

Time Series Analysis

Outline

- ❖ Time series in astronomy
- ❖ Concepts of time series analysis
- ❖ Time domain methods
- ❖ Frequency domain methods
- ❖ Periodicities in irregularly spaced time series
- ❖ Wavelet analysis

Time series in astronomy

- Periodic phenomena: binary orbits (stars, extrasolar planets); stellar rotation (radio pulsars); pulsation (helioseismology, Cepheids)
- Stochastic phenomena: accretion (CVs, X-ray binaries, Seyfert gals, quasars); scintillation (interplanetary & interstellar media); jet variations (blazars)
- Explosive phenomena: thermonuclear (novae, X-ray bursts), magnetic reconnection (solar/stellar flares), star death (supernovae, gamma-ray bursts)

Difficulties in astronomical time series

Gapped data streams:

Diurnal & monthly cycles; satellite orbital cycles;
telescope allocations

Heteroscedastic measurement errors:

Signal-to-noise ratio differs from point to point

Poisson processes:

Individual photon/particle events in high-energy
astronomy

Variety of temporal behaviors

Concepts of time series analysis

Stationarity The temporal behavior, whether deterministic (e.g. orbit) or stochastic, is statistically unchanged by shifts in time. Types of nonstationarity include: *trends* (secular changes in mean value), *heteroscedasticity* (changes in variance), and *change points* (different behaviors before and after t_0). GRS 1915 is very nonstationary.

Periodicity The measured levels repeat themselves deterministically with one or more periods. The signal becomes concentrated in frequency domain study: *spectral analysis*, *harmonic analysis*, *Fourier analysis*. These methods classically use trigonometric sine and cosine functions, but this is not required.

Autocorrelation The measured levels at time t_0 depend on levels measured at previous times. The autocorrelation can be deterministic (trend or periodicity) or can include a stochastic component. For an evenly spaced time series, the *autocorrelation function* (ACF) is the fraction of the total variance due to correlated values at lag k time steps:

$$\hat{\rho}(k) = ACF(k) = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}.$$

If the ACF has significant signal at small k , the time series has *short-term memory*. If the signal extends to large k , it has *long-term memory*. The latter includes *red noise* or $1/f^\alpha$ -type processes that are often seen in (astro)physical systems. A stochastic time series with insignificant ACF values at all k exhibits *white noise*, often assumed to have a Gaussian (normal) distribution.

Other important concepts *Nonlinear* (in the parameters) time series (including *chaotic* systems); *multivariate* time series (including autoregressive with lags); *time-frequency* analysis (for nonstationary periodic behaviors); *wavelet* analysis (for multiscale aperiodic variations); *state space* models (hierarchical deterministic + stochastic models with MLE coefficients updated by the Kalman filter); *unevenly-spaced* time series (methods primarily developed by astronomers).

Nonparametric time domain methods

Autocorrelation function

This sample ACF is an estimator of the correlation between the x_t and x_{t-k} in an evenly-spaced time series. For zero mean and normal errors, the ACF is asymptotically normal with variance $Var \hat{\rho} = [n - k]/[n(n + 2)]$. This allow probability statements to be made about the ACF.

The partial autocorrelation function (PACF) estimates the correlation with the linear effect of the intermediate observations, $x_{t-1}, \dots, x_{t-k+1}$, removed. Calculate with the Durbin-Levinson algorithm based on an autoregressive model.

Density estimation

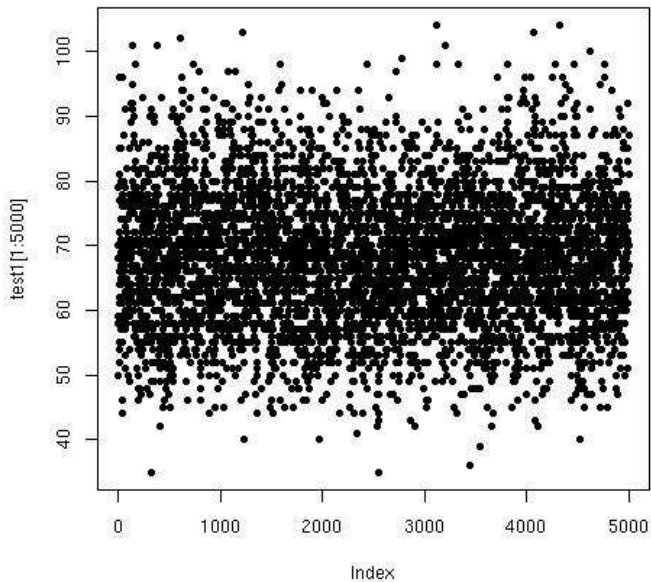
Standard methods of density estimation are often used on time series: kernel density estimation, local regressions, etc.

Ginga observations of X-ray binary GX 5-1

GX 5-1 is a binary star system with gas from a normal companion accreting onto a neutron star. Highly variable X-rays are produced in the inner accretion disk. X-ray binary time series often show 'red noise' and 'quasi-periodic oscillations', probably from inhomogeneities in the disk. We plot below the first 5000 of 65,536 count rates from Ginga satellite observations during the 1980s.

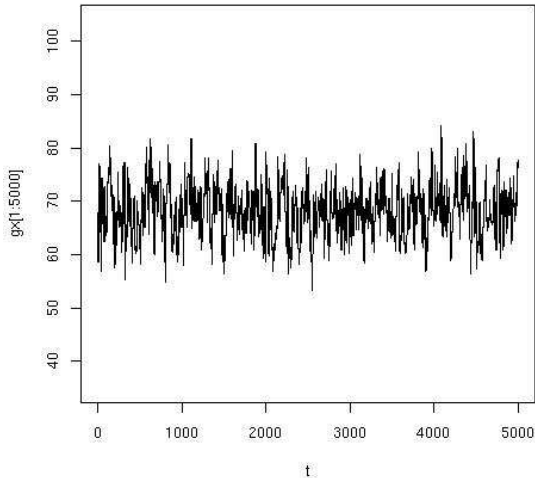
R script:

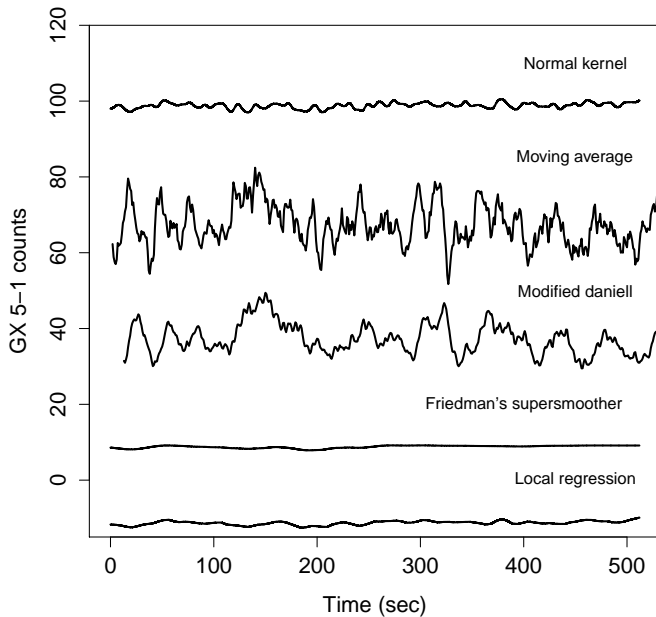
```
gx=scan("GX.dat")  
t=1:5000  
plot(t,gx[1:5000],pch=20)
```



Kernel smoothing of GX 5+1 time series

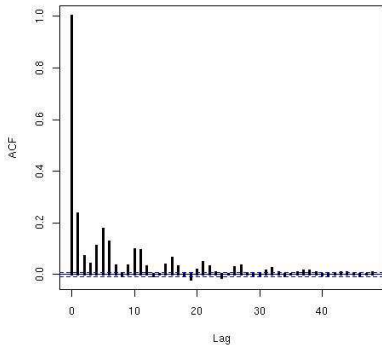
Normal kernel, bandwidth = 7 bins





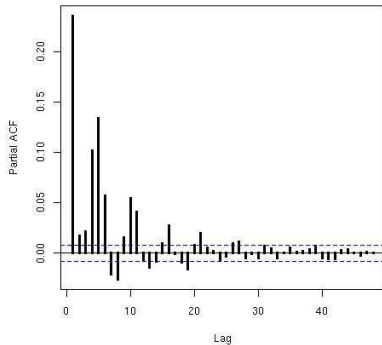
Autocorrelation functions

Series test1



`acf(GX, lwd=3)`

Series test1



`pacf(GX, lwd=3)`

Parametric time domain methods: ARMA models

Autoregressive moving average model

Very common model in human and engineering sciences, designed for aperiodic autocorrelated time series (e.g. 1/f-type 'red noise'). Easily fit by maximum-likelihood. Disadvantage: parameter values are difficult to interpret physically.

$$\text{AR}(p) \text{ model } x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t$$

$$\text{MA}(q) \text{ model } x_t = w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q}$$

The AR model is recursive with memory of past values. The MA model is the moving average across a window of size $q + 1$. ARMA(p,q) combines these two characteristics.

Many extensions to ARMA models:

- VAR (vector autoregressive)
- ARFIMA (ARIMA with long-memory component)
- GARCH (generalized autoregressive conditional heteroscedastic for stochastic volatility)
- Dozens of variants from econometrics: see ftp://ftp.econ.au.dk/creates/rp/08/rp08_49.pdf.

GX 5+1 modeling

```
ar(x = GX, method = "mle")
```

Coefficients:

1 2 3 4 5 6 7 8

0.21 0.01 0.00 0.07 0.11 0.05 -0.02 -0.03

```
arima(x = GX, order = c(6, 2, 2))
```

Coefficients:

ar1 ar2 ar3 ar4 ar5 ar6 ma1 ma2

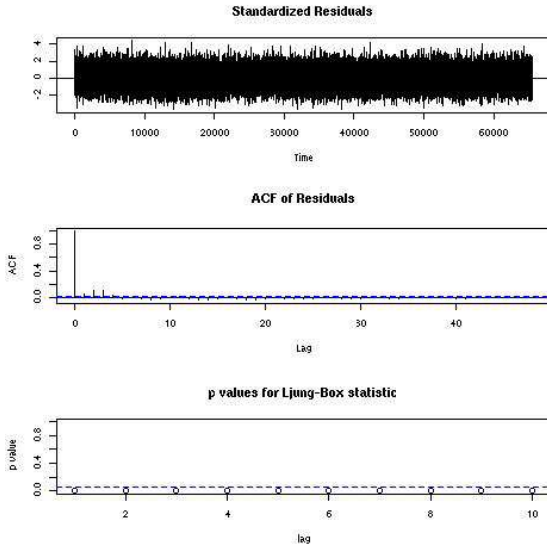
0.12 -0.13 -0.13 0.01 0.09 0.03 -1.93 0.93

Coeff s.e. = 0.004

$\sigma^2 = 102$

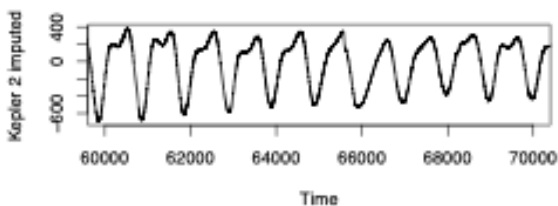
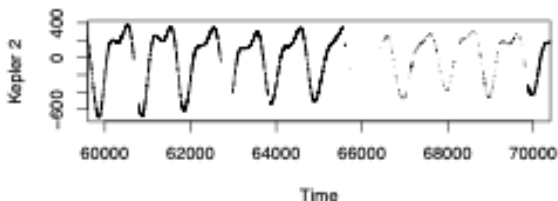
log L = -244446.5

AIC = 488911.1



Although the scatter is reduced by a factor of 30, the chosen model is not adequate: Ljung-Box test shows significant correlation in the residuals. Use AIC for model selection.

ARIMA imputation of missing data



State space models

Often we cannot directly detect x_t , the system variable, but rather indirectly with an observed variable y_t . This commonly occurs in astronomy where y is observed with measurement error (errors-in-variable or EIV model). For AR(1) and errors $v_t = N(\mu, \sigma)$ and $w_t = N(\nu, \tau)$,

$$y_t = Ax_t + v_t \quad x_t = \phi_1 x_{t-1} + w_t$$

This is a *state space model* where the goal is to estimate x_t from y_t , $p(x_t|y_t, \dots, y_1)$. Parameters are estimated by maximum likelihood, Bayesian estimation, Kalman filtering, or prediction. Extended state space models: non-stationarity, hidden Markov chains, etc. MCMC evaluation of nonlinear and non-normal (e.g. Poisson) models

Important Fourier Functions

Discrete Fourier Transform

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t \exp(-2\pi i t \omega_j)$$

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^n x_t \cos(2\pi i \omega_j t) - i n^{-1/2} \sum_{t=1}^n x_t \sin(2\pi i \omega_j t)$$

Classical (Schuster) Periodogram

$$I(\omega_j) = |d(\omega_j)|^2$$

Spectral Density

$$f(\omega) = \sum_{h=-\infty}^{h=\infty} \exp(-2\pi i \omega h) \gamma(h)$$

Fourier analysis reveals nothing of the evolution in time, but rather reveals the variance of the signal at different frequencies.

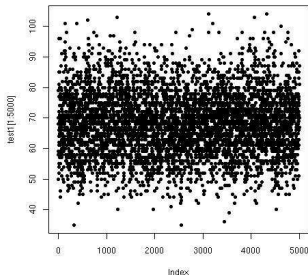
It can be proved that the classical periodogram is an estimator of the spectral density, the Fourier transform of the autocovariance function.

Fourier analysis has restrictive assumptions: an infinitely long dataset of equally-spaced observations; homoscedastic Gaussian noise with purely periodic signals; sinusoidal shape

Formally, the probability of a periodic signal in Gaussian noise is $P \propto e^{d(\omega_j)/\sigma^2}$. But this formula is often not applicable, and probabilities are difficult to infer.

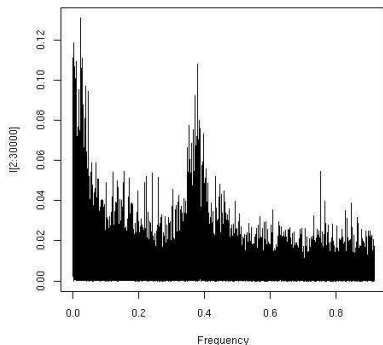
Ginga observations of X-ray binary GX 5-1

GX 5-1 is a binary star system with gas from a normal companion accreting onto a neutron star. Highly variable X-rays are produced in the inner accretion disk. XRB time series often show 'red noise' and 'quasi-periodic oscillations', probably from inhomogeneities in the disk. We plot below the first 5000 of 65,536 count rates from Ginga satellite observations during the 1980s.



```
gx=scan("~/Desktop/CASt/SumSch/TSA/GX.dat")  
t=1:5000  
plot(t,gx[1:5000],pch=20)
```

Fast Fourier Transform of the GX 5-1 time series reveals the 'red noise' (high spectral amplitude at small frequencies), the QPO (broadened spectral peak around 0.35), and white noise.



```
f = 0:32768/65536
I = (4/65536)*abs(fft(gx)/sqrt(65536))^2
plot(f[2:60000],I[2:60000],type="l",xlab="Frequency")
```


Limitations of the spectral density

But the classical periodogram is not a good estimator! E.g. it is formally 'inconsistent' because the number of parameters grows with the number of datapoints. The discrete Fourier transform and its probabilities also depends on several strong assumptions which are rarely achieved in real astronomical data: evenly spaced data of infinite duration with a high sampling rate (Nyquist frequency), Gaussian noise, single frequency periodicity with sinusoidal shape and stationary behavior. Formal statement of strict stationarity:

$$P\{x_{t_1} \leq c_1, \dots, x_{t_K} \leq c_K\} = P\{x_{t_1+h} \leq c_1, \dots, x_{t_K+h} \leq c_K\}.$$

Each of these constraints is violated in various astronomical problems. Data spacing may be affected by daily/monthly/orbital cycles. Period may be comparable to the sampling time. Noise may be Poissonian or quasi-Gaussian with heavy tails. Several periods may be present (e.g. helioseismology). Shape may be non-sinusoidal (e.g. elliptical orbits, eclipses). Periods may not be constant (e.g. QPOs in an accretion disk).

Improving the spectral density I

The estimator can be improved with **smoothing**,

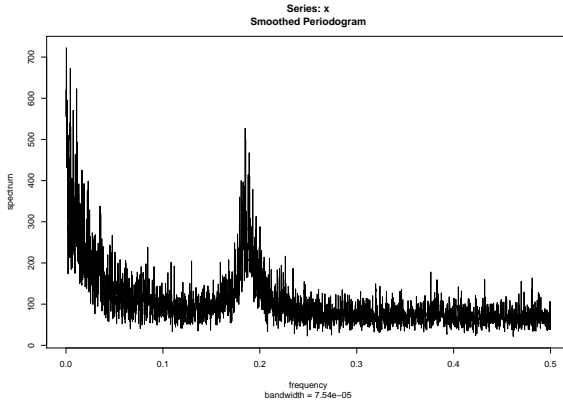
$$\hat{f}(\omega_j) = \frac{1}{2m_1} \sum_{k=-m}^m I(\omega_{j-k}).$$

This reduces variance but introduces bias. It is not obvious how to choose the smoothing bandwidth m or the smoothing function (e.g. Daniell or boxcar kernel).

Tapering reduces the signal amplitude at the ends of the dataset to alleviate the bias due to leakage between frequencies in the spectral density produced by the finite length of the dataset. Consider for example the cosine taper

$$h_t = 0.5[1 + \cos(2\pi(t - \bar{t})/n)]$$

applied as a weight to the initial and terminal n datapoints. The Fourier transform of the taper function is known as the spectral window. Other widely used options include the Fejer and Parzen windows and multitapering. Tapering decreases bias but increases



```
postscript(file="~/Desktop/GX_sm_tap_fft.eps")  
k = kernel("modified.daniell", c(7,7))  
spec = spectrum(gx, k, method="pgram", taper=0.3, fast=TRUE, detrend=TRUE, log="no")  
dev.off()
```

Improving the spectral density II

Pre-whitening is another bias reduction technique based on removing (filtering) strong signals from the dataset. It is widely used in radio astronomy imaging where it is known as the CLEAN algorithm, and has been adapted to astronomical time series (Roberts et al. 1987).

A variety of **linear filters** can be applied to the time domain data prior to spectral analysis. When aperiodic long-term trends are present, they can be removed by spline fitting (high-pass filter). A kernel smoother, such as the moving average, will reduce the high-frequency noise (low-pass filter). Use of a parametric autoregressive model instead of a nonparametric smoother allows likelihood-based model selection (e.g. BIC).

Improving the spectral density III

Harmonic analysis of unevenly spaced data is problematic due to the loss of information and increase in aliasing.

The **Lomb-Scargle periodogram** is widely used in astronomy to alleviate aliasing from unevenly spaced:

$$d_{LS}(\omega) = \frac{1}{2} \left(\frac{[\sum_{t=1}^n x_t \cos \omega(x_t - \tau)]^2}{\sum_{i=1}^n \cos^2 \omega(x_t - \tau)} + \frac{[\sum_{t=1}^n x_t \sin \omega(x_t - \tau)]^2}{\sum_{i=1}^n \sin^2 \omega(x_t - \tau)} \right)$$

where $\tan(2\omega\tau) = (\sum_{i=1}^n \sin 2\omega x_t) (\sum_{i=1}^n \cos 2\omega x_t)^{-1}$

d_{LS} reduces to the classical periodogram d for evenly-spaced data. Bretthorst (2003) demonstrates that the Lomb-Scargle periodogram is the unique sufficient statistic for a single stationary sinusoidal signal in Gaussian noise based on Bayes theorem assuming simple priors.

Conclusions on spectral analysis

For challenging problems, smoothing, multitapering, linear filtering, (repeated) pre-whitening and Lomb-Scargle can be used together. Beware that aperiodic but autoregressive processes produce peaks in the spectral densities. Harmonic analysis is a complicated 'art' rather than a straightforward 'procedure'.

It is extremely difficult to derive the significance of a weak periodicity from harmonic analysis. Do not believe analytical estimates (e.g. exponential probability), as they rarely apply to real data. It is essential to make simulations, typically permuting or bootstrapping the data keeping the observing times fixed. Simulations of the final model with the observation times is also advised.

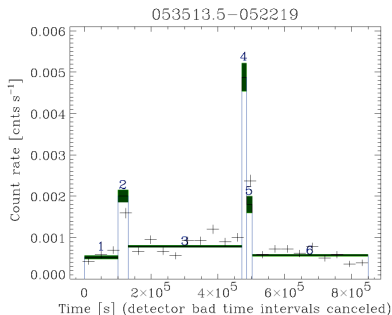
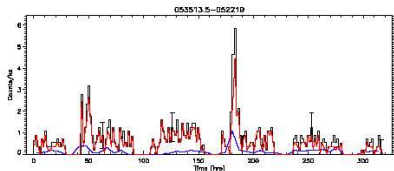
Periodicities in irregularly spaced time series

Several non-Fourier periodograms have been developed by astronomers confronting unevenly sampled data. Most are nonparametric, without assumption of a sinusoidal shape to the periodic variation, particularly adapted to eclipses or other short-duty cycle variations. These include:

- Phase dispersion minimization (Stellingwerf 1978)
- Analysis of variance (Schwarzenberg-Czerny 1989)
- Minimum string length (Dworetsky 1983)
- Rayleigh and Z_n^2 tests for Poisson photon arrival times (Leahy et al. 1983).
- Bayesian modeling (Gregory & Loredó 1992)

Bayesian Blocks

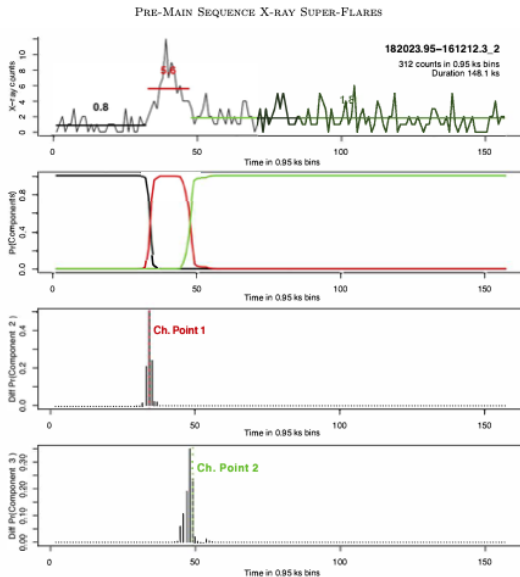
Bayesian Blocks constructs a segmented piecewise-constant (histogram) model for event, integer (binned), or real measurements as a function of time (Scargle 1998, Scargle et al. 2013). An example of *change point analysis* where both the number and location of change points is unknown. Very useful in X-ray and gamma-ray astronomy.



Poisson regression model with multiple changepoints

(similar to Bayesian Blocks)

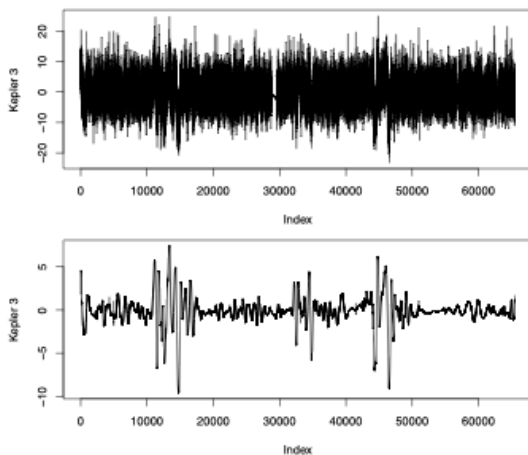
Getman & Feigelson 2021



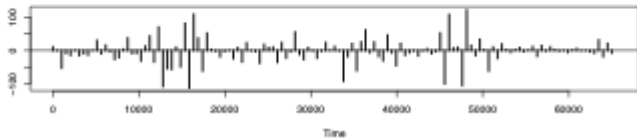
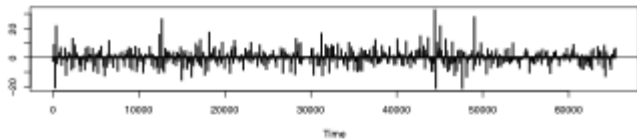
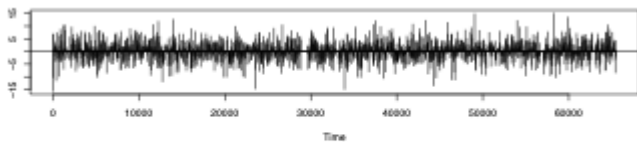
Wavelet analysis

Non-stationary periodic behaviors can be studied using **time-frequency Fourier analysis**. Here the spectral density is calculated in time bins and displayed in a 3-dimensional plot.

Wavelets are now well-developed for non-stationary time series, either periodic or aperiodic. Here the data are transformed using a family of non-sinusoidal orthogonal basis functions with flexibility both in amplitude and temporal scale. The resulting wavelet decomposition is a 3-dimensional plot showing the amplitude of the signal at each scale at each time. Wavelet analysis is often very useful for noise thresholding and low-pass filtering.

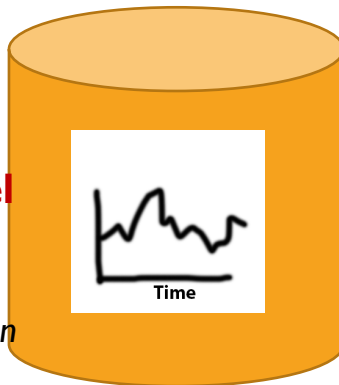


KIC 007609553 lightcurve before (*top*) and after (*bottom*) wavelet denoising.

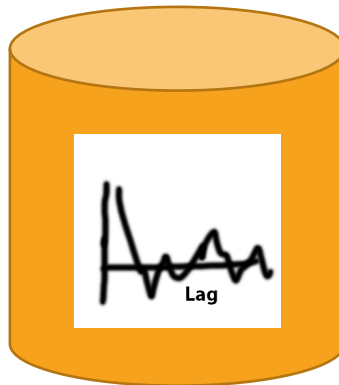


What astronomers do with a time series?

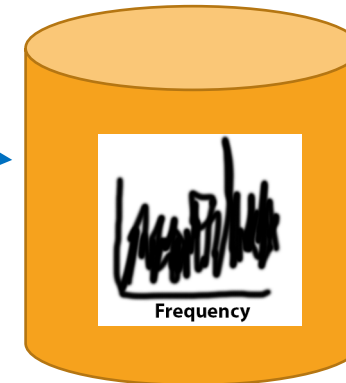
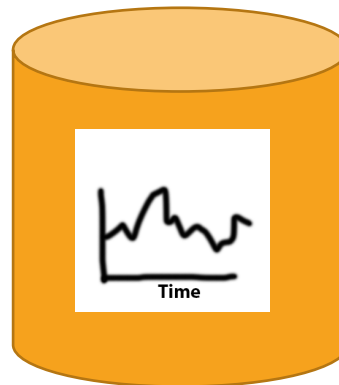
Regression
Astrophysics
Ad hoc model
Autoregressive
 R^2 / AD tests
AIC model selection



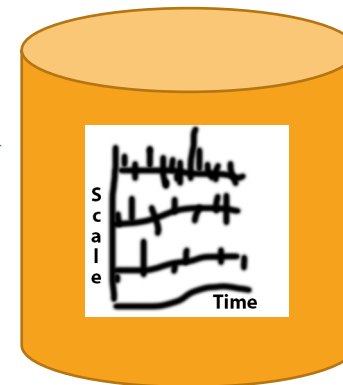
Autocorrelation
ACF / PACF / CCF
DW / LB tests



Original data
Moments
SW / AD tests
ADF / KPSS tests
Change point analysis



Power spectrum
Fourier filtering
PDM / AoV
BLS/TCF

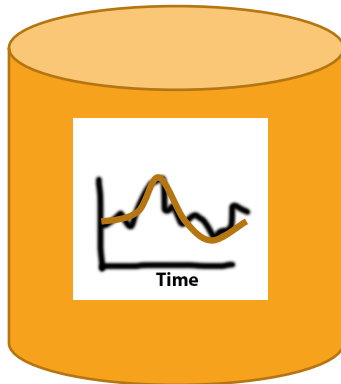


Wavelet transform
Feature detection
Denoising
Compression

What to do with a time series?

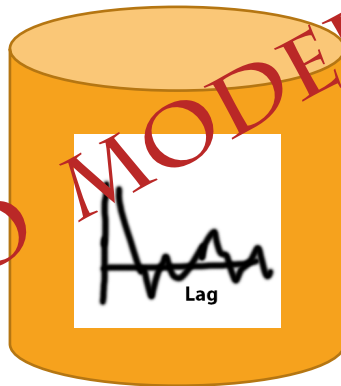
Regression

Parametric
Semi-parametric
(splines, GP, ...)
Autoregressive
(ARMA, ...)
 R^2 / AD tests
AIC model selection



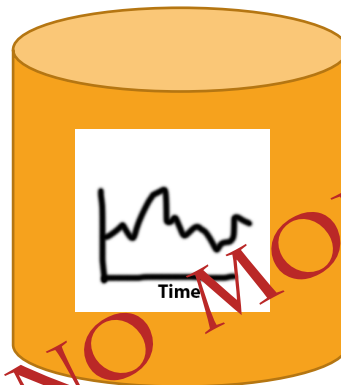
Autocorrelation

ACF / PACF / CCF
DW / LB tests



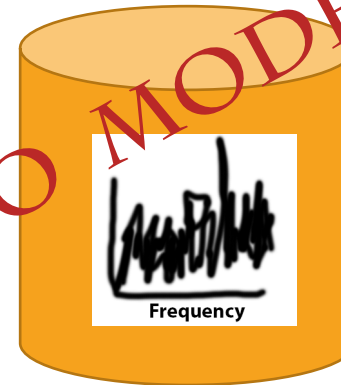
Original time series

Moments
SW / AD tests
ADF / KPSS tests
Change point analysis



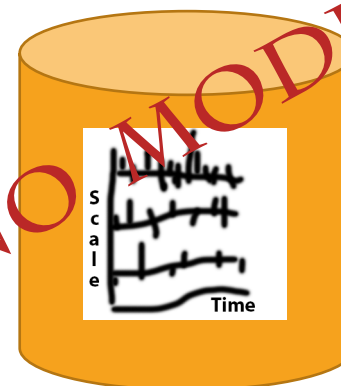
Power spectrum

Fourier filtering
LS / GLS
String length
PDM / AoV
Box LS
Transit Comb
No tests!



Wavelet transform

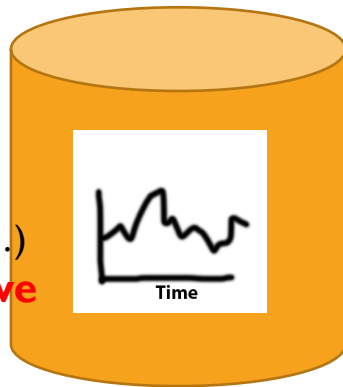
Denoising
Compression
Significance tests (rare)



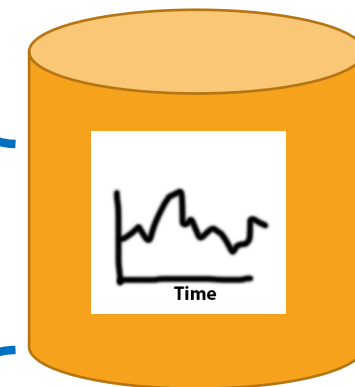
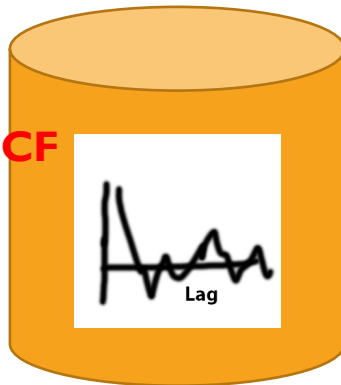
plus signal processing methods like Independent Component Analysis, correntropy, Singular Spectrum Analysis, hidden Markov models, ...

What econometricians do with a time series?

Regression
 Parametric
 Semi-parametric
 (**splines**, GP, ...)
Autoregressive
 (ARMA, ...)
 R^2 / AD tests
 AIC model selection

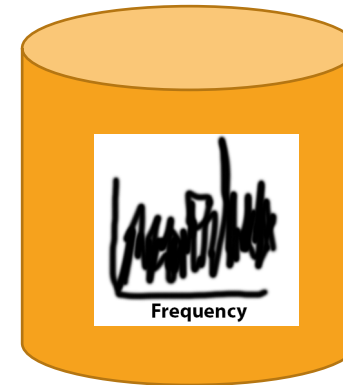


Autocorrelation
ACF / PACF / CCF
 DW / LB tests

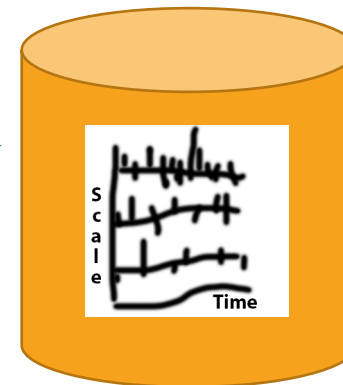


**Original
 time series**

Moments
 SW / AD tests
 ADF / KPSS tests
 Change pt analysis



Power spectrum
 Fourier filtering
 LS / GLS
 String length
 PDM / AoV
 Box LS
 Transit Comb
 No tests!



Wavelet transform
 Denoising
 Compression
 Significance tests (rare)

plus signal processing methods like Independent Component Analysis, correntropy, Singular Spectrum Analysis, hidden Markov models, ...

LESSONS LEARNED FROM STATISTICIANS, ECONOMETRICIANS & SIGNAL PROCESSING ENGINEERS

Theorems establish that the nonparametric Fourier power spectrum, autocorrelation function and the wavelet decomposition, and parametric ARMA models, all reproduce the full time series if an infinite number coefficients are retained. Each transform has its own strengths:

- Use Fourier or related analysis for stationary periodic signals
- Use ARIMA modeling for autocorrelated signals (including $1/f^\alpha$ noise) with (P)ACF for testing
- Use wavelets for transient and (distinct or overlapping) multi-scale signals
- Use ICA, SSA, or combination of methods for overlapping signals with different behaviors
- Use change point analysis, CUSUM charts, sequential analysis, Markov switching models, ... for multi-state behavior

Astronomers overuse Fourier analysis ...

it is poor for characterizing aperiodic or nonstationary phenomena

like trends, autocorrelation ($1/f^\alpha$), transient events, multi-scale, multi-state, etc!