Time series

Francisco Förster Burón

AS4501

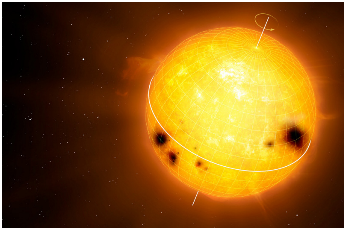
November 15, 2020

Bibliography

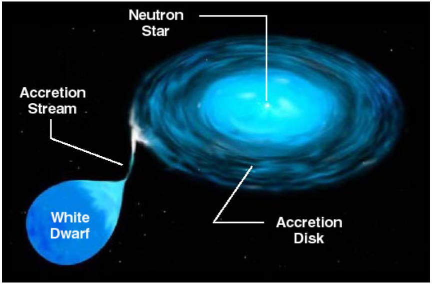
- Modern Statistical Methods for Astronomy, with R Applications. E. Feigelson, G. Babu.
- Statistics, Data Mining, and Machine Learning in Astronomy, A Practical Python Guide for the Analysis of Survey Data. Z. Ivezic, A. Connolly, J. VanderPlas, A. Gray

Variable phenomena

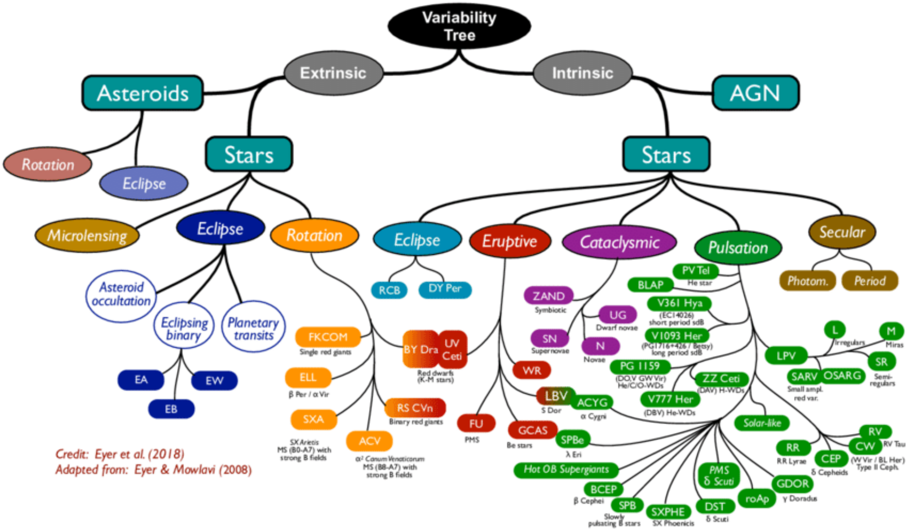
- Apparent: rotation, orbital motions



- Intrinsic: pulsations, explosions, ejections, accretion



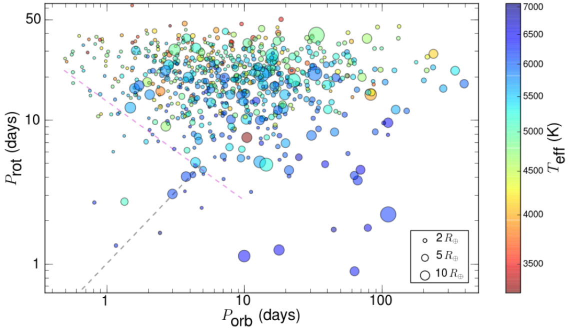
Variable phenomena



See https://www.youtube.com/watch?v=Pcy4U5uvL8I

Timescales

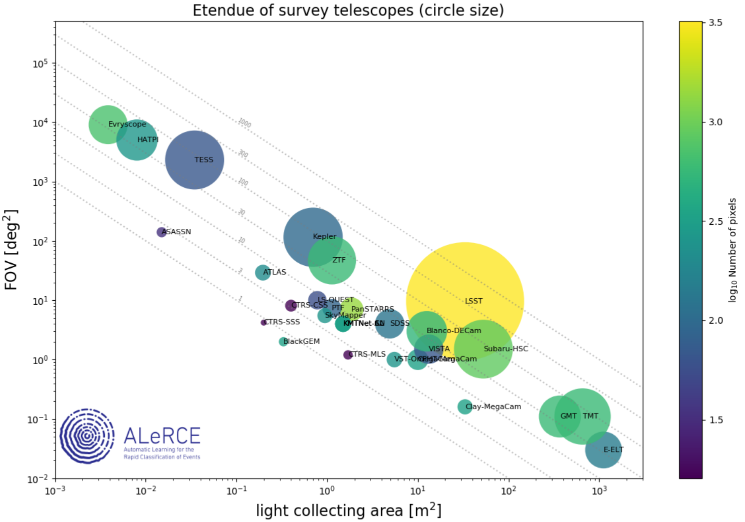
- Binary star orbits: minutes to centuries
- Rotation: msec to months



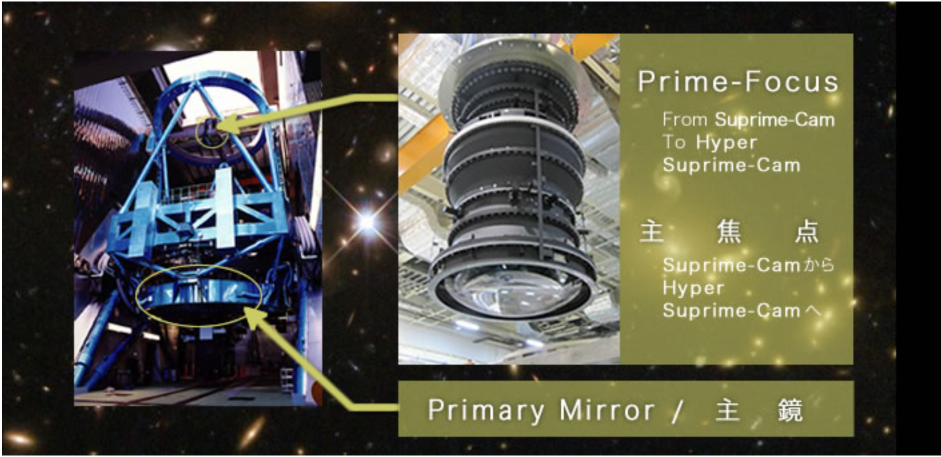
- Accretion: μsec to decades
- Explosions: seconds to years

Astronomical surveys

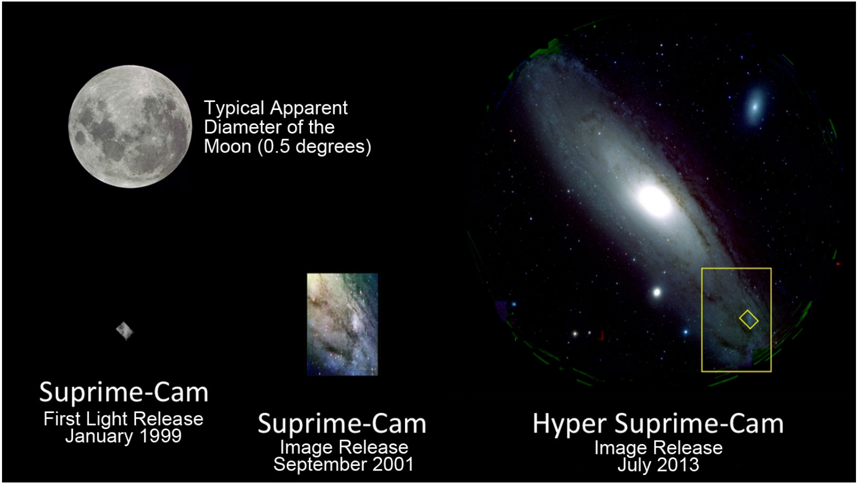
- Main variable: *etendue* = area x field of view

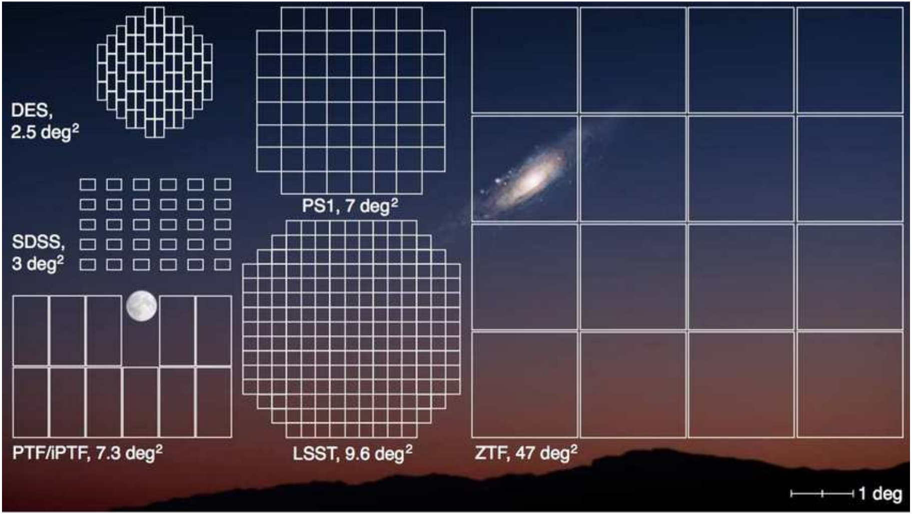


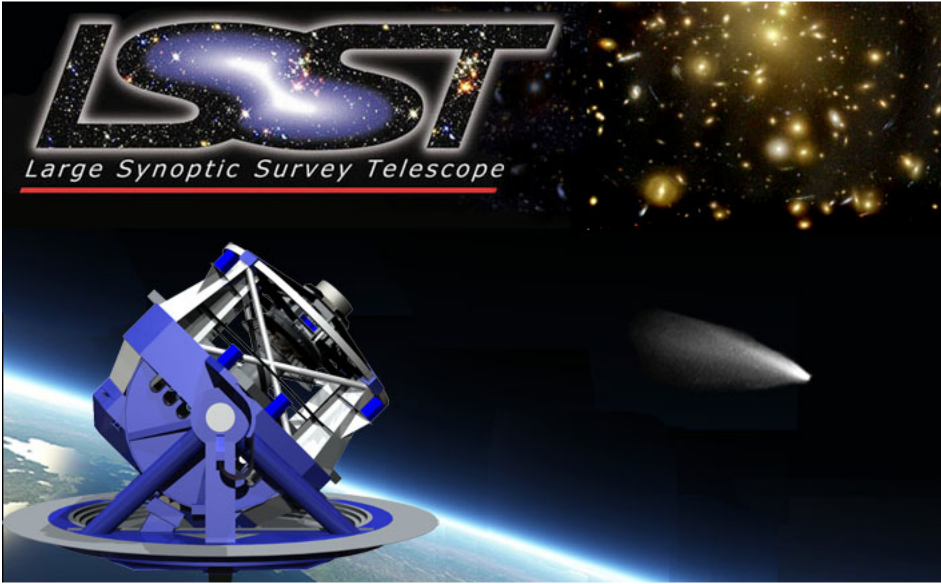
The Dark Energy Camera on the Blanco telescope.



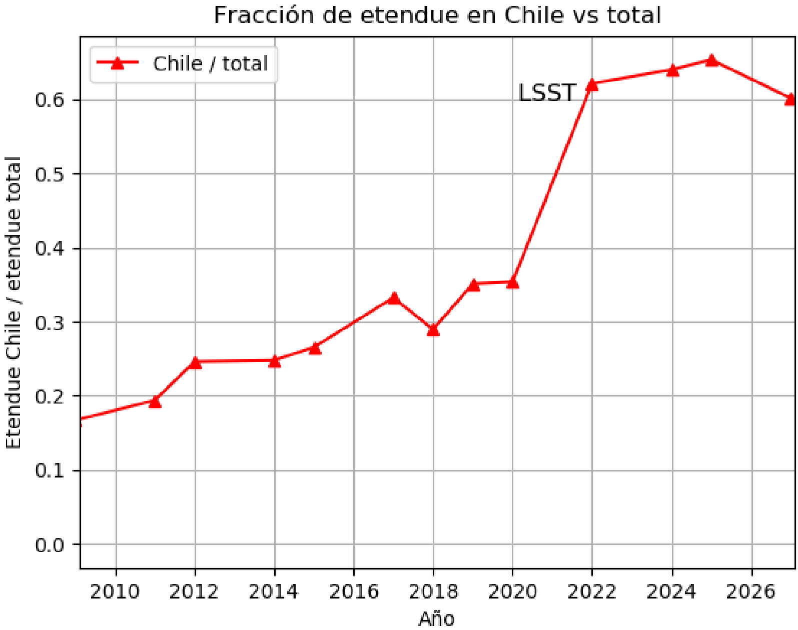
The Hyper Suprime Camera on the Subaru Telescope.

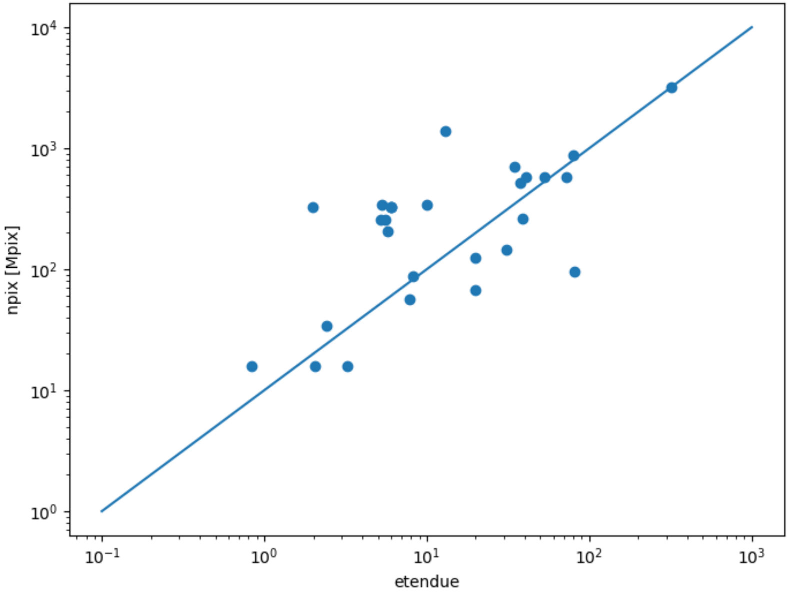

Typical Apparent Diameter of the Moon (0.5 degrees)
Suprime-Cam
First Light Release
January 1999
Suprime-Cam
Image Release
September 2001
Hyper Suprime-Cam
Image Release
July 2013


DES,
2.5 deg^2
PS1, 7 deg^2
SDSS,
3 deg^2
PTF/iPTF, 7.3 deg^2
LSST, 9.6 deg^2
ZTF, 47 deg^2
1 deg



The future Large Synoptic Survey Telescope (LSST).


Fracción de etendue en Chile vs total
Chile / total
LSST
Etendue Chile / etendue total
0.0
0.1
0.2
0.3
0.4
0.5
0.6
2010
2012
2014
2016
2018
2020
2022
2024
2026
Año



npix [Mpix]
etendue
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
10^{4}

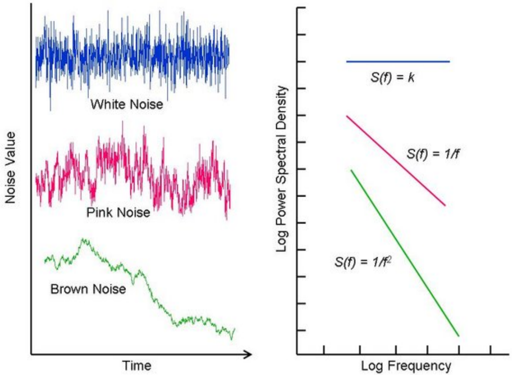
I. Stochastic autoregressive models

Stochastic processes

In probability theory, a **stochastic process**, or sometimes random process, is a collection of random variables representing the evolution of some system of random values over time. This is the probabilistic counterpart to a deterministic process (or deterministic system).

Statistical behaviour - $1/f^{\gamma}$ noise

- Power spectral density $S(\nu) \propto 1/f^{\gamma}$
- white (Gaussian, no correlation in time, f = 0), pink (long memory, f = 1), red or brown (random walk, f = 2)



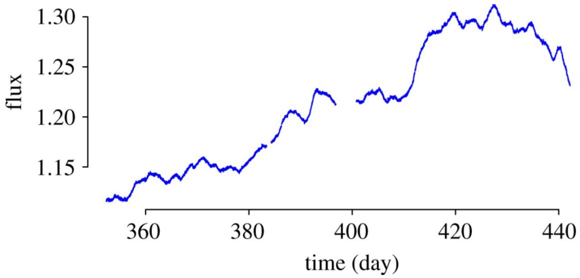
- Existence of 1/f noise is one of the oldest puzzles of physics

Statistical behaviour – Explosive

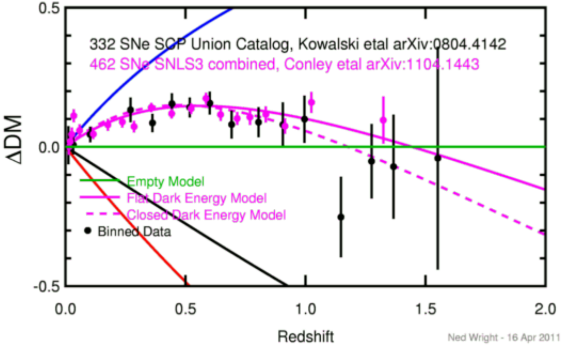
Diversity of behaviours depend on physical mechanism

- Magnetic reconnection
- Scintillation
- Relativistic jets
- Microlensing
- Supernova
- Thermonuclear explosions
- Black hole birth

- Irregular sampling / Gaps

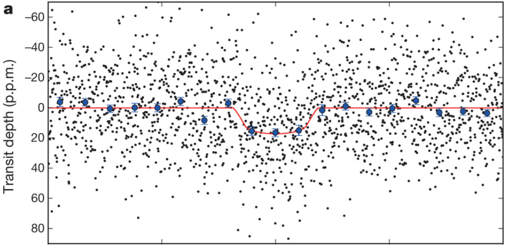


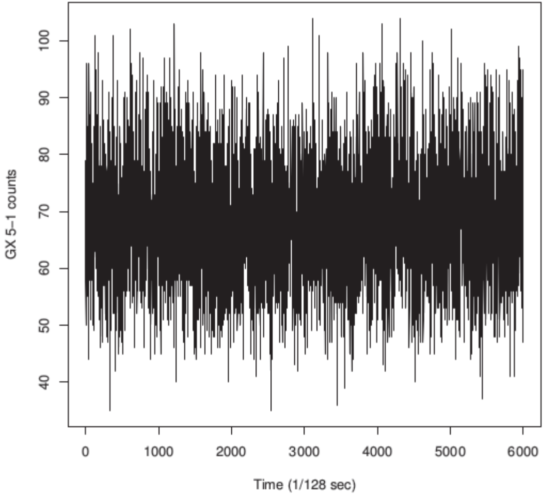
- Heteroscedastic errors



Astronomical Data

- Low signal to noise ratio





Time series of the X–ray emission from the galactic X–ray binary system GX 5–1 measured from the Ginga satellite observatory.

Variability modeling

- Deterministic trends: Regression analysis
- **Stochastic and autocorrelated: Autoregressive models**
- Periodic signal: Fourier transform and related Harmonics analysis

Procedures for all the above are well developed for evenly spaced data. Theory for uneven sampling and heteroscedastic errors is still being developed.

Autocorrelation

- Autocorrelation measures correlated structure
- $ACF(k) = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2}$, where k is the lag time
- a plot of the autocorrelation function is called a **correlogram**

- White noise: $ACF(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$
- Short term autocorrelation: ACF large for small k
- Long term autocorrelation: ACF large for wide range of k
- Periodic signal: periodic ACF

Autocorrelation and autocovariance

- Autocovariance

$$\gamma(k) = E[(x_t - \bar{x})(x_{t+k} - \bar{x})] \tag{1}$$

$$= \frac{1}{N-1} \sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \tag{2}$$

$$\tag{3}$$

- Autocorrelation

$$ACF(k) \equiv \rho(k) = \frac{\gamma(k)}{\gamma(0)} \tag{4}$$

$$= \frac{E[(x_t - \bar{x})(x_{t+k} - \bar{x})]}{E[(x_t - \bar{x})(x_t - \bar{x})]} \tag{5}$$

$$= \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2} \tag{6}$$

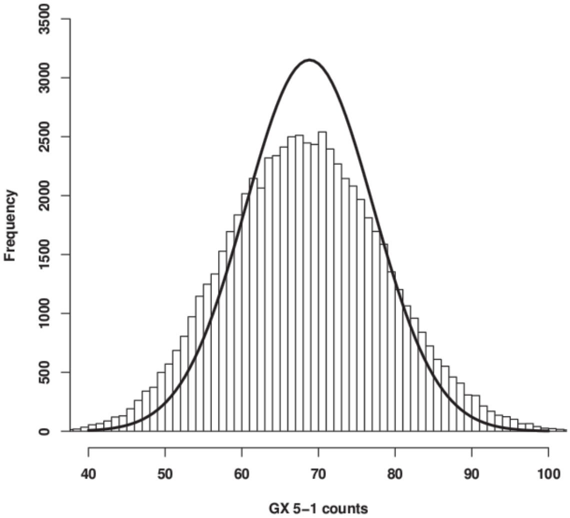
Stationarity

- Stationarity is the property that temporal behaviour, whether stochastic or deterministic, is statistically unchanged by an arbitrary shift in time
- Weakly stationary processes have unchanged moments (e.g. mean or autocovariance)

- The variance of the mean of a stationary process changes when autocorrelation is present

  $\hat{var}(\bar{x}_n) = \frac{\sigma^2}{n}[1 + 2\sum_{k=1}^{n-1}(1 - k/n)ACF(k)]$

  This means that the comparison of means of different objects or the same object at different times must take into account the increase in variance due to the autocorrelation.



Histogram of the timeseries of GX 5–1 compared to the best fit Gaussian distribution. Excess variance with respect to a white noise model is found. Some asymmetry is also observed.

Smoothing

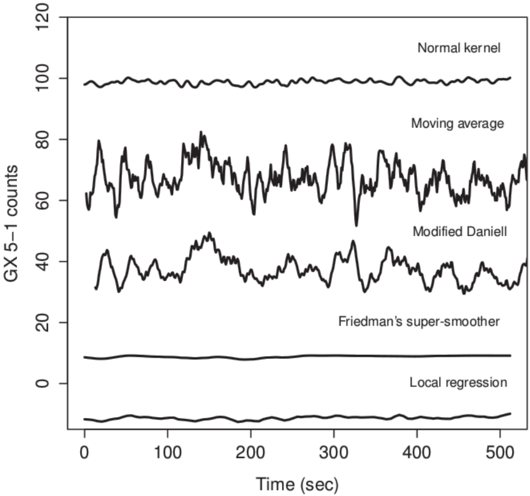
- Useful for visual exploration of a time series
- Central moving average (CMA) with bandwidth of j time intervals:

$$\hat{X}_{i,CMA}(j) = \frac{1}{j+1} \sum_{k=-j/2}^{j/2} X_{i+k} \tag{7}$$

- Exponentially weighted moving average (EWMA)

$$\hat{X}_{i,EWMA} = \alpha X_i + (1-\alpha) X_{i-1}, \tag{8}$$

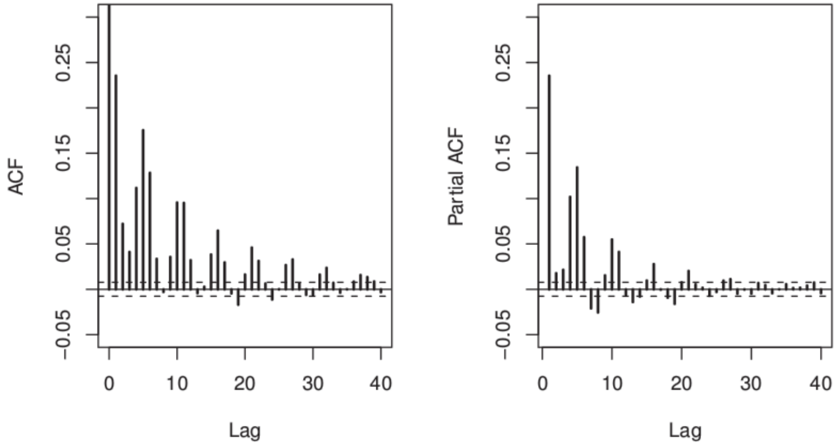
where $0 \leq \alpha \leq 1$



Smoothed timeseries of GX 5–1, showing no obvious non–stationary behaviour. However, the source shows stochastic, possibly autocorrelated and quasi–periodic, variability.

Partial autocorrelation function (PACF)

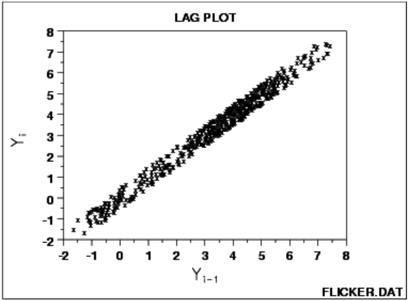
- It gives the autocorrelation at lag time k removing the effects of correlations at shorter lags
- It is found by successively fitting autoregressive models with order 1, 2, ... p and setting the last coefficient of each model to the PACF parameter.
- PACF is useful to understand the timescales responsible for the autocorrelated behaviour



Autocorrelation function (ACF) and partial autocorrelation function (PACF) of the GX 5–1 timeseries. PACF shows that the strongest non–zero peak is located between 4–6 increments (autocorrelation), but there are weaker peaks at larger lags periodically separated by 5–6 increments up to 20–30 increments (quasi–periodic) .

Lag k scatter plot

- Scatter plot of $x(t+k)$ vs $x(t)$
- A useful diagnostic to understand autocorrelation structure
- No structure: uncorrelated noise
- Linear: Stochastic autoregression
- Circular: periodic signal
- Clusters: Non–stationarity



Durbin–Watson statistic

- Useful statistic that is a simple measure of autocorrelation in an evenly spaced time series:

$$d_{DW} = \frac{\sum_{i=2}^{n}(x_i - x_{i-1})^2}{\sum_{i=1}^{n} x_i^2} \qquad \begin{cases} 0 \leq d_{DW} < 2, & \textit{positive autocorrelation} \\ d_{DW} = 2, & \textit{no autocorrelation} \\ 2 < d_{DW} \leq 4, & \textit{negative autocorrelation} \end{cases}$$

Stochastic autoregressive models

- Simplest stochastic autocorrelated model:

  $X_i = X_{i-1} + \epsilon$

  $\Rightarrow$

  $X_i = X_0 + \epsilon_1 + \epsilon_2 + ... + \epsilon_i \quad$ (random walk)

- $ACF_{RW}(i, k) = \frac{1}{1+k/i}$

Autoregressive (AR) models

- $AR(p)$ model:

$$X_i = \alpha_1 X_{i-1} + \alpha_2 X_{i-2} + ... + \alpha_p X_{i-p} + \epsilon_i \tag{9}$$

where $\epsilon_i \sim N(0, \sigma_i^2)$ or white noise

- The effect of the error terms is infinite in time.
- $AR(1)$ model:
  - Autocorrelation function for lag k is time invariant for $|\alpha_1| < 1$:

    $ACF_{AR(1)}(k) = \alpha_1^k$
  - ACF decays rapidly for $\alpha_1 \sim 0$, but remains high for $\alpha_1 \sim 1$

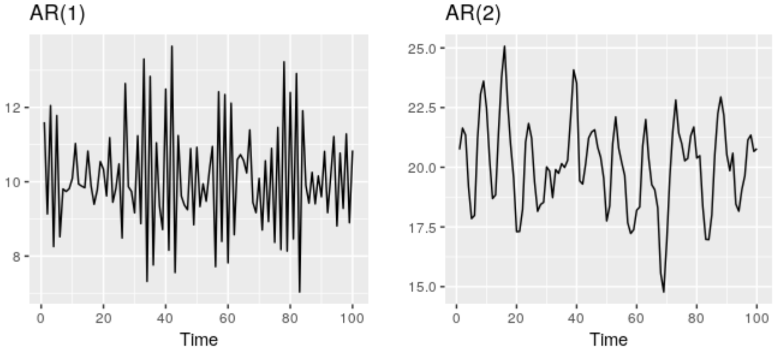
Autoregressive (AR) models

- $AR(p)$ model:

$$X_i = \alpha_1 X_{i-1} + \alpha_2 X_{i-2} + ... + \alpha_p X_{i-p} + \epsilon_i \tag{10}$$

where $\epsilon_i \sim N(0, \sigma_i^2)$ or white noise

- In AR(1), $\alpha_1 = 0$ is equivalent to white noise
- In AR(1), $\alpha_1 = 1$ is equivalent to a random walk
- For stationary data:
  - In AR(1) we need $-1 < \alpha_1 < 1$
  - In AR(2) we need $-1 < \alpha_2 < 1, \quad \alpha_1 + \alpha_2 < 1, \quad \alpha_2 - \alpha_1 < 1$



Example of AR(1) and AR(2) timeseries.

More AR timeseries examples.

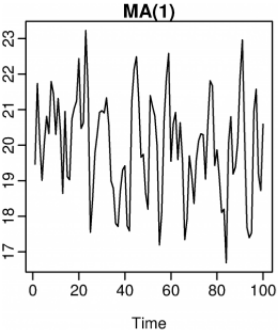
Moving average (MA) models

- $MA(q)$ model:

$$X_i = \epsilon_i + \beta_1 \epsilon_{i-1} + ... + \beta_q \epsilon_{i-q} \tag{11}$$

  where $\epsilon_i \sim N(0, \sigma_i^2)$ or white noise.

- the current value depends on past values of the noise, but the effect of the error terms is finite in time.
- MA models effectively filter high frequency (short timescales) noise
- fitting is more complicated than AR models because the error terms are unobserved.



Example of MA(1) and MA(2) timeseries

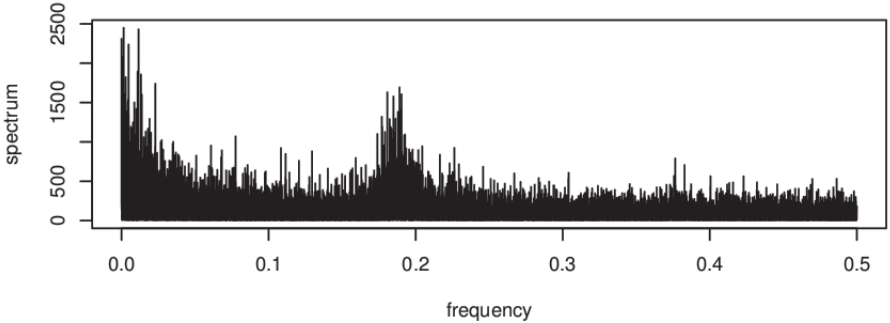
ARMA(p, q) models

- $ARMA(p, q)$ model:

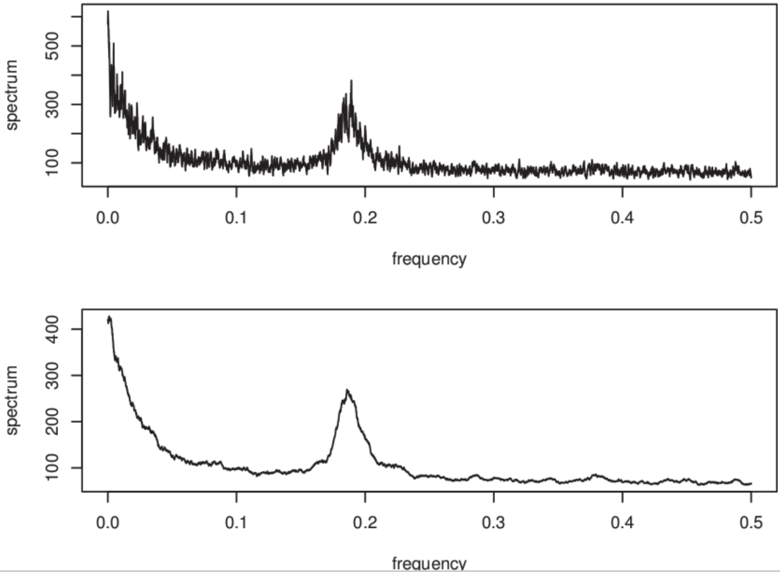
$$X_i = \alpha_1 X_{i-1} + \alpha_2 X_{i-2} + ... + \alpha_p X_{i-p} + \epsilon_i + \beta_1 \epsilon_{i-1} + ... + \beta_q \epsilon_{i-q} \quad (12)$$

  where $\epsilon_i \sim N(0, \sigma_i^2)$

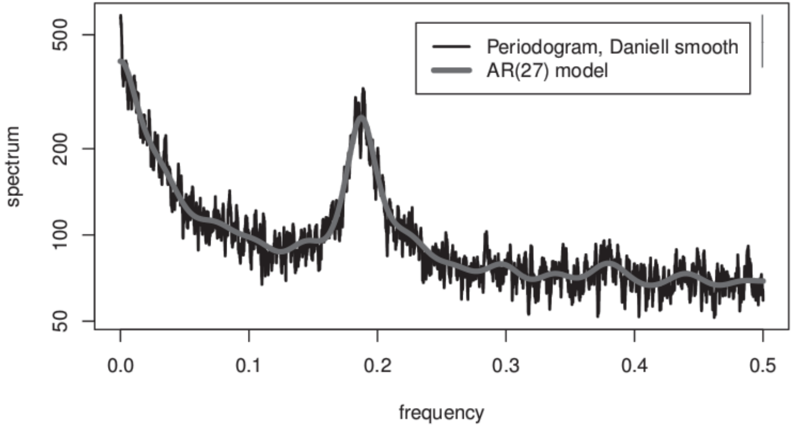
- ARMA models can be fitted via least square minimization
- The difficult part is the choice of p and q



Periodogram of the GX 5–1 timeseries.



Smoothed Periodogram of the GX 5–1 timeseries.



AR(27) best fit model to the GX 5–1 timeseries.

Akaike information criterion (AIC, Akaike 1974)

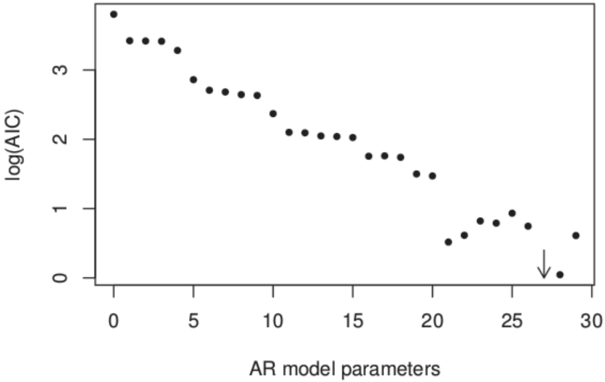
- A model selection tool that takes into account the likelihood of the model under consideration and the number of parameters used in the model.
- For a model with p parameters θ_p:

$$AIC(\theta_p) = -2\ln(L(\theta_p)) + 2p \tag{13}$$

  where $L(\theta_p)$ is the likelihood of the model under consideration
- The model with the lowest AIC should be used.
- For AR(p) models, the AIC reduces to:

$$AIC(p) = N\ln(\hat{\sigma}^2) + 2p \tag{14}$$

  where $\hat{\sigma}_p^2$ is the maximum likelihood estimator (likelihood of the data given the model) of the variance of the ϵ noise terms.



Akaike information criterion for the AR(p) model on the GX 5–1 timeseries. Best fit model was AR(27) based on this criterion.

Bayesian information criterion (BIC, Schwarz 1978)

- BIC penalizes model complexity more heavily. AIC and BIC rely on different assumptions.
- For a model with p parameters θ_p:

$$BIC(\theta_p) = -2\ln(L(\theta_p)) + \ln(n)p \tag{15}$$

  where $L(\theta_p)$ is the likelihood of the model under consideration and n is the sample size.
- The model with the lowest BIC should be used.
- For AR(p) models, the BIC reduces to:

$$BIC(p) = N\ln(\hat{\sigma}^2) + \ln(n)p \tag{16}$$

  where $\hat{\sigma}_p^2$ is the maximum likelihood estimator (likelihood of the data given the model) of the variance of the ϵ noise terms.

Other models

- ARIMA: autoregresive integrated moving average models (Box & Jenkins 1970)
  Can treat non–stationary stochastic processes
- ARCH: autoregressive coditional heteroscedastic
  Useful for datasets where the variance rather than the local mean changes during the observations
- GARCH: generalized ARCH models
  Non–linear dependency on past noise levels

Irregular Sampling

- CARMA: continuous–time autoregressive moving average (Kelly, Becker et al. 2014)
  Accounts for irregular sampling and measurement errors. Scales linearly with the number of data points, i.e. for massive data sets.
- IAR: Irregular Autoregressive model (Eyheramendy et al. 2018)

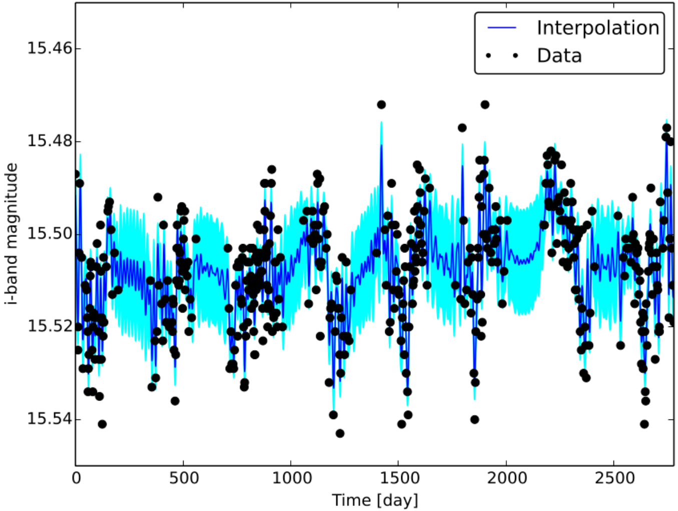
$$y_{t_j} = \phi^{t_j - t_{j-1}} y_{t_{j-1}} + \sigma \sqrt{1 - \phi^{2(t_j - t_{j-1})}} \epsilon_{t_j} \tag{17}$$

Application to astronomy

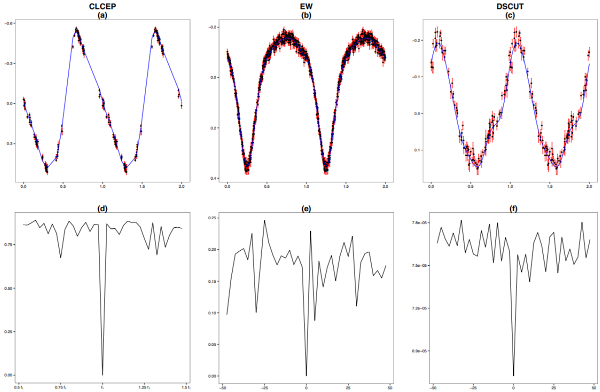
- A timeseries can be represented as

$$z_t = g(t, \theta) + \delta_t \tag{18}$$

  where z_t are the observations, $g(t, \theta)$ is a model, and δ_t are the errors.
- It is common to assume that δ_t are independent and Gaussian distributed, but this is not always the case.
- The solution is to model $z_t - g(t, \theta)$ as an autoregressive model.
- To account for irregular sampling one can use CARMA or IAR



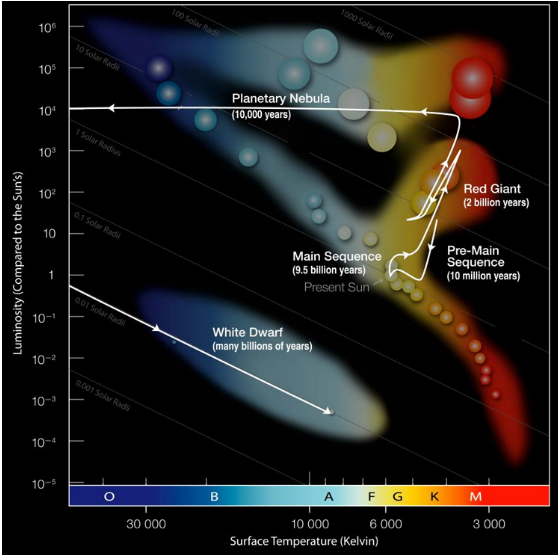
The i-band light curve for a long-period variable star on the red giant branch, from the OGLE-III survey. Also shown is the interpolated light curve and its uncertainty assuming a CARMA(6,0) model (Kelly et al. 2014, ApJ).



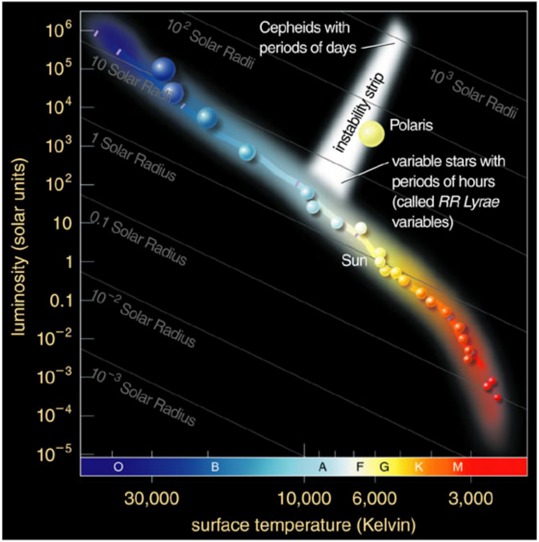
1st row: light curves of a Classical Cepheid, EW and DSCUT, respectively (harmonic best fit as continuous blue line). **2nd row:** Estimate of ϕ vs the % of variation from the correct frequency. (Eyheramendy+2018)

II. Fourier Analysis

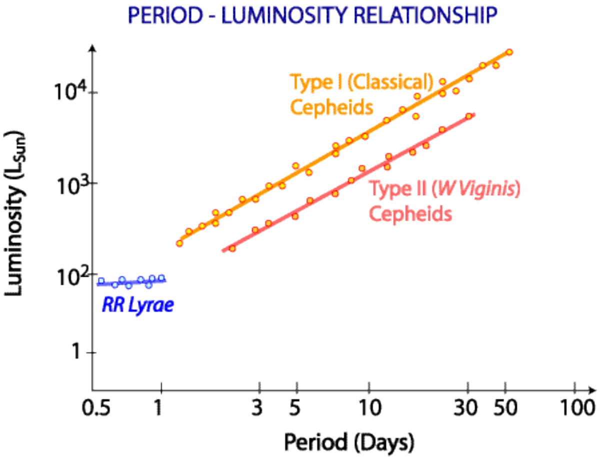
Hertzsprung–Russell diagram



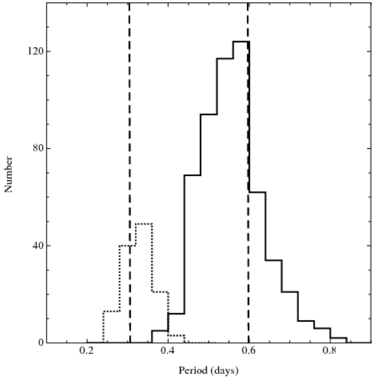
Hertzsprung–Russell diagram



Period–luminosity relation for some periodic stars

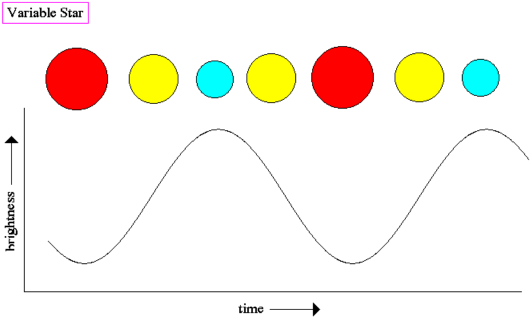


Period–luminosity relation for some periodic stars



The period distributions of ab-type (solid) and c-type (dotted) RR Lyrae variables

Period distribution in RR Lyrae stars



The temperature and size evolution of a pulsating star in the instability strip

Fourier transform

- Fourier transform of $h(t)$:

$$H(f) = \int_{-\infty}^{\infty} h(t) \exp(-i\ 2\pi\ ft) dt \tag{19}$$

- Inverse Fourier transform of $H(f)$:

$$h(t) = \int_{-\infty}^{\infty} H(f) \exp(i\ 2\pi\ ft) df \tag{20}$$

Fourier transform

- Gaussian $N_{(0,\ \sigma)}(t)$ transform (not the same as white noise)

$$H_{Gauss}(f) = \exp(-2\pi^2\sigma^2 f^2) \tag{21}$$

- Translation transform

$$H_{h(t+\Delta t)}(f) = H(f)\exp(i2\pi f \Delta t) \tag{22}$$

- Power spectral density

$$PSD(f) \equiv |H(f)|^2 + |H(-f)|^2, \tag{23}$$

e.g. $PSD_{\sin(2\pi t/T)}(f) = \delta(1/T)$

Fourier transform

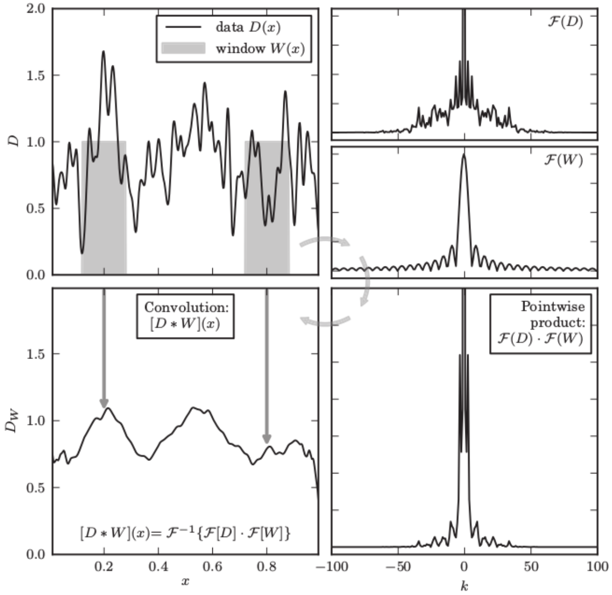
- Parseval's theorem

$$P_{tot} \equiv \int_0^\infty PSD(f)df = \int_{-\infty}^{\infty} |h(t)|^2 dt \tag{24}$$

- Convolution theorem

$$(a * b)(t) \equiv \int a(t')b(t-t')dt' \tag{25}$$

$$H_{a*b}(f) = A(f)B(f) \tag{26}$$



Convolution of two function in real and Fourier space.

Discrete Fourier Transform (even sampling)

- The discrete Fourier transform of the vector of values h_j is a complex vector of length N defined by:

$$H_k = \sum_{j=0}^{N-1} h_j \exp[-i2\pi jk/N], \tag{27}$$

where $k = 0, ..., (N-1)$.

- The corresponding inverse discrete Fourier transform is:

$$h_j = \frac{1}{N} \sum_{k=0}^{N-1} H_k \exp[i2\pi jk/N], \tag{28}$$

where $j = 0, ..., (N-1)$.

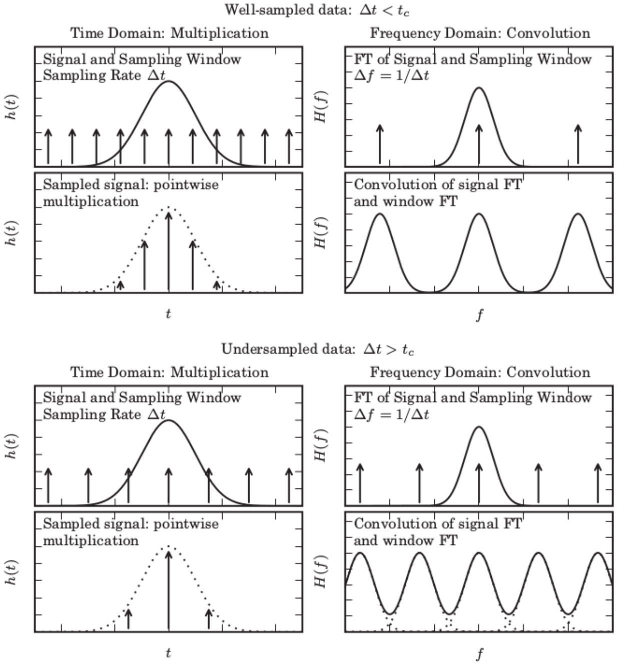
Nyquist sampling theorem

- Let us define $h(t)$ as **band limited** if $H(f) = 0$ for $|f| > f_c$, where f_c is the band limit, or the **Nyquist critical frequency**.
- If $h(t)$ is **band limited**:
  - there is some resolution limit in t space, $t_c = 1/(2fc)$, below which $h(t)$ appears smooth.
  - according to the **Nyquist sampling theorem** we can *exactly* reconstruct $h(t)$ from evenly sampled data when $\Delta t \leq t_c$ as:

$$h(t) = \frac{\Delta t}{t_c} \sum_{-\infty}^{\infty} h_k \frac{\sin[2\pi f_c(t - k\Delta t)]}{2\pi f_c(t - k\Delta t)} = \frac{\Delta t}{t_c} \sum_{-\infty}^{\infty} h_k sinc[2\pi f_c(t - k\Delta t)] \tag{29}$$

Nyquist sampling theorem

- When the sampled function is not band limited or when the sampling rate is not sufficient (i.e. $\Delta t > t_c$) we cannot reconstruct $h(t)$
- In the previous case, the power spectral density from frequencies $|f| > f_c$ is tranferred (or **aliased**) into the $-f_c < f < f_c$ range.
- The **aliasing** effect can be recognized if the Fourier transform is nonzero at $|f| = 1/(2\Delta t)$.


Well-sampled data: $\Delta t < t_c$
Time Domain: Multiplication
Frequency Domain: Convolution
Signal and Sampling Window
Sampling Rate Δt
FT of Signal and Sampling Window
$\Delta f = 1/\Delta t$
Sampled signal: pointwise multiplication
Convolution of signal FT and window FT
$h(t)$
$H(f)$
t
f
Undersampled data: $\Delta t > t_c$
Time Domain: Multiplication
Frequency Domain: Convolution
Signal and Sampling Window
Sampling Rate Δt
FT of Signal and Sampling Window
$\Delta f = 1/\Delta t$
Sampled signal: pointwise multiplication
Convolution of signal FT and window FT
$h(t)$
$H(f)$
t
f

Discrete and true Fourier transform (even sampling)

- The discrete Fourier transform is a good estimate of the true Fourier transform only **for properly sampled and band limited functions.**
- Then, the following relation holds:

$$|H(f_k)| \approx \Delta t |H_k|, \tag{30}$$

where $f_k = k/(N\Delta t)$ for $k \leq N/2$ and $f_k = (k-N)/(N\Delta t)$ for $k \geq N/2$,

Discrete power spectral density (even sampling)

- The discrete power spectral density will be:

$$PSD(f_k) = (\Delta t)^2 (|H_k|^2 + |H_{N-k}|^2). \tag{31}$$

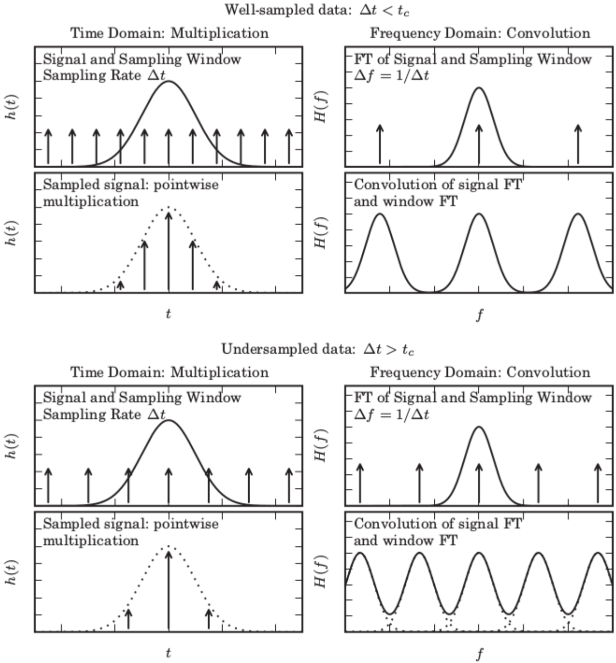
- This can be written as:

$$PSD(f_k) = 2\left(\frac{T}{N}\right)^2 \left[\left(\sum_{j=0}^{N-1} h_j \cos(2\pi f_k t_j) \right)^2 + \left(\sum_{j=0}^{N-1} h_j \sin(2\pi f_k t_j) \right)^2 \right] \tag{32}$$

Note that these formulae are only valid for evenly sampled data and will be strictly valid for noiseless data.

Window function (even sampling)

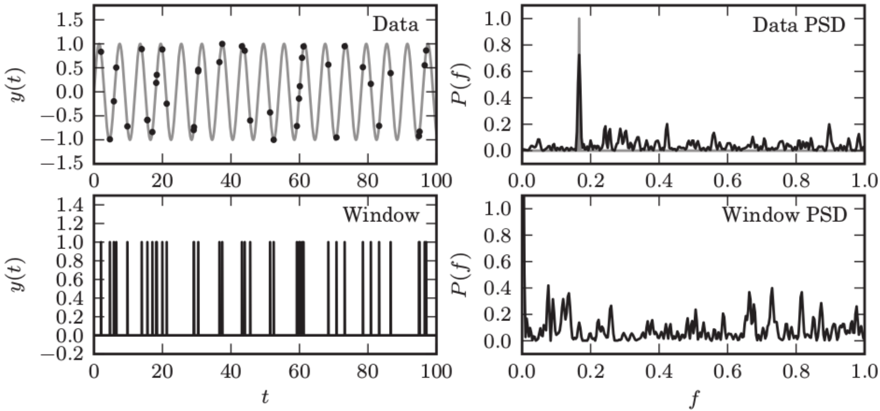
- The **sampling window function** is the sum of delta functions placed at sampled observation times
- When data is evenly sampled (Δt), the Fourier transform of the window function is a set of evenly spaced deltas ($1/\Delta t$).



Well-sampled data: $\Delta t < t_c$
Time Domain: Multiplication
Frequency Domain: Convolution
Signal and Sampling Window
Sampling Rate Δt
FT of Signal and Sampling Window
$\Delta f = 1/\Delta t$
Sampled signal: pointwise multiplication
Convolution of signal FT and window FT
$h(t)$
$H(f)$
t
f
Undersampled data: $\Delta t > t_c$
Time Domain: Multiplication
Frequency Domain: Convolution
Signal and Sampling Window
Sampling Rate Δt
FT of Signal and Sampling Window
$\Delta f = 1/\Delta t$
Sampled signal: pointwise multiplication
Convolution of signal FT and window FT
$h(t)$
$H(f)$
t
f

Window function (uneven sampling)

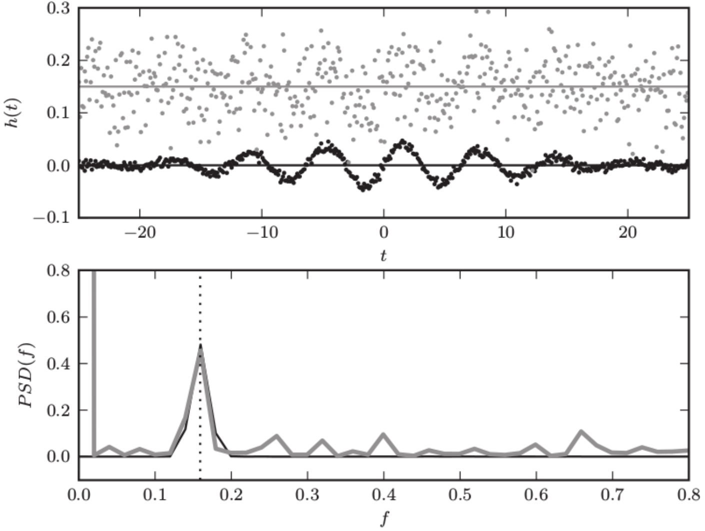
- When data is unevenly sampled, the Fourier transform of the window function has a more complicated structure.
- The power spectral density (PSD) of the window function can be constructed using a fine regular grid of points which are one near the sampling window function and zero elsewhere.
- The resulting PSD is called the **spectral window function**



The data is the multiplication of the true function being sampled with the sampling window function. The data PSD is computed from the convolution of the true function FT and the sampling window function FT.

Fast Fourier transform (even sampling)

- The Fast Fourier Transform (FFT) is an algorithm for computing discrete Fourier transforms in $\mathcal{O}(N \log N)$ time rather than $\mathcal{O}$.
- It is widely used for evenly sampled, high signal to noise ratio, timeseries data.


0.3
0.2
0.1
0.0
−0.1
$h(t)$
−20
−10
0
10
20
t
0.8
0.6
0.4
0.2
0.0
$PSD(f)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
f

Lomb–Scargle periodogram

- The Lomb–Scargle periodogram is a standard method to search for periodicity in unevenly sampled timeseries data. It is defined as:

$$P_{LS}(\omega) = \frac{1}{2}\left[\frac{\left(\sum_{j=1}^{N} X_j \cos[\omega(t_j - \tau)]\right)^2}{\sum_{j=1}^{N} \cos^2[\omega(t_j - \tau)]} + \frac{\left(\sum_{j=1}^{N} X_j \sin[\omega(t_j - \tau)]\right)^2}{\sum_{j=1}^{N} \sin^2[\omega(t_j - \tau)]}\right] \tag{33}$$

where τ is defined by

$$\tan(2\omega\tau) = \frac{\sum_{j=1}^{N} \sin(2\omega t_j)}{\sum_{j=1}^{N} \cos(2\omega t_j)} \tag{34}$$

Lomb–Scargle periodogram

- The Lomb–Scargle periodogram has two main advantages:
  - It has a simple statistical behaviour by design, allowing a simple evaluation of the statistical significance of its peaks
  - It is equivalent to the reduction of the sum of squares in least–squares fitting of sine waves to the data.

Lomb–Scargle periodrogram: statistical significance

- $P_{LS}(\omega)$ is a random variable.
- We must answer the question: *What is the probability that a feature is due to random statistical fluctuations?*
- If $Z \equiv P_{LS}(\omega)$ one can show

$$p_Z(z)dz = Pr(z < Z < z + dz) = \exp(-z)dz \tag{35}$$

$$Pr(Z < z) = \int_0^z p_Z(z')dz' = 1 - \exp(-z) \tag{36}$$

$$Pr(Z > z) = \exp(-z), \tag{37}$$

which is the probability of a large observed power at a given frequency.

- For many peaks one has to apply a penalty for inspecting many values:

$$Pr(Z > z)\Big|_{P_{LS}^{\max}(\omega)} = 1 - (1 - \exp(-z))^N \tag{38}$$

- In fact, the expected maximum of a pure noise spectrum over a set of N frequencies is:

$$< Z_{\max} >= \sum_{k=1}^{N} \frac{1}{k}, \tag{39}$$

  which diverges logarithmically with N.

- The previous point means that if enough frequencies are inspected, a significant peak will eventually be found if no penalty is applied.

Lomb–Scargle periodrogram: statistical significance

- The periodogram threshold z_0 such that a detection is spurious only with a small probability p_0 is:
  $z_0 = -\ln\left(1 - (1 - p_0)^{1/N}\right)$
- Given the threshold z_0, the probability of missing a signal of power P is given by:
  $P_{miss} = (1 - p_0)^{1-1/N}\left\{\{1 - \exp[-(z_0 + P)]\}\phi(z_0, P)\right\}$,
  where $\phi(x, y) = \sum_{m=0}^{\infty}\sum_{k=0}^{\infty} x^k y^m/(k!m!)$

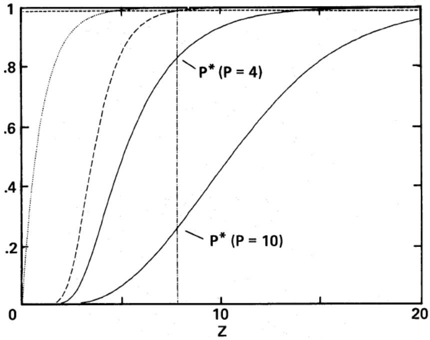


FIG. 1.—The dotted curve is the cumulative distribution function (CDF) for the power at a single, preselected frequency, in the no-signal case (eq. [13]). The point where this curve crosses $1-p_0$ (indicated by the horizontal dashed line) gives the value of z such that the probability of a noise fluctuation exceeding z is p_0. Similarly, the dashed curve is the CDF for the maximum over $N=25$ frequencies, again with no signal present (eq. [14]). The value of z where this curve reaches $1-p_0$ (p_0 is the desired small false alarm rate) is called z_0 (eq. [18]). The reason that this value of z, indicated in the figure by a vertical dot-dash line, is called the detection threshold is that signal powers above this threshold are spurious only a fraction p_0 of the time. The solid lines are CDFs for the maximum power when signals are present, the upper one with $P=4$ and the lower one with $P=10$. Since detection can be claimed only if P exceeds z_0, the probabilities of not detecting these signals are given by the corresponding ordinates of the CDF at z_0 (i.e. p^*; see eq. [22]).

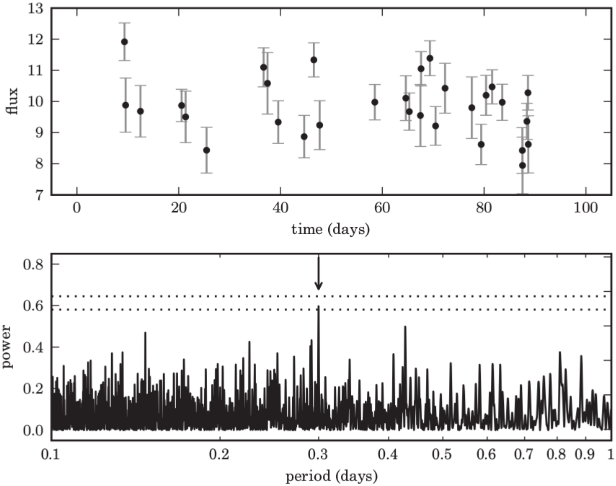
Scargle 1982, ApJ

Lomb–Scargle periodrogram: practical application

- Given that we must choose carefully which frequency peaks to explore, it is important to understand what frequency range will be meaninful
- For the minimum frequency we can choose $\omega_{\text{min}} = 2\pi/(T_{\text{max}} - T_{\text{min}})$
- When the sampling is even, the maximum frequency that can be explored is $\omega_{\text{max}} = 2\pi/\Delta t$.
- For the frequency steps, we can choose $\Delta\omega = \eta\omega_{\text{min}}$, with $\eta = 0.1$

Lomb–Scargle periodrogram: practical application

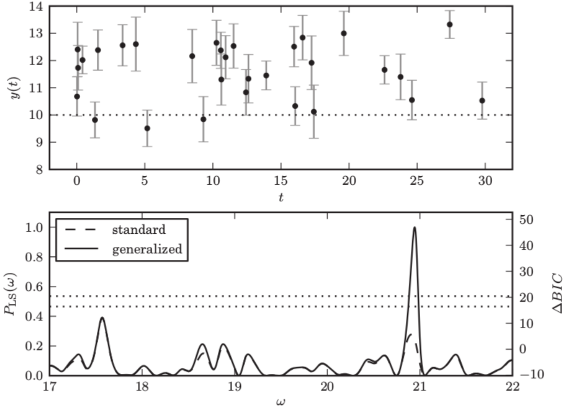
- When the sampling is uneven, the maximum frequency to explore is not obvious.
- e.g. the median value of $2\pi/\Delta t_i$ can be a gross underestimate because sometimes uneven sampling can detect periodicity with frequencies even higher than $2\pi/\Delta t_{\rm min}$ (Eyer & Bartoldi 1999)
- a very conservative hard limit is to use $2\pi/T_{exp}$, where T_{exp} is the exposure time, but in practice one must look at the phase coverage of the maximum frequency chosen.
- See "Understanding the Lomb-Scargle Periodogram" by Jake VanderPlas



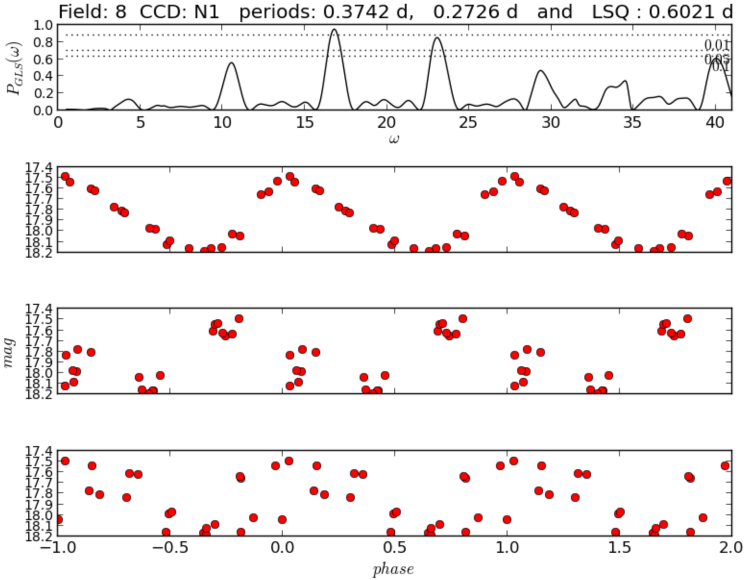
Lomb–Scargle periodogram for $y(t) = 10 + \sin(2\pi t/P)$, with $P = 0.3$ and heteroscedastic errors with a uniform distribution for the standard deviations between 0.5 and 1.

Generalized Lomb–Scargle periodrogram

- A generalization of the Lomb–Scargle method (Zechmeister & Kürster 2009, A& A) includes an offset term to the model
- This results in much better detection efficiencies when the mean of the distribution is overestimated due to an unlucky choice of the sampling window function.
- Both the Lomb–Scargle and generalized Lomb–Scargle methods are implemented in Python library AstroML (lomb_scargle)



Lomb–Scargle vs generalized Lomb–Scargle periodogram for $y(t) = 10 + \sin(2\pi t/P)$, with $P = 0.3$ and heteroscedastic errors with a uniform distribution for the standard deviations between 0.5 and 1. The sampling window function was chosen to overestimate the true mean.



Candidate RR Lyrae star obtained with DECam (credit: G. Medina)

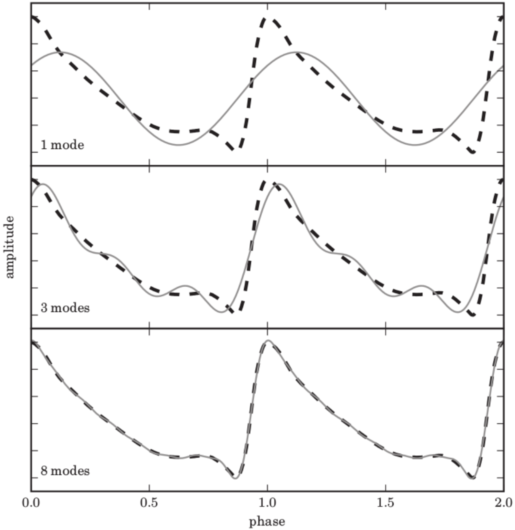


Figure 10.1. An example of a truncated Fourier representation of an RR Lyrae light curve. The thick dashed line shows the true curve; the gray lines show the approximation based on 1, 3, and 8 Fourier modes (sinusoids).

Other methods

- When the underlying variability has a complex spectral structure one needs to use other methods, e.g.:
  - Truncated Fourier Series Model (implemented in AstroML as multiterm_periodogram)
  - Correntropy methods (Huijse, P. et al. 2011, 2012, IEEE Transactions)

Summary

- Astronomical timeseries are unevenly sampled, have heteroscedastic errors and low signal to noise
- Some important concepts: autocovariance and autocorrelation, $1/f^{\gamma}$ noise, power spectral density, sampling and spectral window function, periodogram, Lomb–Scargle periodogram
- For non–periodic autocorrelated variability use autoregressive modelling like ARMA and its derivatives.
- For periodic signals, work in the frequency domain and use the generalized Lomb-Scargle method. When the underlying light curve cannot be represented by a few Fourier terms use more sophisticated methods.