An introduction to time series analysis

Matthew J. Graham
Center for Data-Driven Discovery/ZTF, Caltech
mjg@caltech.edu
A history of anomalous observations

• Who?
  Babylonian Astronomical Diary

• What?
  The comet which previously had appeared in the east in the path of Anu in the area of Pleiades and Taurus

• Where?
  to the west [...] and passed along in the path of Ea in the region of Sagittarius, 1 cubit in front of Jupiter, 3 cubits high toward the north [...]

• When?
  Month VIII, SE 148 (lunar month beg. 21 October 164 BC)

• How?
  By eye

• Why?
  Celestial divination
The first astronomical time series

Thomas Harriott: Dec 1610

Image credit: University of Michigan Special Collections Library
“If the new star were outside the ordinary course of nature, it would tell us little about the constitution of the universe.”
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
- ...

20 May 2019
Matthew J. Graham
What we do ask of time series?

**Population behaviors**
- Characterize, categorize, classify

**Outliers**
- Extreme sources
- Physical models
- Predictions

(Cody & Hillenbrand 2018)
Types of astronomical variability

Credit: L. Eyer & N. Mowlavi (03/2009)
A time series is a set of time-tagged measurements: \( \{X_i(t_i)\} \) with observation errors \( \sigma_i \)

**Non-IID**
- Data is sequential

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

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

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

(Belloni et al. 2000)

**Homoskedasticity**
- All errors drawn from same process
Foundational concepts - II

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

Power spectrum
- Power spectral density tells you everything: \( PSD(\nu) = |F(x)|^2 \)
- PSD is Fourier transform of autocorrelation function:

\[
PSD(\nu) = \int_{-\infty}^{\infty} ACF(\Delta t) e^{-2\pi i \nu \Delta t} \Delta t
\]

\[
ACF(\Delta t) = \mathbb{E}[ (x_t - \mu)(x_{t+\Delta t} - \mu) ] / \sigma^2
\]

- The structure function is related to the autocorrelation function:

\[
SF(\Delta t) = \sqrt{2} \sigma_s \sqrt{1 - ACF(\Delta t)}
\]

\[
SF(\Delta t) = 0.742 \text{ IQR}(x)
\]
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 $C \times R(t)$.

Deterministic chaos vs. stochastic?
Characterization – extracting data features

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

**Amplitude**

**Slope**

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

**Amplitude**

**Slope**
Common statistical features

• **Timescales:**
  • Lomb-Scargle

• **Variability:**
  • von Neumann variability (phase-folded)
  • Stetson K index

• **Morphology:**
  • Skewness
  • Kurtosis
  • IQR
  • Cumulative sum index (phase-folded)
  • Ratio of magnitudes brighter/fainter than mean

• **Trends:**
  • Slope percentiles (phase-folded)

• **Model:**
  • Fourier amplitude ratios
  • Fourier phase differences
  • Fourier amplitude
  • Shapiro-Wilk normality test

“Actually they all look alike to me.”
Categorization

(Cody & Hillenbrand 2018)
Characteristic timescales

![Graph showing characteristic timescales with various labels and data points.](Image)

(Sartori et al. 2018)
## Data-derived classes

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

![ATLAS PULSE variables](image)

(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

\[ x(t + P) = x(t); f = 1/P \]

\[ x(t, f) = A_f \sin 2\pi f (t - \varphi_f) \]

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

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

\[ \varphi(t, f) = tf - \text{int}(tf) \]

\[ \theta(f) = g(\varphi_n, x_n; f) \]

\[ P(f) = h(\theta(f)) \]
Period finding is not a single algorithm

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

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

- **Data set:**
  - 15522 CRTS light curves for all objects in SIMBAD and VSX with a quoted period
  - 50124 ACVS light curves for MACC classification
  - 1500 MACHO light curves for RR Lyrae, EBs and Cepheids

- **Classes:**
  - Eruptive (4194): T Tauri, red supergiants, RS Can Ven
  - Pulsating (45599): semiregulars, RR Lyrae, Mira, δ Scuti, Cepheids
  - Rotating (455): chemically peculiar, BY Dra
  - Cataclysmic (386): S U Ma, U Gem, novalike
  - Eclipsing (14952): eclipsing binaries, AM Her
  - Other (1369)

- 9 different algorithms (Graham et al. 2013)
What can we say about period finding

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

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

\[ dX(t) = -\frac{1}{\tau} X(t) dt + \sigma \sqrt{dt} \epsilon(t) + b dt \quad \tau, \sigma, t > 0 \]

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

- 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)

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

(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

- Gaussian process regression
- Localized chaos measure

(Golestani & Gras 2014)
Deep modelling of time series

- LSTM Autoencoder:

(Naul et al. 2018)
Deep time series features

(Tachibana et al. 2019)
RNNs with QSOs
Summary

• Traditional time series analyses in astronomy involve:
  • (simple) discriminative features as (possible) inputs to machine learning algorithms
  • outlier detections based on Gaussian tails
  • little predictive power

• Data volumes now mean that we can *model individual sources*:
  • capturing full time series behavior
  • better identifying extrema
  • with generative approaches

• Next generation surveys enable real-time validation of predicted behaviors and swift identification of deviance

• Let’s go hunting for technosignatures