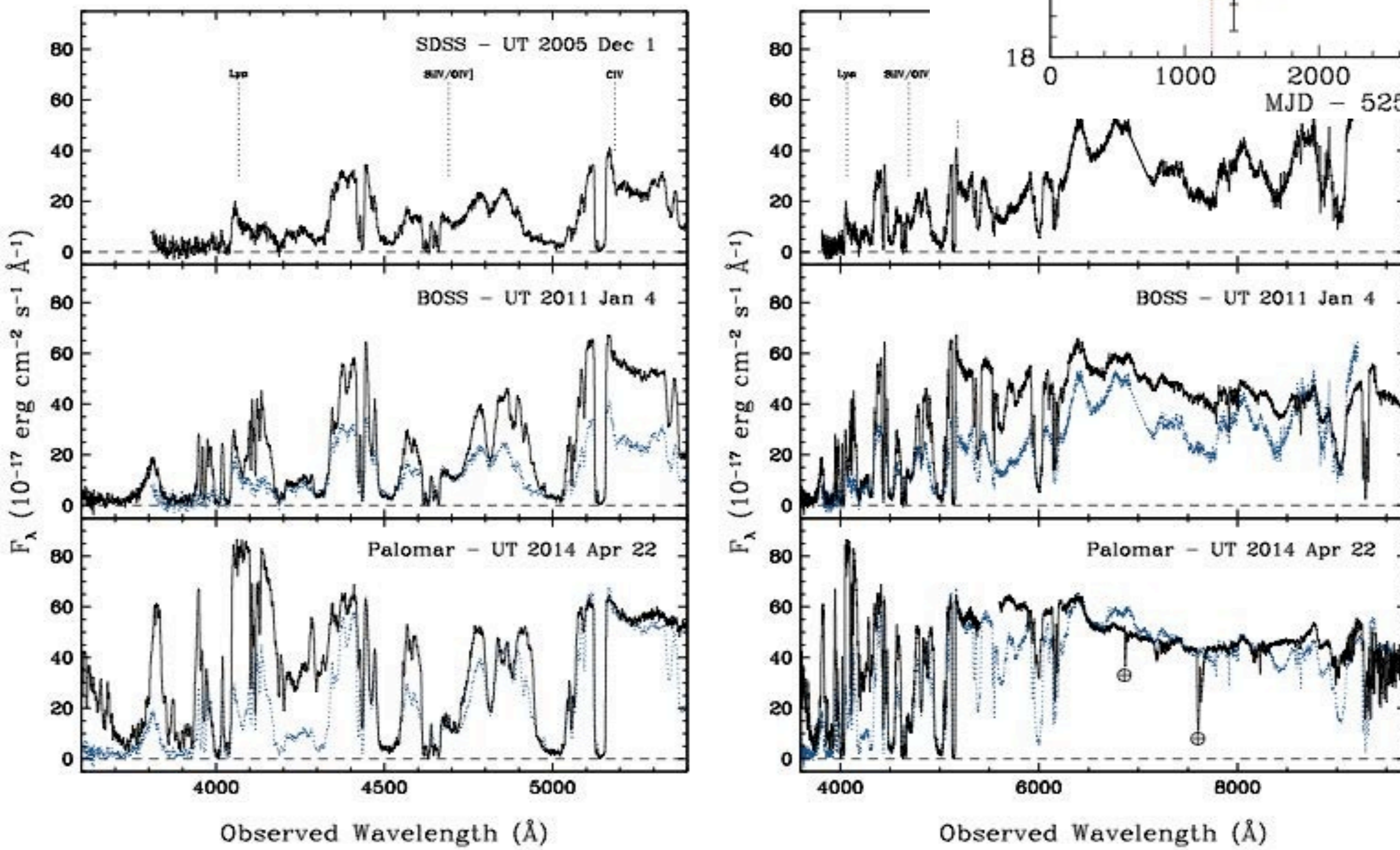




- Correlated photometric and spectroscopic change, Type I to Type II, or v.v.
- Indicative of changes in the accretion rate or obscuration

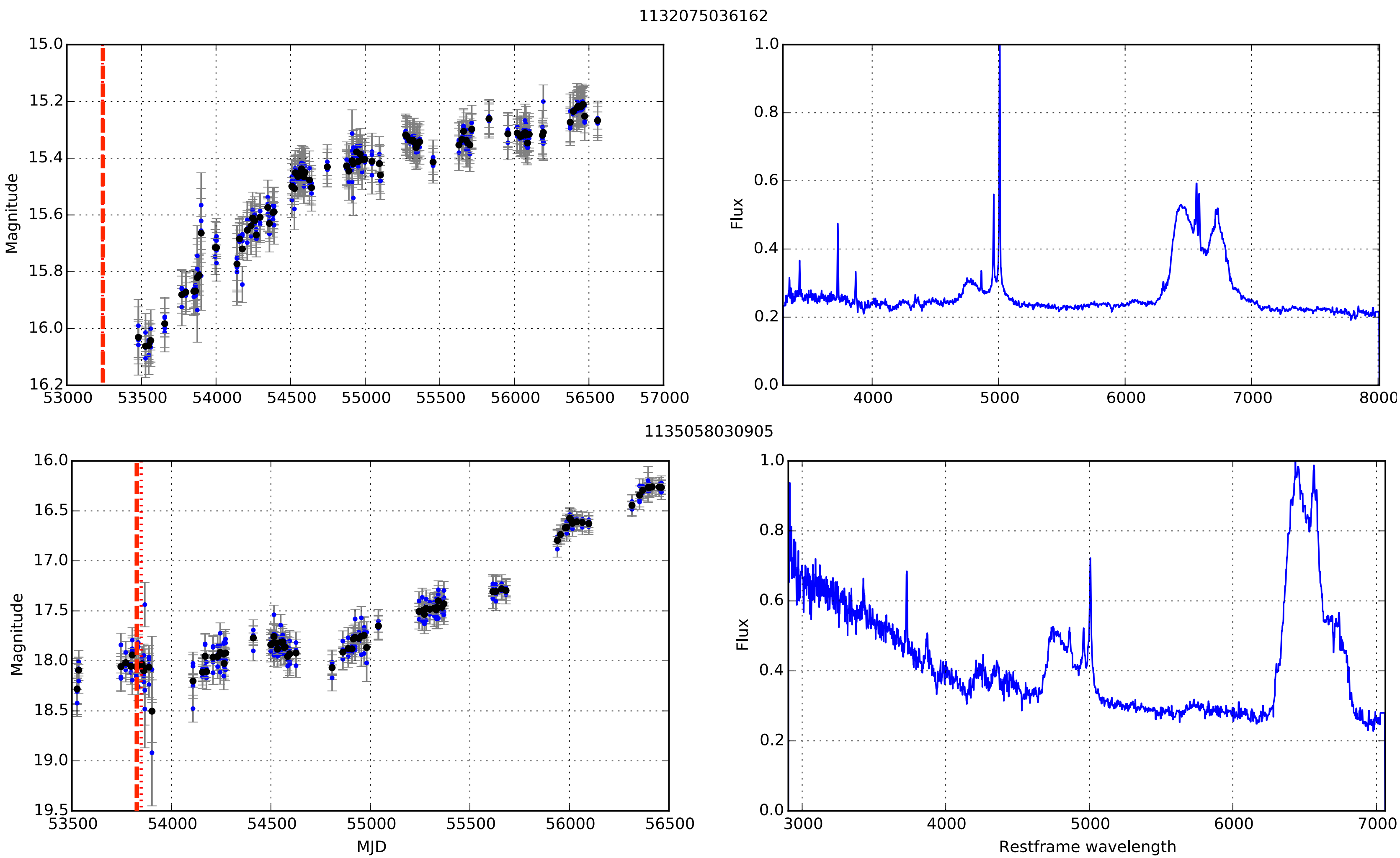
Fe-LoBAL quasar with time-varying absorption trough depths, correlated with a rise in luminosity. Suggests changes in the photoionization equilibrium.





Stern et al. 2017

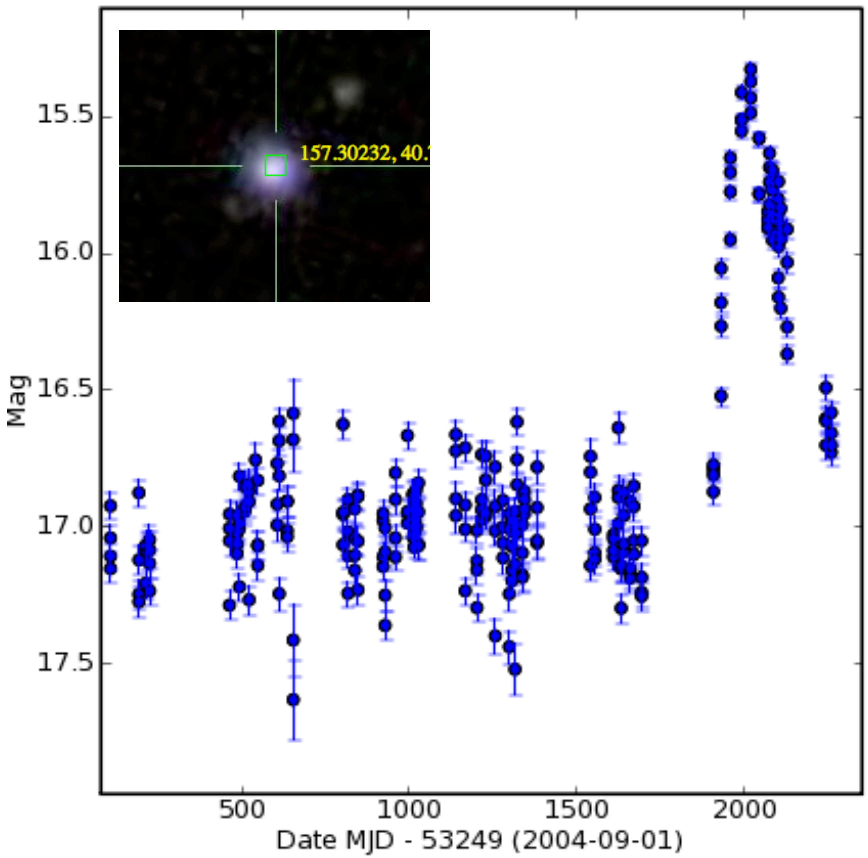
Some Are Double Peak Emitters

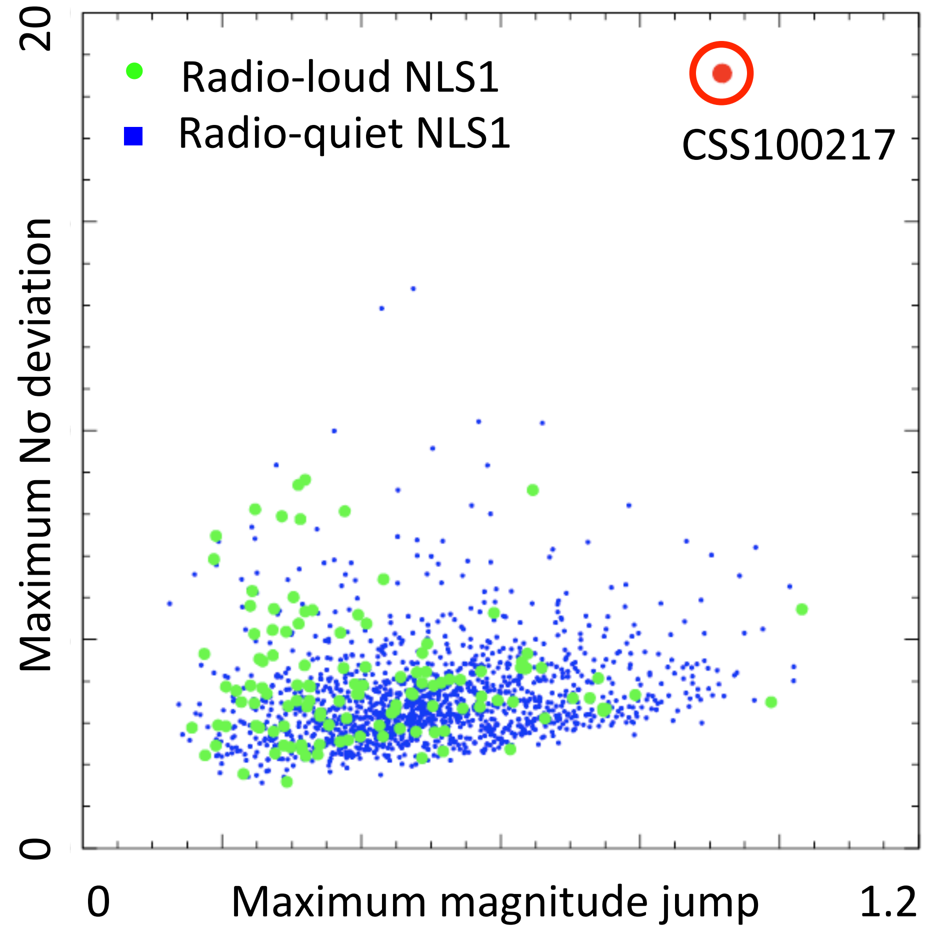


Known to be spectroscopically variable. Believed to be caused by instabilities in the outer regions of the accretion disks

The Case of CSS100217:102913+404220

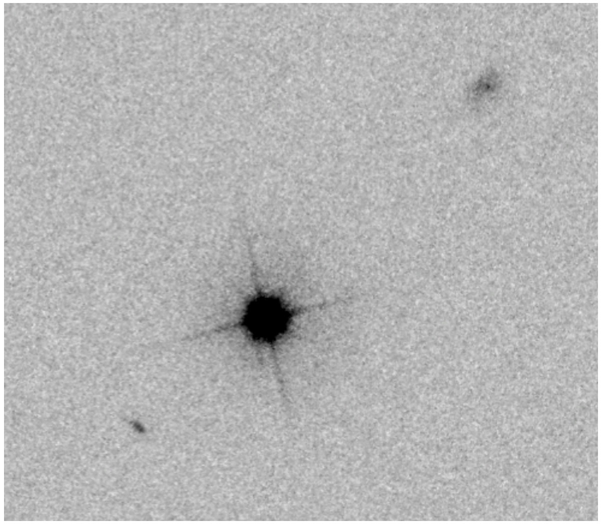
Drake et al. 2011, ApJ 735, 106





- Transient in a narrow-line Seyfert 1 (NLS1) galaxy at $z = 0.147$
- Peak $M_I \approx -23$ mag, integrated visible luminosity > 6×10^{51} erg
- *SWIFT* and *GALEX* ToO obs. exclude a “traditional” TDE

The Nature of CSS100217



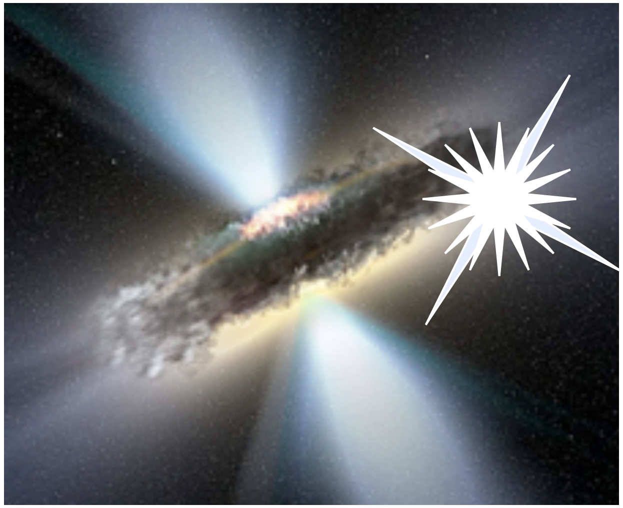
HST ToO and Keck AO+LGS imaging shows a single, unresolved point source:

The event within ~ 150 pc from the AGN

No morphological indications of star forming regions or dust outside of the unresolved nucleus

Vicinity of an AGN is not conducive to star formation, except…

… near the outer edge of the accretion disk, which is shielded from the UVX radiation, and should be violently unstable

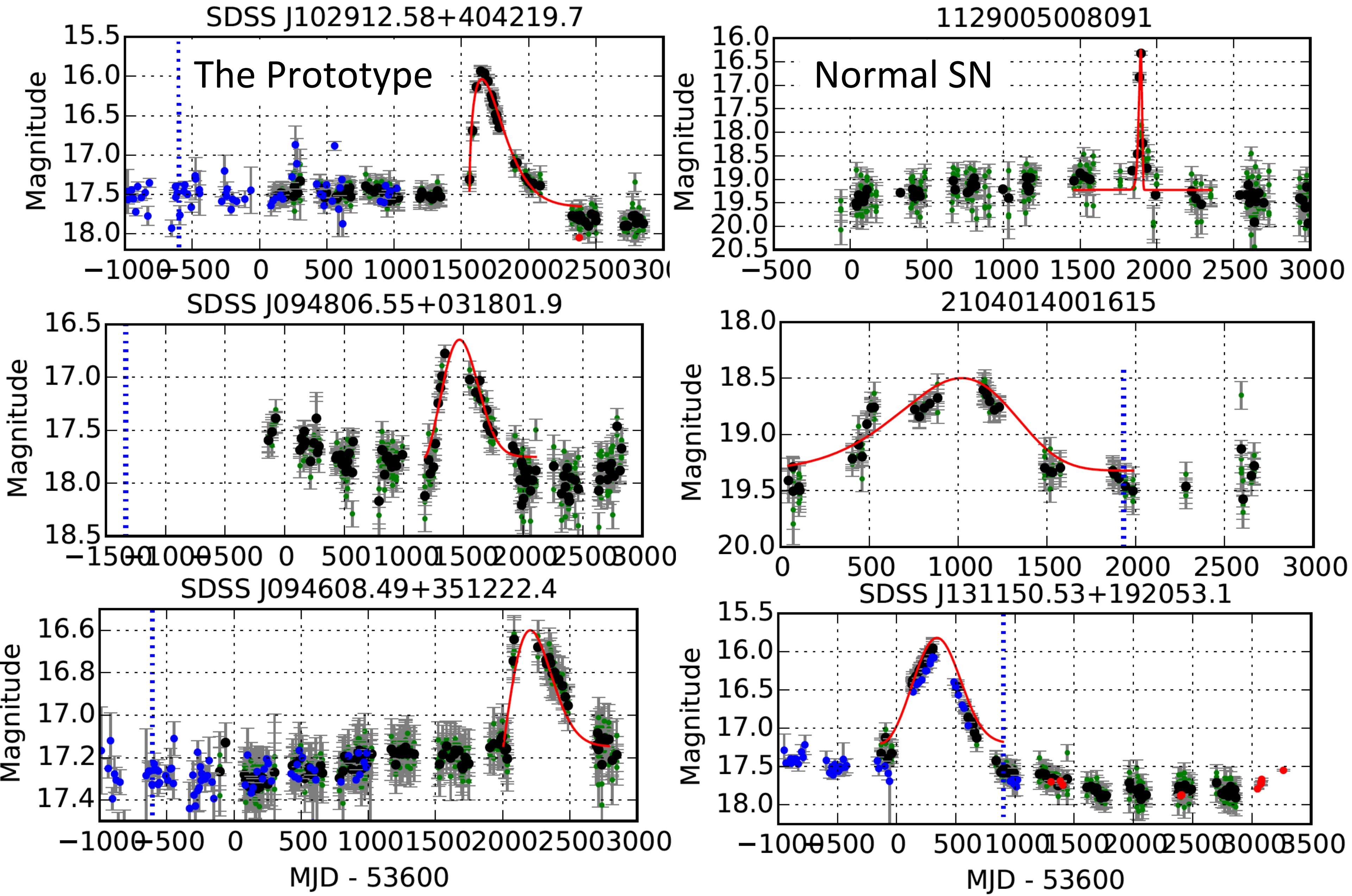


The first case of a SN from an AGN accretion disk?

Predicted by theory but never previously seen

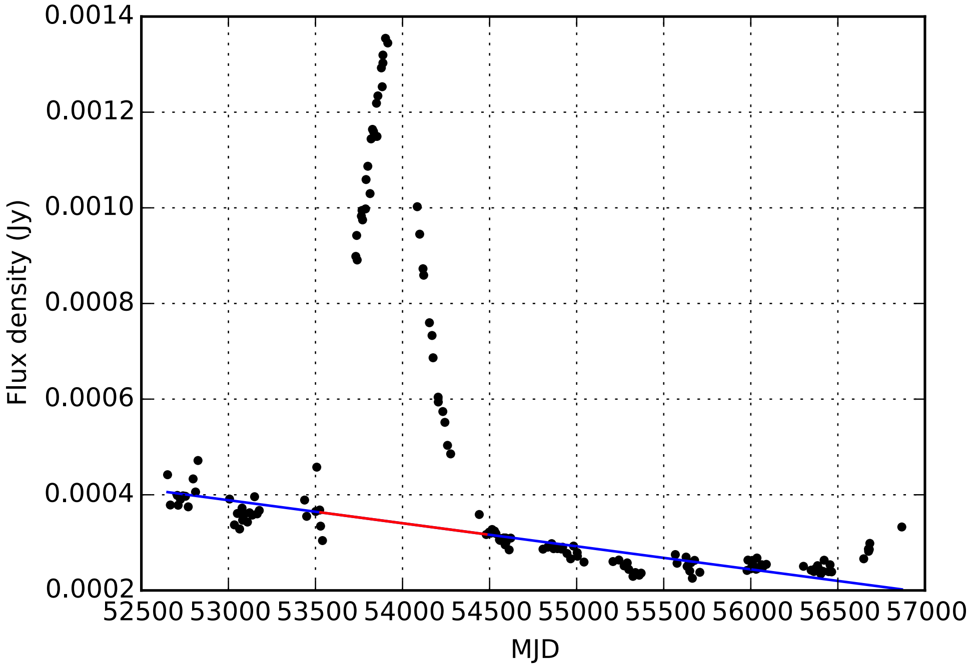
Quasar Megaflares

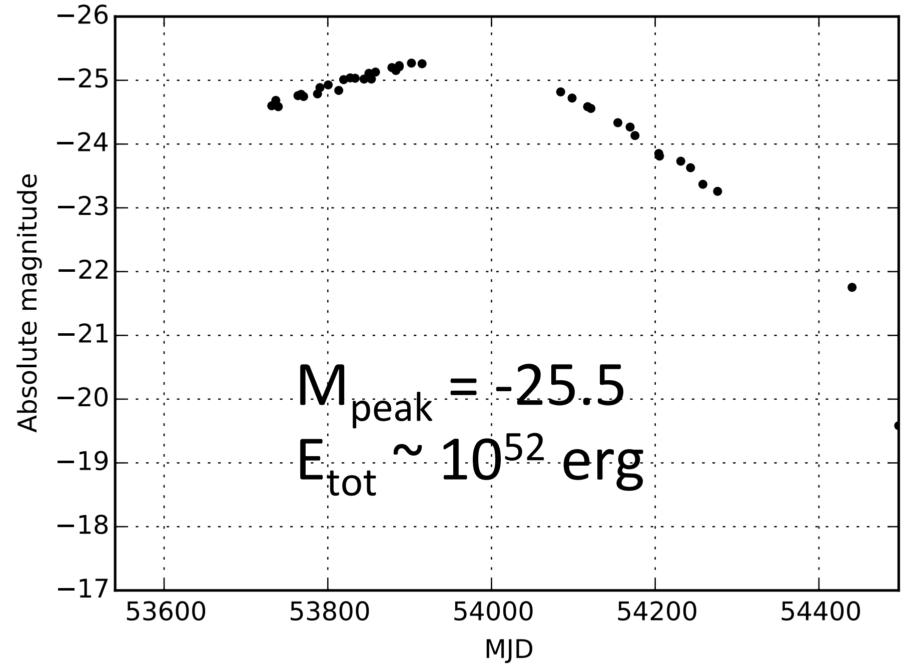
(Graham et al. 2017)



A Mixed Population?

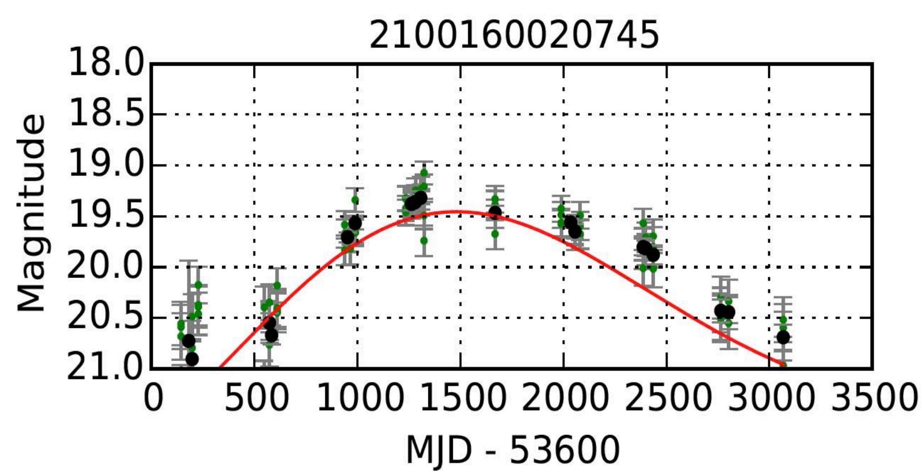
Some reach unprecedented energetics for SNe:





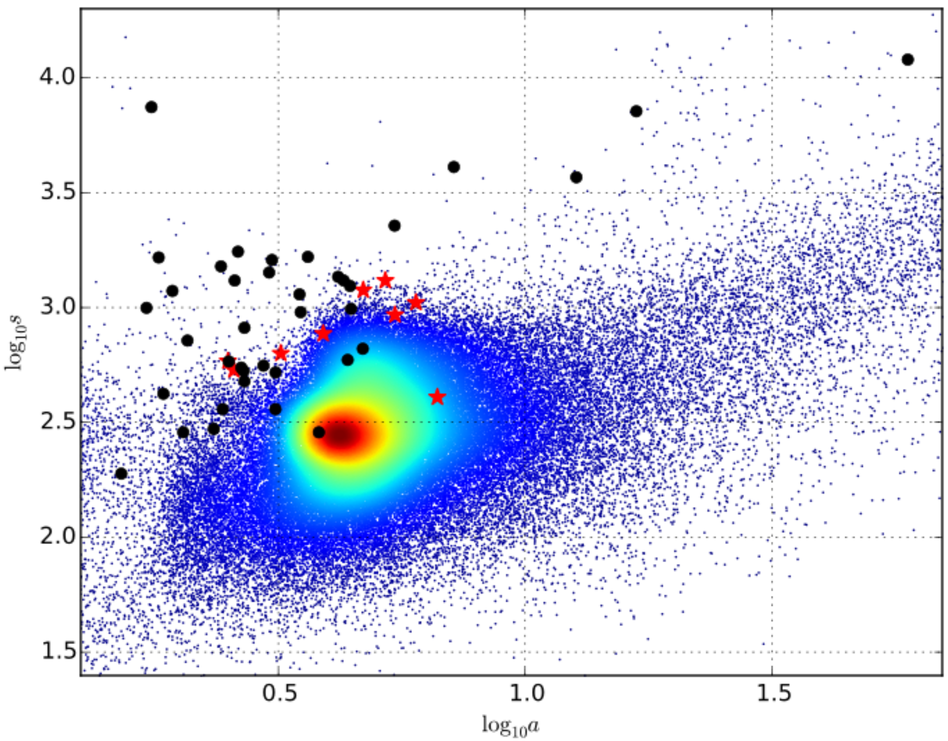
The light curves do not match the traditional TDEs, and some are too broad and/or symmetric for SNe

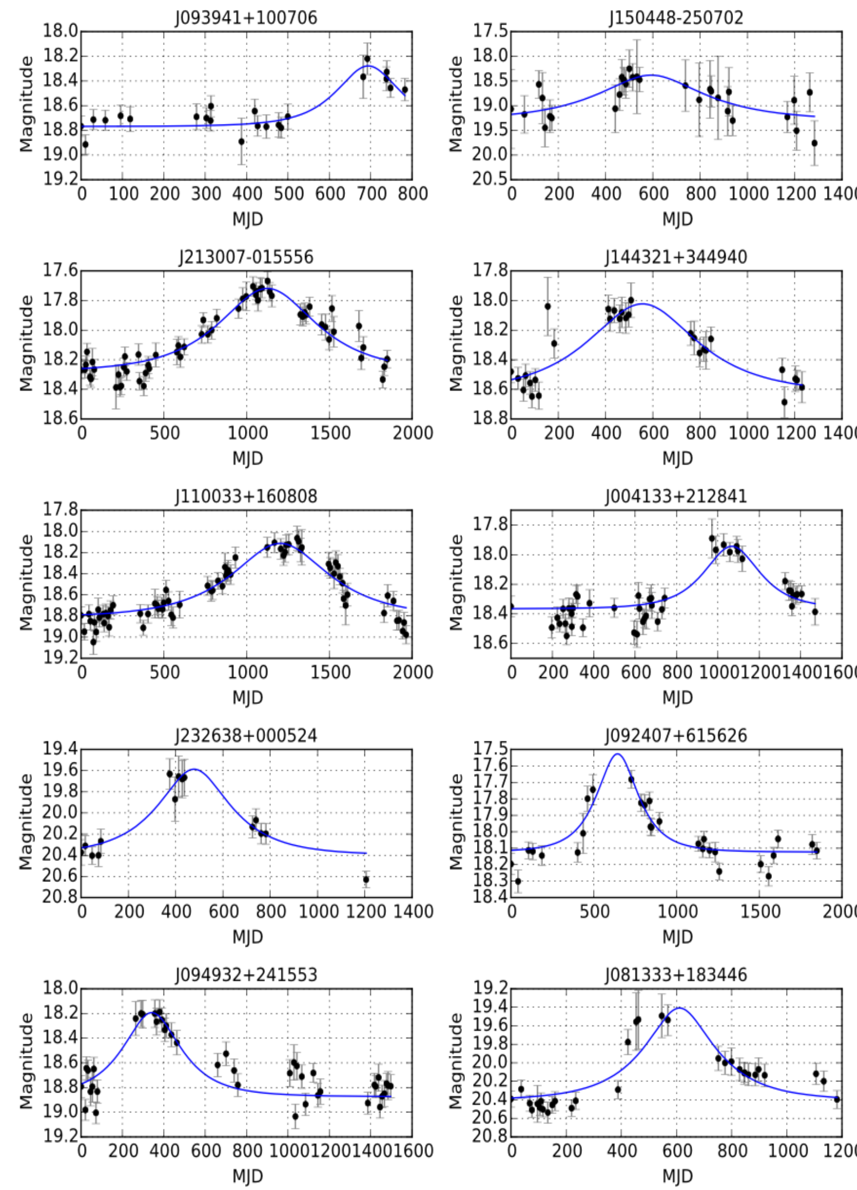
Some may be gravitational microlensing events



Does Lensing Explain All Events?

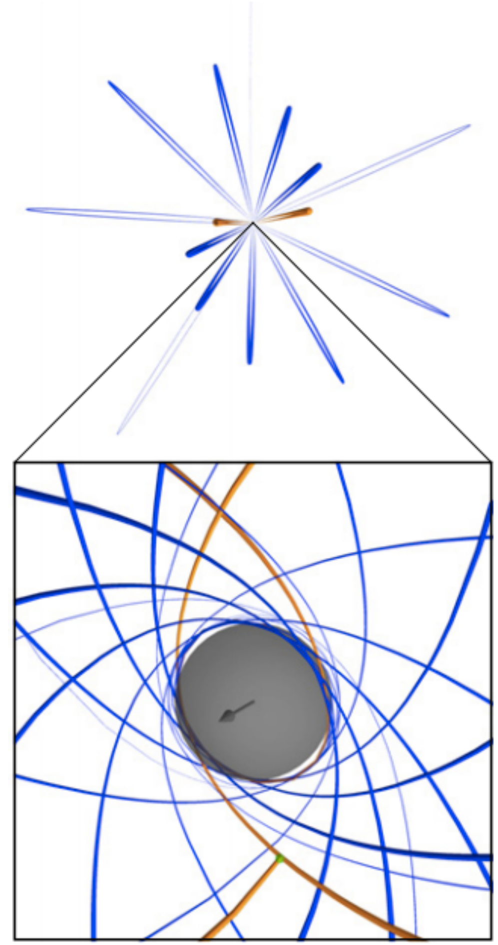
- Weibull characterization for 100,000 simulated single-point single lens model with data priors
- Best-fit MCMC single-point single lens model to detected





Other Possible Explanations

- Superluminous supernova (SLSN-II)
  - J102912+404220 (Drake et al. 2011) within 150pc of the nucleus of NLS1
- Slow TDE (spinning SMBH)
  - Relativistic precession from black hole spin prevents the TDE debris stream from self-interacting until after many windings
  - Not for M > 10^8 solar masses
- Stellar mass black hole merger
  - Potentially important dynamic sub-channel

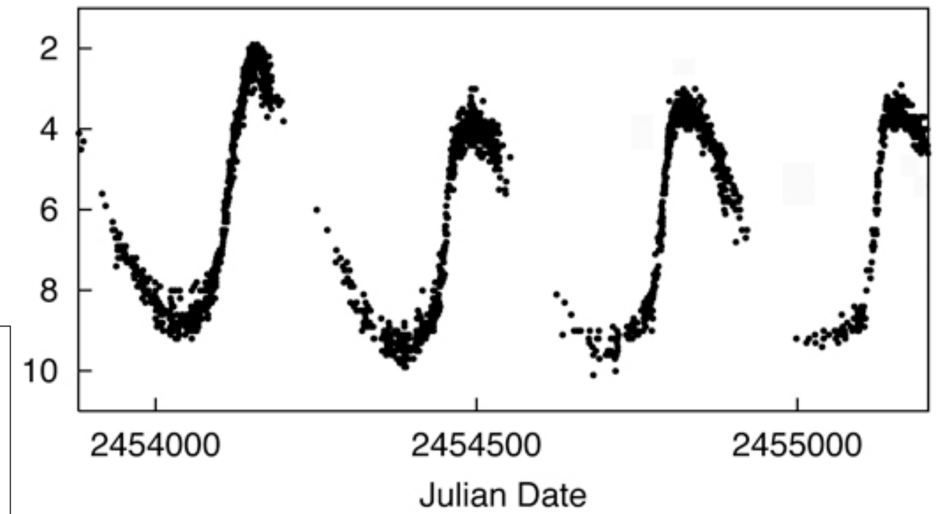


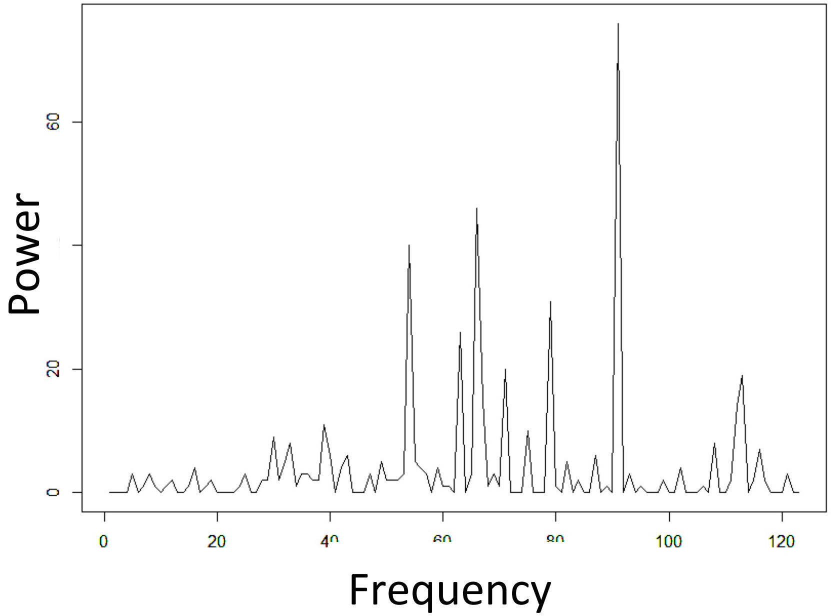
=> Explosive stellar activity in the accretion disk

- Accretion disk wind – BLR interaction?

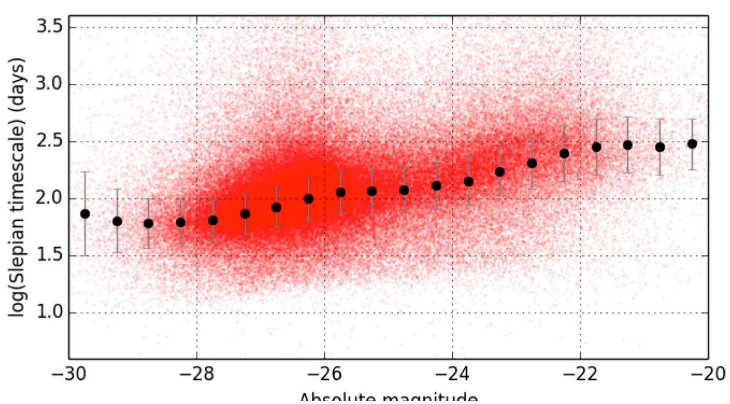
Measuring Periodicity

Early 20th Century:
count the waves





Late 20th Century: periodograms



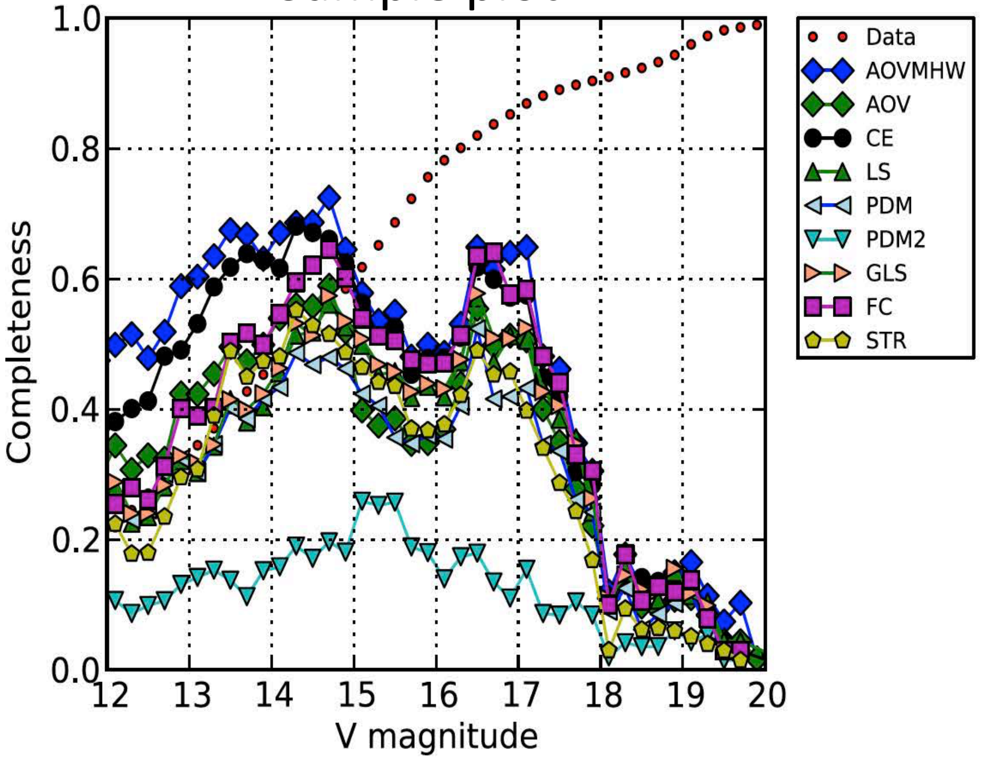
Early 21st Century:
detailed process modeling

$$W(\omega, \tau, X_t) = \sqrt{\omega} \sum_{\alpha=1}^{N} X_{t_\alpha} f^*(\omega(t_\alpha - \tau))$$

Accuracy of Period Estimates

Graham et al. (2013, MNRAS, 434, 3423) did a systematic comparative study of **9 different period finding algorithms** for a variety of periodic variable types from MACHO, CRTS, and ASAS, as a function of magnitude, for different samplings, S/N, etc.

Sample plot



All methods generally measure the periods with a reasonable accuracy over a 10 yr baseline in only **~ 50%** of the cases. If just a detection of periodicity is needed, the success rate is **~70%.**

The best performing method is the ***Conditional Entropy.***

Coditional Entropy Method

Graham et al. (2013, MNRAS, 434, 2629)

A time series, $m(t_i)$, is normalized to occupy a unit square in the (φ, m) plane where φ_i is the phase at t_i for a trial period, P.

The square is then partitioned into k bins, and the Shannon entropy for the distribution, H_0, is given by:

$$H_0 = -\sum_{i=1}^{k} \mu_i \ln(\mu_i)$$

where μ_i is the occupation probability for the non-empty bins.

The Conditional Entropy is:

$$H_c = \sum_{i,j} p(m_i, \phi_j) \ln\left(\frac{p(\phi_j)}{p(m_i, \phi_j)}\right)$$

where where $p(m_i, \varphi_j)$ is the occupation probability for the corresponding bins, and $p(\varphi_j)$ is integrated over all m_i's.

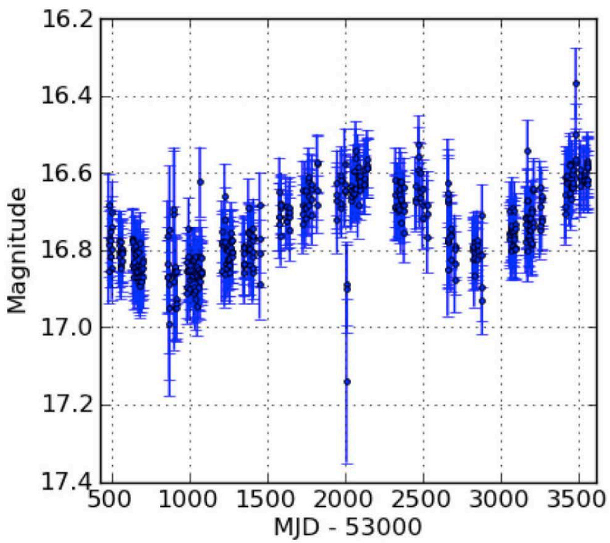
H_c is computed for every trial period P, and the smallest value is interpreted as the true period.

Finding Periodic Signals in a Red Noise

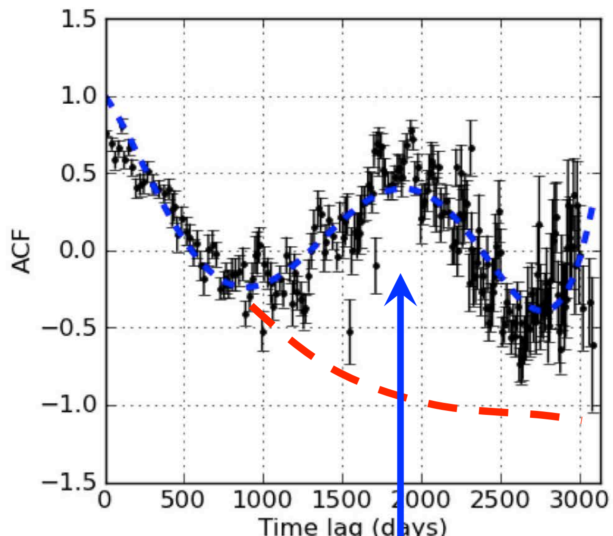
If a periodic variability is present, there will be peaks in the autocorrelation function ACF at the multiples of the period

For the irregularly sampled, gappy data, the best estimator is the z-transform based discrete correlation function (ZDCF) defined by Alexander (2013)

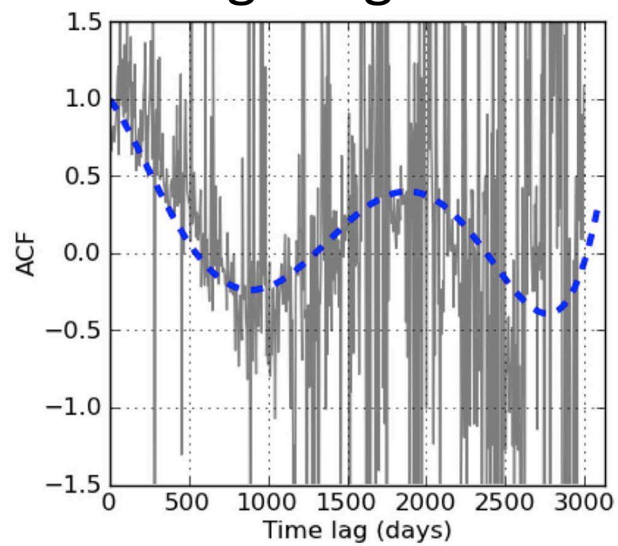
A quasar light curve



ZDCF based ACF



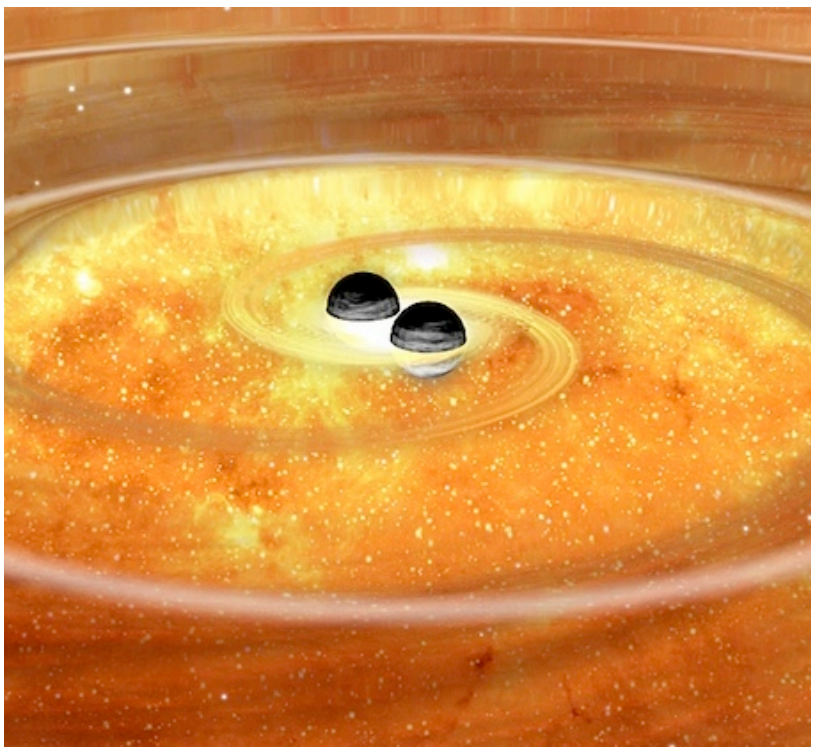
ACF derived using the standard Scargle algorithm

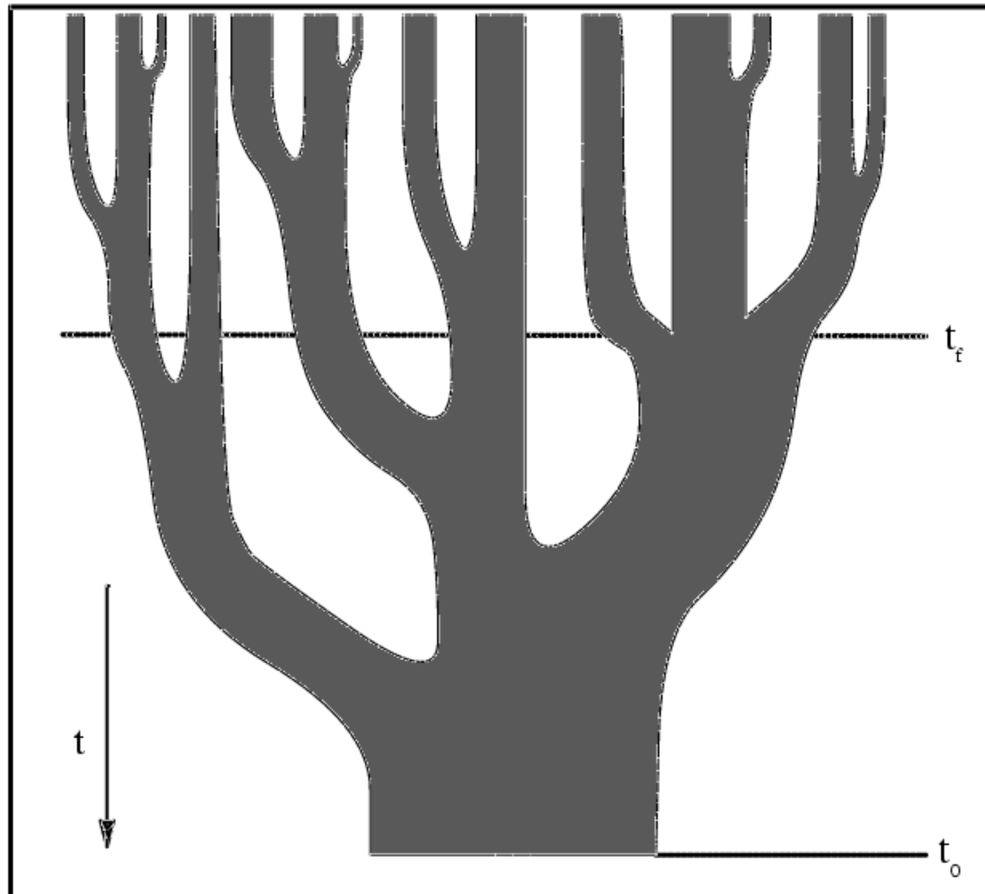


Possible period

SMBH Growth Mechanisms

- In a hierarchical picture, as galaxies merge so will their BH's
- This can naturally lead to the establishment of the SMBH - host galaxy correlations, which may be also sharpened by the AGN feedback





The Physics of SMBH Mergers

Stage I (> 1pc)

- SMBHs dissipate angular momentum through dynamical friction with surrounding stars

Stage II (0.01 – 1pc)

- Stalled phase due to stellar depletion (~10^6 – 10^7 yrs)

Stage III (< 0.01pc)

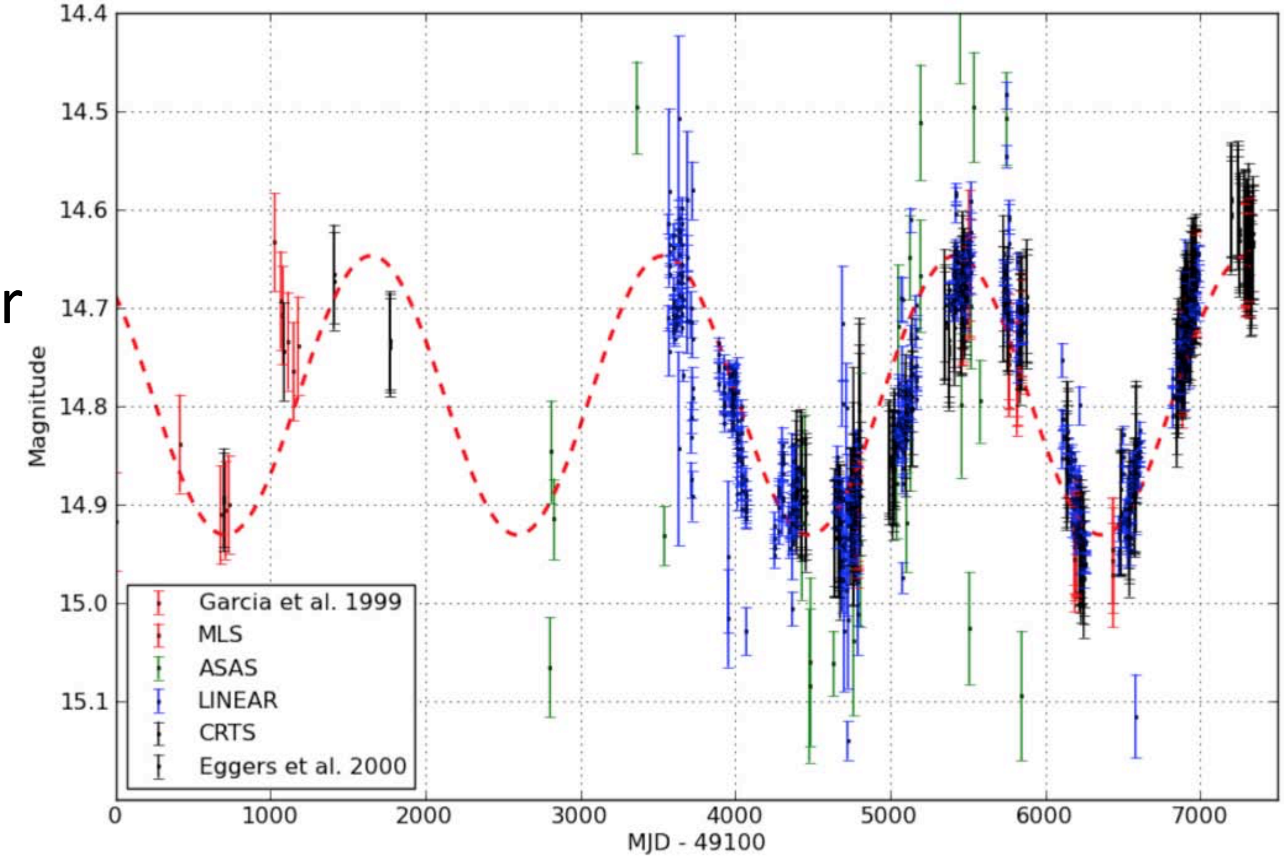
- Orbital angular momentum lost by gravitational radiation

Stage IV

- Coalescence and recoil
- The “final parsec” problem
- ***Subparsec systems are not resolvable***

Periodically Variable Quasars: Evidence for SMBH Binaries?

- Applying a novel technique to CRTS light curves of 247,000 known quasars

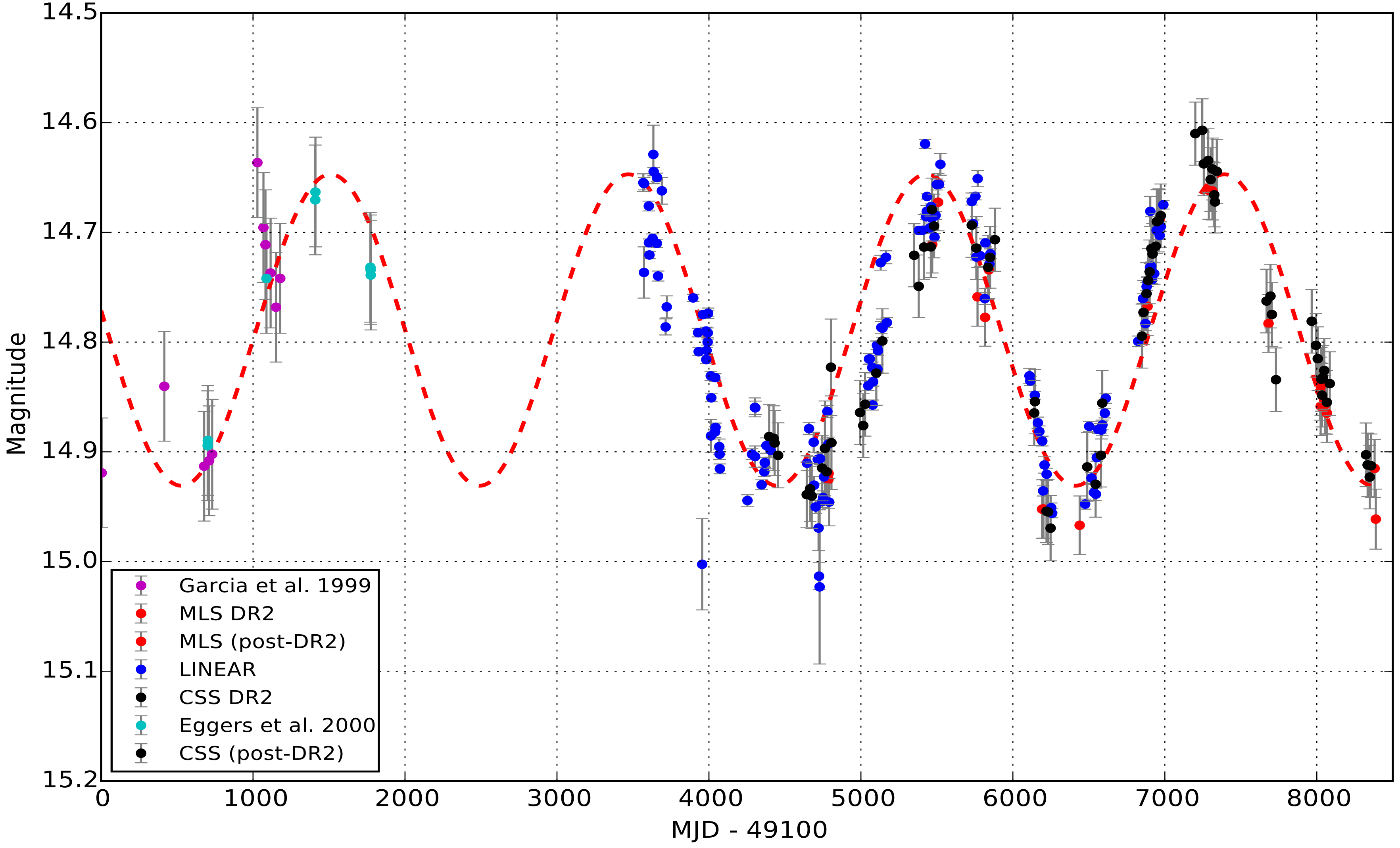


- The best case: **PG 1302–102** P_{rest}= 4.04 ± 0.19 yr
- For $M_{\bullet}$~10^8 M_{sun}, implied separation < 10 millipc
- Additional ~20 candidates ~ 10^{-4} of all quasars, in an agreement with the theoretical predictions

(Graham et al. 2015)

New Data Extend the Light Curve

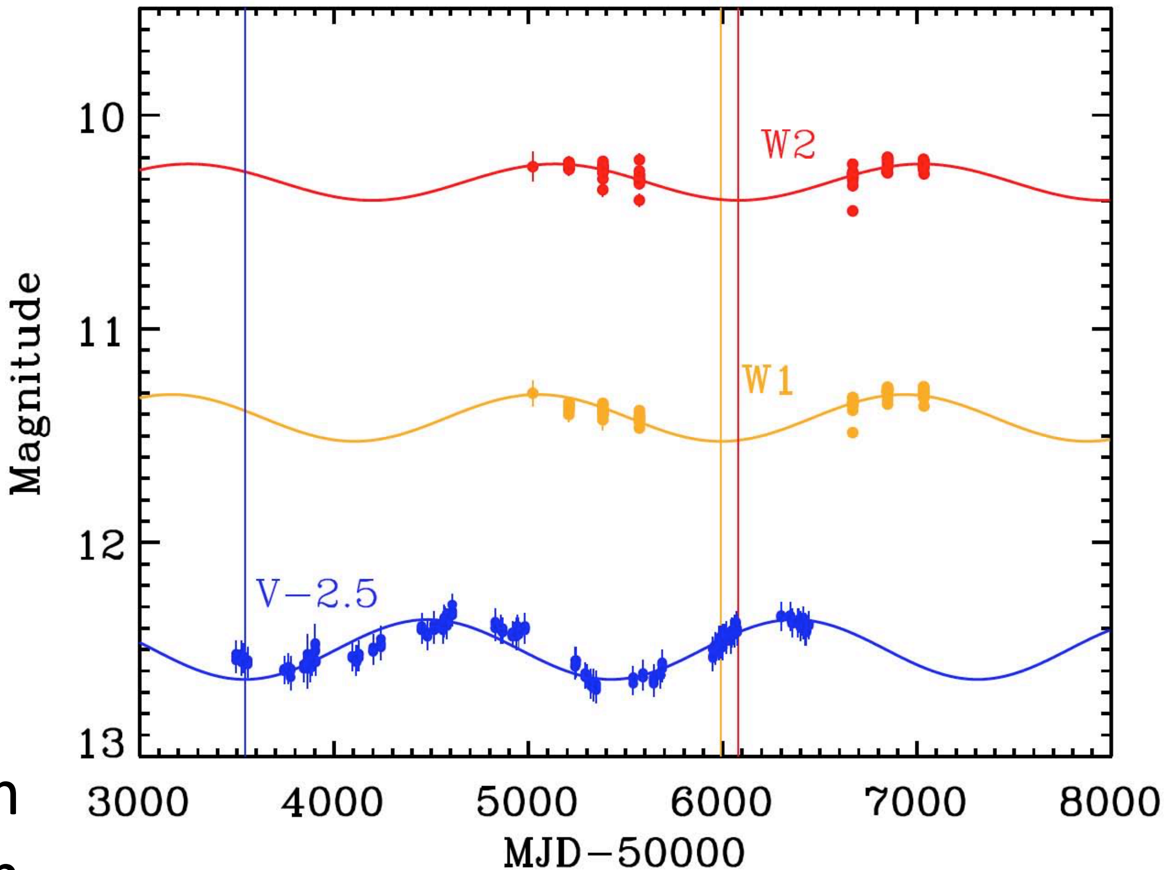
New data



IR Light Curve of PG 1302-102

(H. Jun et al. 2015)

Using WISE data: consistent with the optical period, but with a wavelength-dependent time delay and amplitude



Time lags:
2448 ± 12 days at 3.4 μm
2538 ± 14 days at 4.6 μm

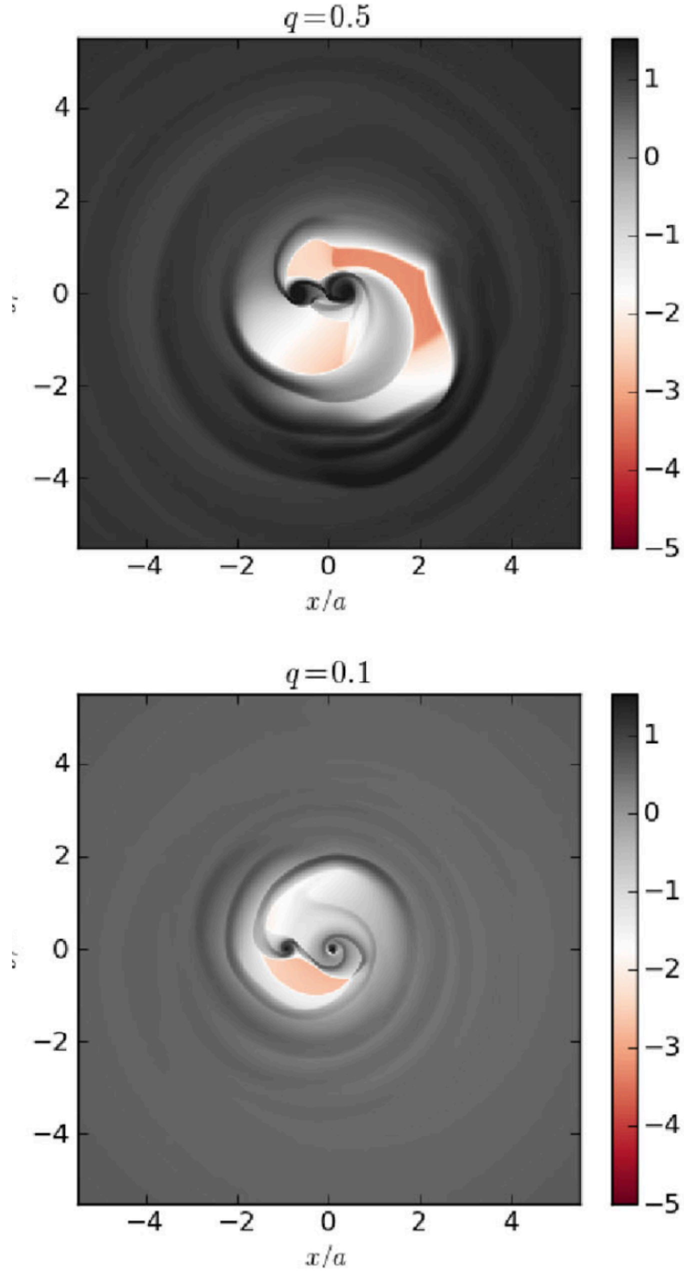
Interpretation: due to the light-travel time from the accretion disk to the surrounding dust "torus". Estimated radius = 2.1 pc, in an excellent agreement with theoretical expectations

Theoretical Models and Inetrpretation

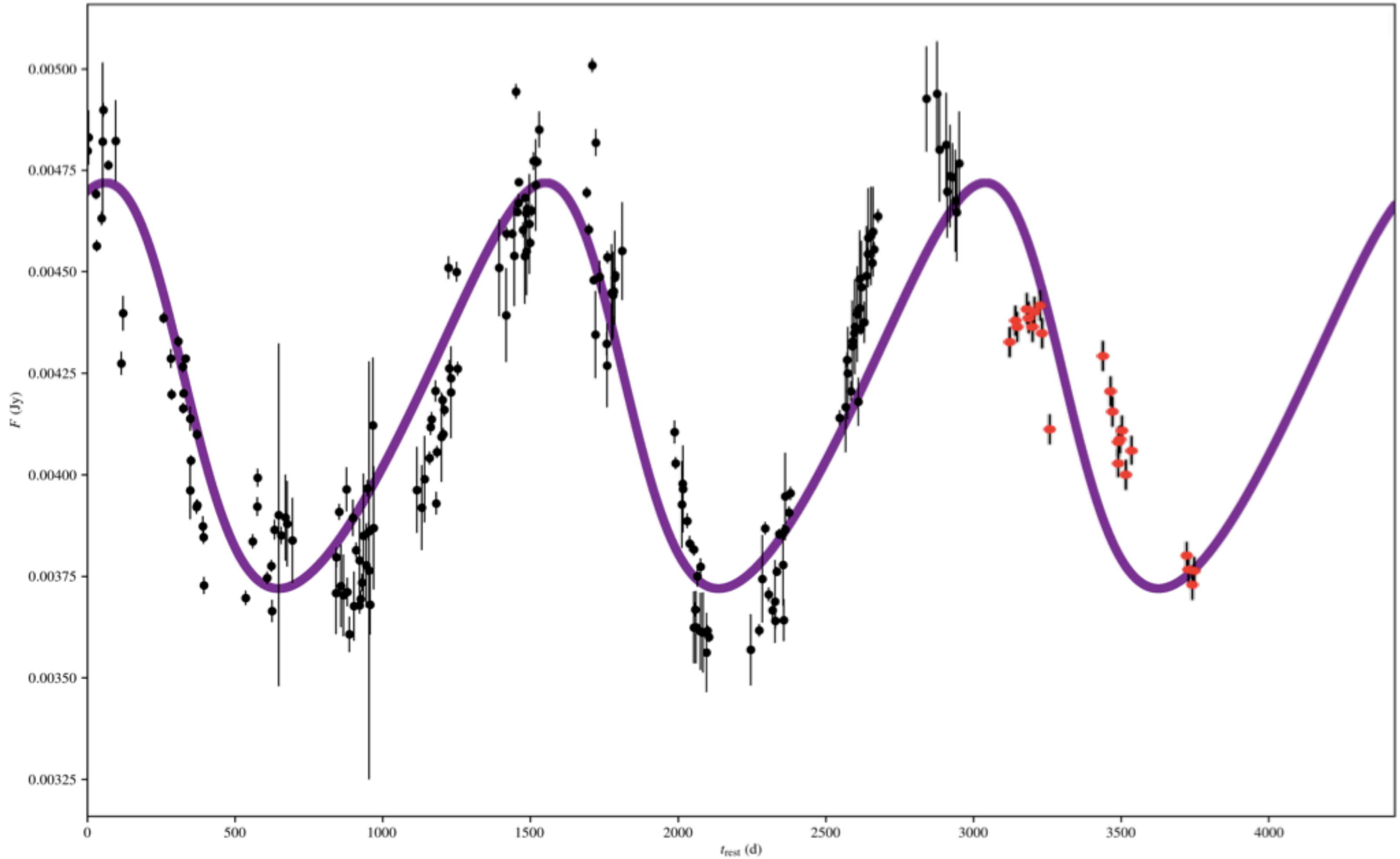
- Several papers by Z. Haiman's group (D'Orazio et al., Charisi et al. 2015)
- Archival UV data support the interpretation as a binary SMBH

The model:

- Hydrodynamical simulations suggest that the strongest periodicity is associated with a cavity in circumbinary disk => true binary period 3-8 times shorter than observed
- Relativistic boosting for line-of-sight motion of minidisk around secondary SMBH orbiting around system barycenter ~ scaled version of QPOs seen in stellar BH binaries



A Relativistic Doppler Boosting Model



It fits reasonably well the shape of the waveform, and predicts correctly the wavelength dependence of the amplitude (larger at the shorter wavelengths)

An Improved and Expanded Search

<u>Wavelets</u> *(Graham et al. 2015, MNRAS, 453, 1562)*

Peak value

Period

Slepian wavelet characteristic timescale

<u>Autocorrelation function</u>

Period

Amplitude of exponentially damped cosine

Decay constant of exponentially damped cosine

<u>Shape and coverage</u>

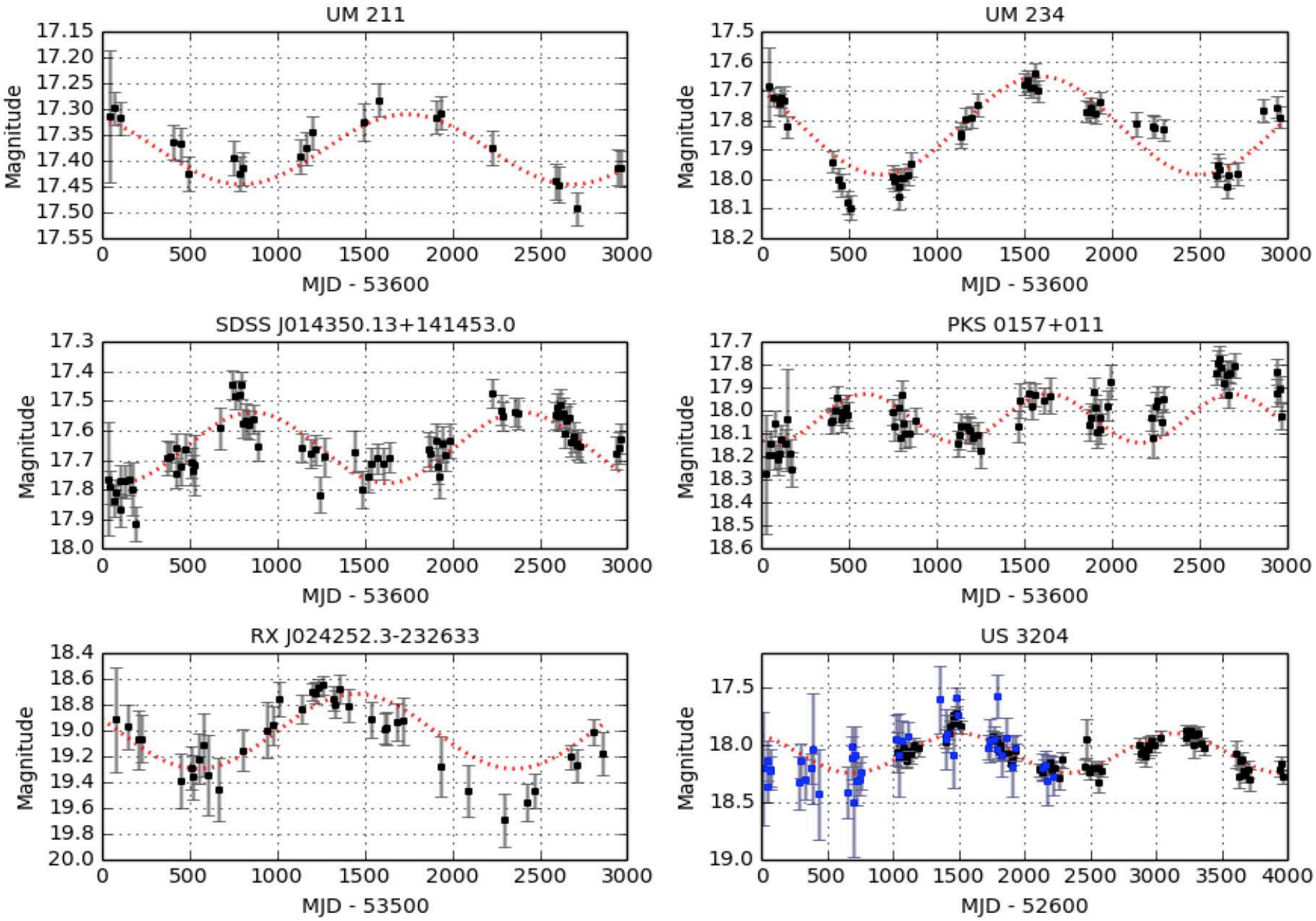
Scatter around best-fit Fourier series

At least 1.5 cycles

Train SVM to better describe discriminating hyperplane

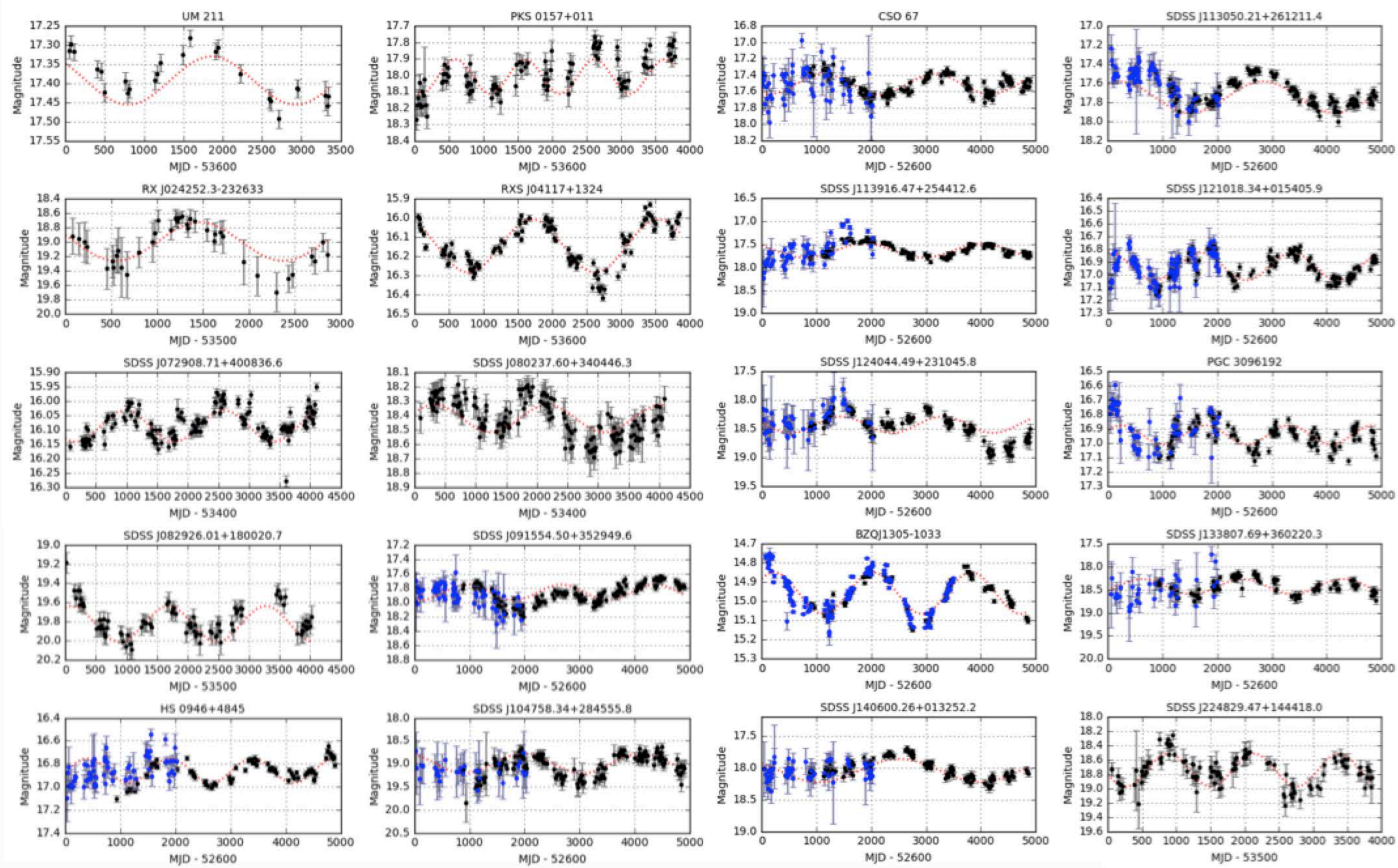
The result: *111 candidates* out of a sample of ~250,000 QSOs

Examples of Light Curves



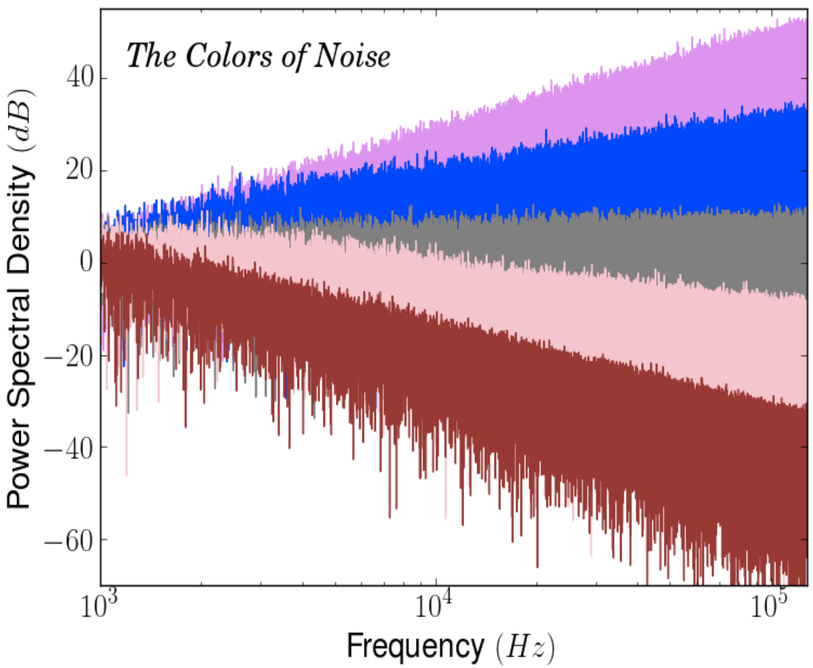
More Data Confirms the Periodicity

Black = CRTS, Blue = LINEAR



How Do We Know the Detections are Real?

- Stochastic variability is a **red noise** process
- Typical periodograms assume a white noise for determining the significance of peaks --> this will overestimate the significance for a red noise model
- Too much white noise (incorrect error model) reduces probability of simulations producing strong, smooth modulations
- Simulated data set of objects following a DRW model with the same sampling as the real data and CRTS errors produces **no candidates** with our selection criteria



Real vs. the Simulated Light Curves

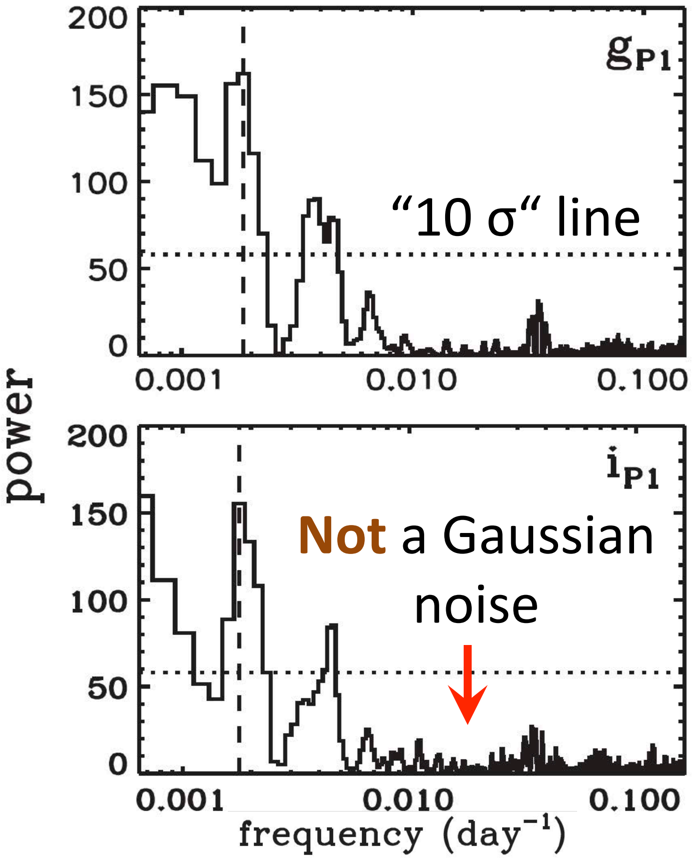
Simulated data from a pure CAR(1) process, with the same sampling and errors as the real data

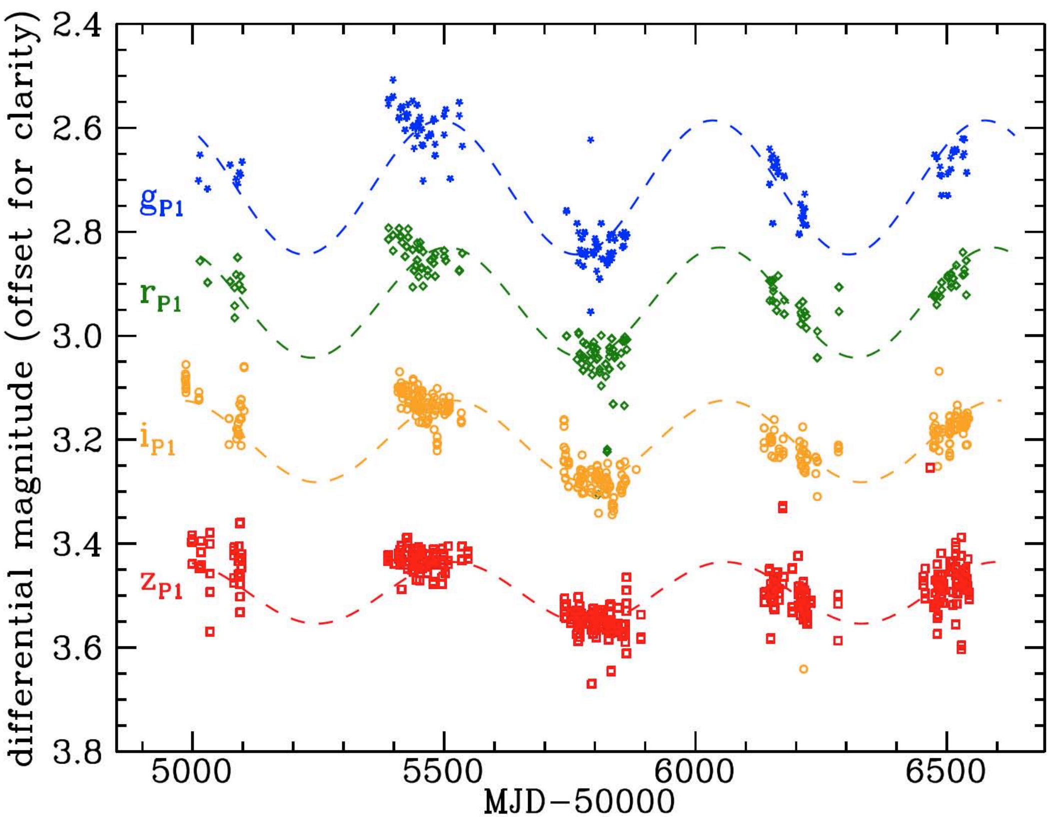
Component	Constraint	Real	Mock
Wavelet peak value	wwzpk > 50	24 437	32 078
Slepian wavelet deviation	$\tau_{\text{slep}} < \tau(M_V) - \sigma_\tau$	37 828	27 746
WWZ–ACF period	$0.9 < p_{\text{ACF}}/p_{\text{WWZ}} < 1.1$	30 330	63 637
ACF amplitude	$A > 0.3$	108 625	143 447
ACF decay	$\lambda < 10^{-3}$	172 049	182 592
Shape	rms/MAD < 0.67	11 794	3944
Temporal coverage	$\tau/p_{\text{ACF}} > 1.5$	245 234	182 257
Number of points	$n > 50$	243 486	243 486
Combined	–	101	0
Final	–	111	0

Graham et al. 2015, MNRAS 453, 1562

How to Find a Fake Periodicity

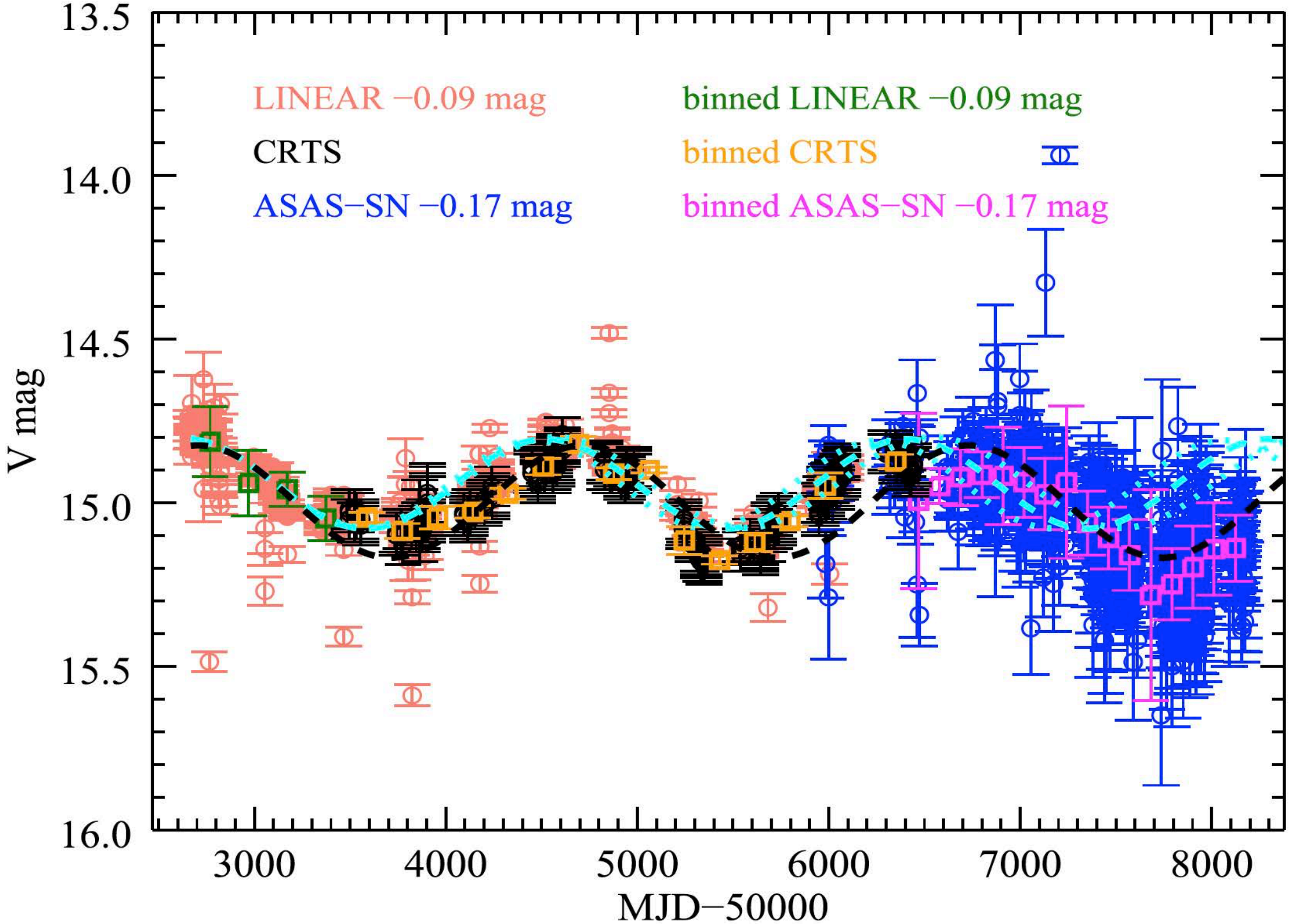
Liu et al. (2015) find the observed period of 542 ± 15 days in PanSTARRS data for PSO J334.2028+01.4075, using periodograms. This is the strongest candidate out of 40 “statistically significant”, out of a parent sample of 320 QSOs. Subsequently they retracted the claim.





The News of PG1302's Death Has Been Greatly Exaggerated

Liu et al. (2018) claim that the inclusion of ASAS-SN data has "killed" the SMBHB in PG 1302-102. Judge for yourself:



Their use of the p-value statistics for a combined data sample is also incorrect (Graham et al. in prep.)

How Many Should We Expect to See?

Using theoretical predictions for a population of SMBHs en route to a merger, in the range of periods we probe:

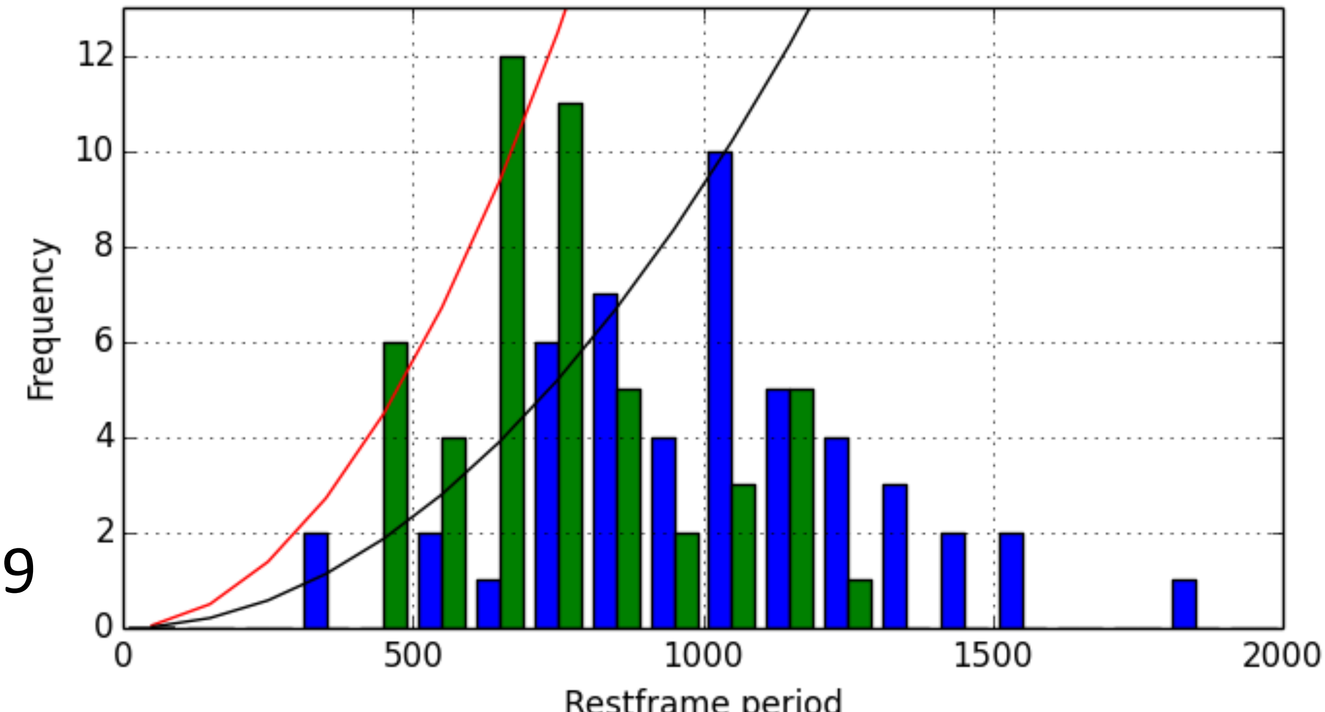
Down to 19th mag: predicted 116 - we find 104

Down to 20th mag: predicted 451 - we find 110

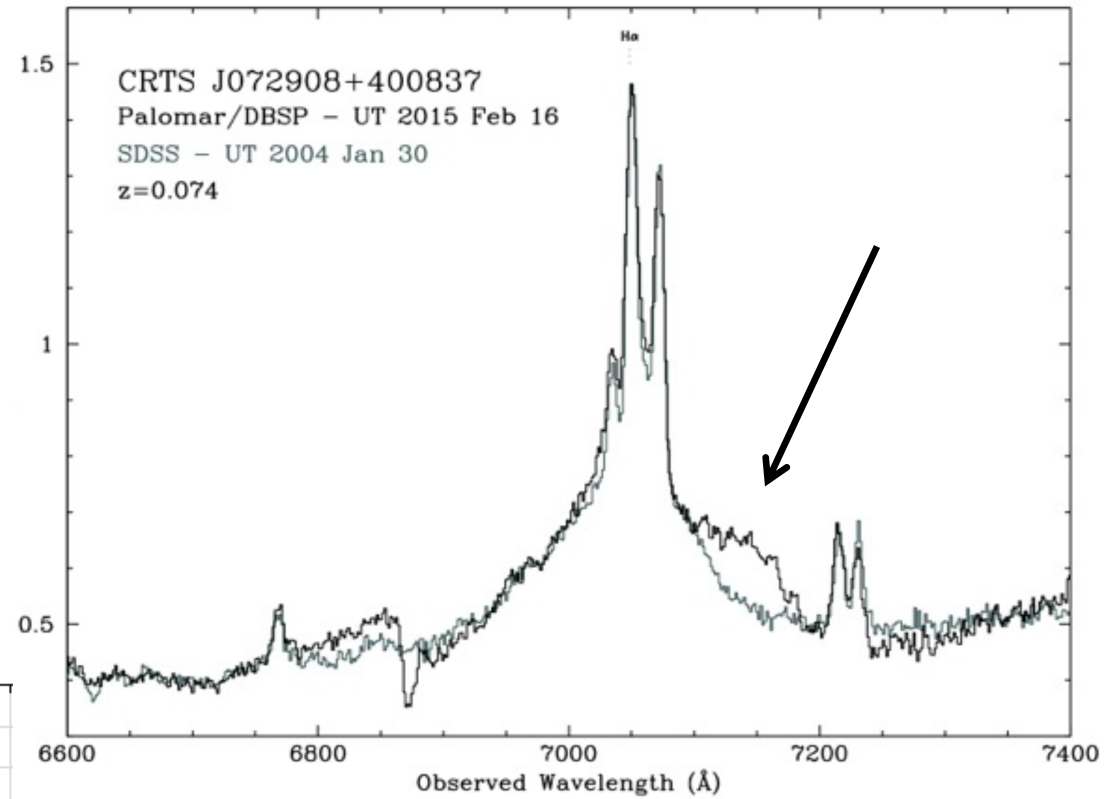
(but we are seriously incomplete)

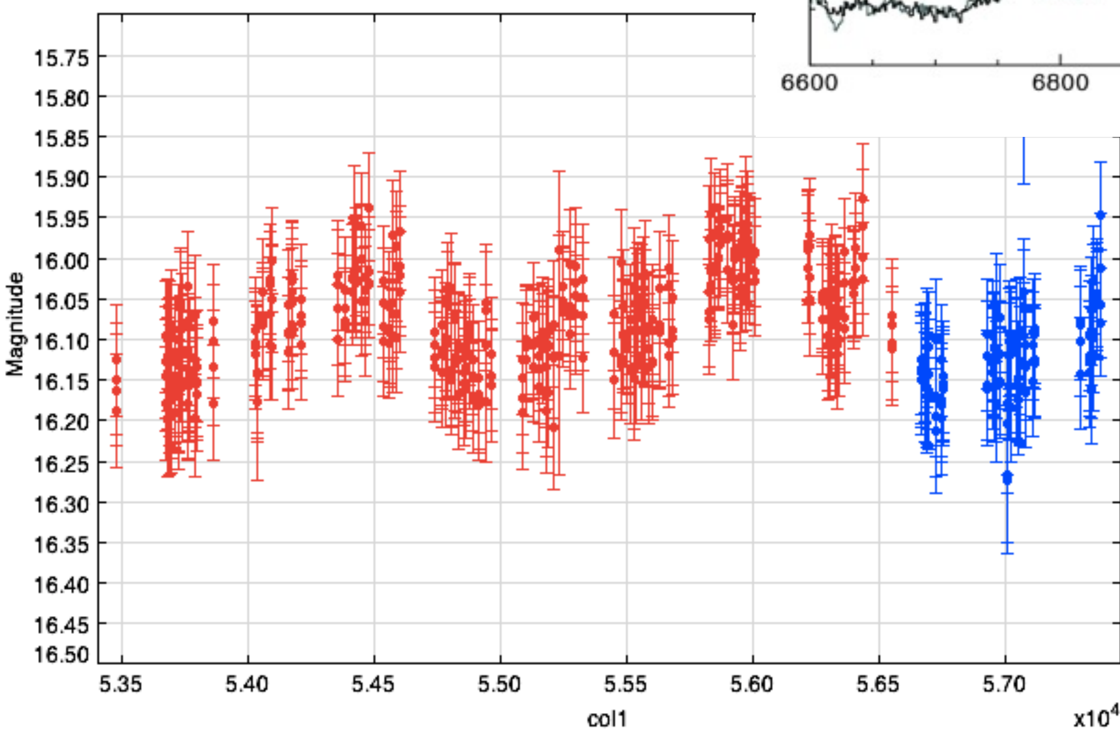
The frequency of binary SMBH as a function of restframe period

Blue: $\log(M_{BH}) < 9$
Green: $\log(M_{BH}) > 9$

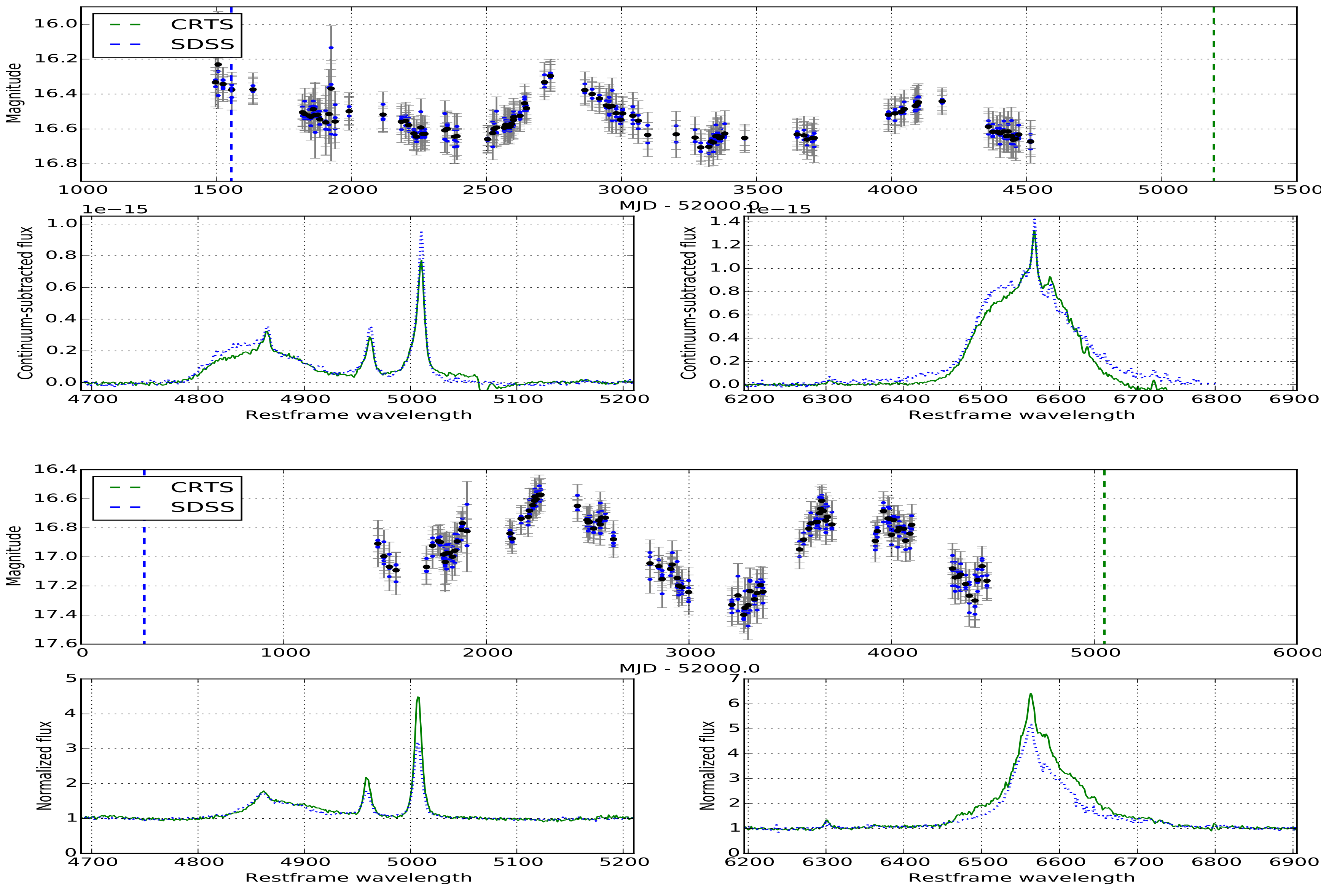


Spectroscopic follow-up: looking for the shape changes in the emission lines



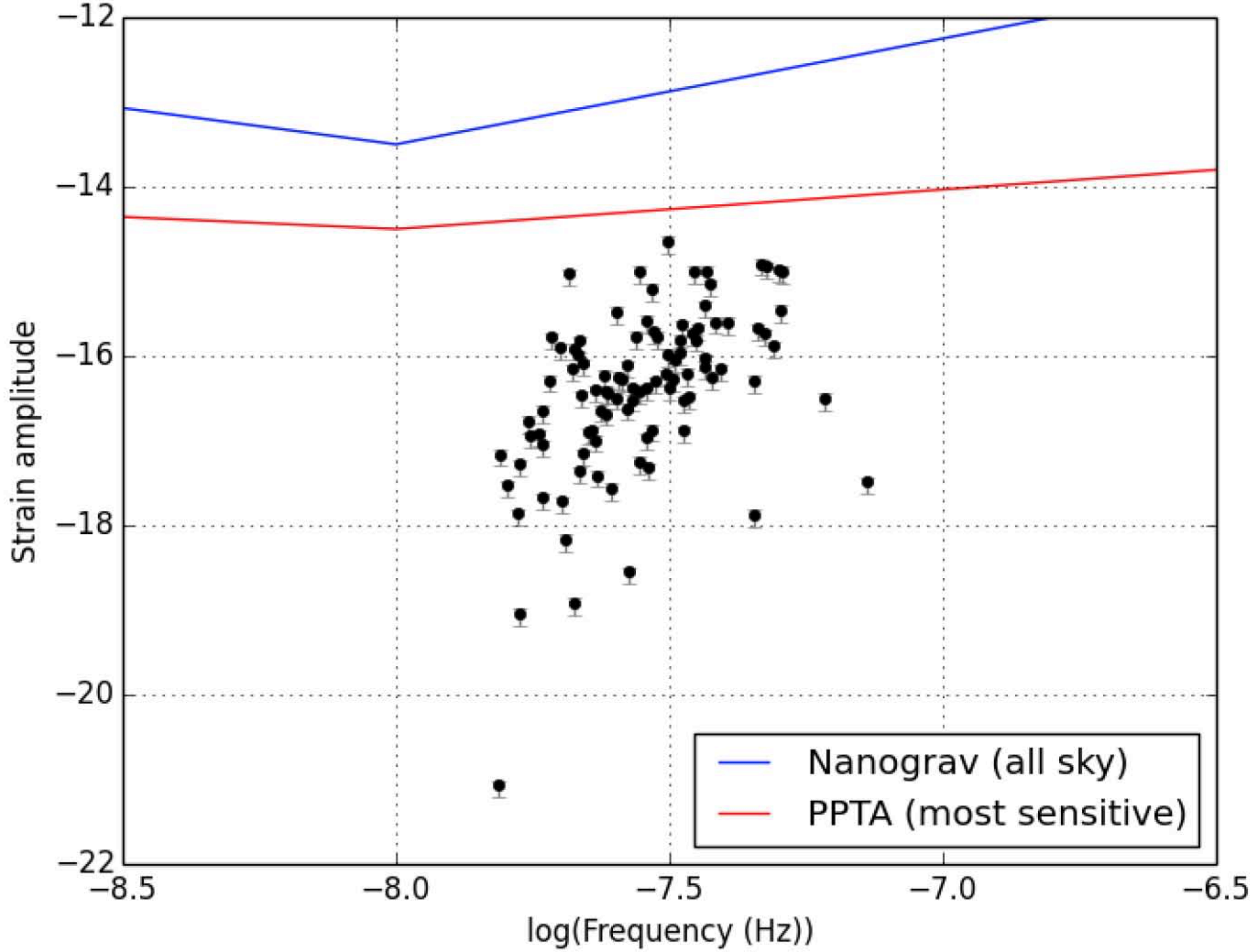


Spectroscopic follow-up: looking for shape changes in the emission lines

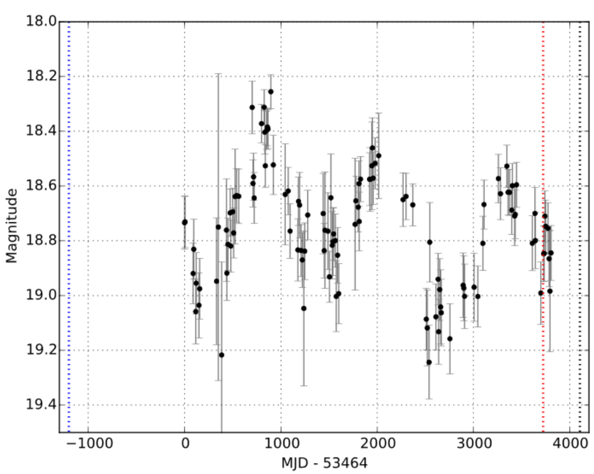


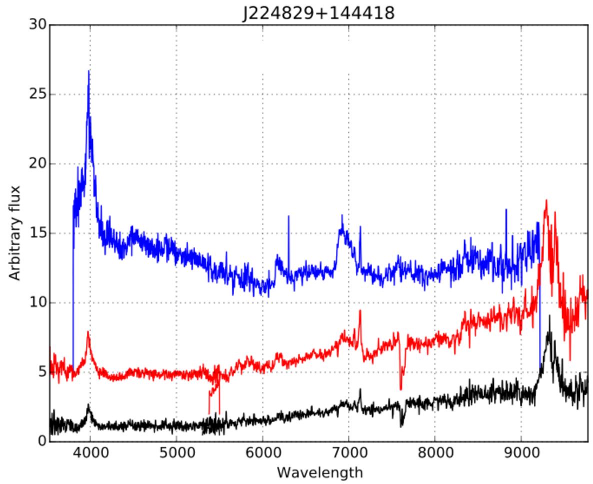
Can These SMBH Binaries Be Detected in Gravitational Waves?

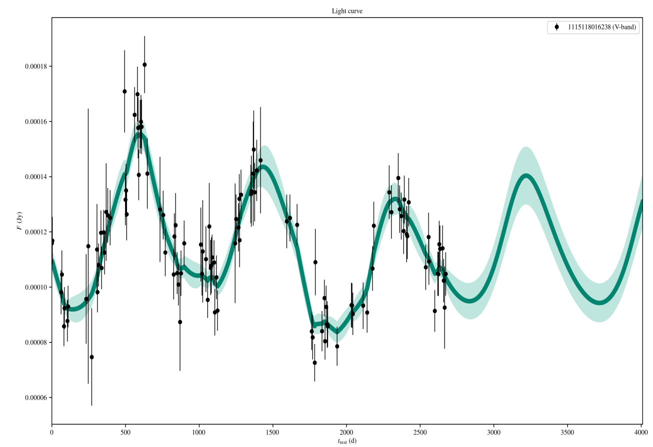
Not yet, but maybe within a decade, with the pulsar timing arrays



SMBH Binaries: Looking for More



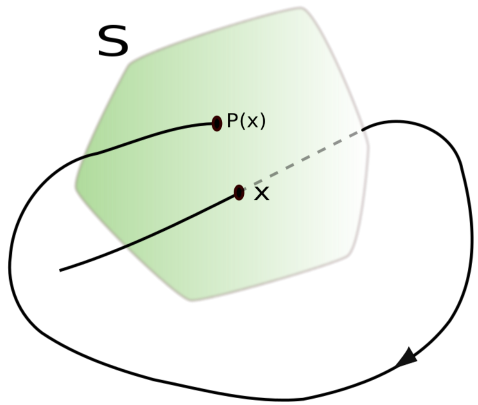




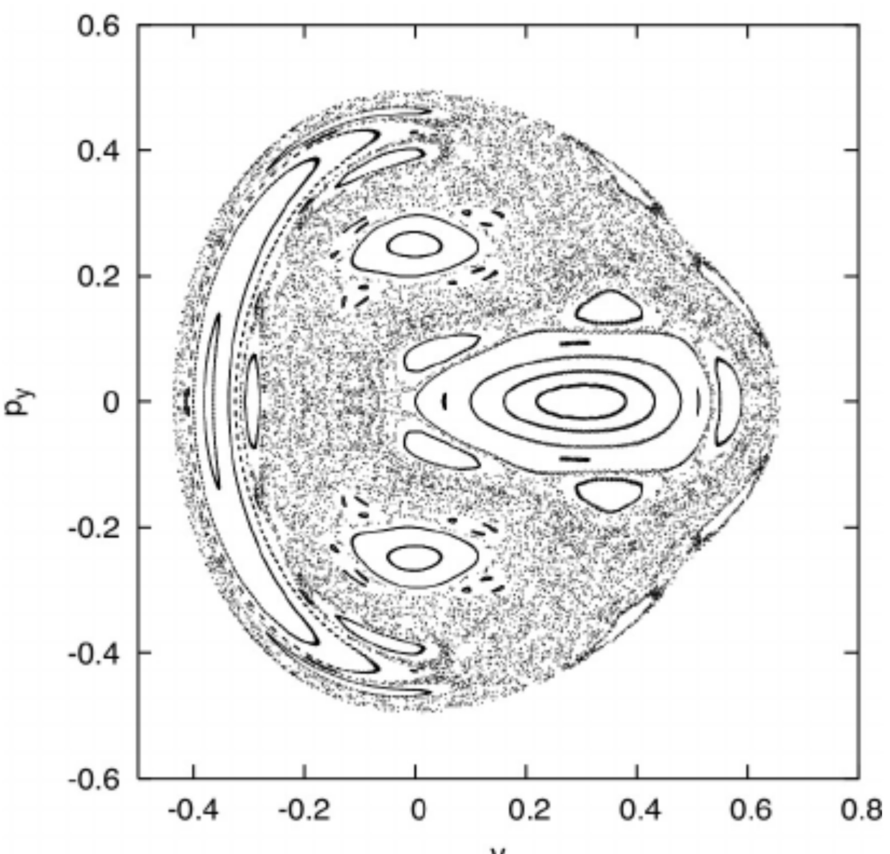
- Extending search with more sophisticated algorithms and combined data sets (LINEAR, PTF, ZTF)
  - Using coregionalized Gaussian process regression
- Understanding the issues of red noise components in the light curve (Vaughan et al. 2016) through both detection algorithms and population simulations

Randomness and Deterministic Chaos

Deterministic Chaos: generation of random, unpredictable behavior from a simple, but nonlinear rule.

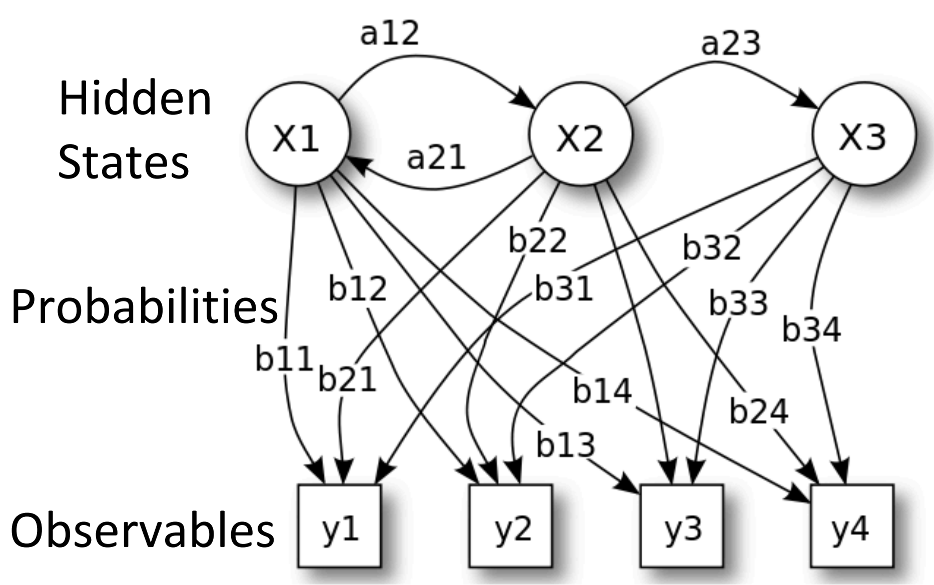


Example: Poincare map, or Surface of Section



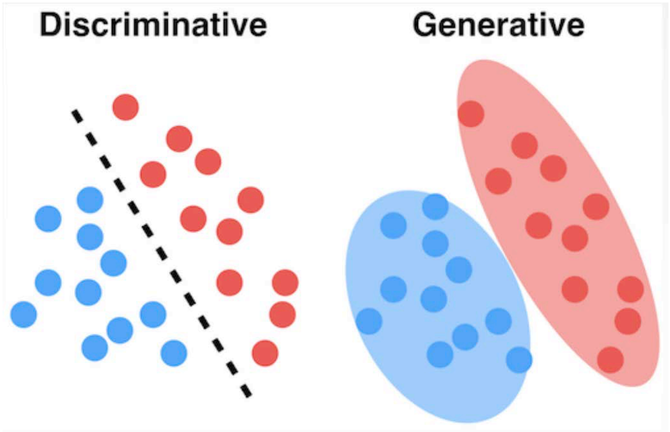
Henon-Heiles potential

Hidden Markov Models:



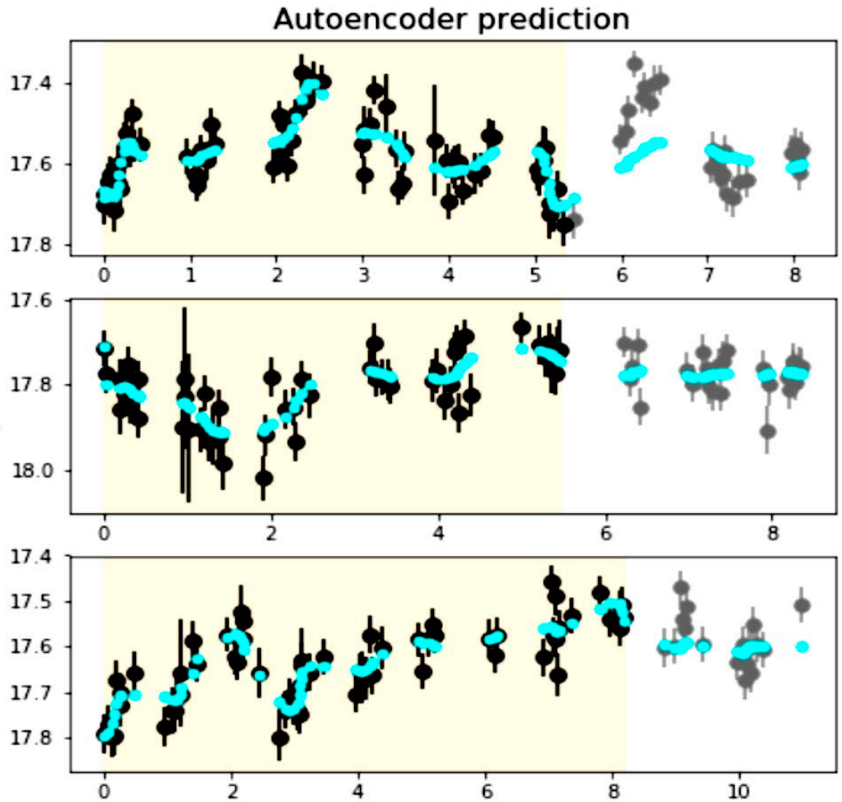
Predicting Stochastic Behavior

Discriminative models (e.g., most supervised ML methods) learn the boundaries between classes.
Generative models find a probabilistic model describing the structure of the data.



They may be used to **predict** (within some time scale) the stochastic behavior.

Predictions using an Autoencoder ANN



Work still in progress…

Summary

- The new large samples allow systematic studies of AGN on an unprecedented scale, and discovery of rare or unusual types of objects or events
- Correct modeling of the stochastic variability is *essential*
- Quasar variability studies and results so far include:
  - ✧ The best ever method for quasar discovery
  - ✧ Discovery of a characteristic time scale for the stochastic variability, a possible probe of the accretion disk physics
  - ✧ Insights into the physics of unusual populations of quasars (LoBAL, DPE, …)
  - ✧ Quasar Megaflares, including a new population of luminous SNe from accretion disks and microlensing events
  - ✧ Discovery of a population of binary SMBH, a key predictions of the hierarchical formation models

Big Data + Novel Analytics = New Discoveries