Astroinformatics 2019
Caltech, 2019 June 24

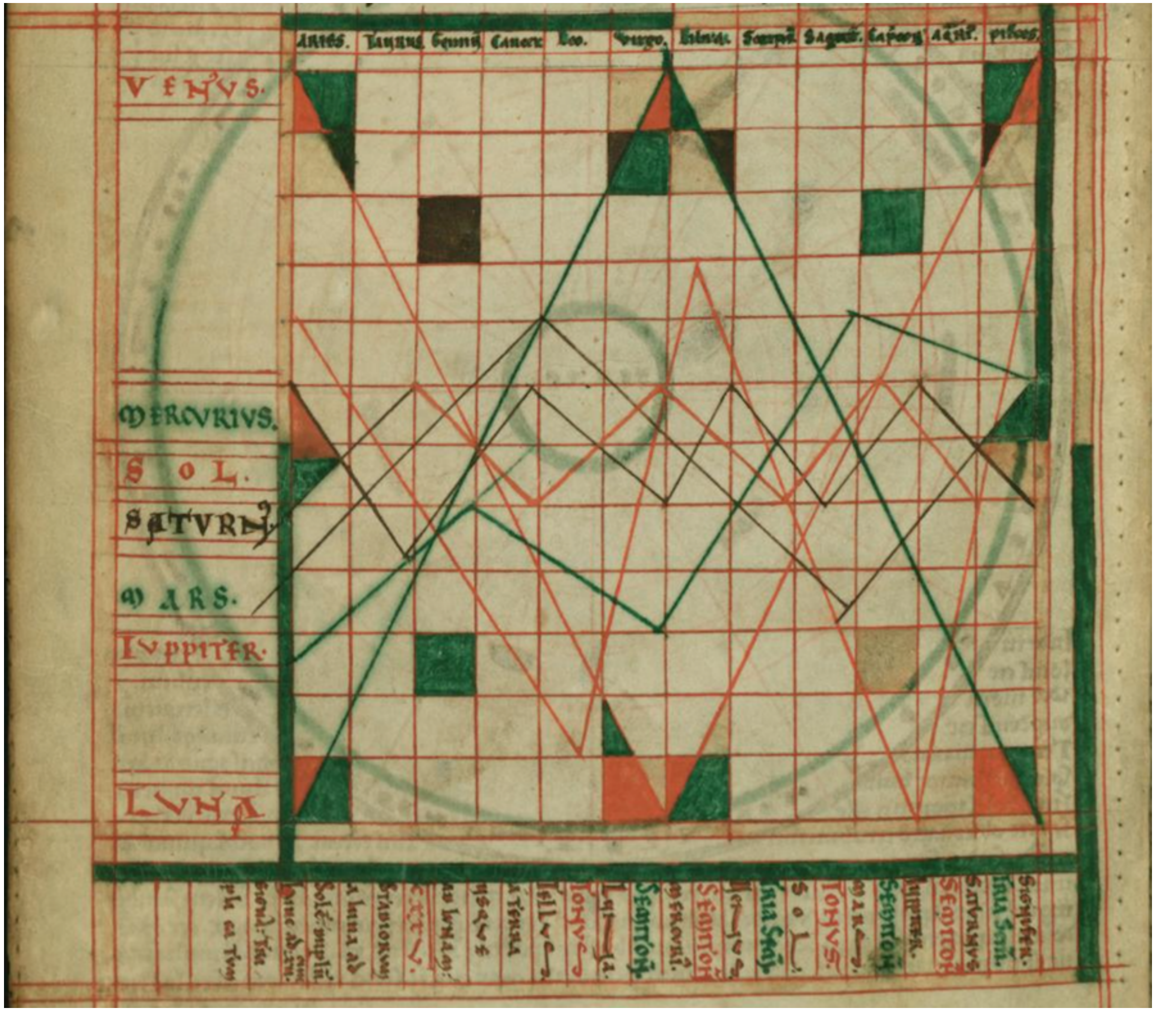
An introduction to time series analysis

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

CENTER FOR DATA-DRIVEN DISCOVERY

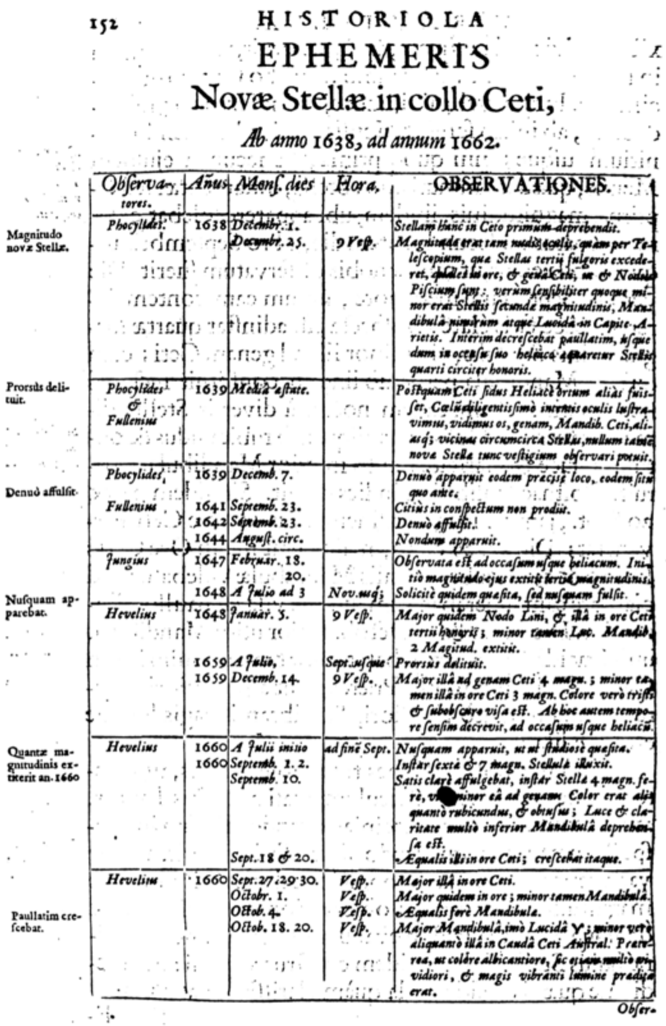
ZTF

The first astronomical time series analysis



Abbo of Fleury, 10th century CE

A wondrous star in the neck of the Whale



152 HISTORIOLA

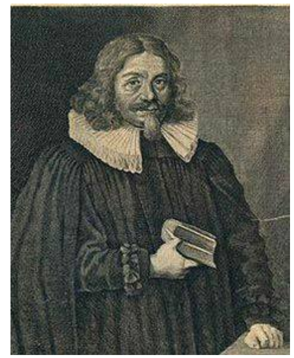
EPHEMERIS

Novæ Stellæ in collo Ceti,

Ab anno 1638, ad annum 1662.

	Observatores.	Ânus	Mens. dies	Hora.	OBSERVATIONES.
Magnitudo novæ Stellæ.	Phocylides.	1638	Decembr. 1.		Stellam hanc in Ceto primum deprehendit.
			Decembr. 25.	9 Vesp.	Magnitudo erat tam nudis oculis, quam per Telescopium, qua Stellas tertij fulgoris excederet, quales in ore, & genâ Ceti, ut & Nodo Piscium sunt; verùm sensibiliter quoque minor erat Stellis secundæ magnitudinis, Mandibulâ puta 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, Cœlum diligentissimè intentis oculis lustravimus, vidimus os, genam, Mandib. Ceti, aliasq; vicinas circumcirca Stellas, nullum tamen nova Stella tunc vestigium observari potuit.
Denuò affulsit.	Phocylides,	1639	Decemb. 7.		Denuò apparuit eodem præcisè loco, eodem situ quo ante.
	Fullenius	1641	Septemb. 23.		Citiùs in conspectum non prodiit.
		1642	Septemb. 23.		Denuò affulsit.
		1644	August. circ.		Nondum apparuit.
	Jungius	1647	Februar. 18.		Observata est ad occasum usque heliacum. Initiò magnitudo ejus extitit tertiæ magnitudinis.
			20.		
Nusquam apparebat.		1648	A Julio ad 3	Nov. usq;	Solicitè quidem quæsita, sed nusquam fulsit.
	Hevelius	1648	Januar. 5.	9 Vesp.	Major quidem Nodo Lini, & illâ in ore Ceti tertii honoris; minor tamen Luc. Mandib. 2 Magitud. extitit.
		1659	A Julio,	Sept. usque	Prorsùs delituit.
		1659	Decemb. 14.	9 Vesp.	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	A Julii initio	ad finẽ Sept.	Nusquam apparuit, ut ut studiosè quæsita.
		1660	Septemb. 1. 2.		Instar sextæ & 7 magn. Stellulæ illuxit.
			Septemb. 10.		Satis clarè affulgebat, instar Stellæ 4 magn. ferè, [illegible] minor eâ ad genam. Color erat aliquantò rubicundus, & obtusus; Luce & claritate multò inferior Mandibulâ deprehensa est.
			Sept. 18 & 20.		Æqualis illi in ore Ceti; crescebat itaque.
Paullatim crescebat.	Hevelius	1660	Sept. 27. 29. 30.	Vesp.	Major illâ in ore Ceti.
			Octobr. 1.	Vesp.	Major quidem in ore; minor tamen Mandibulâ.
			Octob. 4.	Vesp.	Æqualis ferè Mandibulæ.
			Octob. 18. 20.	Vesp.	Major Mandibulâ, imò Lucidâ ♈; minor verò aliquantò illâ in Caudâ Ceti Austral. Præterea, ut colôre 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. "

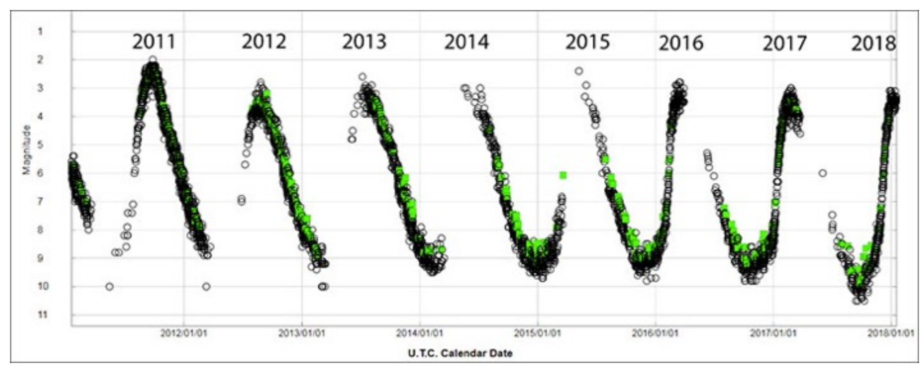


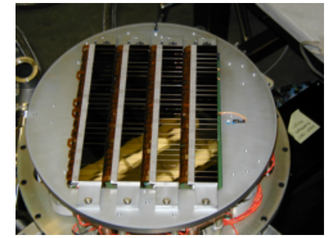
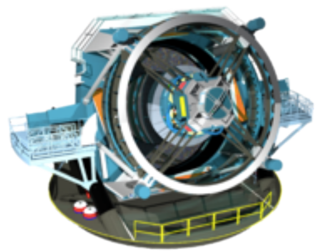
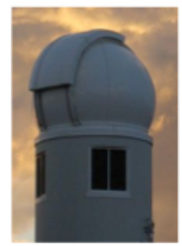
Image credit: AAVSO

A billion time series and counting

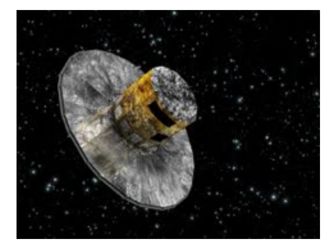
- Palomar-Quest Synoptic Sky Survey
- SDSS (Stripe 82)
- Catalina Real-time Transient Survey
- Palomar Transient Factory
- Zwicky Transient Factory
- Pan-STARRs
- SkyMapper
- ASKAP
- ThunderKat (MeerKAT)
- KEPLER
- GAIA
- LIGO
- IceCUBE
- LOFAR
- LSST
- SKA
- TESS
- ASAS-SN
- MASTER
- DES
- ATLAS
- BlackGEM

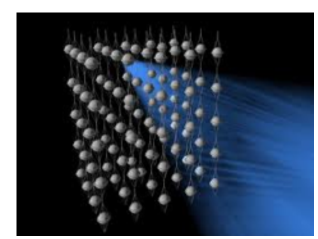
- GoTo
- MeerKAT
- ASKAP
- WISE
- OGLE
- DESI
- SDSS-V
- LAMOST

…



What we do ask of time series?

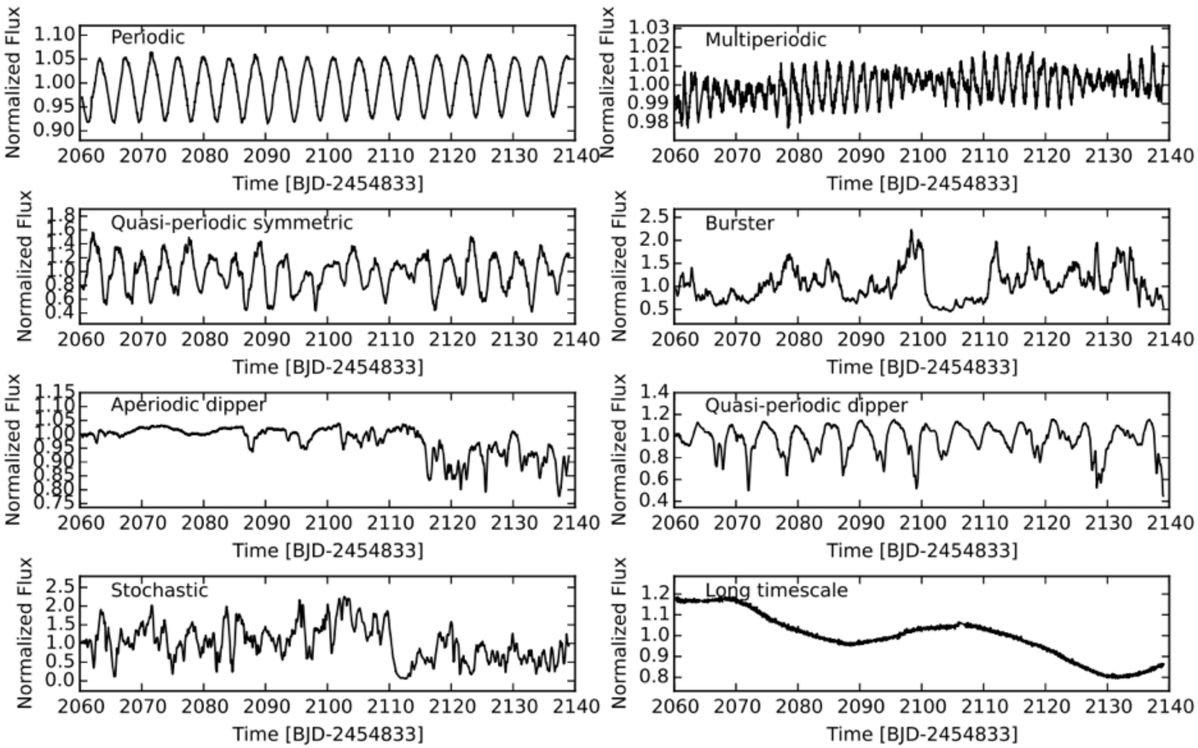
Population behaviors

- Characterize, categorize, classify

Outliers

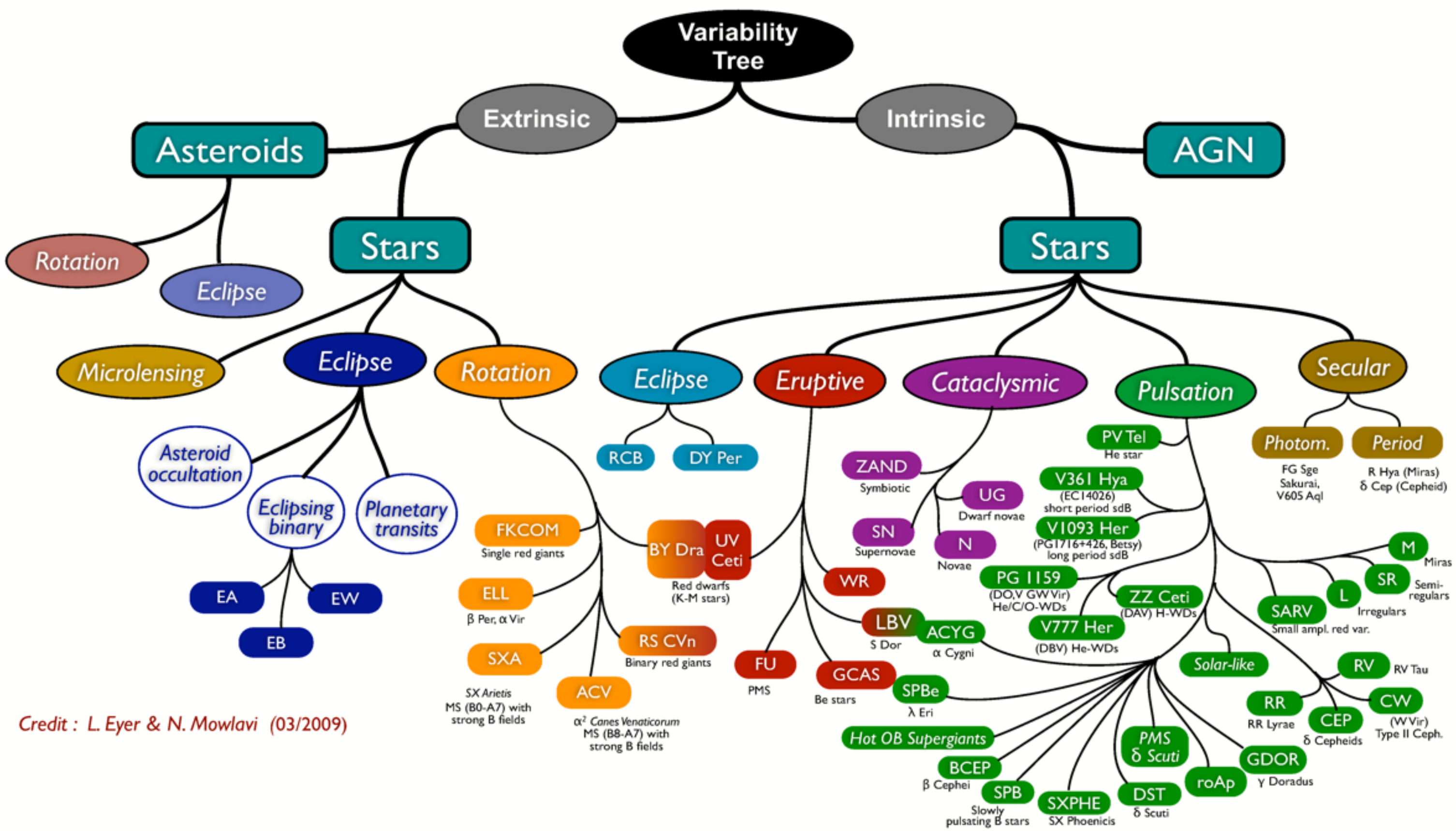
- Extreme sources

- **Physical models**
  - Predictions



(Cody & Hillenbrand 2018)

Types of astronomical variability



Foundational concepts

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

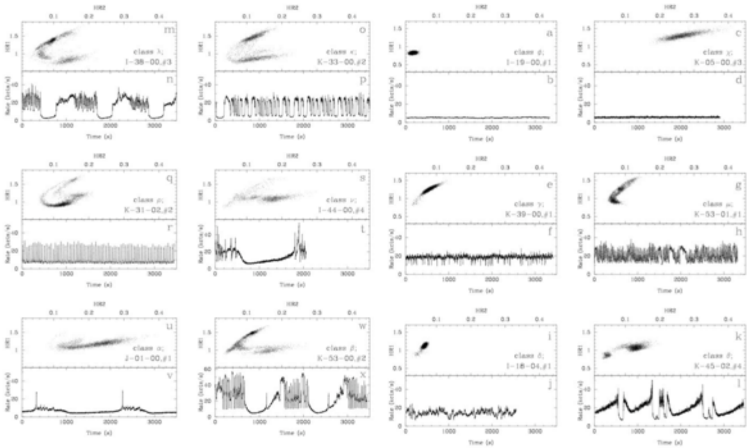
Ergodicity

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

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

Foundational concepts - stationarity

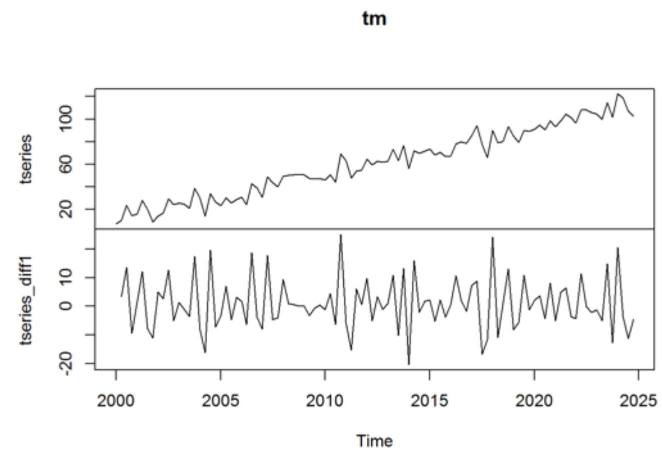
- The generating process is time independent:
  - Joint probability distribution is translationally invariant (strong)
  - Mean, variance, autocorrelation are constant (weak)
- Examples:
  - White noise is stationary
  - GSR 1915+215 has ~20 variability states
  - GARCH models where variance is a stochastic function of time
- Nonstationary time series do not have to be stationary in any limit



(Belloni et al. 2000)

Foundational concepts - stationarity

- Transformations to achieve stationarity (constant location and scale):
  - Difference the data:
    $$Z_i = X_i - X_{i-1}$$
  - Detrend the data:
    $$Z(t) = X(t) - f(t)$$
  - Stabilize the variance:
    $$Z(t) = \sqrt{(X(t) + A)} \text{ or } \log(X(t) + A)$$



Test with ACF

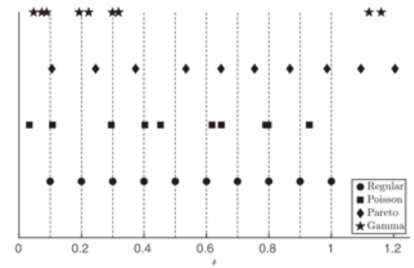
Foundational concepts - sampling

- Even or regular sampling:

$$y(t) = x(t_0 + n\Delta t) \text{ where } n = 0,1,\ldots,m$$

- Uneven or irregular sampling:

$$y(t) = x(t_0),\ldots,x(t_m)$$



- Regularization/resampling:
  - Bin data onto regular grid: $y(t) = \frac{\sum_i w_i x_i}{\sum_i w_i}$ for $t_i \in [t_a, t_b]$
  - Interpolate: linear, spline, Gaussian process
- Continuous time process:
  - Observations are a random sample drawn from a continuous process described by some differential equation:

$$dX(t) = -\frac{1}{\tau}X(t)dt + \sigma\sqrt{dt}\epsilon(t) + bdt$$

Foundational concepts – power spectrum

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

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

Discrete FT:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / N}$$

Nonuniform Discrete FT:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i p_n \omega_k}$$

- 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))

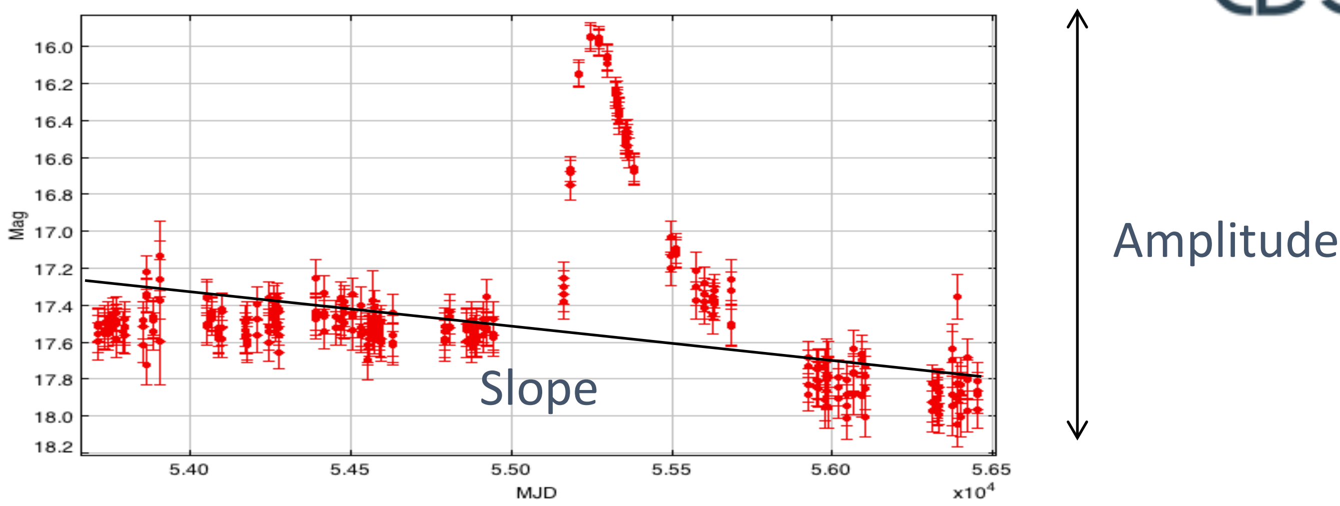
Different physical processes contribute to deterministic dominance *D(t)* or stochastic dominance *C x R(t).*

Deterministic chaos vs. stochastic?

Characterization – extracting data features

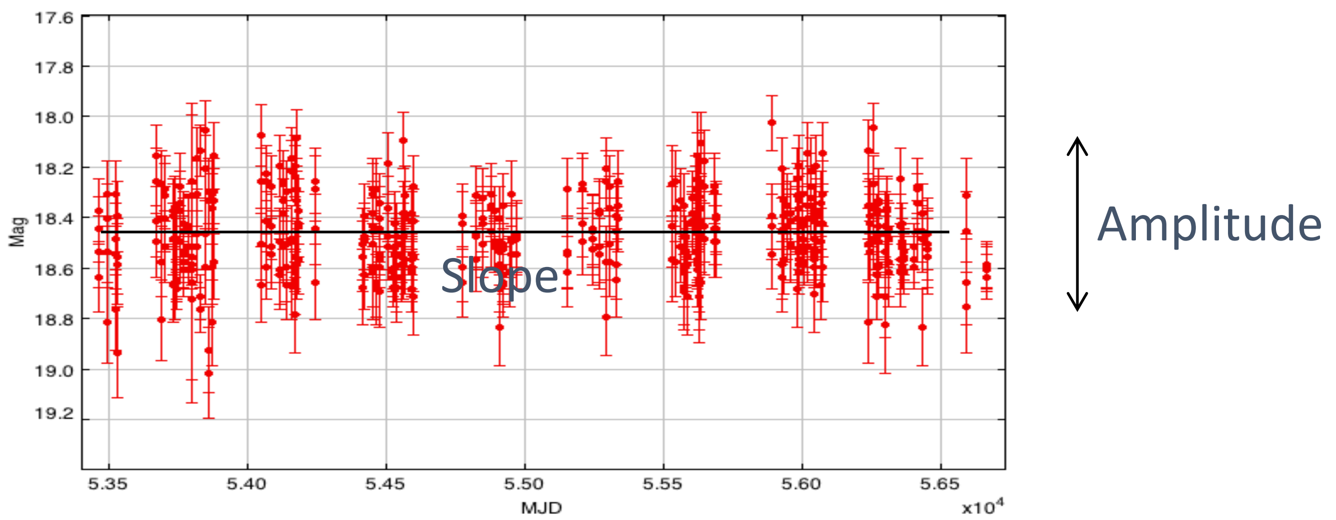
$$\sum_{i=1}^{n} A_i \sin(\omega t + \phi_i)$$

Fourier



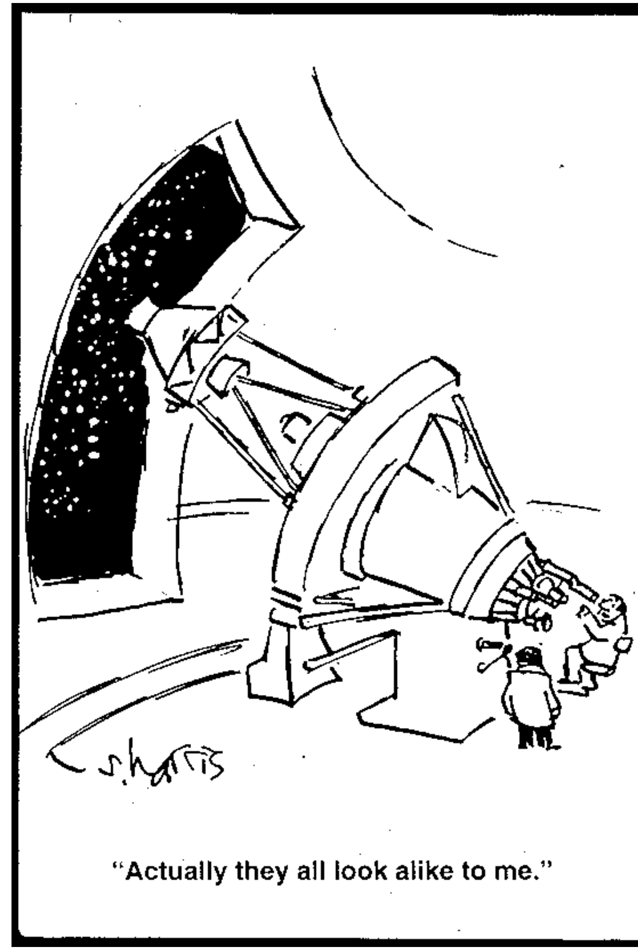
$$\sum_{i=1}^{n} A_i \sin(\omega t + \phi_i)$$

Fourier

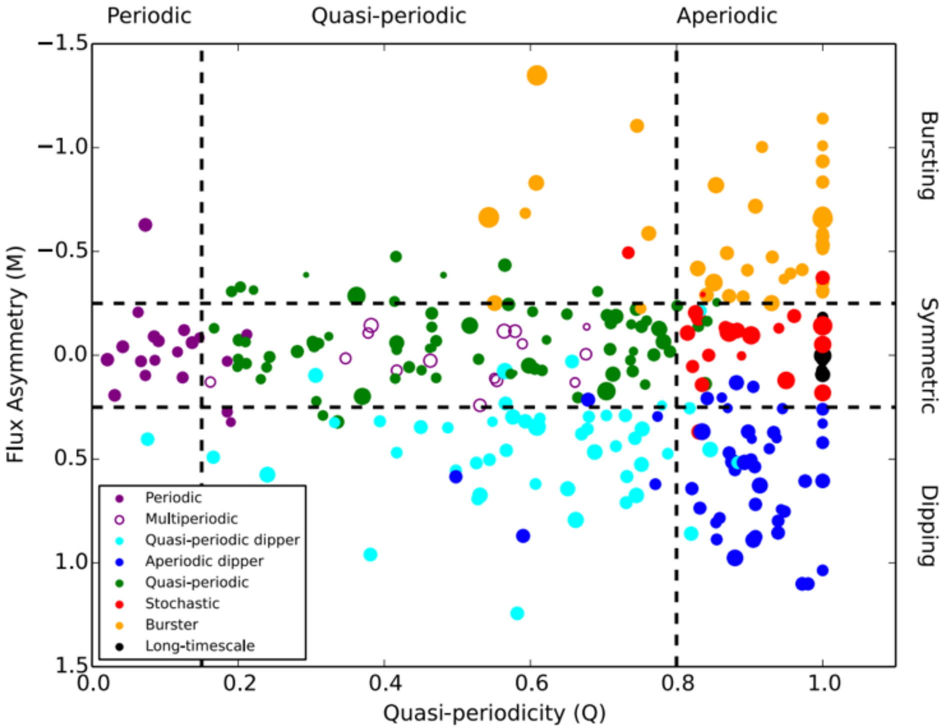


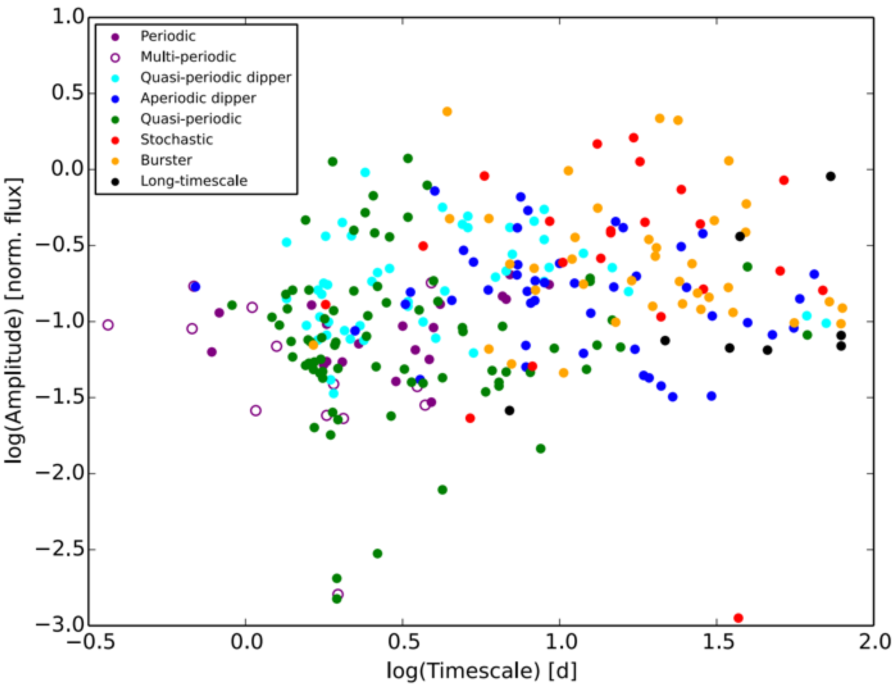
Common statistical features

- 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



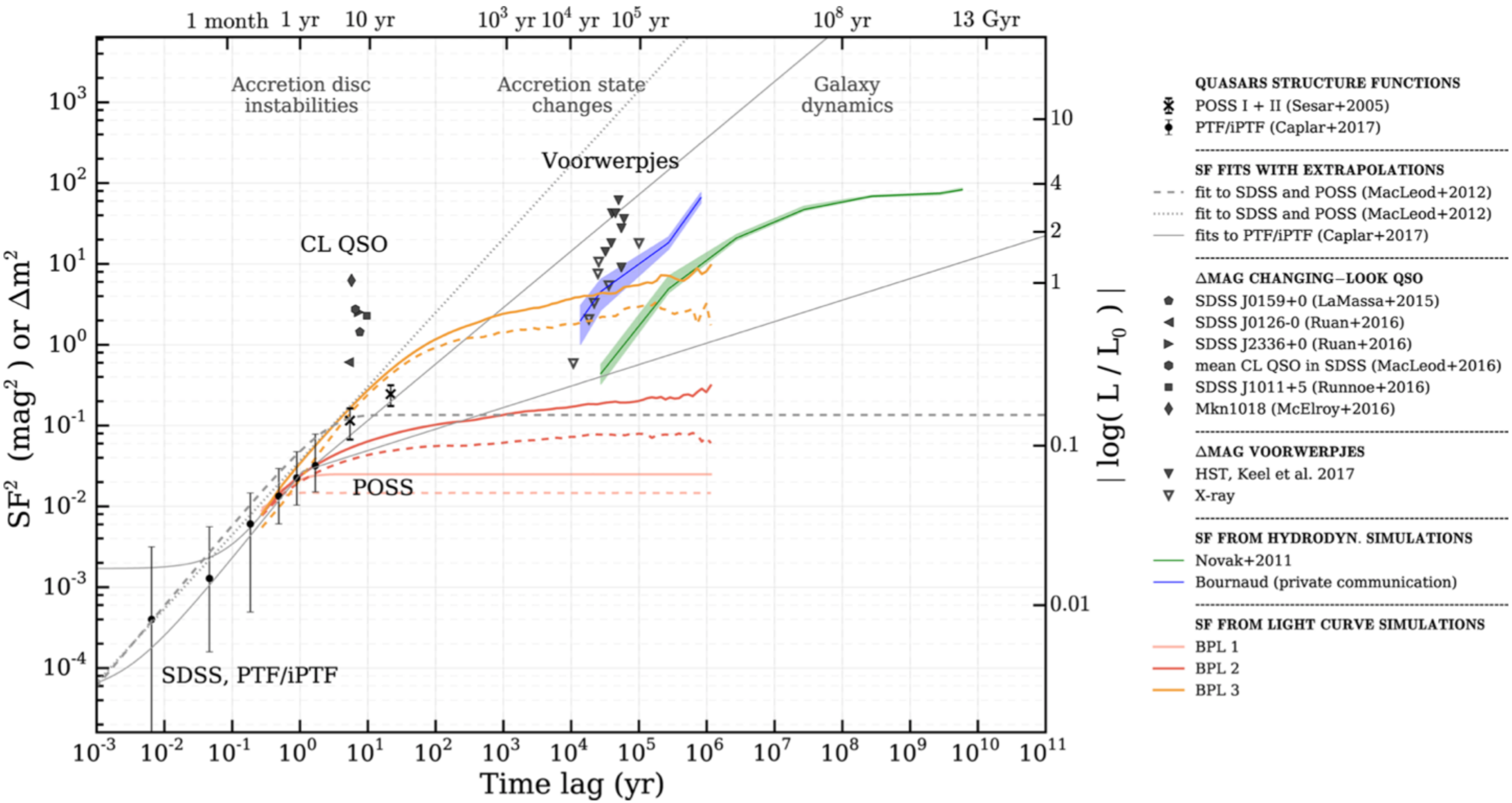
Categorization





(Cody & Hillenbrand 2018)

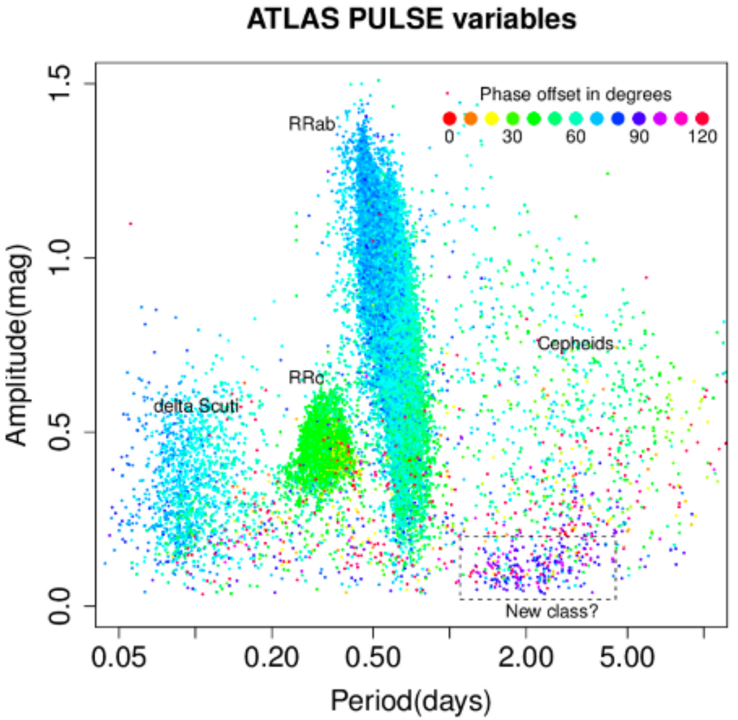
Characteristic timescales



(Sartori et al. 2018)

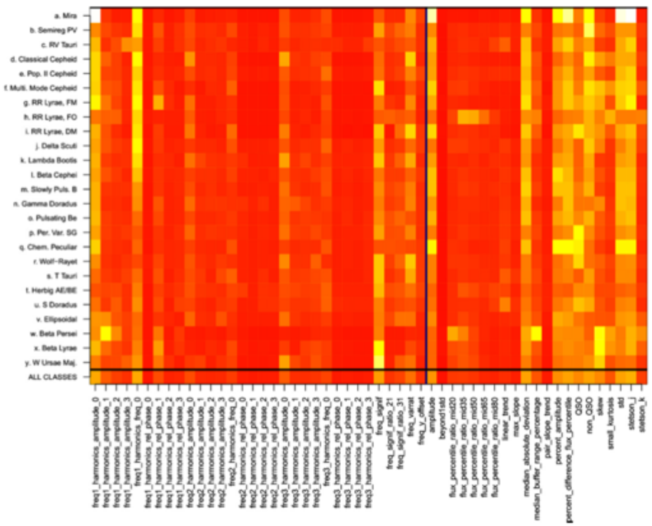
Data-derived classes

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

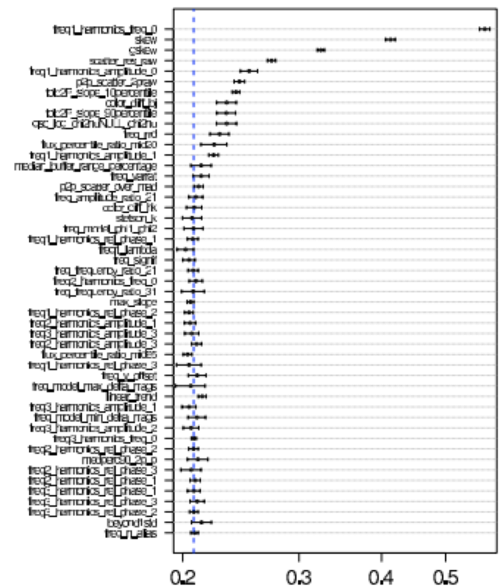


(Heinze et al. 2018)

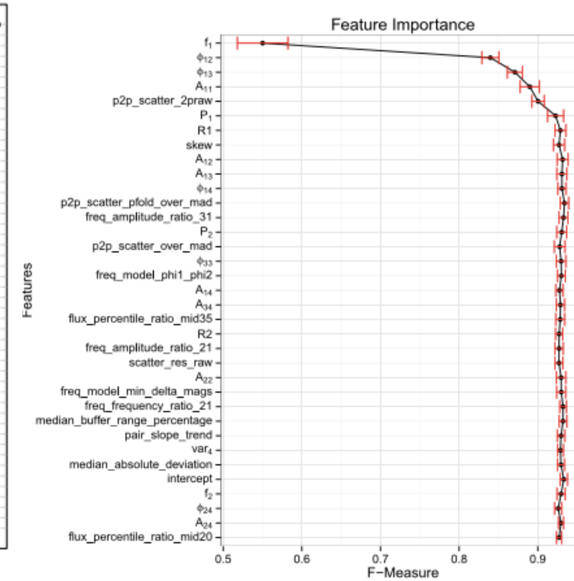
Not all features are equal



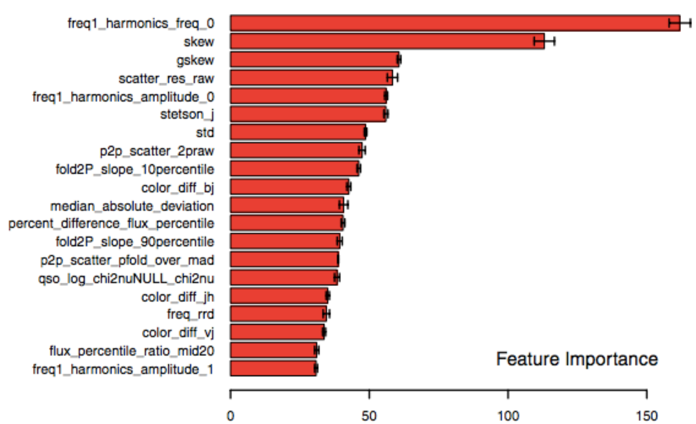
Richards et al. 2011



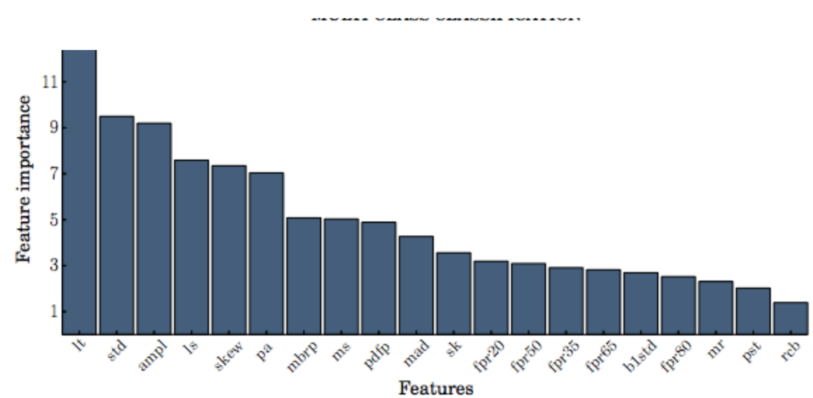
Dubath et al. 2012



Elorietta et al. 2016



Richards et al. 2012



D’Isanto et al. 2016

Periodicity

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

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

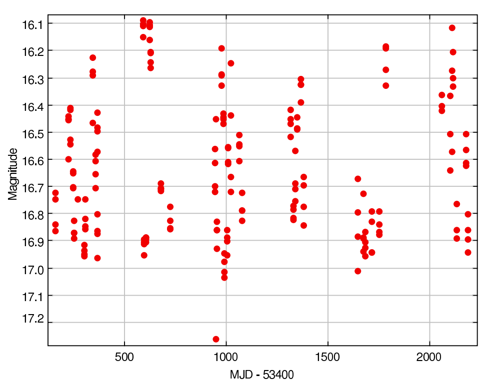
$$\chi^2(f) = \sum_n \left(\frac{x_n - x(t_n; f)}{\sigma_n}\right)^2$$

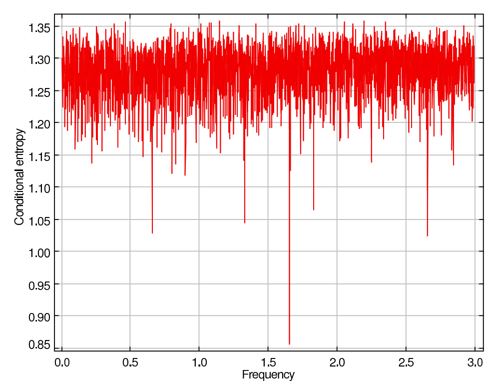
$$P(f) = \frac{1}{2}\left[\hat{\chi}_0^2 - \hat{\chi}^2(f)\right]$$

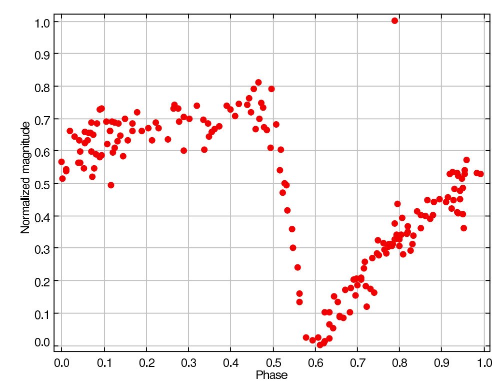
$$\varphi(t,f) = tf - \mathrm{int}(tf)$$

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

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





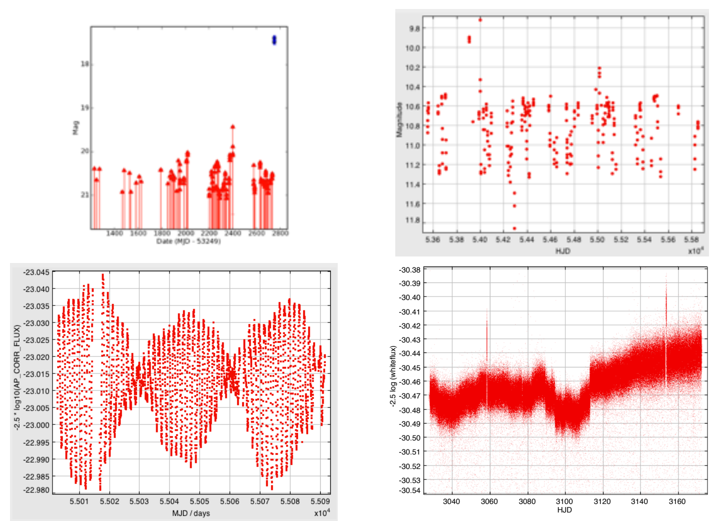


Period finding is not a single algorithm

- 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



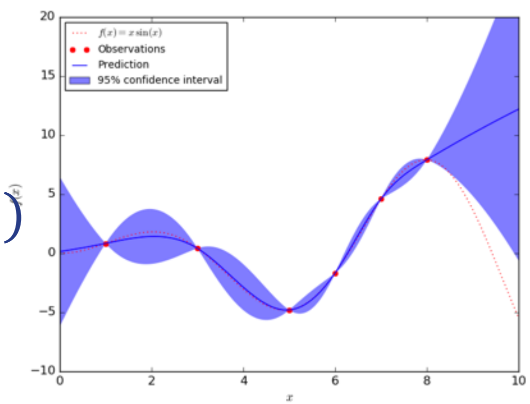
What can we say about period finding

- 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 ($\Delta\varphi = 0.1$)
- 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

Gaussian processes

- Fundamental idea:

$$P\big(\boldsymbol{y}|\boldsymbol{X},(\boldsymbol{\theta},\boldsymbol{\varphi})\big) = \mathcal{N}[\mu(\boldsymbol{X},\boldsymbol{\varphi}),\mathbf{K}]$$
$$K_{nm} \equiv \mathrm{cov}[\boldsymbol{x}_n,\boldsymbol{x}_m] = k(\boldsymbol{x}_n,\boldsymbol{x}_m,\boldsymbol{\theta})$$



- Hyperparameter estimation:

$$\log p(\boldsymbol{y}|X,\boldsymbol{\theta}) = -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})^T K_y^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) - \frac{1}{2}\log|K_y| - \frac{n}{2}\log 2\pi$$

- Prediction:

$$p(\boldsymbol{y}_*) = \mathcal{N}[\boldsymbol{m}_*,\mathbf{C}_*]$$
$$\boldsymbol{m}_* = \boldsymbol{\mu}(\boldsymbol{x}_*) + \mathbf{K}(\boldsymbol{x}_*,\boldsymbol{x})\mathbf{K}(\boldsymbol{x},\boldsymbol{x})^{-1}(\boldsymbol{y}(\boldsymbol{x}) - \boldsymbol{\mu}(\boldsymbol{x}))$$
$$\mathbf{C}_* = \mathbf{K}(\boldsymbol{x}_*,\boldsymbol{x}_*) - \mathbf{K}(\boldsymbol{x}_*,\boldsymbol{x})\mathbf{K}(\boldsymbol{x},\boldsymbol{x})^{-1}\mathbf{K}(\boldsymbol{x}_*,\boldsymbol{x})^T$$

Popular kernels

- Squared exponential:

$$K_{SE}(x,x') = \exp\left(-\frac{r^2}{2l^2}\right), \qquad r = \|x - x'\|$$

- Ornstein-Uhlenbeck:

$$K_{OU}(x,x') = \exp\left(-\frac{|r|}{l}\right)$$

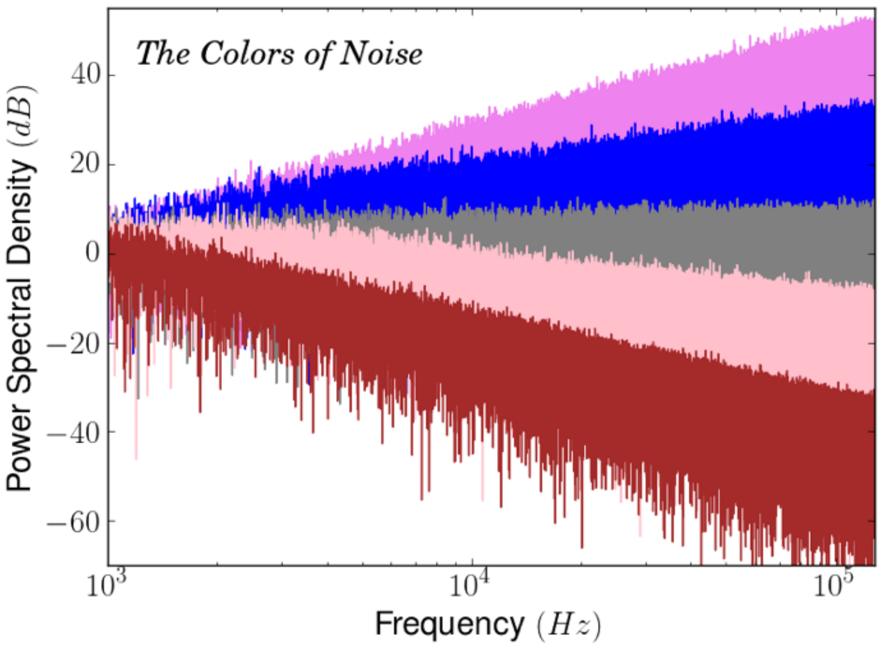
- Periodic:

$$K_P(x,x') = \exp\left(-\frac{2\sin^2\left(\frac{r}{2}\right)}{l^2}\right)$$

$$K_{\text{celerite}} = \sum_{j=1} J\left[a_j exp\left(-c_j r\right)\cos d_j r + b_j \exp\left(-c_j r\right)\sin d_j r\right]$$

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):
- $(1 - B)^d x_t = z_t$



Autoregressive GPs

- A process is said to be autoregressive if the psd of the kernel can be written in the form:

$$S(\omega) = \frac{1}{\left|\sum_{k=0}^{m} \alpha_k (i\omega)^k\right|^2}$$

- Matern kernel:

$$C_\upsilon(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{d}{\rho}\right)^\upsilon K_\upsilon\left(\sqrt{2\nu}\frac{d}{\rho}\right)$$

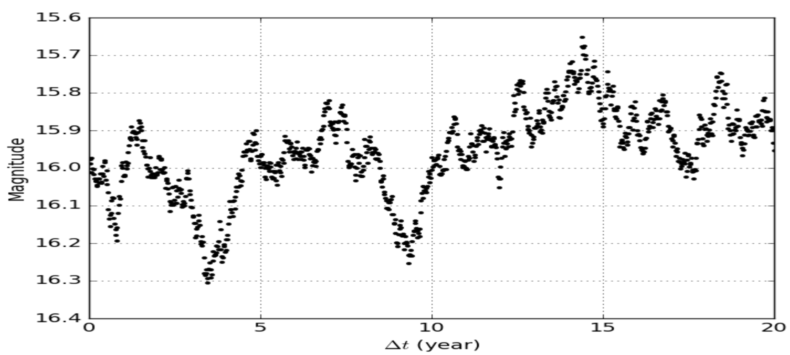
$$S_\upsilon(\omega) = \frac{1}{\left|\sqrt{\frac{\Gamma(\nu)\theta^{2\nu}}{2\sigma^2\sqrt{\pi}\Gamma(\nu+\frac{1}{2})(2\nu)^\nu}}\left(\frac{\sqrt{2\nu}}{\theta} + i\omega\right)^{(\nu+\frac{1}{2})}\right|^2}$$

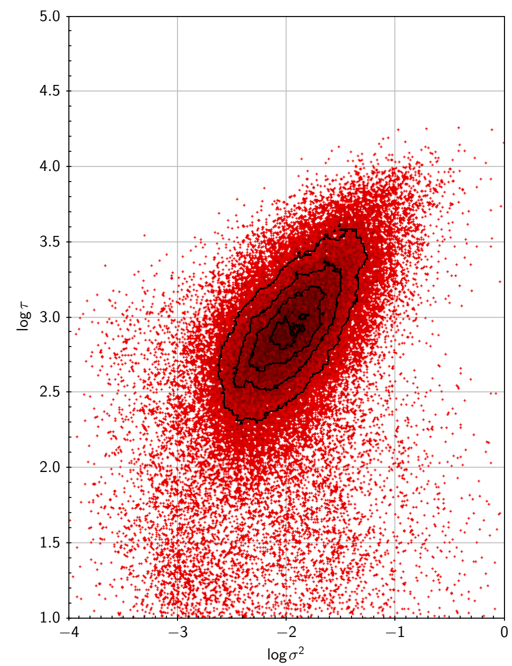
Quasar variability as a damped random walk

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

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

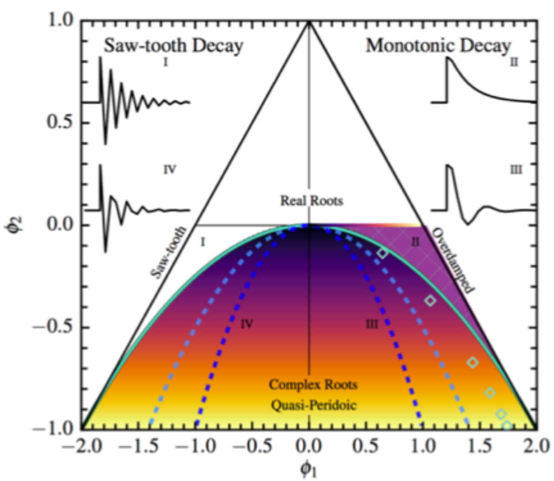
- Characterized by variability amplitude and timescale
- Basis for stochastic models of variability
- Deviations noted (e.g., Mushotzky 2011, Zu et al. 2013, Graham et al. 2014)
- Degenerate model – can be best fit for a non-DRW process (Kozlowski 2016)

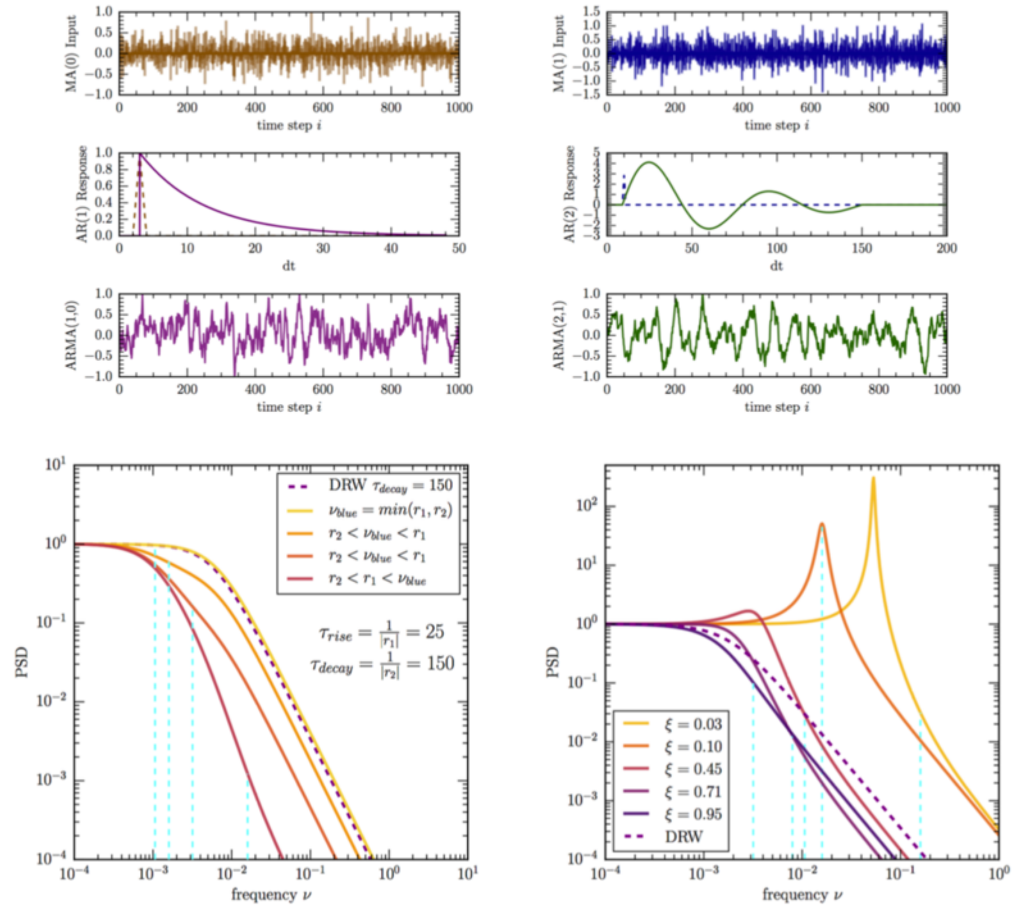




More autoregressive – CARMA(2,1)

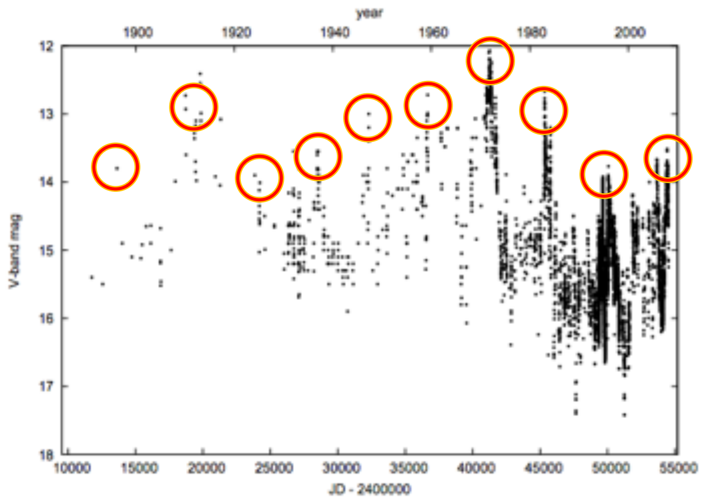
$$d^2x + \alpha_1 d^1 x + \alpha_2 x = \beta_0 z_t + \beta_1 z_{t-1}$$

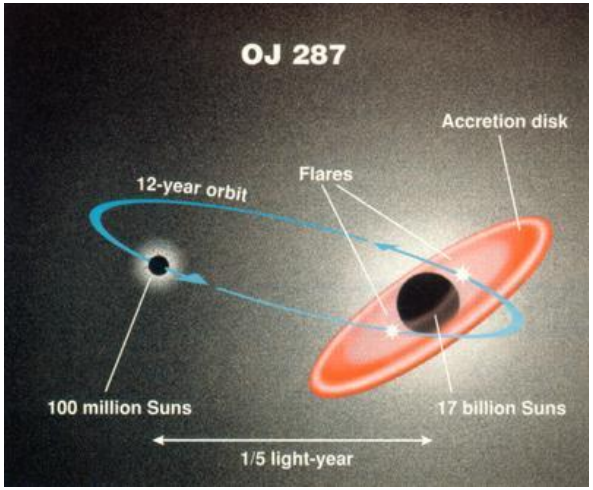


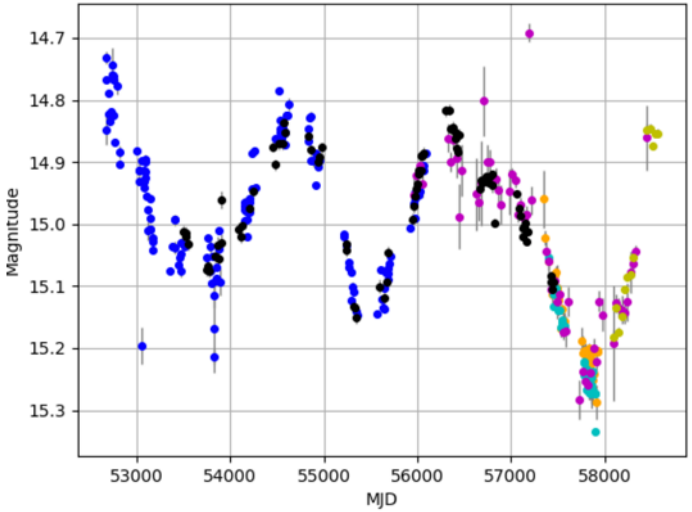


(Moreno et al. 2019)

Periodic quasars?

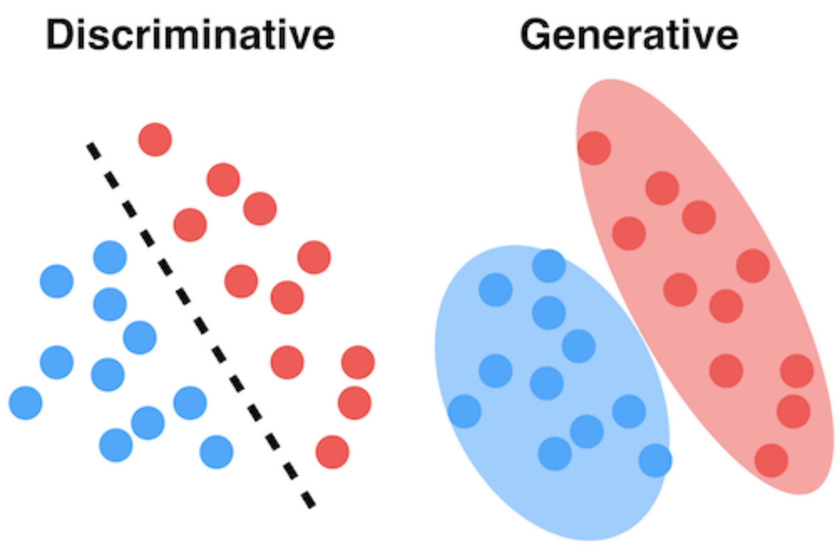






Generative vs. discriminative

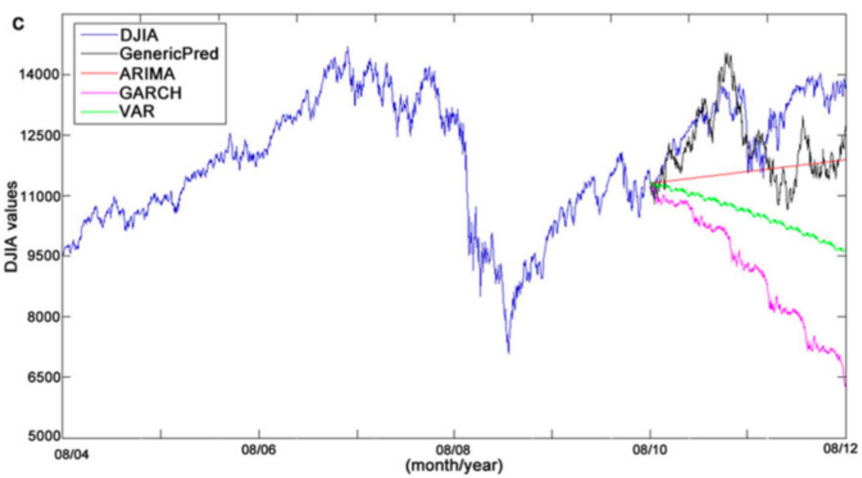
- Current statistical models of variability are designed to discriminate between classes, e.g. stars/galaxies – p(y|x)
- Better to learn time series (shape) rather than determining some parameterizable form – p(y, x)
- Generative approach that supports predictions

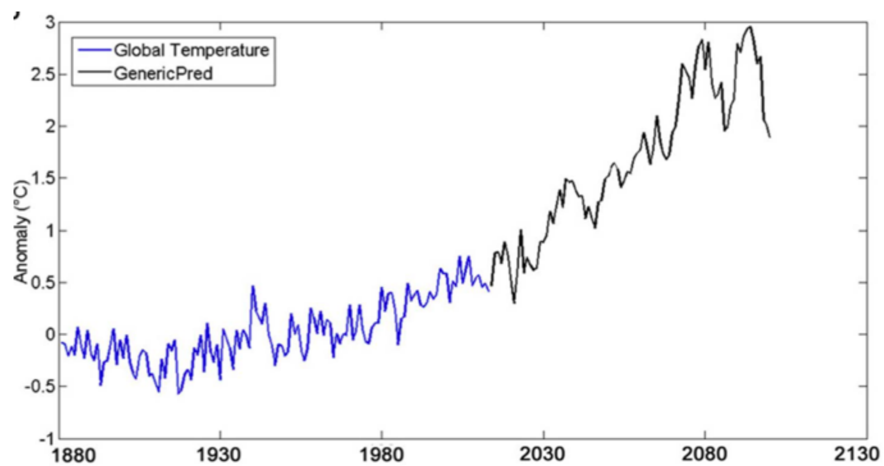


Forecasting

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

- ARIMA, ARFIMA, GARCH models
- Gaussian processes

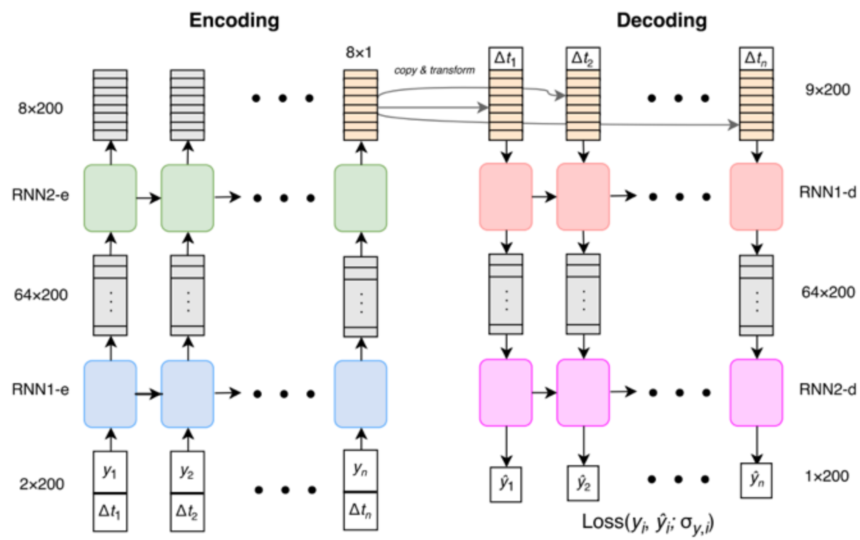


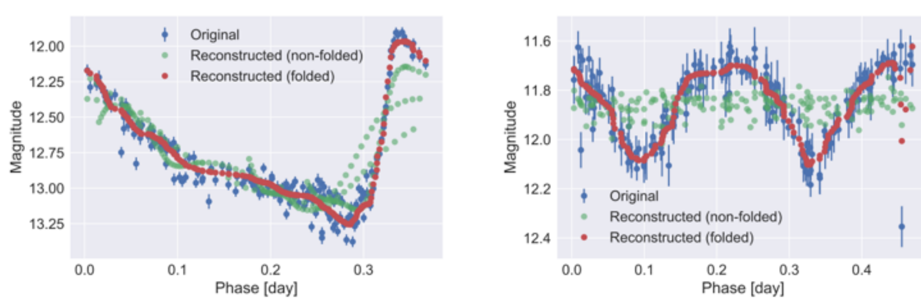


(Golestani & Gras 2014)

Deep time series

- Learn features directly from the data
- Networks for sequential data





(Naul et al. 2018)

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