Data-Driven Approaches to Searches for
The Technosignatures of Advanced Civilizations
KISS, Caltech, 2019 May 20

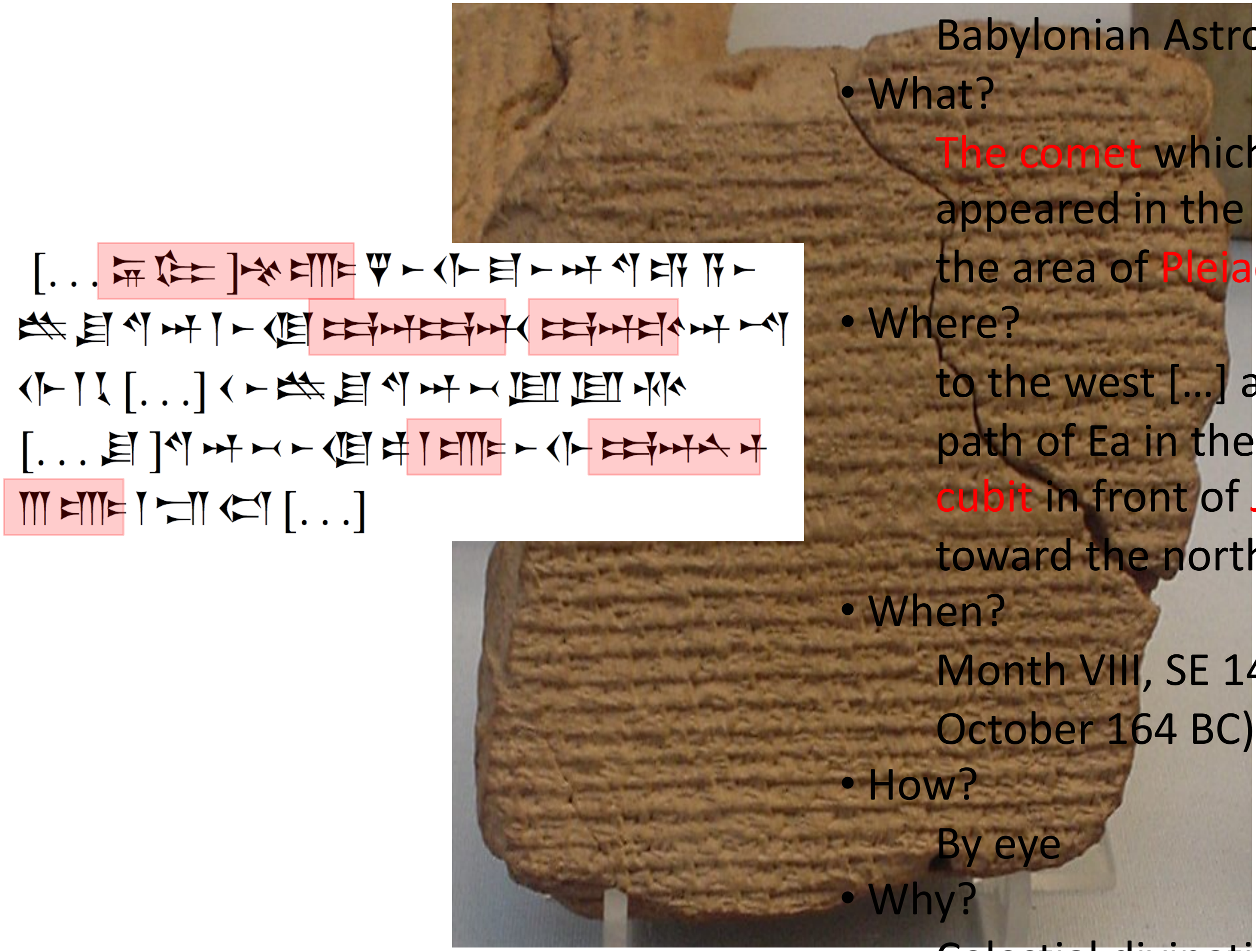
An introduction to time series analysis

Matthew J. Graham
Center for Data-Driven Discovery/ZTF, Caltech
mjg@caltech.edu

CENTER FOR DATA-DRIVEN DISCOVERY

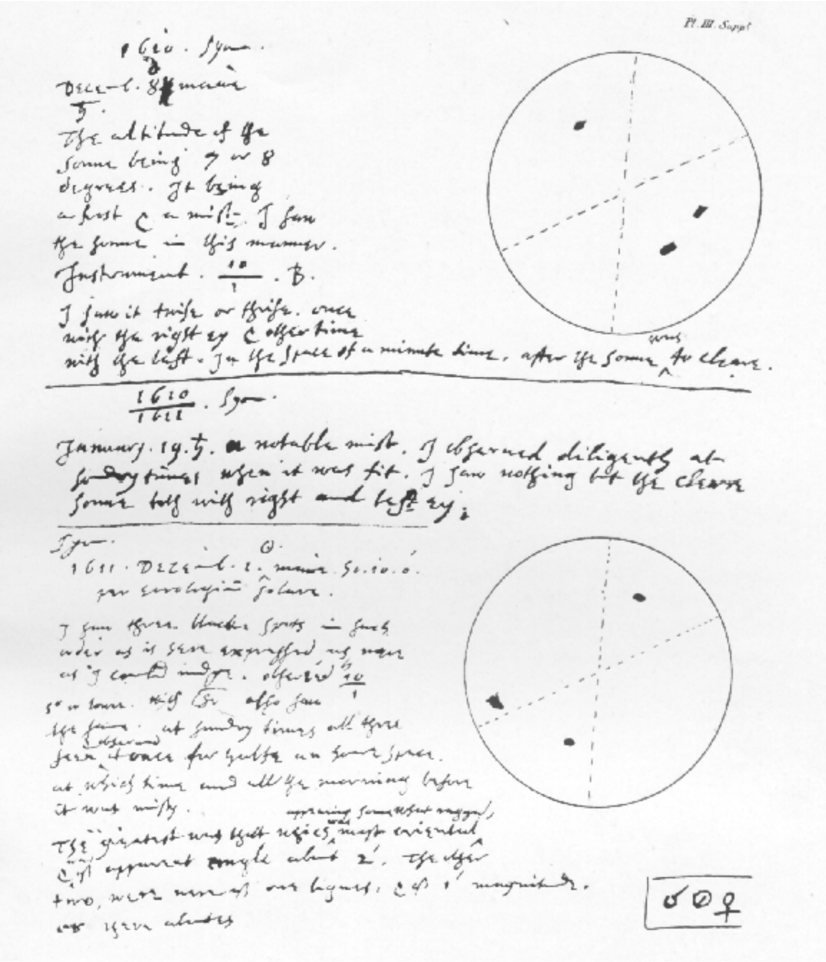
ZTF

A history of anomalous observations



- 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

The first astronomical time series



Pl. III. Suppl

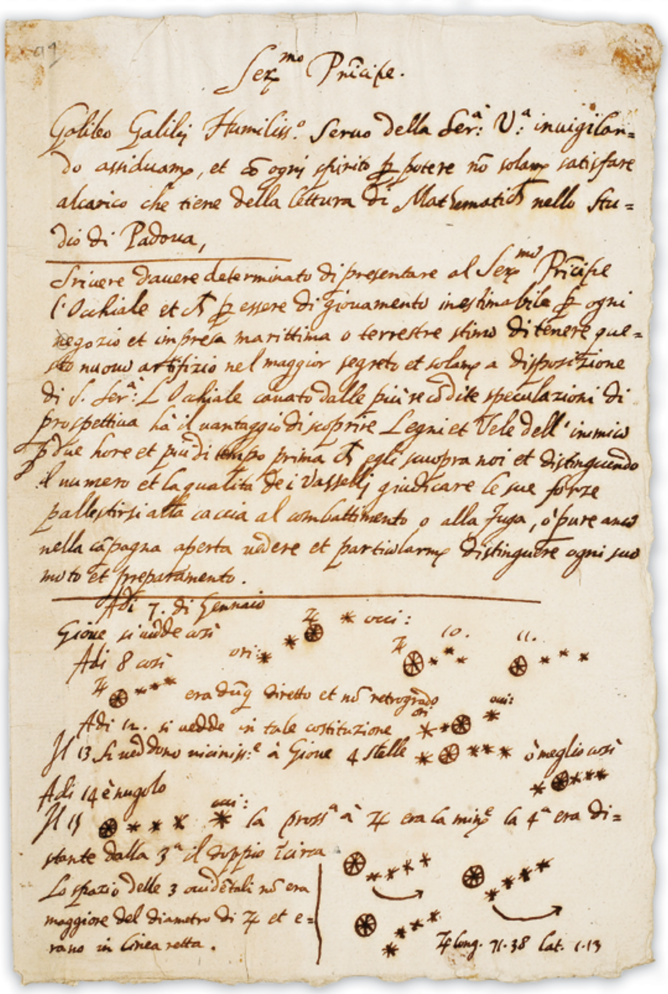
1610. Syon.
Decemb. 8 mane
The altitude of the sonne being 7 or 8 degrees. It being a frost & a mist. I saw the sonne in this manner. Instrument. 10/1. B.
I saw it twise or thrise. once with the right ey & other time with the left. In the space of a minute time, after the sonne was to cleare.

1610/1611. Syon.
January. 19. a notable mist. I observed diligently at sundry times when it was fit. I saw nothing but the cleare sonne both with right and left ey.

Syon.
1611. Decemb. 1. mane 50.10.0.
per horologium solare.
I saw three blacke spots in such order as is here expressed as neare as I could iudge. observed 10/1 or 20/1 also saw the same at sundry times all three I observed seene at once for halfe an howre space. at which time and all the morning before it was misty.
The greatest was that which was most orientall appearing somewhat ragged, his apparent angle about 2'. the other two, were neare of one bignes, each 1' magnitude. or there abouts

♂ ⊙ ♀

Thomas Harriott: Dec 1610



Ser.mo Principe.

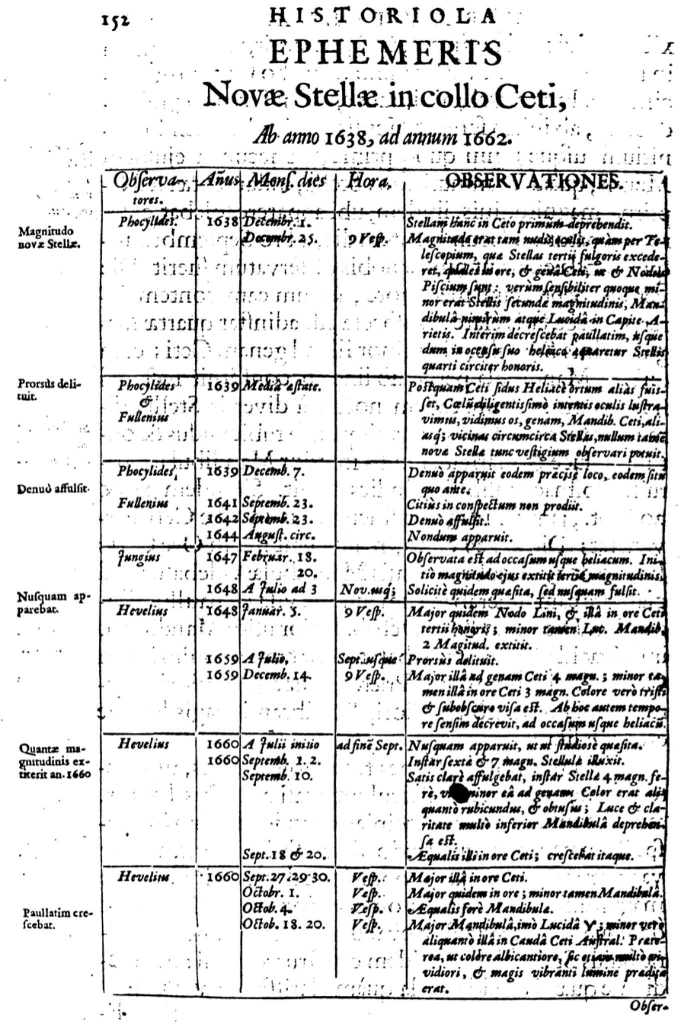
Galileo Galilei Humiliss.o Servo della Ser.tà V.a inuigilando assiduam.te, et con ogni spirito per potere non solam.te satisfare al carico che tiene della Lettura di Matematica nello Studio di Padova,

Hora avere determinato di presentare al Ser.mo Principe l'Occhiale et per essere di giovamento inestimabile per ogni negozio et impresa marittima o terrestre stimo di tenere questo nuovo artifizio nel maggior segreto et solamente a disposizione di V. Ser.tà L'Occhiale cavato dalle più recondite speculazioni di prospettiva ha il vantaggio di scoprire Legni et Vele dell'inimico due hore et più di tempo prima che egli scuopra noi et distinguendo il numero et la qualità dei Vasselli giudicare le sue forze pallestirsi alla caccia al combattimento o alla fuga, o pure ancora nella campagna aperta vedere et particolarmente distinguere ogni suo moto et preparamento.

Adi 7. di Gennaio
Giove si vedde così ** ⊛ * occ:
Adi 8 così ori: ⊛ *** occ. 10. ⊛ *** 11. ⊛ ***
era dunq diretto et non retrogrado
Adi 12. si vedde in tale costituzione ori: * * ⊛ * occ:
Il 13 si veddono vicinissime à Giove 4 stelle * ⊛ *** ò meglio così * ⊛ ***
Adi 14 è nugolo
Il 15 ⊛ * * * * occ: la pross.a à ♃ era la min.e la 4.a era distante dalla 3.a il doppio in circa
Lo spatio delle 3 occidentali non era maggiore del diametro di ♃ et erano in linea retta.
♃ long. 71.38 lat. 1.13

Image credit: University of Michigan Special Collections Library

A wondrous star in the neck of the Whale



152 HISTORIOLA

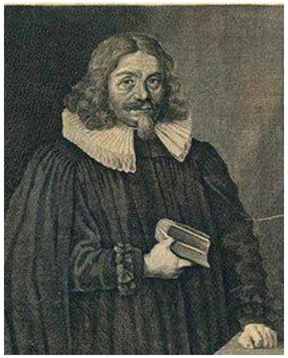
EPHEMERIS

Novæ Stellæ in collo Ceti,

Ab anno 1638, ad annum 1662.

	Obſervatores.	*Añus*	*Menſ. dies*	*Hora.*	OBSERVATIONES.
Magnitudo novæ Stellæ.	*Phocylides.*	1638	*Decembr. 1.* *Decembr. 25.*	 *9 Veſp.*	*Stellam hanc in Ceto primùm deprehendit.* *Magnitudo erat tam nudis oculis, quàm per Teleſcopium, qua Stellas tertii fulgoris excederet, quales in ore, & genâ Ceti, ut & Nodo Piſcium ſunt: verùm ſenſibiliter quoque minor erat Stellis ſecundæ magnitudinis, Mandibulâ nimirùm atque Lucidâ in Capite Arietis. Interim decreſcebat paullatim, uſque dum in occaſu ſuo heliaco æquaretur Stellis quarti circiter honoris.*
Prorsùs delituit.	*Phocylides & Fullenius*	1639	*Mediâ æſtate.*		*Poſtquam Ceti ſidus Heliacè ortum aliàs fuiſſet, Cœlũ diligentiſſimè intentis oculis luſtravimus, vidimus os, genam, Mandib. Ceti, aliaſq; vicinas circumcirca Stellas, nullum tamẽ nova Stella tunc veſtigium obſervari potuit.*
Denuò affulſit.	*Phocylides* *Fullenius*	1639 1641 1642 1644	*Decemb. 7.* *Septemb. 23.* *Septemb. 23.* *Auguſt. circ.*		*Denuò apparuit eodem præciſè loco, eodem ſitu quo ante.* *Citiùs in conſpectum non prodiit.* *Denuò affulſit.* *Nondum apparuit.*
Nuſquam apparebat.	*Jungius*	1647 1648	*Februar. 18.* *20.* *A Julio ad 3*	 *Nov. uſq;*	*Obſervata eſt ad occaſum uſque heliacum. Initiò magnitudo ejus extitit tertiæ magnitudinis.* *Solicitè quidem quæſita, ſed nuſquam fulſit.*
	Hevelius	1648 1659 1659	*Januar. 5.* *A Julio,* *Decemb. 14.*	*9 Veſp.* *Sept. uſque* *9 Veſp.*	*Major quidem Nodo Lini, & illâ in ore Ceti tertii honoris; minor tamen Luc. Mandib. 2 Magitud. extitit.* *Prorſus delituit.* *Major illâ ad genam Ceti 4 magn.; minor tamen illâ in ore Ceti 3 magn. Colore verò triſti & ſubobſcuro viſa eſt. Ab hoc autem tempore ſenſim decrevit, ad occaſum uſque heliacũ.*
Quantæ magnitudinis extiterit an. 1660	*Hevelius*	1660 1660	*A Julii initio* *Septemb. 1. 2.* *Septemb. 10.* *Sept. 18 & 20.*	*ad finẽ Sept.*	*Nuſquam apparuit, ut ut ſtudioſè quæſita.* *Inſtar ſextæ & 7 magn. Stellulæ illuxit.* *Satis clarè affulgebat, inſtar Stellæ 4 magn. ferè, vix minor eâ ad genam. Color erat aliquantò rubicundus, & obtuſus; Luce & claritate multò inferior Mandibulâ deprehenſa eſt.* *Æqualis illi in ore Ceti; creſcebat itaque.*
Paullatim creſcebat.	*Hevelius*	1660	*Sept. 27. 29. 30.* *Octobr. 1.* *Octob. 4.* *Octob. 18. 20.*	*Veſp.* *Veſp.* *Veſp.* *Veſp.*	*Major illâ in ore Ceti.* *Major quidem in ore; minor tamen Mandibulâ.* *Æqualis ferè Mandibula.* *Major Mandibulâ, imò Lucidâ ♈; minor verò aliquantò illâ in Caudâ Ceti Auſtral. Præterea, ut colore albicantiore, ſic etiam multò vividiori, & magis vibranti lumine prædita erat.*

Obſer-



"If the new star were outside the ordinary course of nature, it would tell us little about the constitution of the universe. "

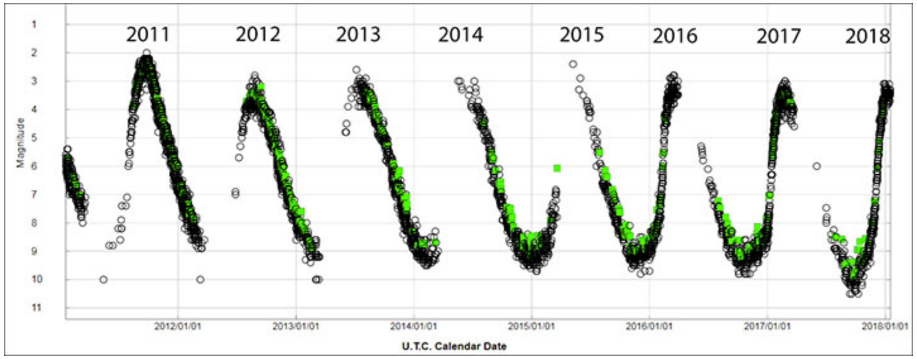


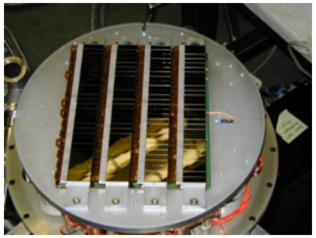
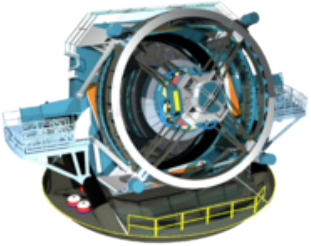
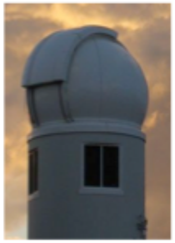
Image credit: AAVSO

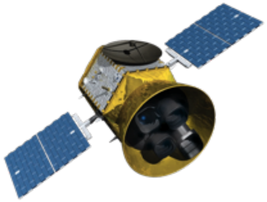
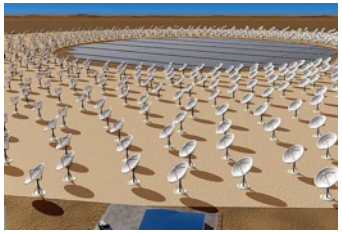
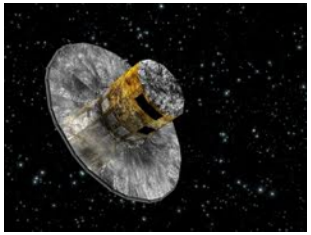
A billion time series and counting

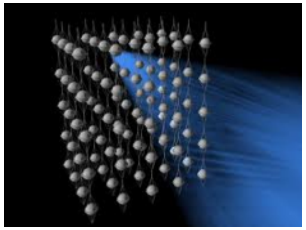
- Palomar-Quest Synoptic Sky Survey
- SDSS (Stripe 82)
- Catalina Real-time Transient Survey
- Palomar Transient Factory
- Zwicky Transient Factory
- Pan-STARRs
- SkyMapper
- ASKAP
- ThunderKat (MeerKAT)
- KEPLER
- GAIA
- LIGO
- IceCUBE
- LOFAR
- LSST
- SKA
- TESS
- ASAS-SN
- MASTER
- DES
- ATLAS
- BlackGEM

- GoTo
- MeerKAT
- ASKAP
- WISE
- OGLE
- DESI
- SDSS-V
- LAMOST

…



What we do ask of time series?

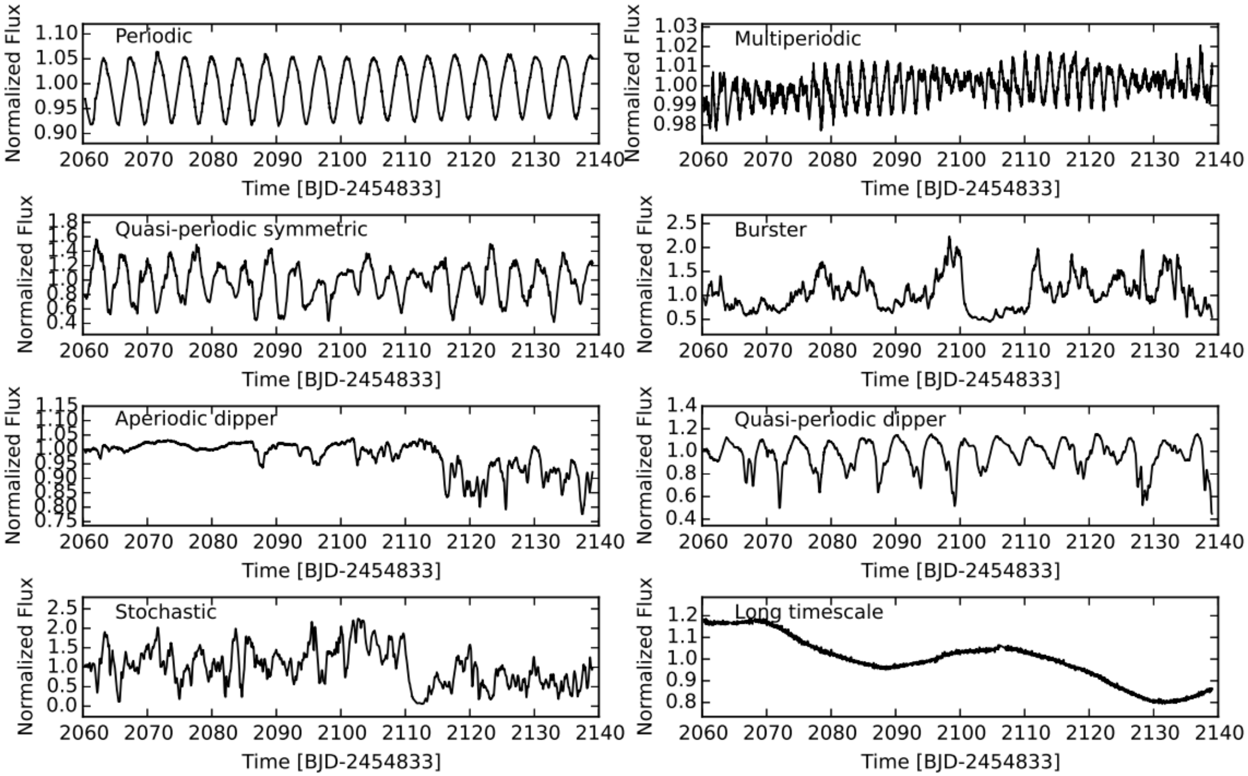
Population behaviors

- Characterize, categorize, classify

Outliers

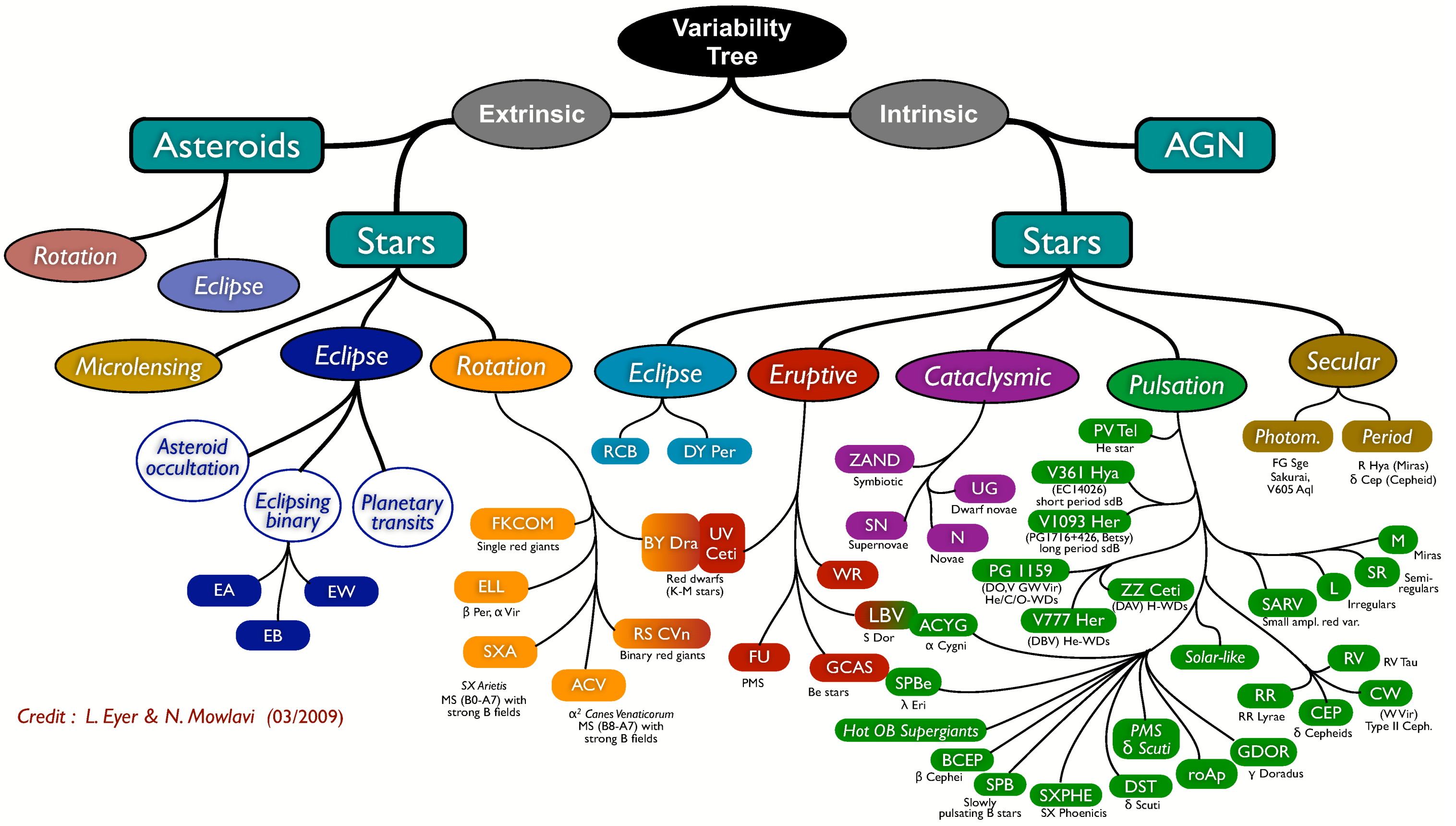
- Extreme sources

- Physical models
  - Predictions



(Cody & Hillenbrand 2018)

Types of astronomical variability



Credit : L. Eyer & N. Mowlavi (03/2009)

Foundational concepts - I

A time series is a set of time-tagged measurements: $\{\boldsymbol{X}_i(t_i)\}$ with observation errors $\boldsymbol{\sigma}_i$

Non-IID

- Data is sequential

Homoskedasticity

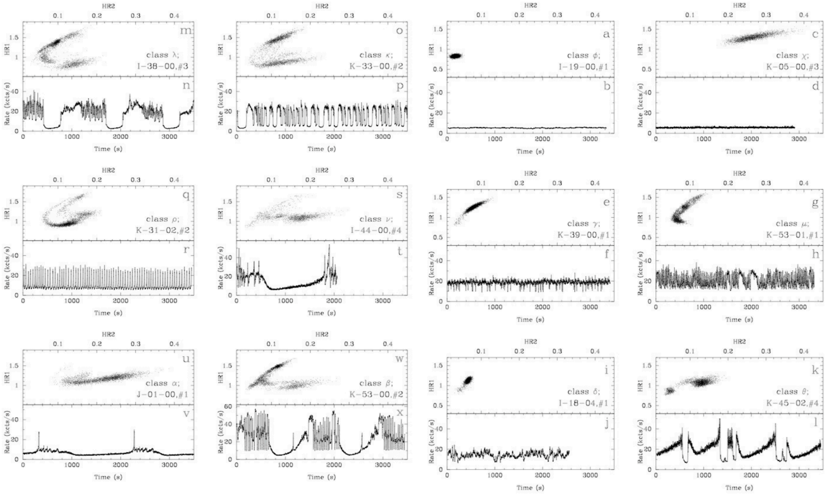
- All errors drawn from same process

Stationarity

- The generating distribution is time independent
- GSR 1915+215 has ~20 variability states
- GARCH models: variance is a stochastic function of time
- Nonstationary time series do not have to be stationary in any limit

- Ergodicity
  - The time average for one sequence is the same as the ensemble average:

$$\hat{f}(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f\left(T^k x\right).$$



(Belloni et al. 2000)

Foundational concepts - II

Sampling

- Even or regular sampling: $y(t) = x(t_0 + n\Delta t)$ where $n = 0,1,\dots,m$
- Uneven or irregular sampling: $y(t) = x(t_0),\dots,x(t_m)$

Power spectrum

- Power spectral density tells you everything: $PSD(v) = |\mathcal{F}(x)|^2$
- PSD is Fourier transform of autocorrelation function:

$$PSD(\nu) = \int_{-\infty}^{\infty} ACF(\Delta t)\, e^{-2\pi i\nu\Delta t} \Delta t$$
$$ACF(\Delta t) = \mathbb{E}[(x_t - \mu)(x_{t+\Delta t} - \mu)]/\sigma^2$$

- The structure function is related to the autocorrelation function:

$$SF(\Delta t) = \sqrt{2}\sigma_s\sqrt{1 - ACF(\Delta t)}$$
$$SF(\Delta t) = 0.742\, IQR(x)$$

Time series decomposition

Given any stationary process, *Y*, there exist:

- a linearly deterministic process, *D*
- an uncorrelated zero mean noise process, *R*
- a moving average filter, *C*

such that:

$$Y(t) = C \times R(t) + D(t)$$

(Wold’s Decomposition Theorem (1938))

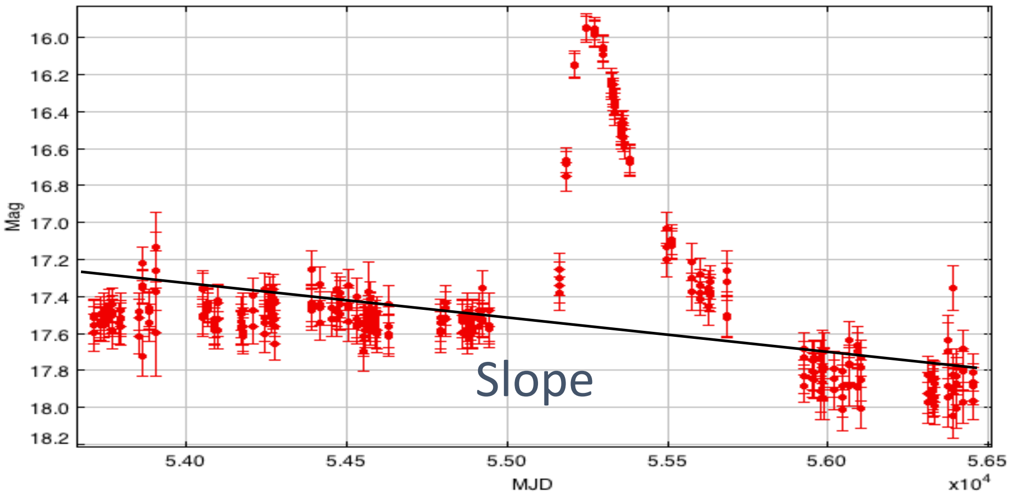
Different physical processes contribute to deterministic dominance *D(t)* or stochastic dominance *C x R(t).*

Deterministic chaos vs. stochastic?

Characterization – extracting data features

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

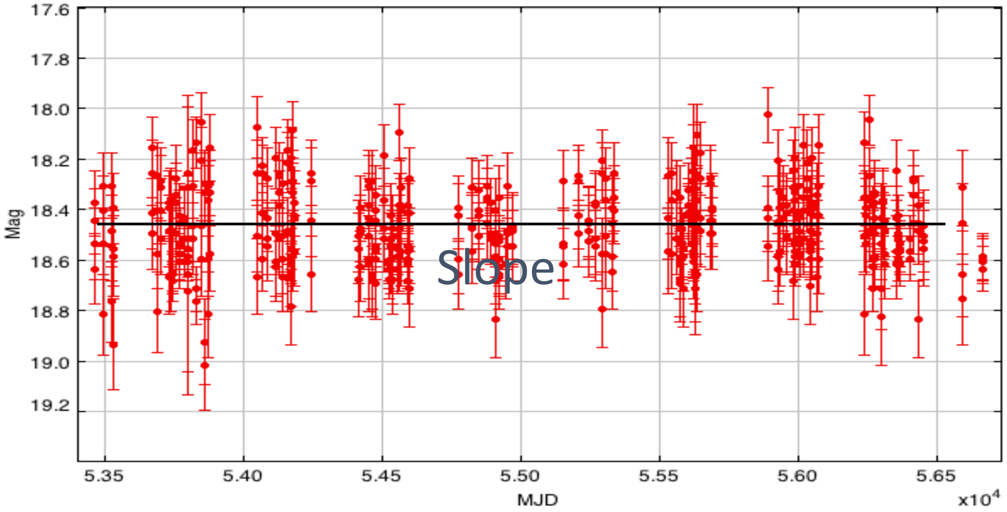
Fourier



Amplitude

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

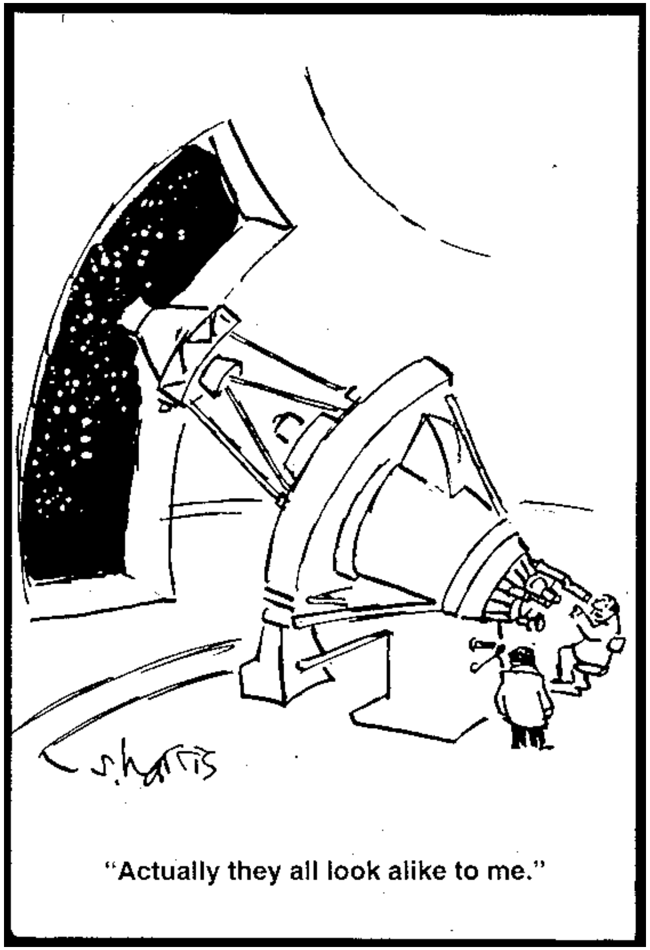
Fourier



Amplitude

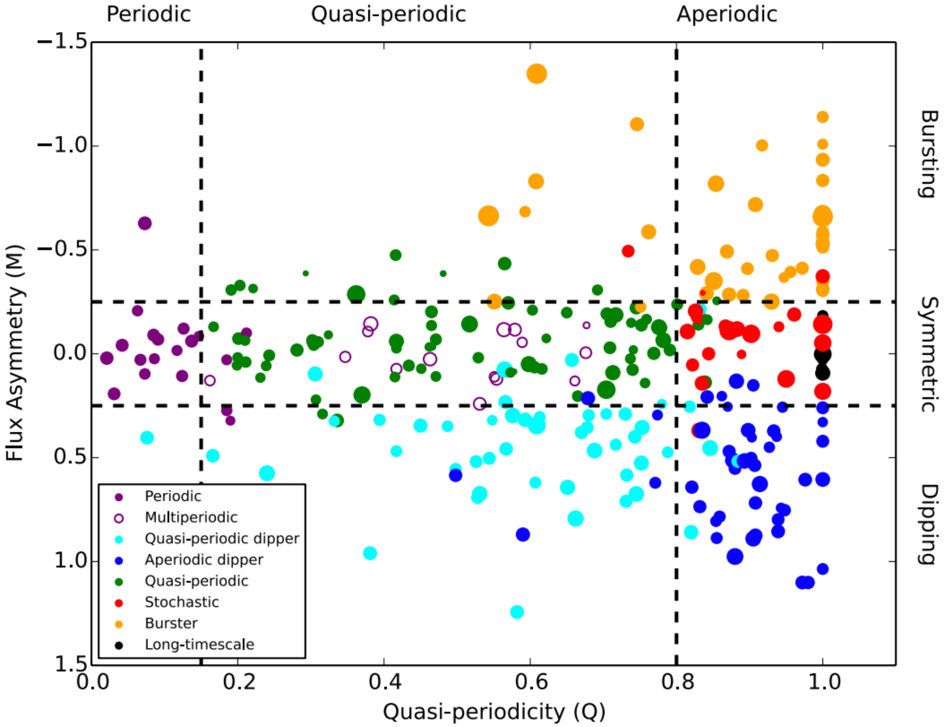
Common statistical features

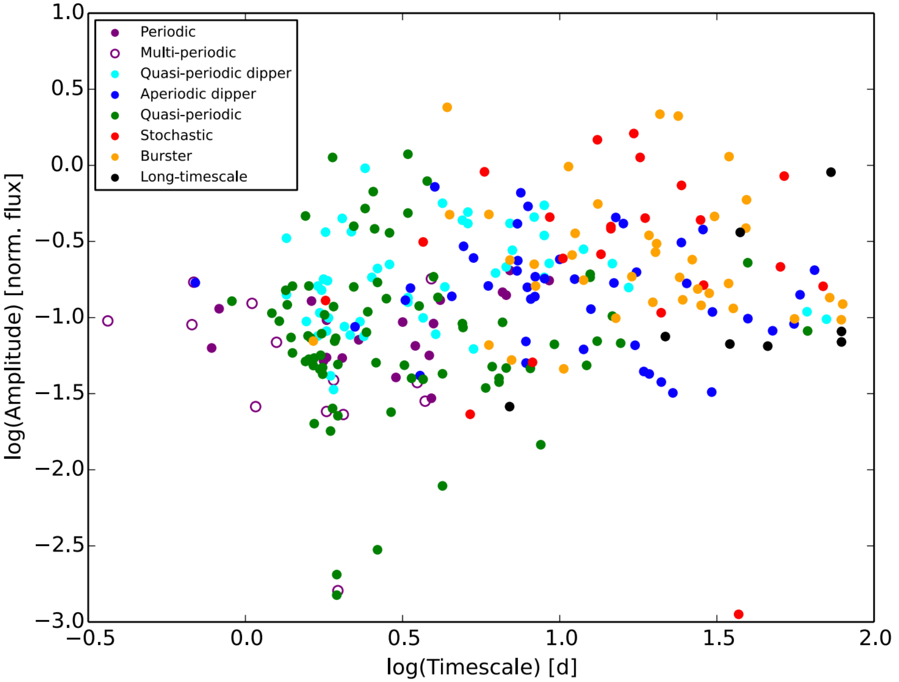
- Timescales:
  - Lomb-Scargle
- Variability:
  - von Neumann variability (phase-folded)
  - Stetson K index
- Morphology:
  - Skewness
  - Kurtosis
  - IQR
  - Cumulative sum index (phase-folded)
  - Ratio of magnitudes brighter/fainter than mean
- Trends:
  - Slope percentiles (phase-folded)
- Model:
  - Fourier amplitude ratios
  - Fourier phase differences
  - Fourier amplitude
  - Shapiro-Wilk normality test



"Actually they all look alike to me."

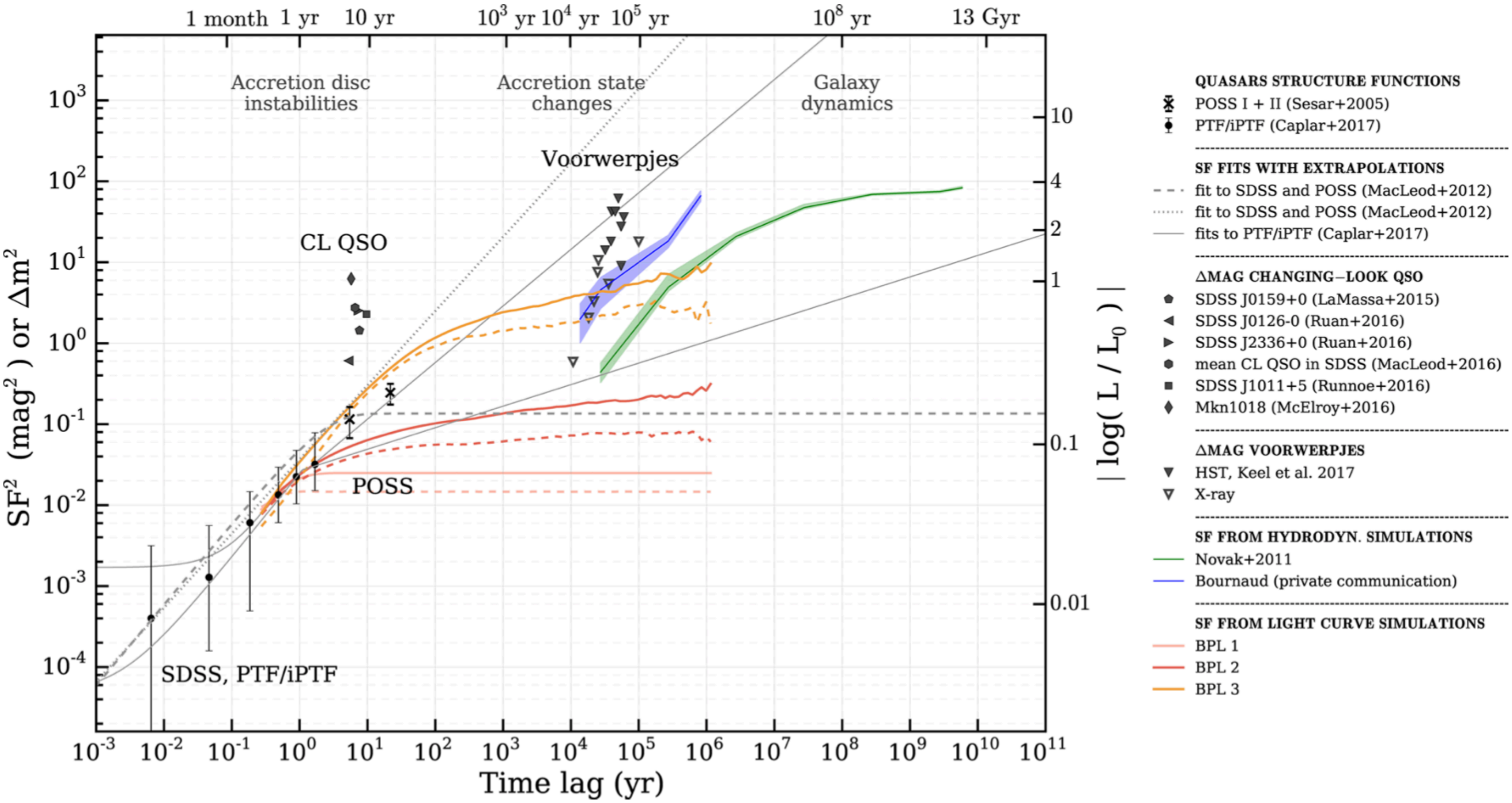
Categorization





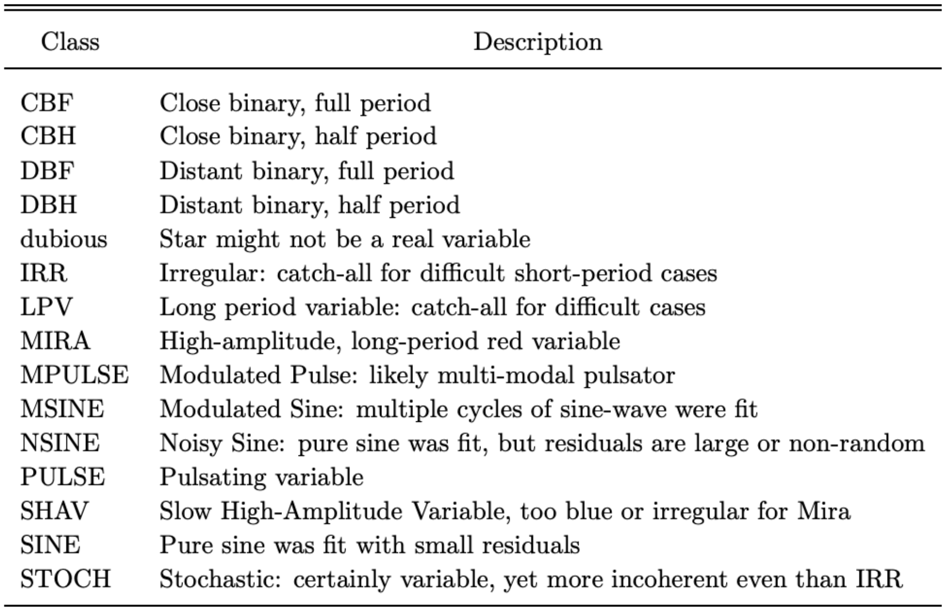
(Cody & Hillenbrand 2018)

Characteristic timescales

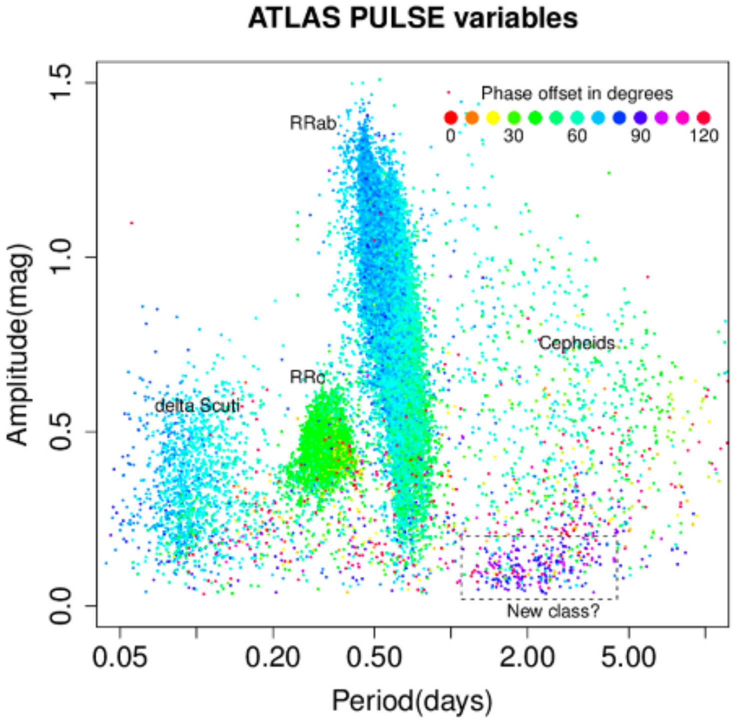


(Sartori et al. 2018)

Data-derived classes

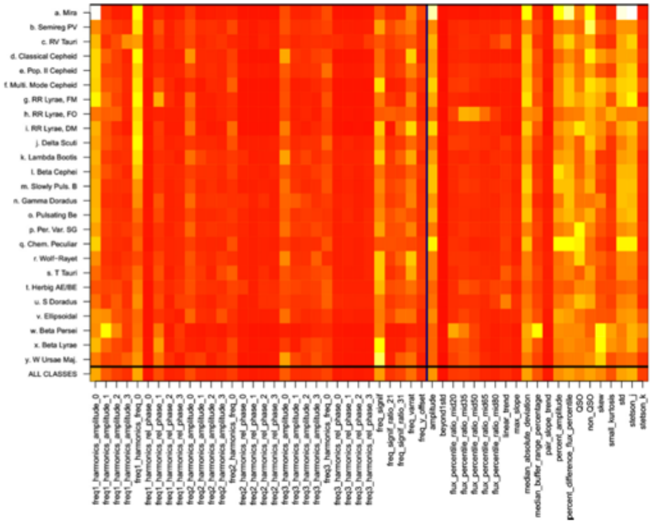


Class	Description
CBF	Close binary, full period
CBH	Close binary, half period
DBF	Distant binary, full period
DBH	Distant binary, half period
dubious	Star might not be a real variable
IRR	Irregular: catch-all for difficult short-period cases
LPV	Long period variable: catch-all for difficult cases
MIRA	High-amplitude, long-period red variable
MPULSE	Modulated Pulse: likely multi-modal pulsator
MSINE	Modulated Sine: multiple cycles of sine-wave were fit
NSINE	Noisy Sine: pure sine was fit, but residuals are large or non-random
PULSE	Pulsating variable
SHAV	Slow High-Amplitude Variable, too blue or irregular for Mira
SINE	Pure sine was fit with small residuals
STOCH	Stochastic: certainly variable, yet more incoherent even than IRR

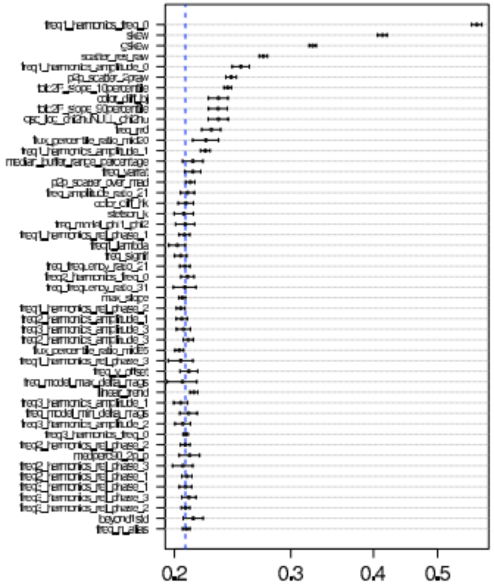


(Heinze et al. 2018)

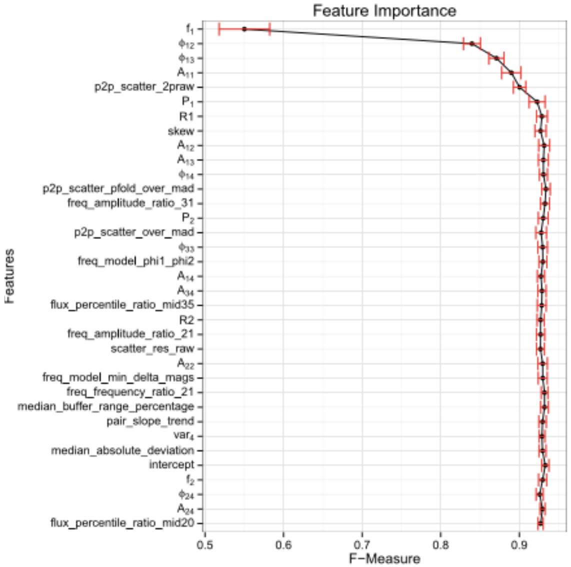
Not all features are equal



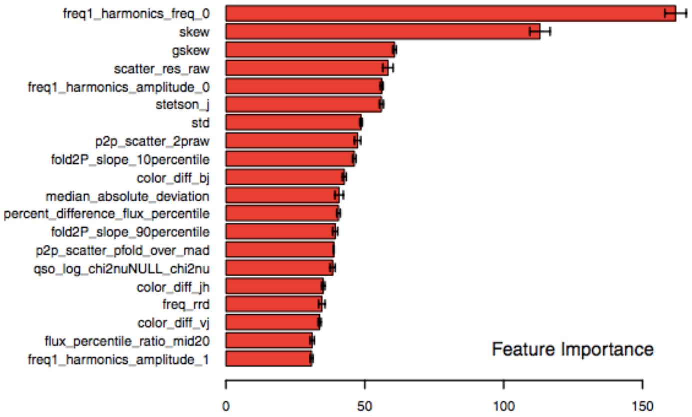
Richards et al. 2011



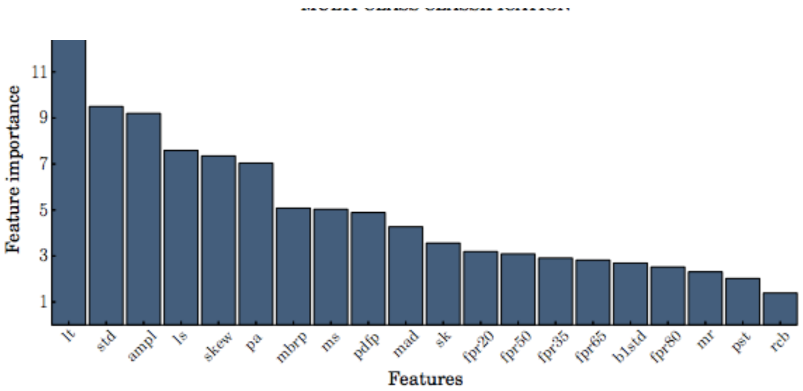
Dubath et al. 2012



Elorietta et al. 2016



Richards et al. 2012



D'Isanto et al. 2016

Periodicity

$$x(t+P) = x(t); f = 1/P$$

$$x(t,f) = A_f \sin 2\pi f\left(t - \varphi_f\right)$$

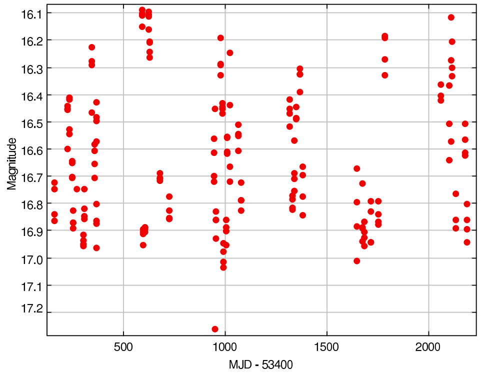
$$\chi^2(f) = \sum_n \left(\frac{x_n - x(t_n; f)}{\sigma_n}\right)^2$$

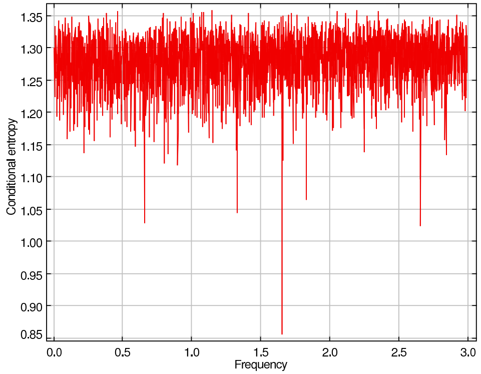
$$P(f) = \frac{1}{2}[\hat{\chi}_0^2 - \hat{\chi}^2(f)]$$

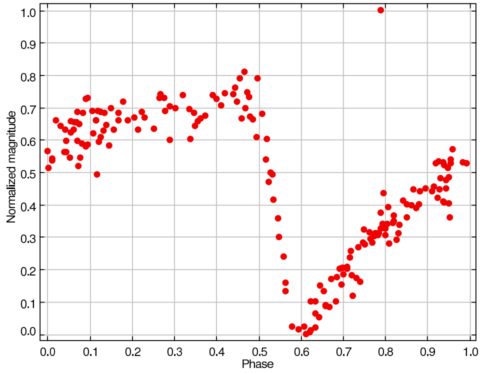
$$\varphi(t,f) = tf - \mathrm{int}(tf)$$

$$\theta(f) = g(\varphi_n, x_n; f)$$

$$P(f) = h(\theta(f))$$





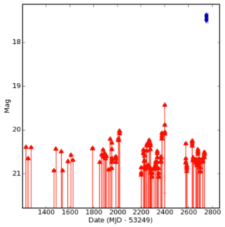


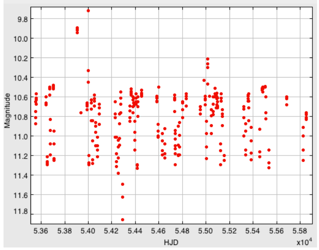
Period finding is not a single algorithm

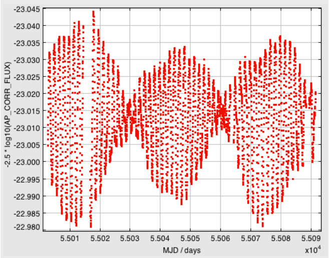
- Minimized (least-squares) fit to a set of basis functions:
  - Lomb-Scargle and its variants
  - Wavelets
- Minimize dispersion measure in phase space:
  - Means (PDM)
  - Variance (AOV)
  - String length
  - Entropy
- Rank ordering (in phase space)
- Bayesian
- Neural networks
- Gaussian process regression
- Convolved algorithms

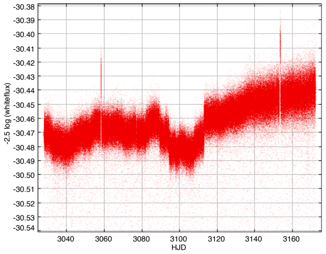
The most important feature: period

- Many features used to characterize light curves rely on a derived period:
  - Dubath et al. (2011) show a 22% misclassification error rate for non-eclipsing variable stars with an incorrect period
  - Richards et al. (2011) estimate that periodic feature routines account for 75% of computing time used in feature extraction
  - Deep learning still applied to folded light curves
- Domain knowledge constraints
  - RR Lyrae: Blazho behavior (30%), small amplitude cycle-to-cycle modulations (RRabs)
  - Close binaries, LPVs: cyclic period changes over multidecade baselines
  - Semi-regular variables: double periods, multiperiodicity
  - ARMA models: quasi-periodicity
- Trustworthiness of quoted periods

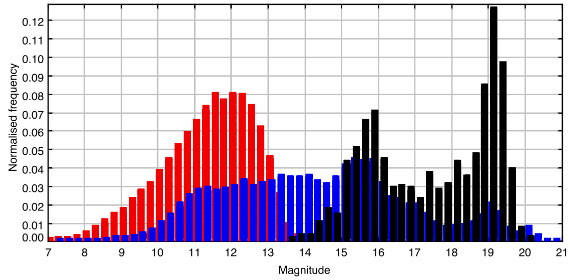


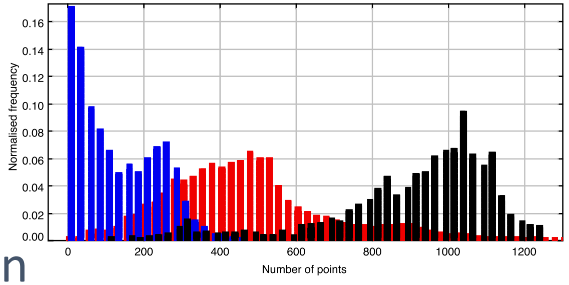






Investigating period finding accuracies





- Data set:
  - 15522 CRTS light curves for all objects in SIMBAD and VSX with a quoted period
  - 50124 ACVS light curves for MACC classification
  - 1500 MACHO light curves for RR Lyrae, EBs and Cepheids
- Classes:
  - Eruptive (4194): T Tauri, red supergiants, RS Can Ven
  - Pulsating (45599): semiregulars, RR Lyrae, Mira, δ Scuti, Cepheids
  - Rotating (455): chemically peculiar, BY Dra
  - Cataclysmic (386): S U Ma, U Gem, novalike
  - Eclipsing (14952): eclipsing binaries, AM Her
  - Other (1369)
- 9 different algorithms

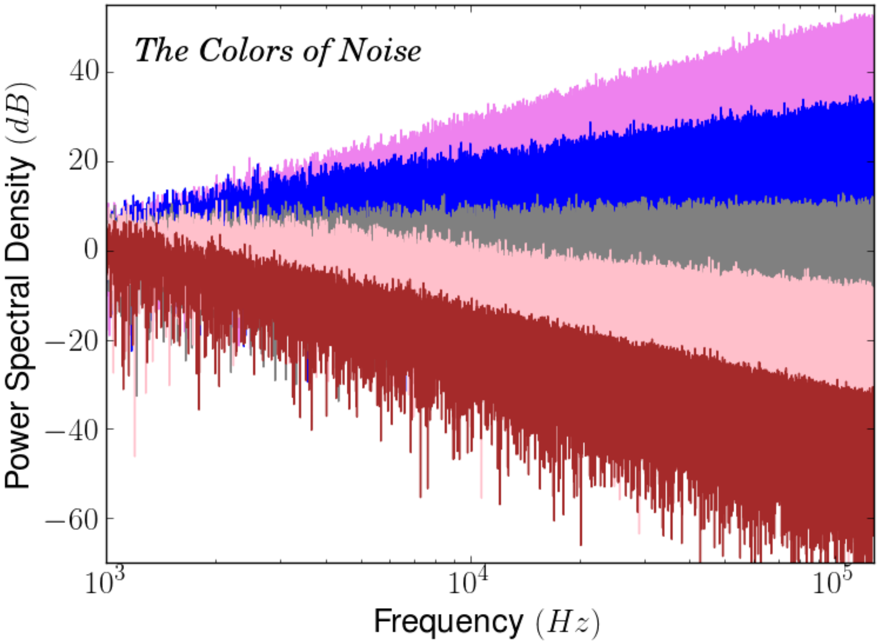
(Graham et al. 2013)

What can we say about period finding

- No algorithm is generally better than ~60% accurate
- All methods are dependent on the quality of the light curve and show a decline in period recovery with lower quality light curves as a consequence of:
  - fewer observations
  - fainter magnitudes
  - noisier data and an increase in period recovery with higher object variability;
- All algorithms are stable with a minimum bin occupancy of ~10 (Δφ = 0.1)
- A bimodal observing strategy consisting of pairs (or more) of short Δt observations per night and normal repeat visits is better
- The algorithms work best with pulsating and eclipsing variable classes
- LS/GLS are strongly effected by half-period issue (eclipsing binaries)
- Specific algorithms work better with irregular sampling, bright magnitudes (containing saturated values), or with performance constraints

Autoregressive models

- Purely random: $x_t = z_t$ where $\{z_t\}$ are iid
- Random walk (Brownian motion): $x_t = x_{t-1} + z_t$
- Autoregressive: $x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + z_t$
- Moving average: $x_t = z_t + \beta_1 z_{t-1} + \cdots + \beta_{t-q} z_{t-q}$
- ARMA(p,q): $x_t = \alpha_1 x_{t-1} + \cdots + \alpha_{t-p} x_{t-p} + z_t + \beta_1 z_{t-1} + \cdots + \beta_q z_{t-q}$
- ARIMA(p, d, q), ARFIMA(p,d, q):
- $(1 - B)^d x_t = z_t$

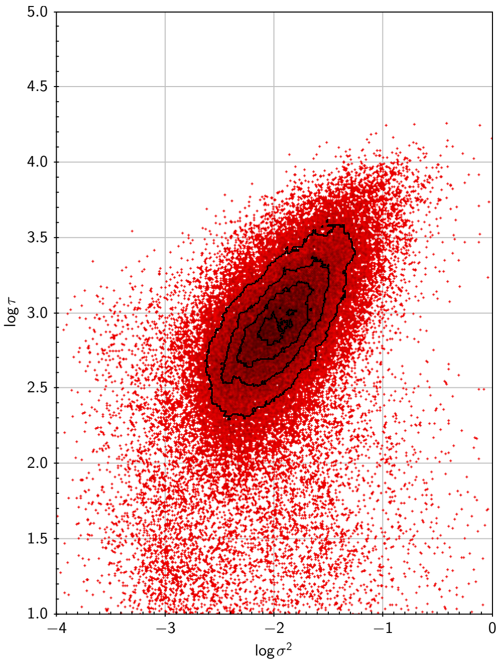


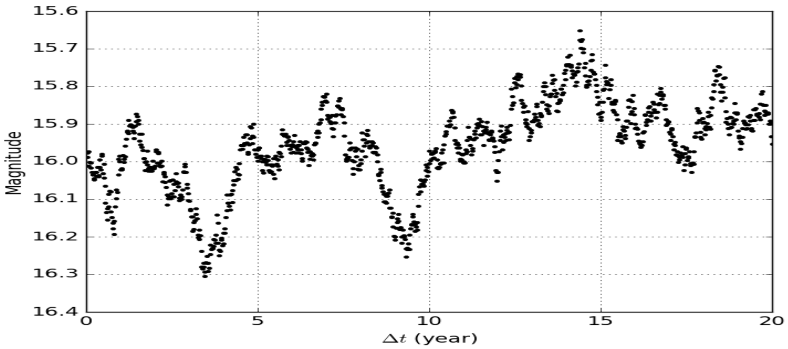
Quasar variability as a damped random walk

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

$$X_{i+1} = X_i e^{-\Delta t/\tau} + G\left[\sigma^2\left(1 - e^{-2\Delta t/\tau}\right)\right] + b$$

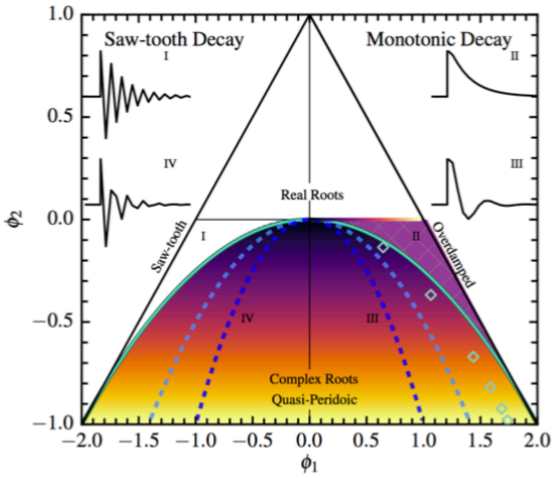
- Characterized by variability amplitude and timescale
- Basis for stochastic models of variability
- Deviations noted (e.g., Mushotzky 2011, Zu et al. 2013, Graham et al. 2014)
- Degenerate model – can be best fit for a non-DRW process (Kozlowski 2016)

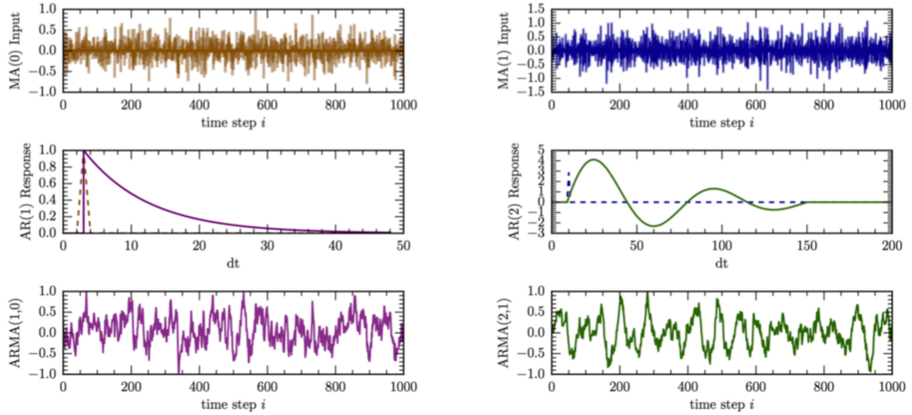


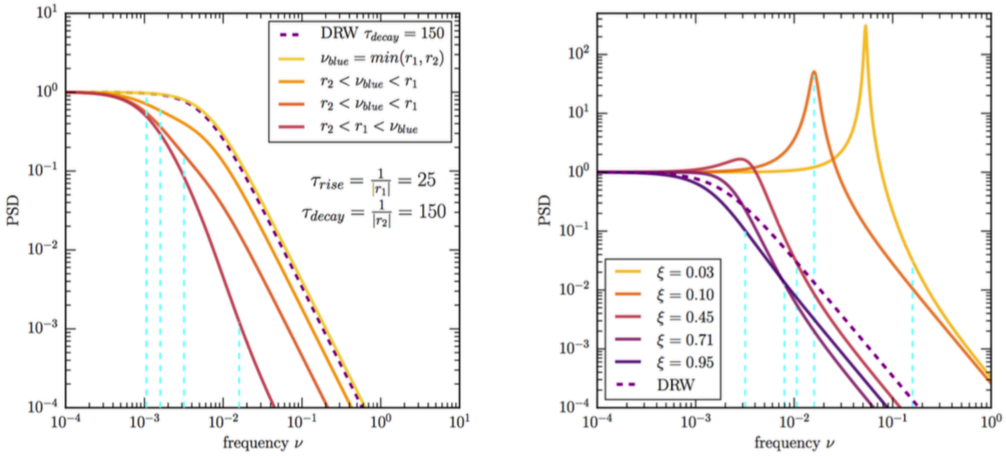


More autoregressive – CARMA(2,1)

$$d^2x + \alpha_1 d^1x + \alpha_2 x = \beta_0 z_t + \beta_1 z_{t-1}$$

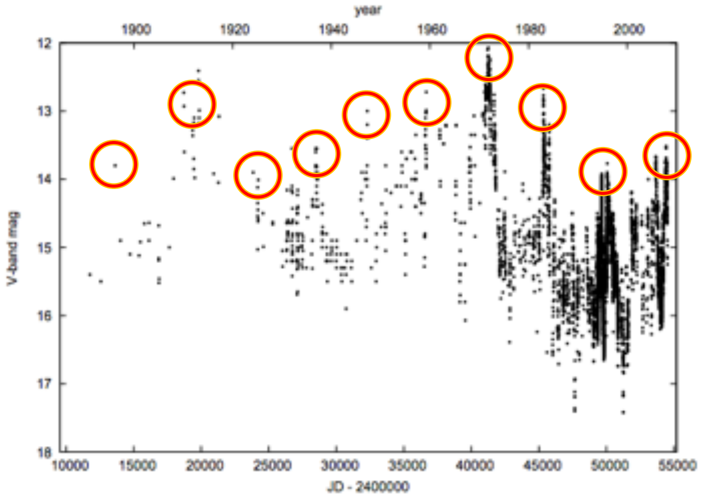


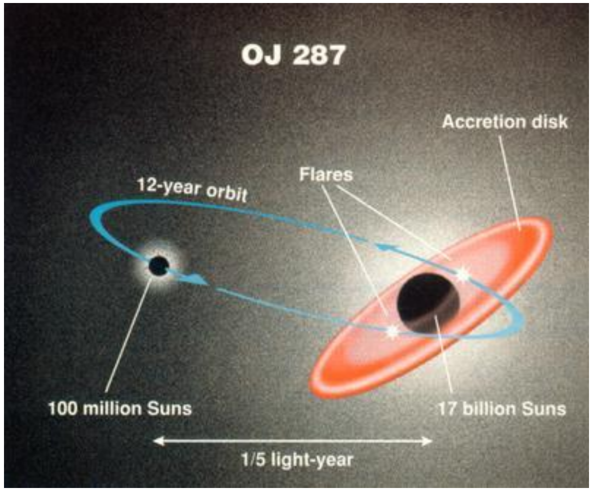


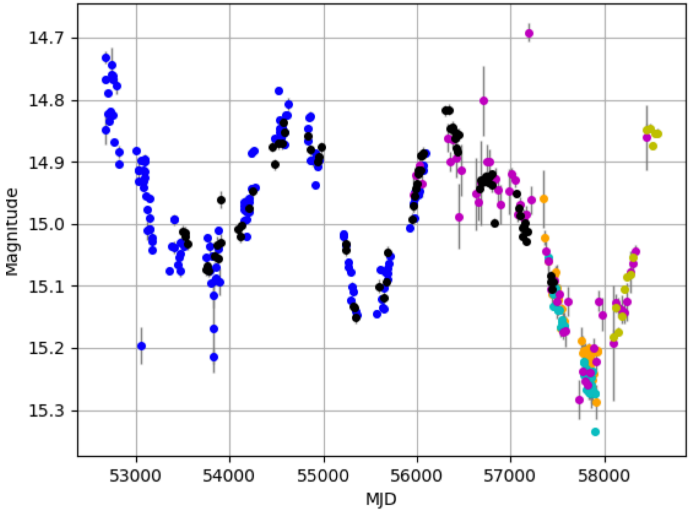


(Moreno et al. 2019)

Periodic quasars?

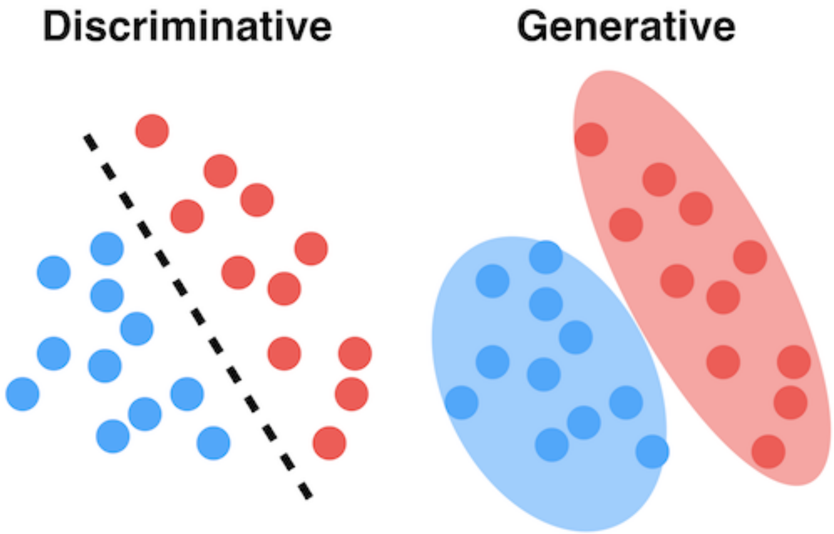






Generative vs. discriminative

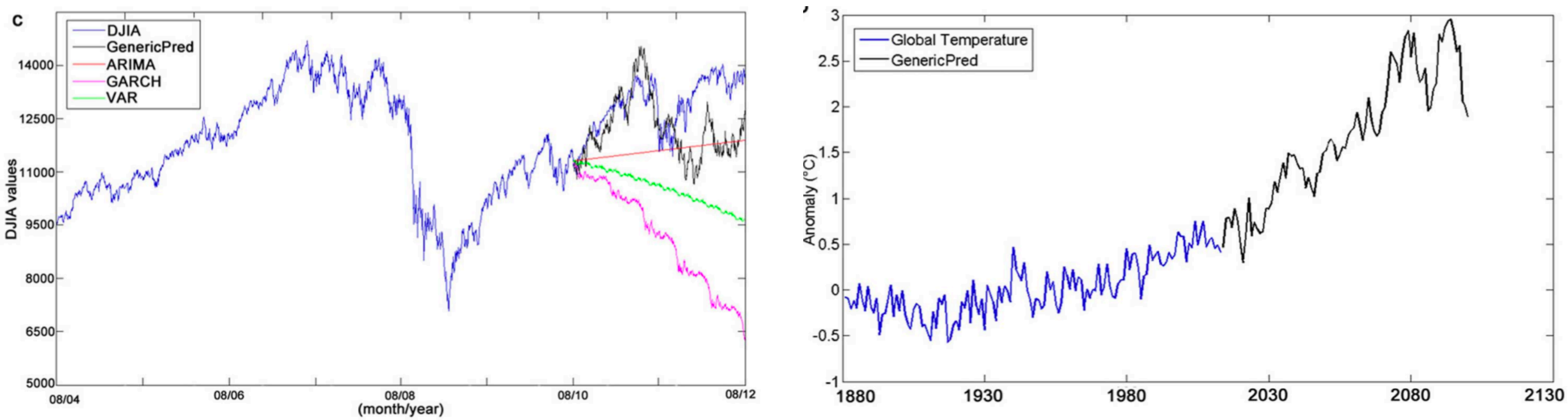
- Current statistical models of variability are designed to discriminate between classes, e.g. stars/galaxies – p(y|x)
- Better to learn time series (shape) rather than determining some parameterizable form – p(y, x)
- Generative approach that supports predictions



Forecasting

- Predicting periodic behavior is trivial
- Predict aperiodic (chaos or stochastic) behavior:
  - Stock market
  - Climate change
  - Epileptic seizures
  - Earthquakes

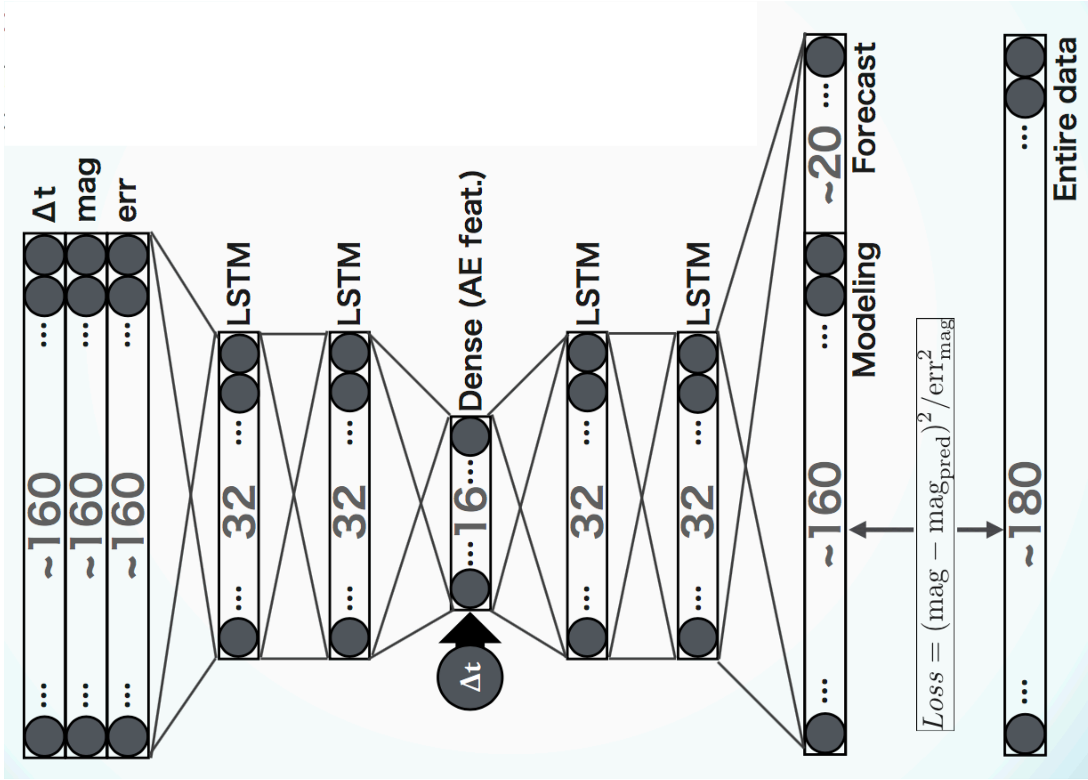
- Gaussian process regression
- Localized chaos measure



(Golestani & Gras 2014)

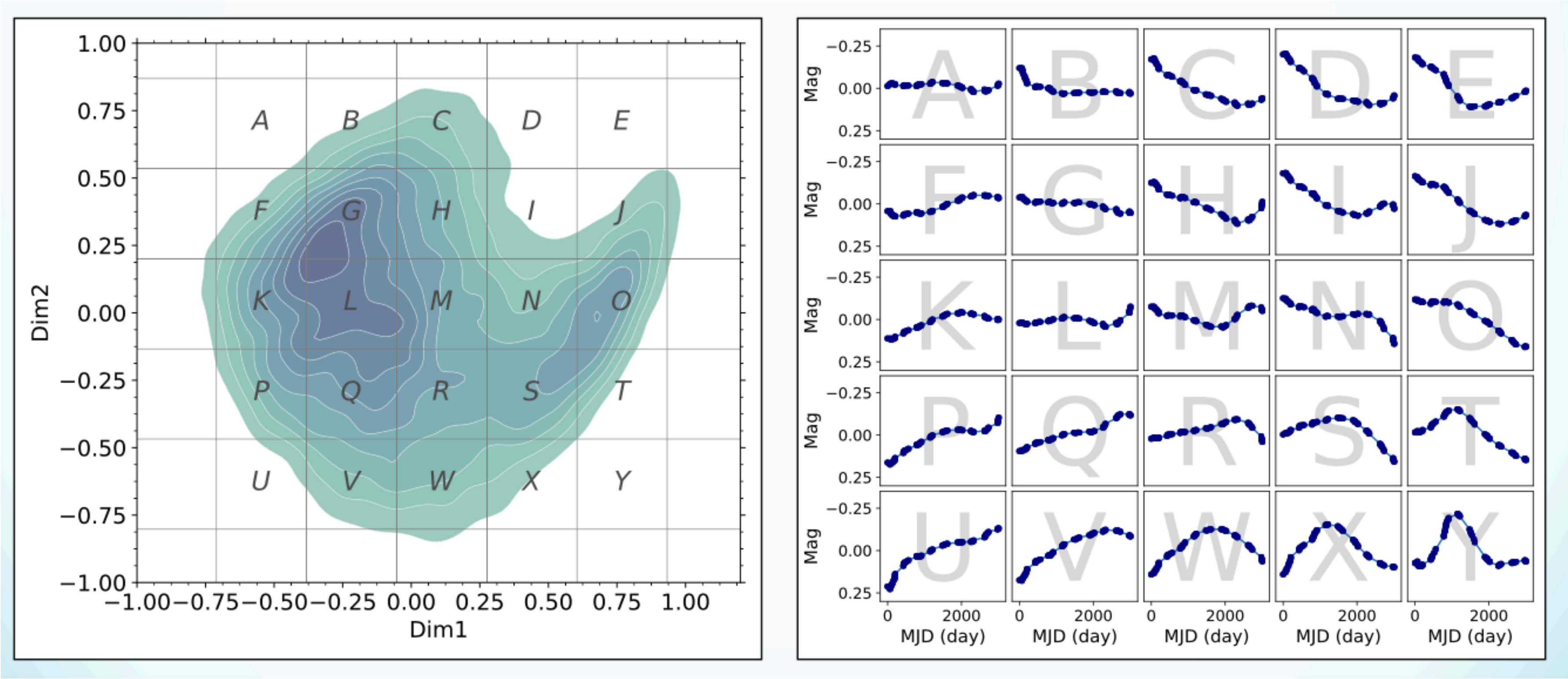
Deep modelling of time series

- LSTM Autoencoder:



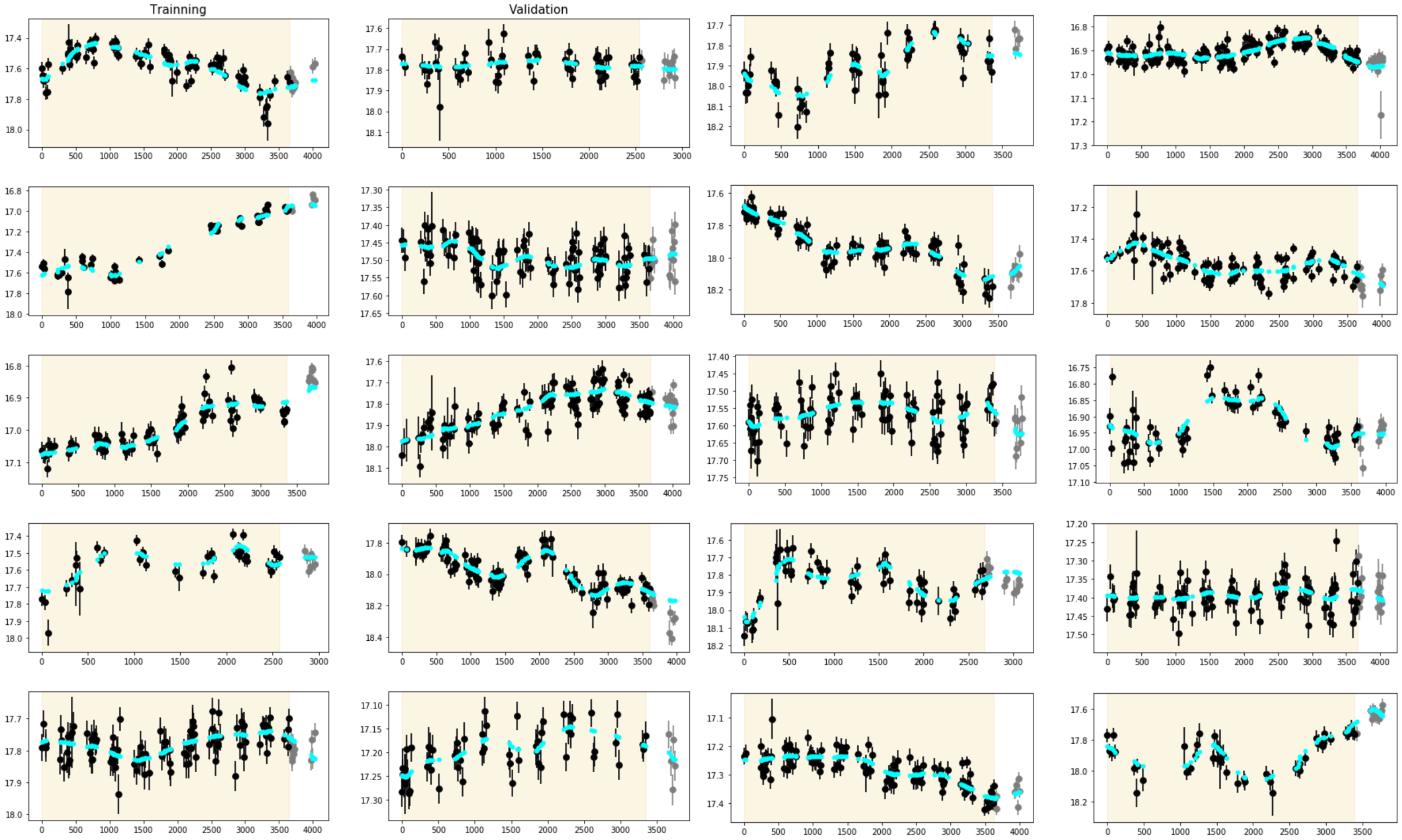
(Naul et al. 2018)

Deep time series features



(Tachibana et al. 2019)

RNNs with QSOs



Summary

- Traditional time series analyses in astronomy involve:
  - (simple) discriminative features as (possible) inputs to machine learning algorithms
  - outlier detections based on Gaussian tails
  - little predictive power
- Data volumes now mean that we can *model individual* sources:
  - capturing full time series behavior
  - better identifying extrema
  - with generative approaches
- Next generation surveys enable real-time validation of predicted behaviors and swift identification of deviance
- Let’s go hunting for technosignatures