

AGA 0505 - Análise de Dados em Astronomia

12. Séries Temporais

Laerte Sodré Jr.

1o. semestre, 2022

aula de hoje:

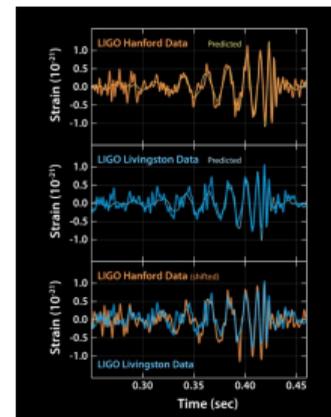
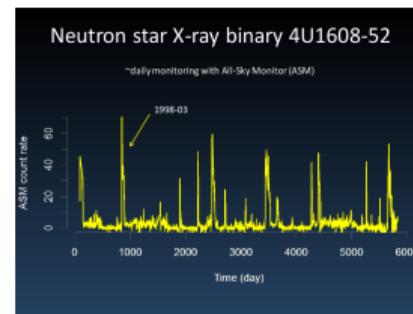
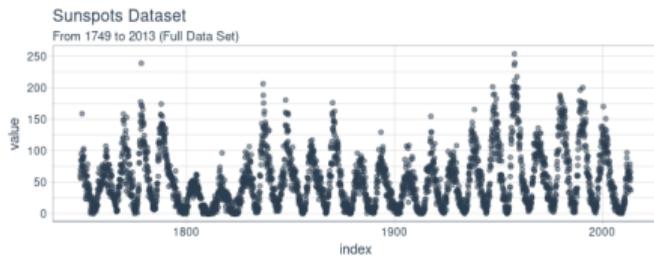
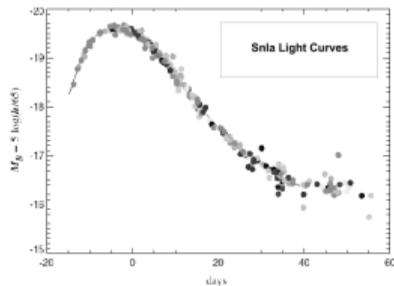
1. séries temporais em astronomia
2. fenômenos que dependem do tempo
3. algumas técnicas de análise de séries temporais:
 - 3.1 autocorrelação
 - 3.2 filtragem
 - 3.3 análise harmônica
 - 3.4 wavelets
 - 3.5 modelos autoregressivos
 - 3.6 *machine learning*

A única razão para o tempo é que as coisas não aconteçam de uma vez.

Albert Einstein

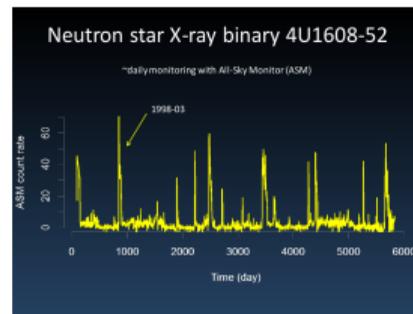
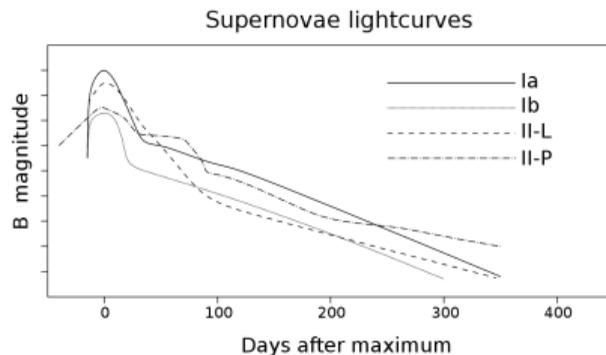
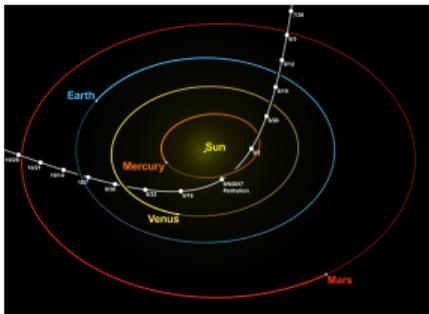
séries temporais (ST) em astronomia

- fenômenos dependente do tempo estão por todo lado em Astronomia



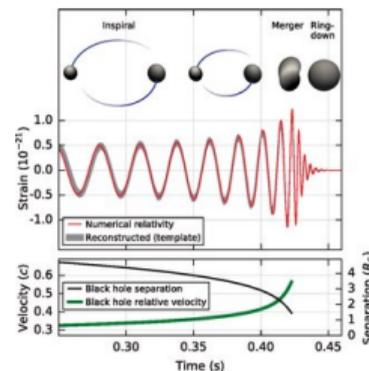
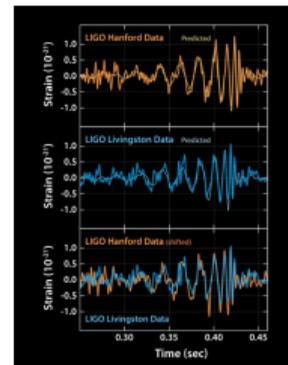
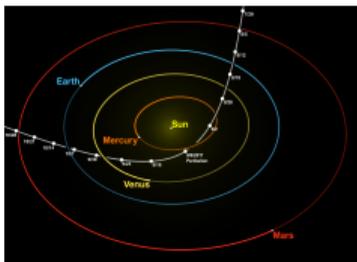
processos dependentes do tempo

- 3 tipos de fenômenos que dependem do tempo:
 - processos **periódicos ou determinísticos**
 - * órbitas
 - processos **transientes ou explosivos**
 - * supernovas
 - processos **estocásticos ou caóticos**
 - * discos de acreção



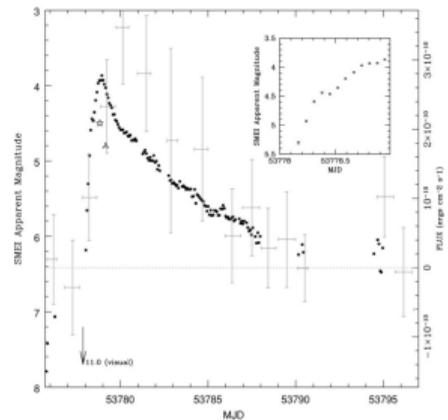
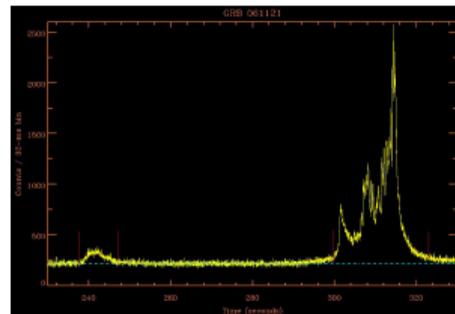
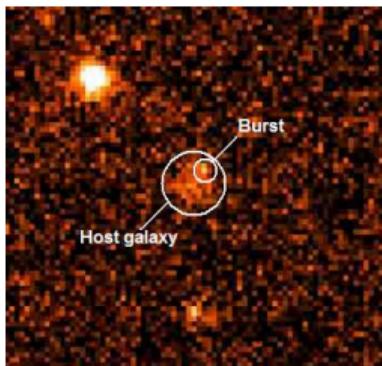
processos dependentes do tempo

- processos **periódicos** ou **determinísticos**:
 - órbitas - planetas, Omuamua, estrelas binárias, ...
 - rotação/ciclos - ciclo solar, pulsares, cefeidas, ...
- o que se pode fazer:
 - modelagem dos processos
 - estimativa dos parâmetros do modelo: órbitas, massas ...



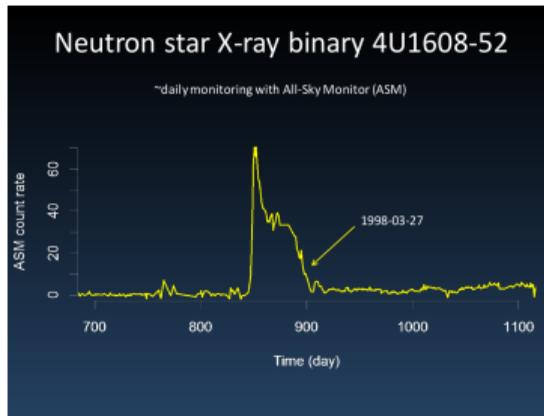
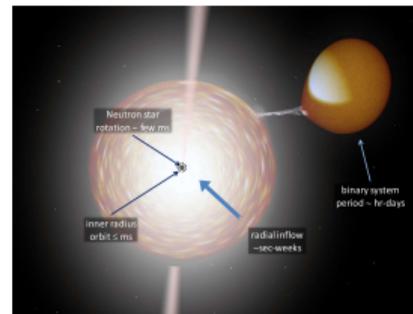
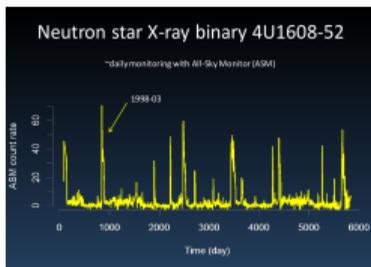
processos dependentes do tempo

- processos **transientes ou explosivos**:
 - supernovas, atividade estelar, GRBs (*gamma ray bursts*), FRBs (*fast radio bursts*)
- o que se pode fazer:
 - detecção, identificação, classificação
 - modelagem e estimativa de parâmetros



processos dependentes do tempo

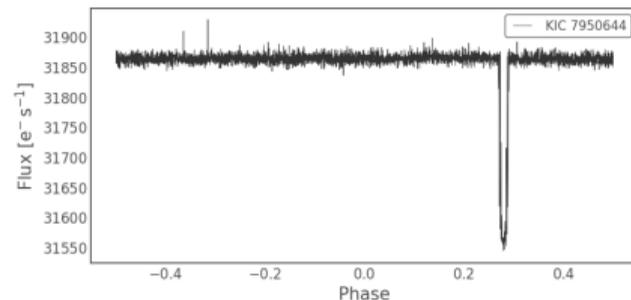
- processos **estocásticos ou caóticos**:
 - flutuações de brilho em discos de acreção, jatos, ...
 - fenômenos quasi-periódicos
 - não se pode prever com precisão a série temporal
- o que se pode fazer:
 - comparação estatística entre dados e modelos para se inferir os parâmetros físicos do sistema



séries temporais em astronomia

- ST em astronomia não são como as encontradas em outras áreas:
 - muitas vezes a série não é igualmente espaçada
 - as observações individuais estão sujeitas a erros heterocedásticos
 - as escalas de tempo relevantes podem ser muito variadas
 - as variações das fontes e da contribuição do fundo que precisa ser subtraído são frequentemente correlacionadas
- em geral, em astronomia estamos mais interessados em modelagem que em previsão

- algumas técnicas de análise de ST:
 - autocorrelação
 - análise harmônica
 - modelos autoregressivos
 - wavelets
 - *machine learning*
 - ...



curva de luz obtida pelo satélite Kepler

autocorrelação

- medida de estruturas correlacionadas numa série temporal:

- ST: $\{X_1, X_2, \dots, X_t, \dots, X_n\}$
 X_i : valor da variável X no instante i

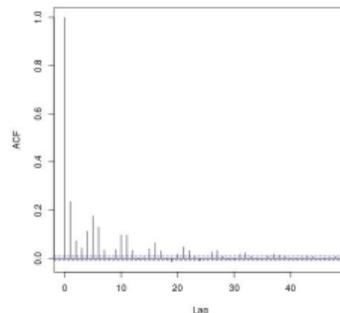
- função de autocorrelação (ACF): é uma função do *lag*/atraso k

$$ACF(k) = \frac{\sum_{t=k+1}^n (X_t - \bar{X})(X_{t-k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2}$$

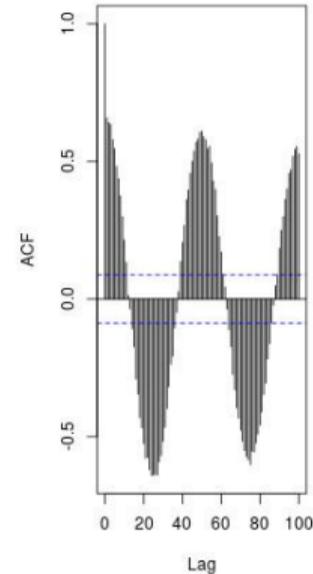
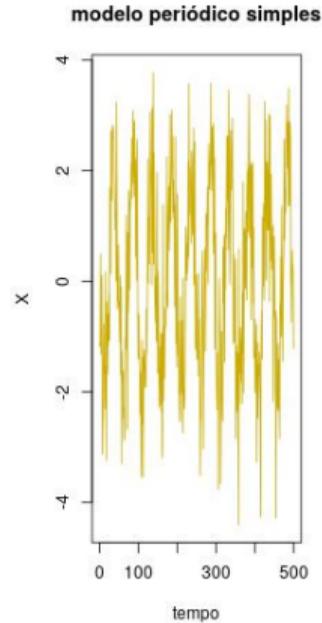
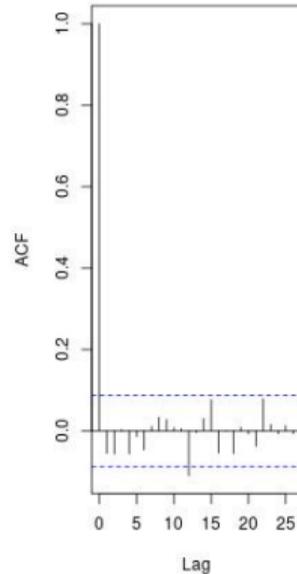
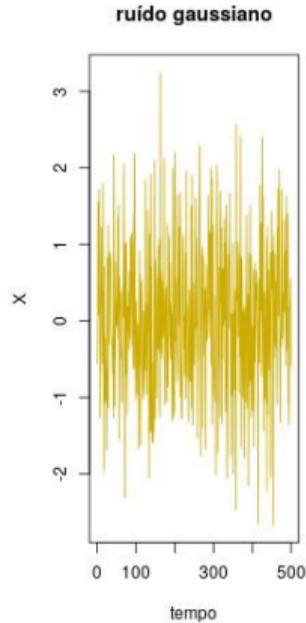
mede a correlação média normalizada entre X_t e X_{t-k} em uma ST igualmente espaçada

- k : número inteiro > 0 , *lag time*
- \bar{X} : valor médio da série temporal
- note que $ACF(0) = 1$

- comportamentos complicados na série temporal podem ter um comportamento mais simples na ACF
- ruído aleatório não-correlacionado (ruído branco) produz valores perto de zero.
- às vezes ACFs são difíceis de se interpretar, principalmente quando se tem tendências e processos estocásticos juntos

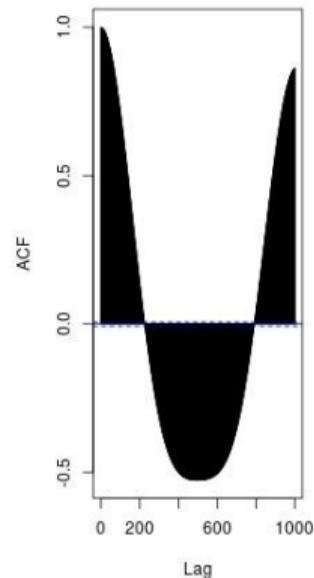
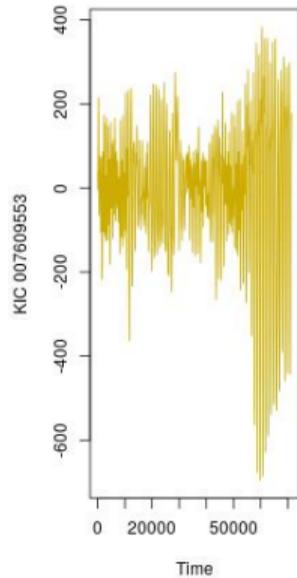
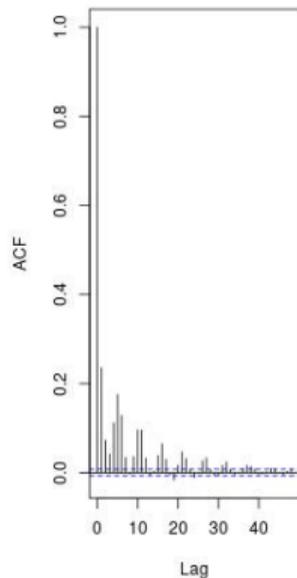
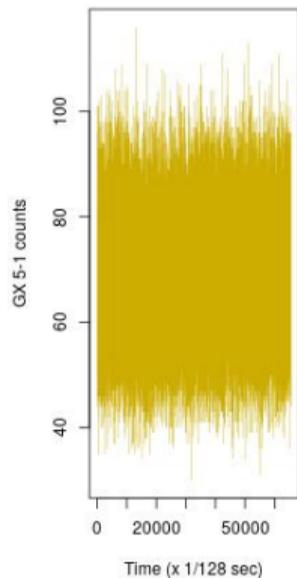


autocorrelação



as linhas tracejadas azuis na ACF delimitam o intervalo de confiança de 95%

autocorrelação

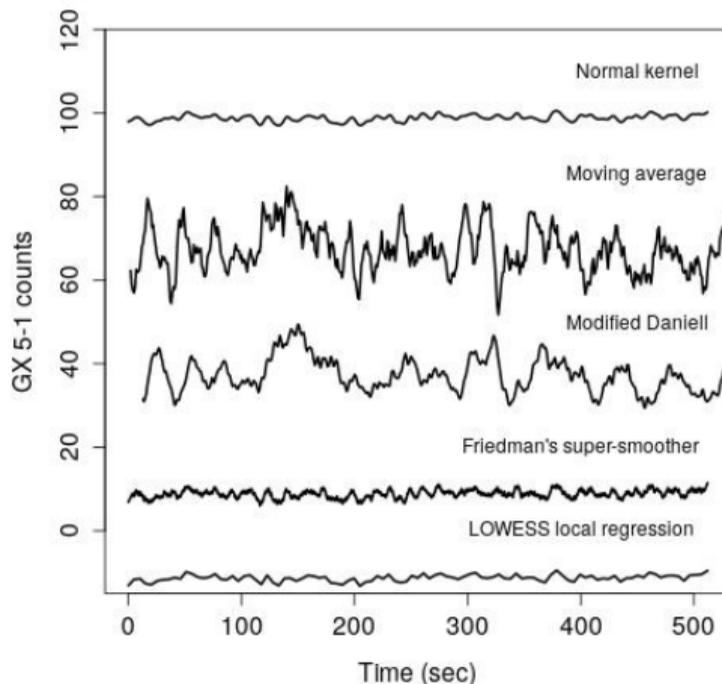


as linhas tracejadas azuis na ACF delimitam o intervalo de confiança de 95%
GX 5-1 é um sistema binário de raios-X

KIC 007609553 é uma estrela que mostra variabilidade quase-periódica devido à presença de rotação e manchas estelares

filtragem

- objetivo: melhorar a visualização do comportamento da ST removendo variações de alta-frequência
- exemplo: média móvel de comprimento 3 substitui-se X_t pela média de X_{t-1} , X_t , X_{t+1}
- filtro de Daniell modificado: média móvel ponderada
- Friedman's super smoother: baseado numa regressão linear local com largura de banda (bandwidth) adaptativa
- LOWESS: regressão polinomial local
- *a filtragem reduz a variância mas introduz viés*



análise harmônica: a transformada de Fourier discreta

- objetivo: identificar ciclos dominantes numa ST
- qualquer ST pode ser expressa como uma super-posição de senos e cossenos com frequências ω_i (ou períodos $T_i = 1/\omega_i$), amplitudes A_i e fases ϕ_i diferentes
- exemplo de uma série periódica:

$$X_t = A \cos(2\pi\omega t + \phi)$$

- uma ST pode ser representada por uma transformada de Fourier discreta (DFT):

$$Y_k = \sum_{j=0}^{n-1} X_j e^{-2\pi i(jk/n)}, \quad k = 1, \dots, n$$

ou

$$Y_k = \sum_{j=0}^{n-1} X_j \left[\cos\left(\frac{2\pi jk}{n}\right) + i \operatorname{sen}\left(\frac{2\pi jk}{n}\right) \right]$$

- transformada inversa:

$$X_j = \frac{1}{n} \sum_{k=0}^{n-1} Y_k e^{2\pi i(jk/n)}$$

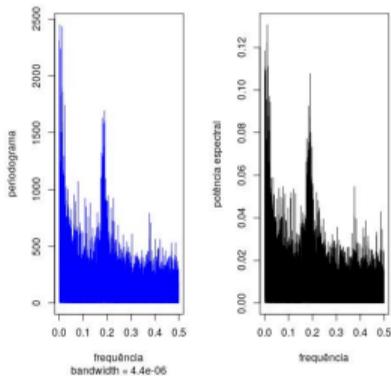
- FFT: *Fast Fourier Transform*: algoritmo eficiente para calcular a transformada de Fourier discreta

análise harmônica: o periodograma

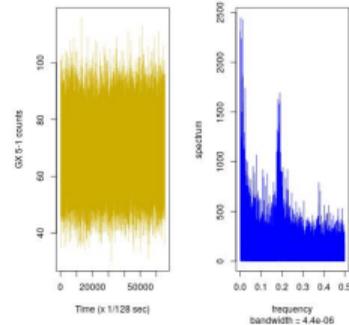
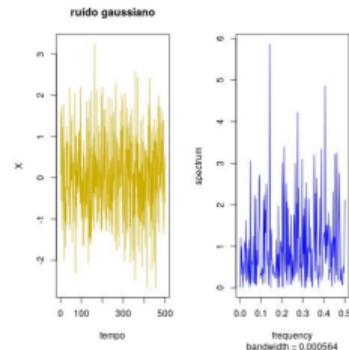
- espectro de potência: mede como a potência do sinal varia com a frequência:

$$P(k) = \frac{4}{n^2} |Y_k|^2$$

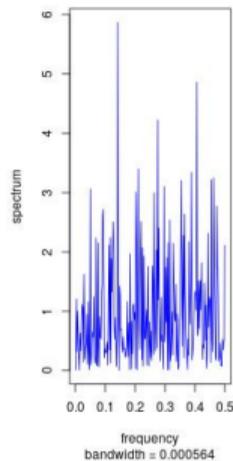
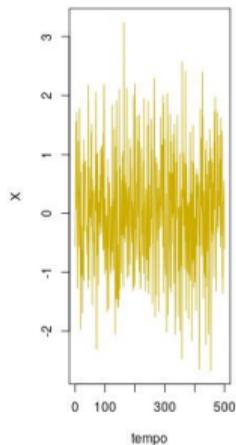
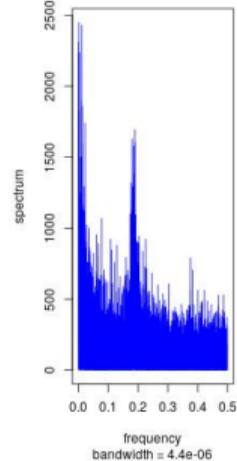
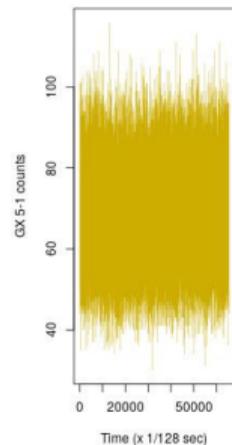
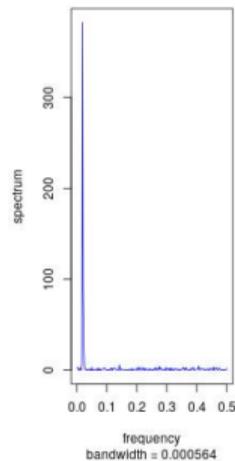
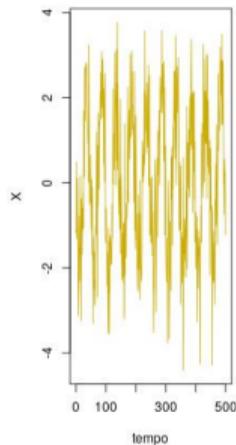
- o periodograma é equivalente (a menos da normalização) ao espectro de potências



periodogramas:



análise harmônica: o periodograma

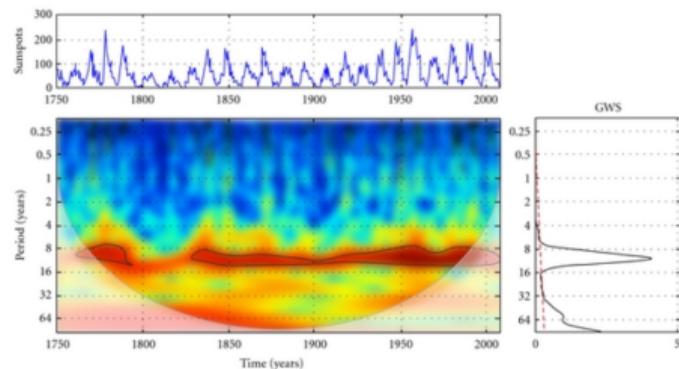
ruído gaussiano**modelo periódico simples**

problemas com a análise harmônica

- E. Feigelson: *Harmonic analysis is a complicated 'art' rather than a straightforward 'procedure'*
- a DFT depende de hipóteses raramente encontradas na prática: séries estacionárias infinitas igualmente espaçadas e com alta taxa de amostragem (frequência de Nyquist), ruído gaussiano, etc
- frequentemente deve-se usar filtragem, *tapering* (“afunilamento”, que reduz a amplitude do sinal nos extremos da série) e outros procedimentos
- processos autoregressivos não periódicos podem produzir picos no espectro de potência ou periodograma
- ...

wavelets

- a transformada de Fourier converte uma série temporal para o espaço de frequências
- a wavelet descreve a série temporal em termos de tempo e frequência
- considere um sinal cuja frequência varia com o tempo: uma TF daria apenas a informação sobre as frequências presentes, mas não sobre quando ocorrem
- exemplo: séries com o número de manchas solares



modelos autoregressivos

- regressão baseada em valores anteriores da série (os lags)
- *séries estacionárias*: o comportamento temporal, estocástico ou determinístico, é invariante estatisticamente por deslocamentos arbitrários no tempo
- modelos $AR(p)$: X_t depende linearmente de seus p valores anteriores mais um termo estocástico:

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

- exemplo: modelo $AR(0)$ - “random walk”:

$$X_i = \alpha_1 X_{i-1} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma)$$

por substituição:

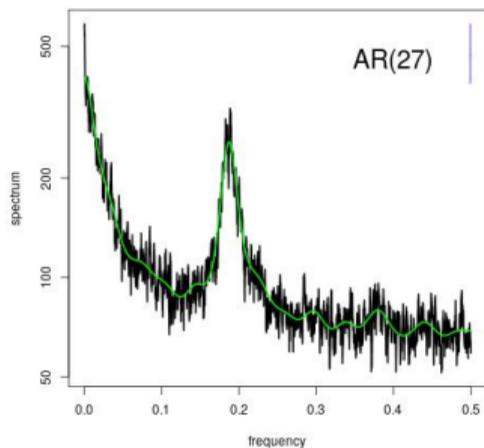
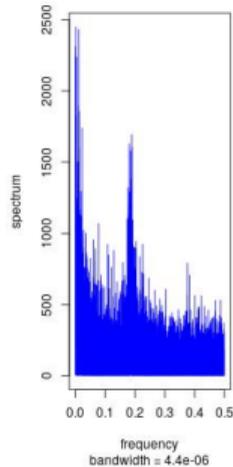
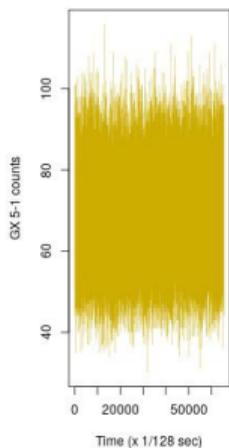
$$X_i = \alpha_1^i X_0 + \{\alpha_1^{i-1} \epsilon_1 + \alpha_1^{i-2} \epsilon_2 + \dots + \epsilon_i\}$$

o termo entre chaves é *ruído branco*

- os coeficientes de um modelo $AR(p)$ podem ser obtidos por máxima verossimilhança e os erros por bootstrap
- seleção de modelos: qual o melhor p ? AIC ou BIC
- a regressão em princípio pode ir até o começo da série
($t-1, t-2, \dots, 0$): modelos de *memória longa*

modelos autoregressivos

- exemplo: modelo autoregressivo de GX 5-1
- melhor modelo: $p = 27$
- comparação do periodograma observado e modelado por AR(27):



modelos autoregressivos

- modelos MA: *moving average*- previsão baseada nos *erros das previsões anteriores*
- exemplo: MA(1)

$$X_t = \beta_1 e_{t-1} + \epsilon_t$$

e_{t-1} : diferença entre o valor observado e o previsto em $t - 1$

- são modelos de *memória curta*:

$$X_{t-1} = \beta_1 e_{t-2} + \epsilon_{t-1}$$

$$X_t = \beta_1 e_{t-1} + \epsilon_t$$

note que X_t não depende de e_{t-2} (e termos anteriores)

- MA(q):

$$X_t = \beta_1 e_{t-1} + \dots + \beta_q e_{t-q} + \epsilon_t$$

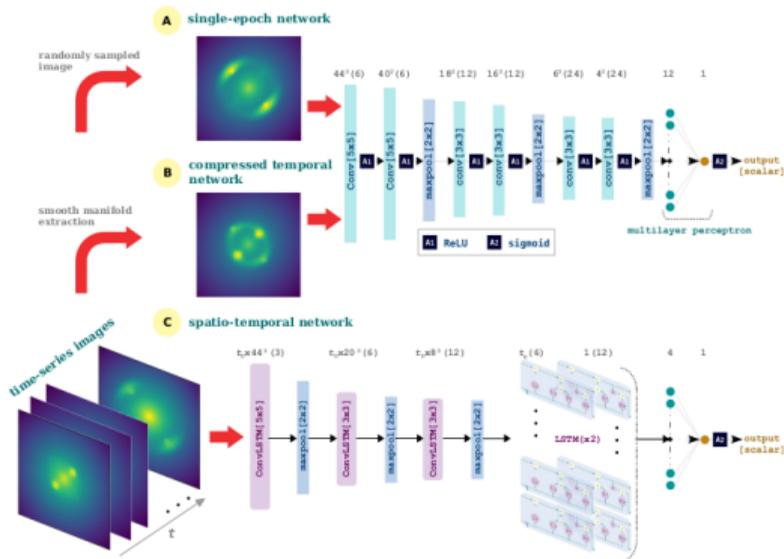
- generalização: ARIMA
autoregressive integrated moving average
- exemplo: ARIMA($p, 0, q$)

$$X_t = \alpha_1 X_{i-1} + \dots + \alpha_p X_{i-p} +$$

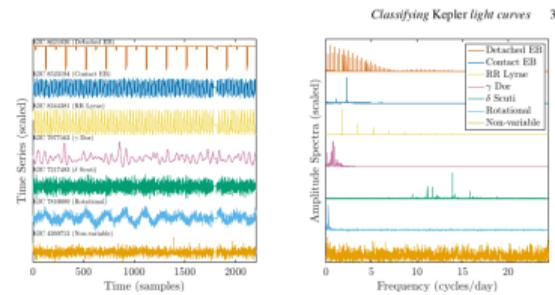
$$+ \beta_1 e_{t-1} + \dots + \beta_q e_{t-q} + \epsilon_t$$

métodos de aprendizado de máquina

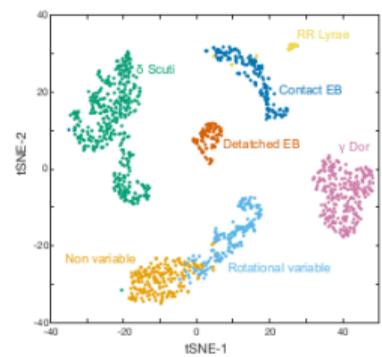
8 D. Kodi Ramanah, N. Arendse, R. Wojtak



sistema para descoberta de supernovas amplificadas gravitacionalmente em imagens ópticas (arXiv:2107.12399)



Classifying Kepler light curves



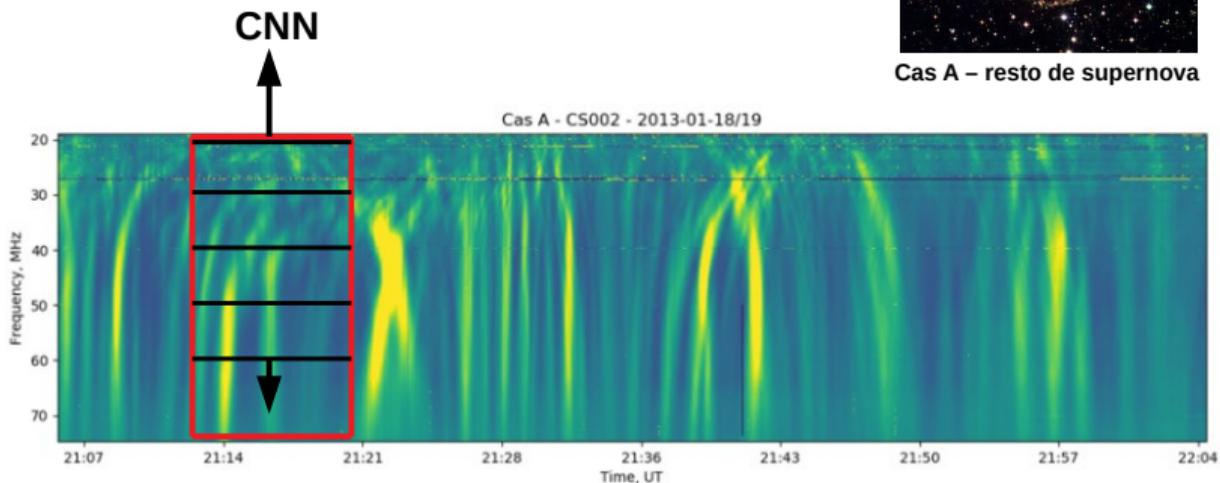
Barbara et al. (arXiv:2205.03020)

métodos de aprendizado de máquina

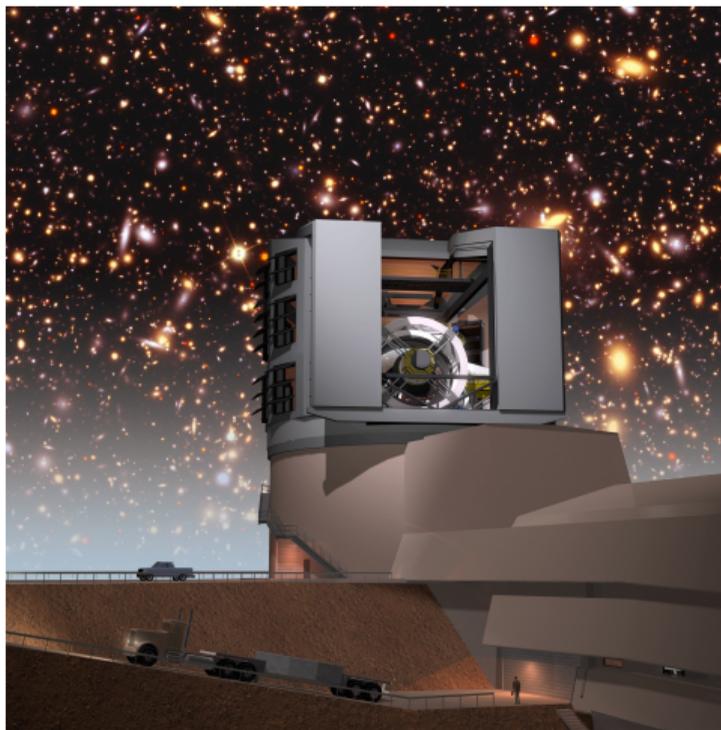
espectrogramas podem ser tratados como imagens



Cas A – resto de supernova



Rubin Observatory/LSST



- *Probing dark energy and dark matter.*
- *Taking an inventory of the solar system.*
- *Exploring the transient optical sky.*
- *Mapping the Milky Way.*

referências

- Feigelson, E., https://sites.psu.edu/astrostatistics/files/2021/02/Tutorial_4_Time_Series.html, 2021
- Modern Statistical Methods for Astronomy: With R Applications , Feigelson & Babu, 2012
- Statistics, Data Mining, and Machine Learning in Astronomy, Ivezić, Connolly, VanderPlas & Gray, 2014
- Vaughan, S., *Random time series in astronomy*, Phil. Trans. Roy. Soc., 371, 20110549, 2012