Lecciones en Astroinformática Avanzada (Semester 1 2025)

Automatic Classification of Variable Stars (I)

Nina Hernitschek
Centro de Astronomía CITEVA
Universidad de Antofagasta

April 15, 2025

Motivation

Nowadays large-scale astronomical surveys enable us to see the universe in much more detail:

deeper (fainter objects), **wider** (larger on-sky footprint), **faster** (higher temporal resolution, cadence)

Motivation

Nowadays large-scale astronomical surveys enable us to see the universe in much more detail:

deeper (fainter objects), **wider** (larger on-sky footprint), **faster** (higher temporal resolution, cadence)

⇨ methodological approaches to analyze the data changes

Modern All-Sky Surveys

SDSS (Sloan Digital Sky Survey) as a pioneer of this technique:

- operating since 2000 (ongoing)
- 2.5 meter telescope (Apache Point Observatory)
- five phases (SDSS I - SDSS V) with multiple sub-surveys
- $>$ 100 Terabyte imaging of $\sim$ 1 Billion objects

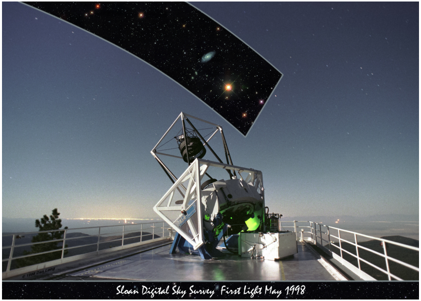


Image Credit: Dan Long (Apache Point Observatory)

Modern All-Sky Surveys

SDSS (Sloan Digital Sky Survey) as a pioneer of this technique:

exciting discoveries over 20 years - from our Solar System to cosmological distances:

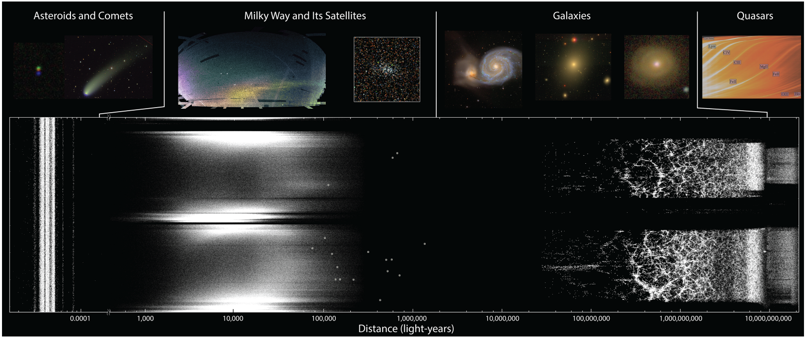
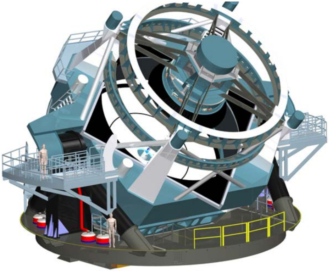


Image Credit: V. Belokurov, M. R. Blanton, A. Bonaca, X. Fan, M. C. Geha, R. H. Lupton, the SDSS Collaboration

Modern All-Sky Surveys

The LSST Survey (Legacy Survey of Space and Time):

- 8.4-meter (6.7 m equivalent) telescope at Rubin Observatory
- 10-year photometric survey *ugrizy*
- 1000 images/night = 15 TB/night, 10 million transients/night
- first light: July 2024



Simonyi Survey Telescope at Vera Rubin Observatory,
Image Credit: NOAO

Modern All-Sky Surveys

What are *all-sky surveys* looking for?

faint objects far away

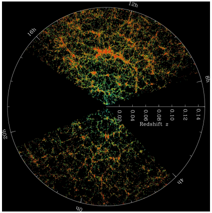
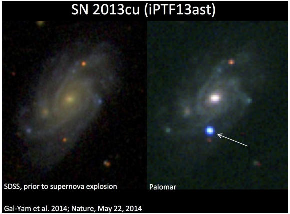


Image Credit: SDSS

variable and transient objects



stellar overdensities

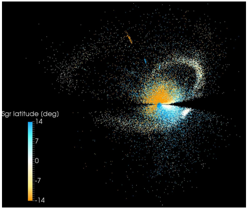


Image Credit: B. Sesar, N. Hernitschek

Modern All-Sky Surveys

From supernovae to stars with exoplanets, variable stars are of high interest in astronomical research.

Modern All-Sky Surveys

From supernovae to stars with exoplanets, variable stars are of high interest in astronomical research.

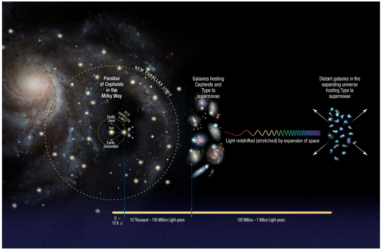
Research on variable stars **provides information about stellar properties**, such as mass, radius, luminosity, temperature, internal and external structure, composition, and evolution.

Modern All-Sky Surveys

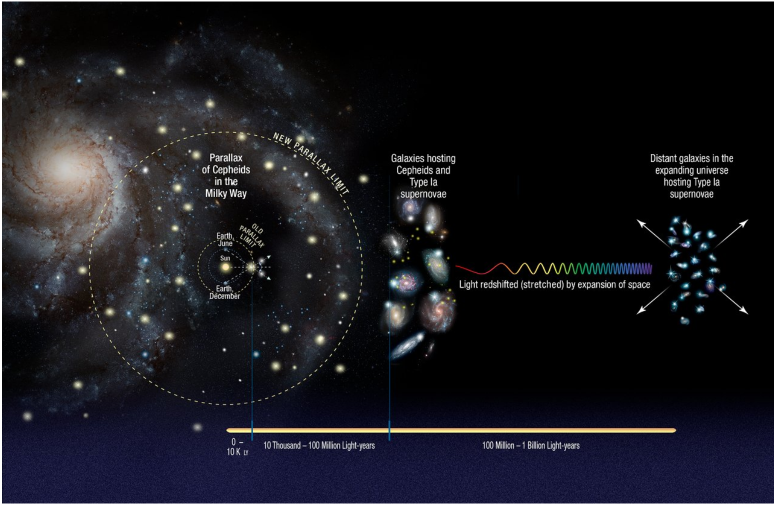
From supernovae to stars with exoplanets, variable stars are of high interest in astronomical research.

Research on variable stars **provides information about stellar properties**, such as mass, radius, luminosity, temperature, internal and external structure, composition, and evolution.

In addition, variable stars provide **distance information** (keyword: *distance ladder*) in our galactic neighborhood.



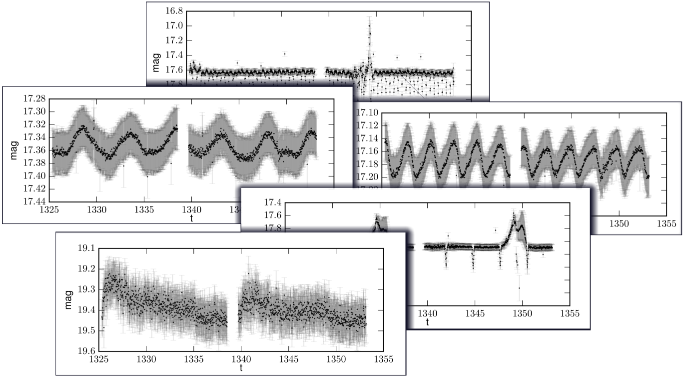
Modern All-Sky Surveys



Credit: NASA, ESA, A. Feild (STScI), and A. Riess (STScI/JHU)

Variable Stars

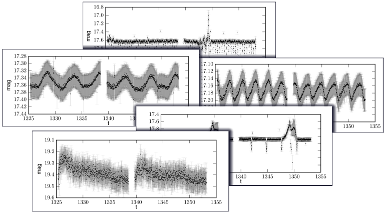
Variable stars are stars showing a change in brightness.



A selection of variable star light curves from the TESS survey.

Variable Stars

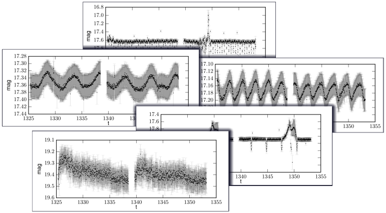
Variable stars are stars showing a change in brightness.



few parts per million — **change in luminosity** — factor 1000

Variable Stars

Variable stars are stars showing a change in brightness.

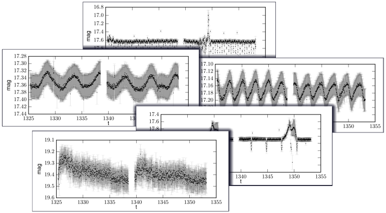


few parts per million — **change in luminosity** — factor 1000

seconds — **temporal baseline** — centuries

Variable Stars

Variable stars are stars showing a change in brightness.



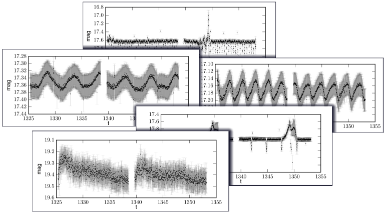
few parts per million — **change in luminosity** — factor 1000

seconds — **temporal baseline** — centuries

periodic — **signal shape** — aperiodic/ random

Variable Stars

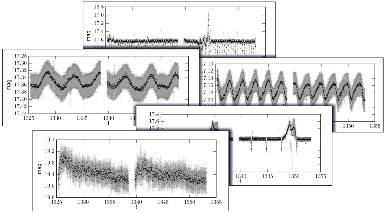
Variable stars are stars showing a change in brightness.



variations provide important and often unique information about the **nature and evolution of stars**

Variable Stars

Variable stars are stars showing a change in brightness.

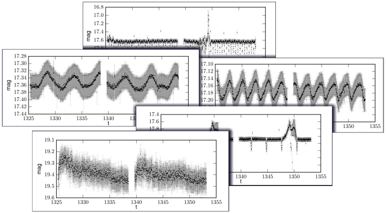


variations provide important and often unique information about the **nature and evolution of stars**

and the **galaxies** that host them

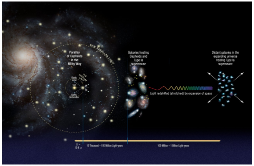
Variable Stars

Variable stars are stars showing a change in brightness.



variations provide important and often unique information about the **nature and evolution of stars**

and the **galaxies** that host them



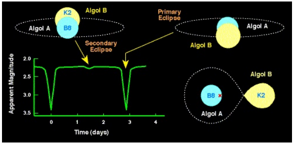
and our **universe** in general

Variable Stars

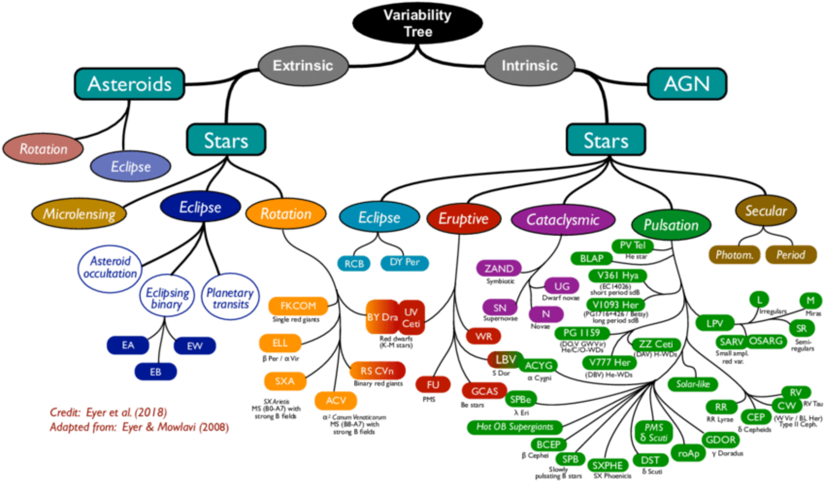
Variable stars are stars showing a change in brightness.

Intrinsic Variables
Stars whose energy output actually varies (pulsating stars, erupting or explosive stars)

Extrinsic Variables
Stars that only appear to vary due to geometric/ external effects (eclipses in binary systems, etc.)

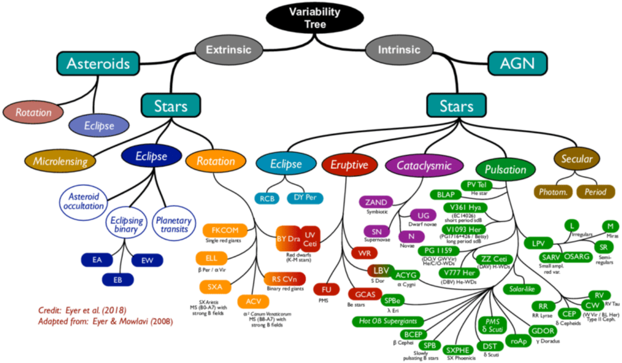


Variable Stars



Credit: Eyer et al. (2018)
Adapted from: Eyer & Mowlavi (2008)

Variable Stars



many astronomical sources vary - describe and classify astronomical sources by their variability

Discovery of Variable Stars

historically: discovered sporadically

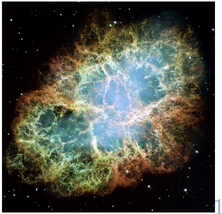
periodic variable stars:
1596: David Fabricius noted that the star o Ceti (now known as Mira) was sometimes visible, sometimes not
1638: Johannes Holwarda found a visibility cycle of 11 months for Mira
1700s: William Herschel discovered the variability of α Herculis and 44i Bootis

Discovery of Variable Stars

historically: discovered sporadically

periodic variable stars:
1596: David Fabricius noted that the star o Ceti (now known as Mira) was sometimes visible, sometimes not
1638: Johannes Holwarda found a visibility cycle of 11 months for Mira
1700s: William Herschel discovered the variability of α Herculis and 44i Bootis

supernovae:
oldest mention of a "new star": 185 AD a "guest star" was observed by Chinese astronomers
1054: supernova mentiones by Chinese astronomers
1572: Tycho's supernova
1604: Johannes Kepler's supernova



The Crab Nebula is a pulsar wind nebula associated with the 1054 supernova.

Discovery of Variable Stars

historically: discovered sporadically

1850s: $\sim$18 variable stars known
1890: establishment of the Variable Star Section of the British Astronomical Association (BAAVSS)
1911: founding of the American Association of Variable Star Observers (AAVSO)

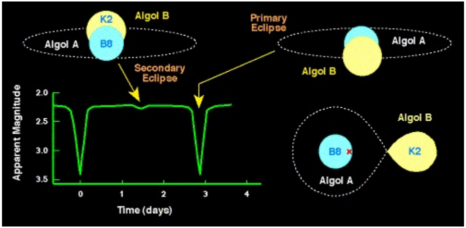
Discovery of Variable Stars

causes of variability?

John Goodricke and Edward Pigott: proposed the theory that Algol's variability might be caused by eclipses of the star by a planetary companion

we know today:

Algol is a three-star system, consisting of β Persei (Per) A, β Per B and β Per C. They regularly pass in front of each other, causing eclipses. This is an **eclipsing binary star**.



(Early) Classification of Variable Stars

systematic observation of variable stars reveiled **differences in their light curves**

Pigott (1780s): variable stars: nova, long-period variables, short-period variables

(Early) Classification of Variable Stars

systematic observation of variable stars reveiled **differences in their light curves**

Pigott (1780s): variable stars: nova, long-period variables, short-period variables

Pickering (1880s): a more detailed scheme:

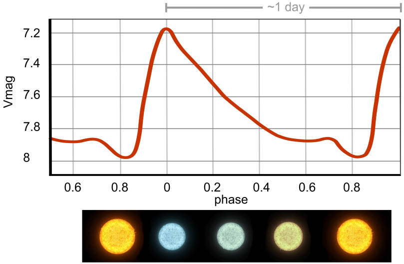
- (Ia) normal novae: now known to be nearby ones in our own galaxy;
- (Ib) novae in nebulae: now known to be supernovae in other galaxies;
- (IIa) long-period variables: cool, large-amplitude pulsating variables;
- (IIb) U Geminorum stars: dwarf novae;
- (IIc) R Coronae Borealis stars: stars which suddenly and unpredictably decline in brightness;
- (III) irregular variables: a motley collection;
- (IVa) short-period variables such as Cepheids and RR Lyrae stars;
- (IVb) Beta Lyrae type eclipsing variables; and
- (V) Algol type eclipsing variables.

Pulsating Variable Stars

underlying physics of variability:

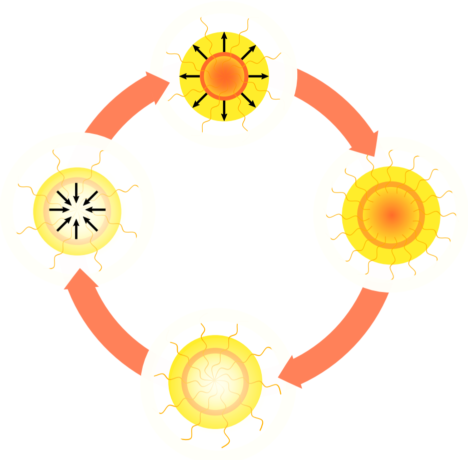
idea: (at least some) periodic variability might be caused by pulsations (A. Ritter 1873)

Observational studies by Harlow Shapley and others around 1915, and the concurrent theoretical studies by Eddington, established the pulsational nature of the Cepheids, cluster type variables (RR Lyrae stars), and long-period variables.



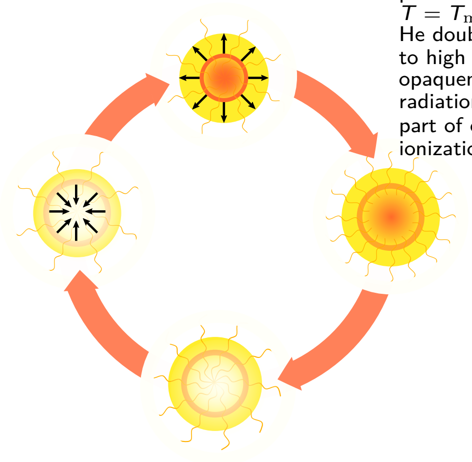
Cause of Pulsation

cause of pulsation: Lack of hydrostatic equilibrium beneath the surface drives the pulsation cycle with expansion and contraction of the outer layers of a star and subsequent change in brightness:



Cause of Pulsation

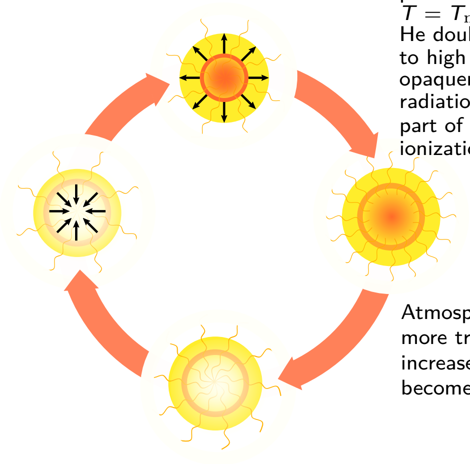
cause of pulsation: Lack of hydrostatic equilibrium beneath the surface drives the pulsation cycle with expansion and contraction of the outer layers of a star and subsequent change in brightness:



point of greatest compression: $T = T_{\max}$
He doubly ionized (HeIII) due to high T
opaqueness of HeIII causes radiation absorption (dimmest part of cycle) $\Rightarrow$ increase of ionization, T and pressure.

Cause of Pulsation

cause of pulsation: Lack of hydrostatic equilibrium beneath the surface drives the pulsation cycle with expansion and contraction of the outer layers of a star and subsequent change in brightness:



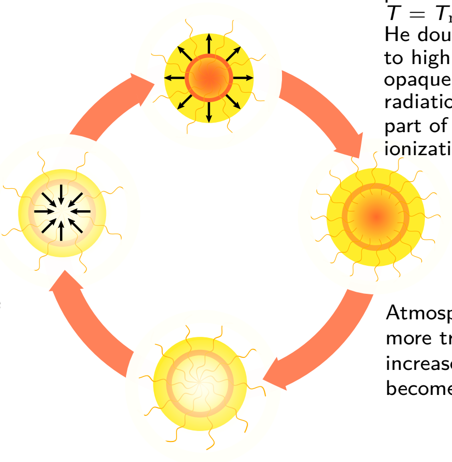
point of greatest compression: $T = T_{\max}$
He doubly ionized (HeIII) due to high T
opaqueness of HeIII causes radiation absorption (dimmest part of cycle) $\Rightarrow$ increase of ionization, T and pressure.

Atmosphere expands, becomes more transparent (brightness increases) and cools. HeIII becomes HeII.

Cause of Pulsation

cause of pulsation: Lack of hydrostatic equilibrium beneath the surface drives the pulsation cycle with expansion and contraction of the outer layers of a star and subsequent change in brightness:

point of greatest compression: $T = T_{\max}$
He doubly ionized (HeIII) due to high T
opaqueness of HeIII causes radiation absorption (dimmest part of cycle) $\Rightarrow$ increase of ionization, T and pressure.



Atmosphere expands, becomes more transparent (brightness increases) and cools. HeIII becomes HeII.

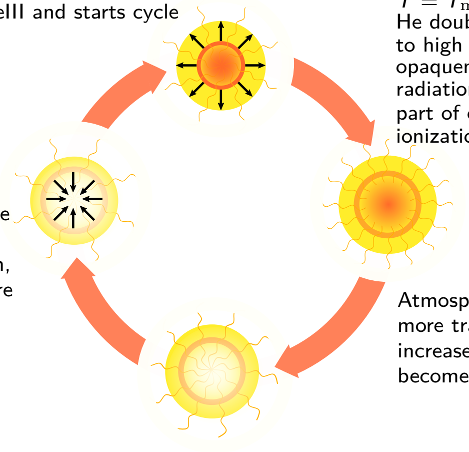
Atmosphere falls back, contraction, temperature rises.

Cause of Pulsation

cause of pulsation: Lack of hydrostatic equilibrium beneath the surface drives the pulsation cycle with expansion and contraction of the outer layers of a star and subsequent change in brightness:

Before it reaches equilibrium, HeII becomes HeIII and starts cycle over again.

point of greatest compression: $T = T_{\max}$
He doubly ionized (HeIII) due to high T
opaqueness of HeIII causes radiation absorption (dimmest part of cycle) $\Rightarrow$ increase of ionization, T and pressure.



Atmosphere falls back, contraction, temperature rises.

Atmosphere expands, becomes more transparent (brightness increases) and cools. HeIII becomes HeII.

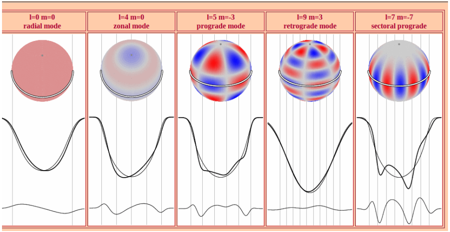
Cause of Pulsation

This is called **radial mode** pulsation. It is found in large-amplitude pulsating variables in the HR-diagram *instability strip*: Cepheids, Miras and RR Lyrae stars.

Cause of Pulsation

This is called **radial mode** pulsation. It is found in large-amplitude pulsating variables in the HR-diagram *instability strip*: Cepheids, Miras and RR Lyrae stars.

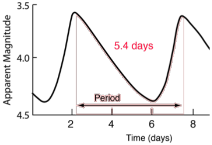
There are stars whose pulsation is **non-radial**: a star changes shape, but not volume. Non-radial pulsation leads to smaller amplitudes of variation. Some stars – β Cephei, δ Scuti stars and to a small amount also RR Lyrae stars – pulsate in both radial and non-radial modes.



models of stars with non-radial pulsations (copyright Coen Schrijvers)
http://staff.not.iac.es/ jht/science/

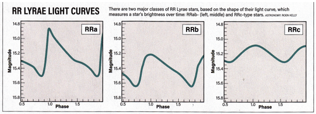
Types of Pulsating Variable Stars

we are looking for **bright, strictly periodic*** stars
*caveat: this condition cannot always be met



Cepheids

- brightness enables us to observe them in other galaxies in our Local Group (such as the Magellanic Clouds, M31 and M33)
- period-luminosity relation makes them important standard candles ⇒ distance ladder



RR Lyrae stars

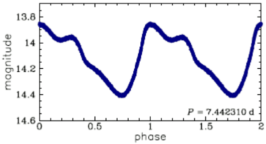
- numerous in the Milky Way halo (globular clusters), thus once called *cluster variables*
- less bright than Cepheids
- period-luminosity relation and their age makes them important tracers of the old Milky Way halo substructure

Cepheids

Classical (Type I) Cepheids

bright yellow, highly luminous, supergiant pulsating variables

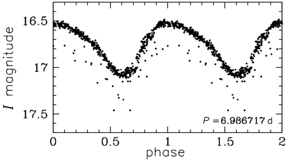
- amplitudes: ~0.01 - 2 mag_V
- periods: 1 - 135 days
- variability is strictly regular
- spectral type: F at maximum light, G to K at minimum light; the longer the period, the later the spectral type



Population II (Type II) Cepheids

similar light curve than Type I, but different evolutional history

- older, low mass stars
- important fossils of the first generation of stars in our galaxy



Soszyński et al. (2018), OGLE data

RR Lyrae stars

prototype: RR Lyrae (variability discovered by Williamina Fleming, $\sim$ 1900); RR Lyrae stars are old helium-burning variable stars of spectral type A5 to F5 with 0.5 $M_\odot$

RRab
asymmetrical light curves with steep ascend

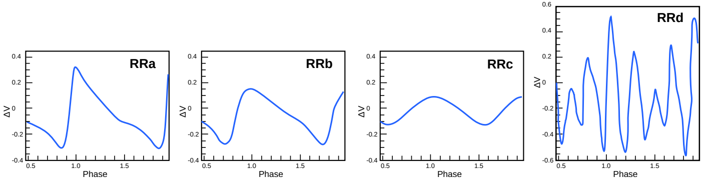
- periods: 0.3 - 1.2 days
- amplitudes: 0.5 - 2 mag_V

RRc
nearly symmetrical light curves

- periods: 0.2 - 0.5 days
- amplitudes: < 0.8 mag_V

RRd
double-mode RR Lyrae stars, fundamental and first overtone

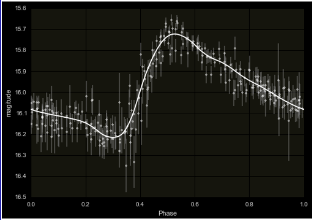
- fundamental period: 0.5 days
- period ratio: 0.74 days



Periodic Variable Stars

periodic variable stars allow for distance calculation:

Cepheid and RR Lyrae stars are variable stars with the period being directly related to their true (absolute) brightness.

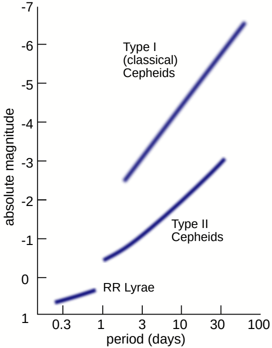


- measure apparent mean brightness m
- measure period P

Periodic Variable Stars

periodic variable stars allow for distance calculation:

Cepheid and RR Lyrae stars are variable stars with the period being directly related to their true (absolute) brightness.



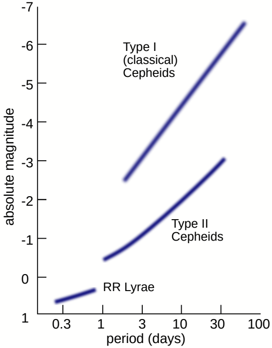
- measure apparent mean brightness m
- measure period P
- using period-luminosity relation, get absolute brightness M
- solve for distance using equation $d = 10^{(m-M+5)}/5$ parsec

  where 1 parsec $= 3.086^{16}$ m $= 3.26156$ light years

Periodic Variable Stars

periodic variable stars allow for distance calculation:

Cepheid and RR Lyrae stars are variable stars with the period being directly related to their true (absolute) brightness.



- measure apparent mean brightness m
- measure period P
- using period-luminosity relation, get absolute brightness M
- solve for distance using equation $d = 10^{(m-M+5)}/5$ parsec

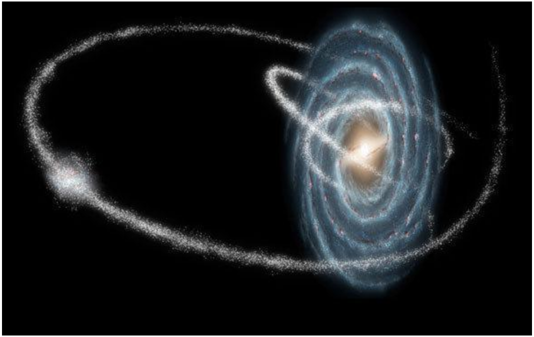
  where 1 parsec $= 3.086^{16}$ m $= 3.26156$ light years

$\Rightarrow$ creating *3D maps* of structures in our Milky Way

Periodic Variable Stars

RR Lyrae stars as tracers for old Milky Way substructure:

- old: $\sim 10^9$ years
- high-precision 3D mapping of the (old) Milky Way



artistic image, www.spitzer.caltech.edu

RR Lyrae stars

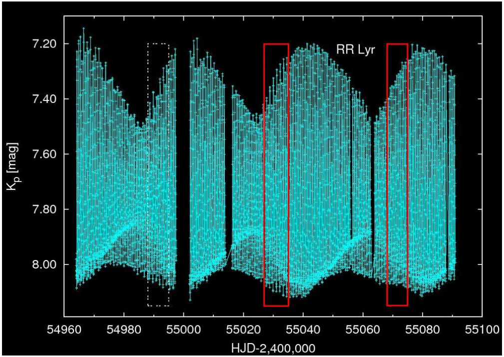
discovery by Harlow Shapley and Richard Prager (1916) independently:

RR Lyrae (the prototype's) light curve is modulated in amplitude and shape

RR Lyrae stars

discovery by Harlow Shapley and Richard Prager (1916) independently:

RR Lyrae (the prototype's) light curve is modulated in amplitude and shape



observation of a RR Lyrae star with Blazhko effect from the Kepler survey

RR Lyrae stars

⇨ period-amplitude-shape modulation is today known as the **Blazhko effect**
its explanation remains one of the enduring mysteries in astrophysics to this day

RR Lyrae stars

⇨ period-amplitude-shape modulation is today known as the **Blazhko effect**
its explanation remains one of the enduring mysteries in astrophysics to this day

two promising **theories** for explaining the Blazhko effect:

(i) resonance between the radial fundamental period of pulsation, and a non-radial period; or

(ii) a deformation or splitting of the radial period by a magnetic field in the star

Pulsating Stars as Distance Estimators

The **distance modulus** equation alone is not enough:

$$d = 10^{(m - M + 5)}/5 \text{ parsec}$$

Pulsating Stars as Distance Estimators

The **distance modulus** equation alone is not enough:

$$d = 10^{(m - M + 5)}/5 \text{ parsec}$$

?

Pulsating stars are a powerful tool for determining distances in astronomy, because the period of pulsation is correlated with the luminosity of the star, and this relation can be calibrated

⇨ the **period-luminosity(-metallicity) relation**

Pulsating Stars as Distance Estimators

The best-known relation between period and absolute magnitude is the direct proportionality law for **Classical Cepheid variables** (Henrietta Swan Leavitt (1908)).

foundation for scaling **galactic and extragalactic distances**

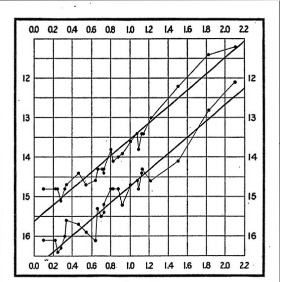


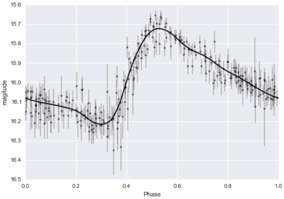
Fig. 2.

Plot from Leavitt's 1912 paper. The horizontal axis is the logarithm of the Cepheid's period, and the vertical axis is its apparent magnitude.

Pulsating Stars as Distance Estimators

Cepheid and RR Lyrae stars are variable stars with the period being directly related to their true (absolute) brightness.

basic concept:

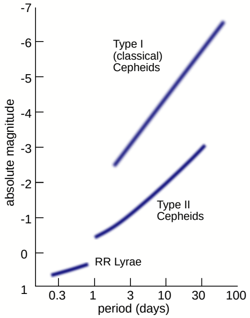


- measure apparent mean brightness m
- measure period P

Pulsating Stars as Distance Estimators

Cepheid and RR Lyrae stars are variable stars with the period being directly related to their true (absolute) brightness.

basic concept:

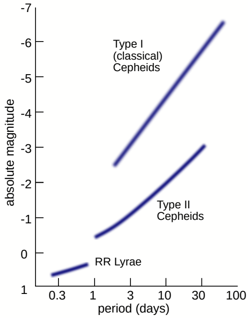


- measure apparent mean brightness m
- measure period P
- using **period-luminosity relation**, get absolute brightness M
- solve for distance using **distance modulus** equation $d = 10^{(m-M+5)}/5$ parsec where 1 parsec $= 3.086^{16}$ m $= 3.26156$ lyr

Pulsating Stars as Distance Estimators

Cepheid and RR Lyrae stars are variable stars with the period being directly related to their true (absolute) brightness.

basic concept:



- measure apparent mean brightness m
- measure period P
- using **period-luminosity relation**, get absolute brightness M
- solve for distance using **distance modulus** equation
  $d = 10^{(m-M+5)}/5$ parsec
  where 1 parsec $= 3.086^{16}$ m $= 3.26156$ lyr

$\Rightarrow$ allow us to create *3D maps* of structures within and beyond our Milky Way

The Period-Luminosity(-Metallicity) Relation

Globular clusters have only little depth - we can treat all the stars in a cluster as being at $\sim$ the same distance from Earth

color-magnitude diagram of stars in a globular cluster (M3):

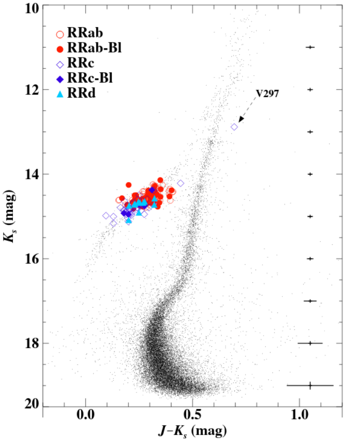
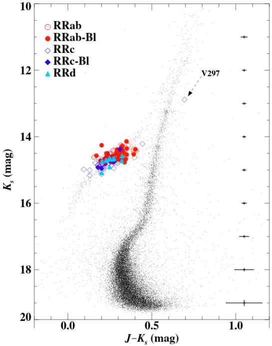


Figure 3 taken from Bhardwaj et al., AJ 160, 220 (2020)

The Period-Luminosity(-Metallicity) Relation

Globular clusters have only little depth - we can treat all the stars in a cluster as being at $\sim$ the same distance from Earth

color-magnitude diagram of stars in a globular cluster (M3):



all RR Lyrae stars have $\sim$ the same apparent magnitude

$\Rightarrow$ as the distance must be $\sim$ the same, they also have the same absolute magnitude

once we know the value of that absolute magnitude, we can compute the distance to each star from the distance modulus

The Period-Luminosity(-Metallicity) Relation

A closer look:

for each RR Lyrae star in the cluster, plot the apparent magnitude as function of its period

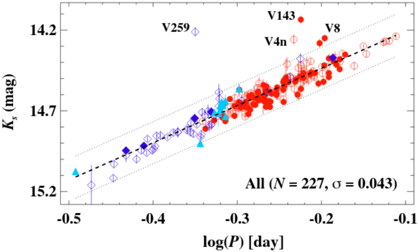


Figure 10 (slightly modified) taken from Bhardwaj et al., AJ 160, 220 (2020)

slight **trend**: stars with longer periods are a bit brighter

The Period-Luminosity(-Metallicity) Relation

To put everything together:

1. we know the **distance modulus** equation:
$d = 10^{(m-M+5)}/5$ parsec

The Period-Luminosity(-Metallicity) Relation

To put everything together:

1. we know the **distance modulus** equation:
$d = 10^{(m-M+5)}/5$ parsec

2. stars at approximately the same distance show a slight **trend**: stars with longer periods are a bit brighter

The Period-Luminosity(-Metallicity) Relation

To put everything together:

1. we know the **distance modulus** equation:
$d = 10^{(m-M+5)}/5$ parsec

2. stars at approximately the same distance show a slight **trend**: stars with longer periods are a bit brighter

(3. There is also a small trend on metallicity Z.)

The Period-Luminosity(-Metallicity) Relation

To put everything together:

1. we know the **distance modulus** equation:
$d = 10^{(m-M+5)}/5$ parsec

2. stars at approximately the same distance show a slight **trend**: stars with longer periods are a bit brighter

(3. There is also a small trend on metallicity Z.)

metallicity is the abundance of elements present in star that are heavier than hydrogen and helium

? For $d(m, P, Z)$, we need to **calibrate** the Period-Luminosity(-Metallicity) Relation.

The Period-Luminosity(-Metallicity) Relation

calibrate the Period-Luminosity(-Metallicity) Relation:

The following methods can be used to determine absolute magnitudes, e.g.:

- Statistical study of the motions of field RR Lyrae stars: statistical parallax. This gives values of M_V ranging from $+0.9$ for short-period, high-metallicity stars, to $+0.5$ for longer-period, lower-metallicity stars. As a statistical method, it must be applied to a large sample of stars, which might not be homogeneous.

The Period-Luminosity(-Metallicity) Relation

calibrate the Period-Luminosity(-Metallicity) Relation:

The following methods can be used to determine absolute magnitudes, e.g.:

- Statistical study of the motions of field RR Lyrae stars: statistical parallax. This gives values of M_V ranging from $+0.9$ for short-period, high-metallicity stars, to $+0.5$ for longer-period, lower-metallicity stars. As a statistical method, it must be applied to a large sample of stars, which might not be homogeneous.
- Fitting of the main sequence of globular clusters containing RR Lyrae stars to a standard main sequence determined for nearby Population II stars with known distances; there are, however, very few of these. This method gives a mean of about $+0.4$ for the RR Lyrae stars in several clusters.

The Period-Luminosity(-Metallicity) Relation

calibrate the Period-Luminosity(-Metallicity) Relation:

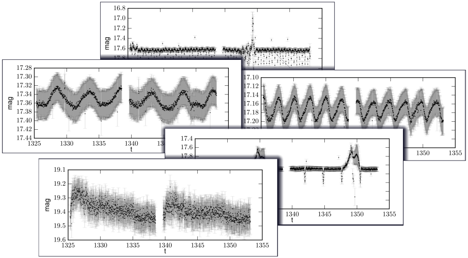
The following methods can be used to determine absolute magnitudes, e.g.:

- Statistical study of the motions of field RR Lyrae stars: statistical parallax. This gives values of M_V ranging from $+0.9$ for short-period, high-metallicity stars, to $+0.5$ for longer-period, lower-metallicity stars. As a statistical method, it must be applied to a large sample of stars, which might not be homogeneous.
- Fitting of the main sequence of globular clusters containing RR Lyrae stars to a standard main sequence determined for nearby Population II stars with known distances; there are, however, very few of these. This method gives a mean of about $+0.4$ for the RR Lyrae stars in several clusters.
- The Baade-Wesselink method (infer distance from measurement of change in radius (from velocity) and angular diameter) has been applied to some of the brightest RR Lyrae stars; it gives an absolute magnitude of about $+0.5$.

Time Series Analysis

Variable stars can be classified by their **time series**:

Different types of variable stars vary characteristically.

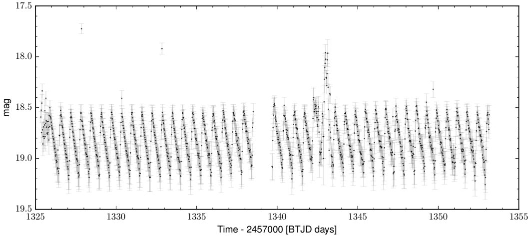


Transforming light curves into **features** will enable automatic classification of variable sources:

calculating features such as colors, amplitudes, periods...

example light curves with a high **cadence** ($\Delta t = 30$ min) from TESS:

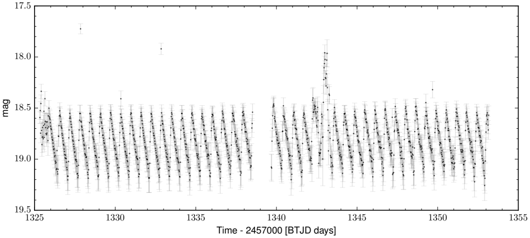
RRab (RR Lyrae type ab):

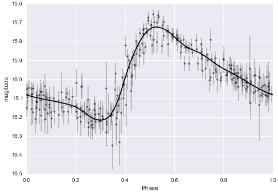


Astronomical Light Curves

example light curves with a high **cadence** ($\Delta t = 30$ min) from TESS:

RRab (RR Lyrae type ab):

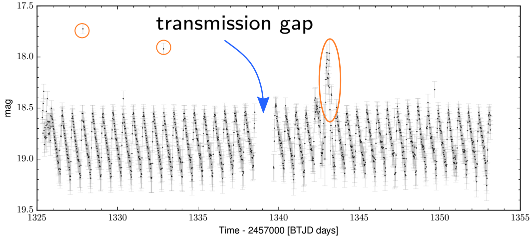


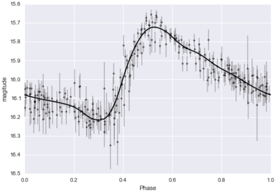


Astronomical Light Curves

example light curves with a high **cadence** ($\Delta t = 30$ min) from TESS:

RRab (RR Lyrae type ab):

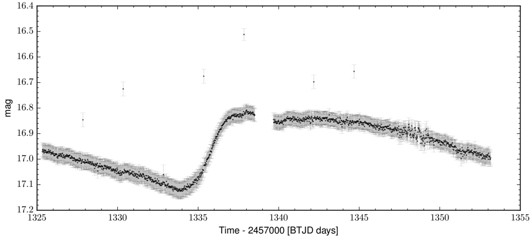




Astronomical Light Curves

example light curves with a high **cadence** ($\Delta t = 30$ min) from TESS:

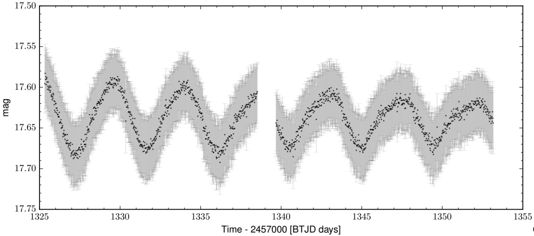
Cepheid:

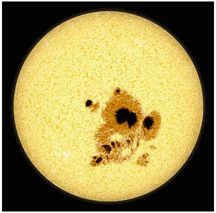


Astronomical Light Curves

example light curves with a high **cadence** ($\Delta t = 30$ min) from TESS:

rotational variable star:



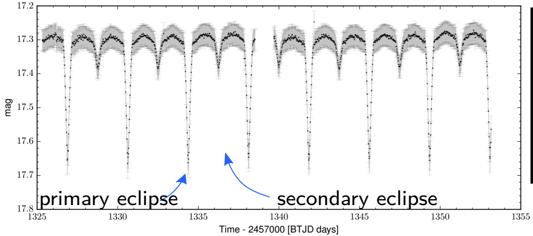


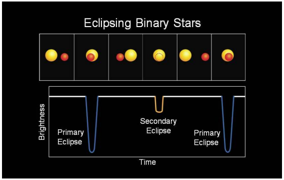
credit: Observer's Guide to Variable Stars, M. Griffiths

Astronomical Light Curves

example light curves with a high **cadence** ($\Delta t = 30$ min) from TESS:

eclipsing binary star:





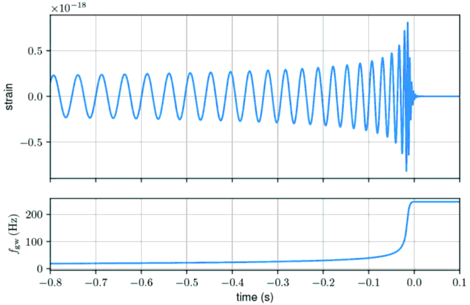
credit: Wikimedia, NASA

When the smaller star partially blocks the larger star, a primary eclipse occurs, and a secondary eclipse occurs when the smaller star is occulted.

Other Astronomical Time Series Data

Light curves show variability from electromagnetic sources.
In addition: gravitational-wave variability as time series data

Gravitational Wave Signal:



Typical GW signal of a compact binary coalescence. The GW strain (above) and the GW frequency (below) are plotted as function of the time before merging. credit: Vallisneri et al. (2015).

Time Series Data

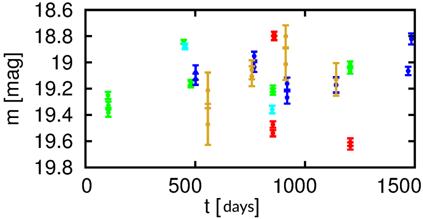
A time series is a sequence of random variables $\{\mathbf{X}_t\}_{t=1,2,\cdots}$.

Thus, a time series is a **series of data points ordered in time**. The time of observations provides a source of additional information to be analyzed.

Astronomical time series are typically assumed to be generated at irregularly spaced interval of time (**irregular time series**).

Time series can have one or more variables that change over time. If there is only one variable varying over time, we call it **univariate time series**. If there is more than one variable it is called **multivariate time series**.

example: light curve from multi-band survey

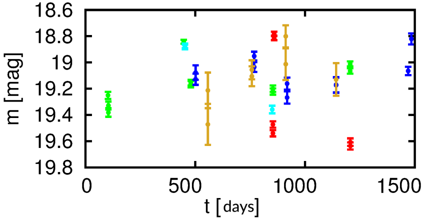


Characteristics of Astronomical Time Series Data

Astronomical time series data in general is:

- irregularly sampled
- multivariate
- not sampled to fully characterize the variability process
- not an independent random variable in their y values: often $y_{i+1} = f(y_i)$

example: light curve from multi-band survey



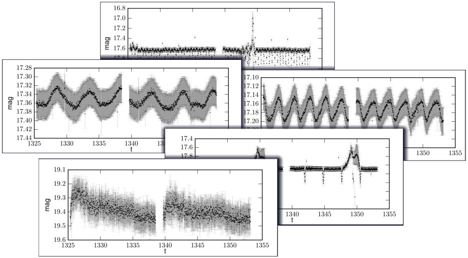
Goals of Time Series Analysis

Time series analysis extracts meaningful statistics and other characteristics of the dataset in order to understand it.

The main tasks of time series analysis are:

- **characterize** the temporal correlation between different values of y, including its significance

  example: classification of variable sources



- **forecast** (predict) future values of y

  example: transient detection, e.g. early supernovae detection

Goals of Time Series Analysis

When dealing with time series data, the first question we ask is ***Does the time series vary over some timescale?*** (if not, there is no point doing time series analysis)

Variability does not mean necessarily periodicity.
Stochastic processes are variable over some timescale, but are distinctly aperiodic through the inherent randomness.

Goals of Time Series Analysis

When dealing with time series data, the first question we ask is ***Does the time series vary over some timescale?*** (if not, there is no point doing time series analysis)

Variability does not mean necessarily periodicity.
Stochastic processes are variable over some timescale, but are distinctly aperiodic through the inherent randomness.

If we find that a source is variable (almost all astronomical sources are), then time-series analysis has two main goals:

1. Characterize the temporal correlation between different values of y (i.e., characterize the light curve), e.g. by learning the parameters for a model.
2. Predict future values of y.

Detecting Variability

For known and Gaussian uncertainties, we can compute χ^2 and the corresponding p values for variation in a signal.
For a sinusoidal variable signal $A\sin(\omega t)$, with homoescedastic measurement uncertainties, the data model would be

$$y(t) = A\sin(\omega t) + \epsilon$$

where $\epsilon \sim N(0, \sigma)$. The overall data variance is then $V = \sigma^2 + A^2/2$.

Detecting Variability

For known and Gaussian uncertainties, we can compute χ^2 and the corresponding p values for variation in a signal.
For a sinusoidal variable signal $A \sin(\omega t)$, with homoescedastic measurement uncertainties, the data model would be

$$y(t) = A \sin(\omega t) + \epsilon$$

where $\epsilon \sim N(0, \sigma)$. The overall data variance is then $V = \sigma^2 + A^2/2$.

If $A = 0$ (no variability, with $\overline{y} = 0$):

- $\chi^2_{\rm dof} = N^{-1} \sum_j (y_j/\sigma)^2 \sim V/\sigma^2$
- $\chi^2_{\rm dof}$ has expectation value of 1 and std dev of $\sqrt{2/N}$

Detecting Variability

If $|A| > 0$ (variability):

- $\chi^2_{\rm dof}$ will be larger than 1.
- probability that $\chi^2_{\rm dof} > 1 + 3\sqrt{2/N}$ is about 1 in 1000 (i.e., $> 3\sigma$ above 1, where 3σ is 0.997).

Detecting Variability

If $|A| > 0$ (variability):

- $\chi^2_{\rm dof}$ will be larger than 1.
- probability that $\chi^2_{\rm dof} > 1 + 3\sqrt{2/N}$ is about 1 in 1000 (i.e., $> 3\sigma$ above 1, where 3σ is 0.997).

If this false-positive rate (1 in a 1000) is acceptable (because even without variability 1 in 1000 will be above this threshold) then the minimum detectable amplitude is $A > 2.9\sigma/N^{1/4}$ (from $V/\sigma^2 = 1 + 3\sqrt{2/N}$, so that $A^2/2\sigma^2 = 3\sqrt{2/N}$).

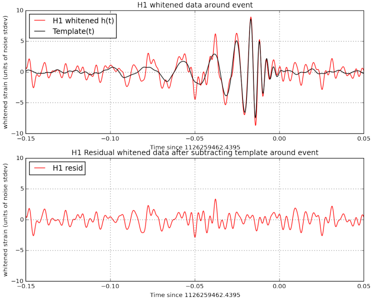
Depending on how big your sample is, you may want to choose a higher threshold. E.g., for 1 million non-variable stars, this criterion would identify 100 as variable.

1. For $N = 100$ data points (not 100 objects), the minimum detectable amplitude is $A_{\rm min} = 0.92\sigma$
2. For $N = 1000$, $A_{\rm min} = 0.52\sigma$

Detecting Variability

We do this under the assumption of the null hypothesis of no variability. If instead we have a model, we can perform a **matched filter analysis** by correlating a known template with an unknown signal to detect the presence of the template in the unknown signal

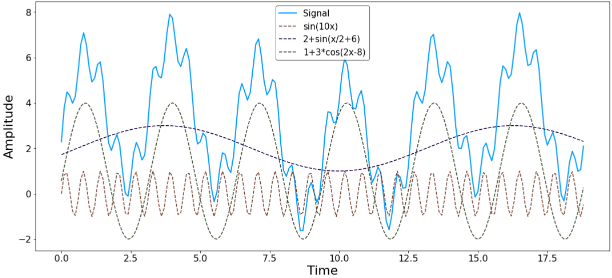
example: gravitational wave event GW150914



credit: https://www.gw-openscience.org/tutorials/

Fourier Analysis

Fourier analysis plays a **major role** in the analysis of time series data. In Fourier analysis, general **functions are approximated by integrals or sums of trigonometric functions**.



For periodic functions, such as periodic light curves in astronomy, often a relatively small number of terms (less than 10) suffices to reach an approximation precision level similar to the measurement precision.

Fourier Analysis

The **Fourier transform (FT)** $H(f)$ of function $h(t)$ is defined as

$$H(f) = \int_{-\infty}^{\infty} h(t) \exp(-i2\pi ft) dt$$

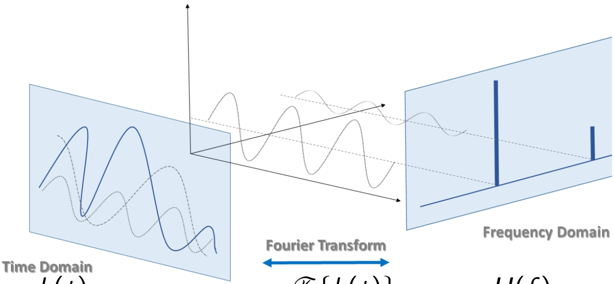
with inverse transformation

$$h(t) = \int_{-\infty}^{\infty} H(f) \exp(-i2\pi ft) df$$

where t is time and f is frequency (for time in seconds, the unit for frequency is hertz, or Hz).

Fourier Analysis

In other words, FT transforms a periodic function in **Time Domain** to a function in **Frequency Domain**:



$h(t)$ $\mathscr{F}\{h(t)\}$ $H(f)$

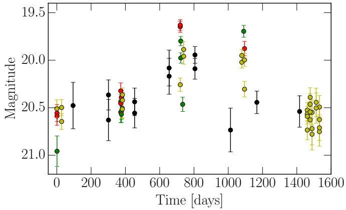
Detecting Periodic Signals

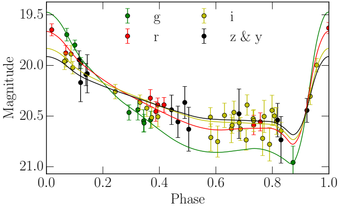
many objects/ systems have periodic signals: e.g., pulsars, RR Lyrae, Cepheids, eclipsing binaries

For a periodic signal, if the period is known

- we can write $y(t+P) = y(t)$, where P is the period.
- we can create a **phased light curve** that plots the data as function of phase: $\phi = \frac{t}{P} - \mathrm{int}\left(\frac{t}{P}\right)$ with $\mathrm{int}(x)$ being the integer part of x.

phase folding

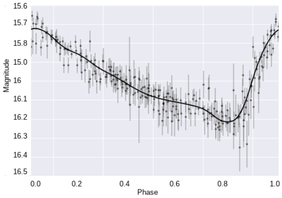




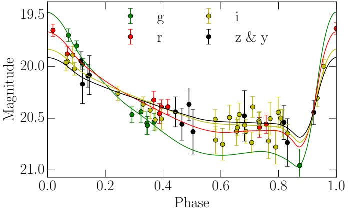
Detecting Periodic Signals

for well-sampled, high-cadence data: easy, standard methods can be applied

for sparse, low-cadence data: harder, specialized methods like template fitting necessary



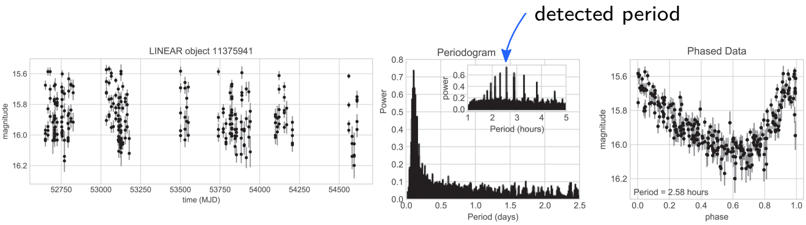
vs.



measure the period and amplitude in the face of both noisy and incomplete data

Detecting Periodic Signals

A **periodogram** is a plot of the *power* in the time series at each possible period (as illustrated below):



left panel: observed light curve from LINEAR object ID 11375941
middle panel: periodogram computed from the light curve
right panel: light curve folded over the detected 2.58 hr period
credit: VanderPlas (2018)

Detecting Periodic Signals

The periodogram is defined as

$$P(\omega) = \frac{1}{N}\left[\left(\sum_{j=1}^{N} y_j \sin(\omega t_j)\right)^2 + \left(\sum_{j=1}^{N} y_j \cos(\omega t_j)\right)^2\right]$$

The **best value** ω is given by

$$\chi^2(\omega) = \chi_0^2 \left[1 - \frac{2}{N\,V}\,P(\omega)\right],$$

where $P(\omega)$ is the periodogram, V the variance of the data y, and χ_0^2 is the χ^2 for the null-hypothesis model $y(t) = \text{const}$:

$$\chi_0^2 = \frac{1}{\sigma^2}\sum_{j=1}^{N} y_j^2 = \frac{N\,V}{\sigma^2}$$

Detecting Periodic Signals

We can renormalize the periodogram, defining the **Lomb-Scargle periodogram** as

$$P_{\mathrm{LS}}(\omega) = \frac{2}{NV} P(\omega),$$

where $0 \leq P_{\mathrm{LS}}(\omega) \leq 1$.

Detecting Periodic Signals

How to determine if our source is variable or not:

- compute Lomb-Scargle periodogram $P_{\mathrm{LS}}(\omega)$
- model the odds ratio for our variability model vs. a non-variability model.

If our variability model is correct, then the **peak** of $P(\omega)$ (found by grid search) gives the best period ω.

⇨ The Lomb-Scargle periodogram (Lomb 1976; Scargle 1982) is the **standard method** to search for periodicity in unevenly-sampled time-series data.

Outlook

Detected **features** are what classification relies on.

?

How does classification happen?

Data products?

Science cases?

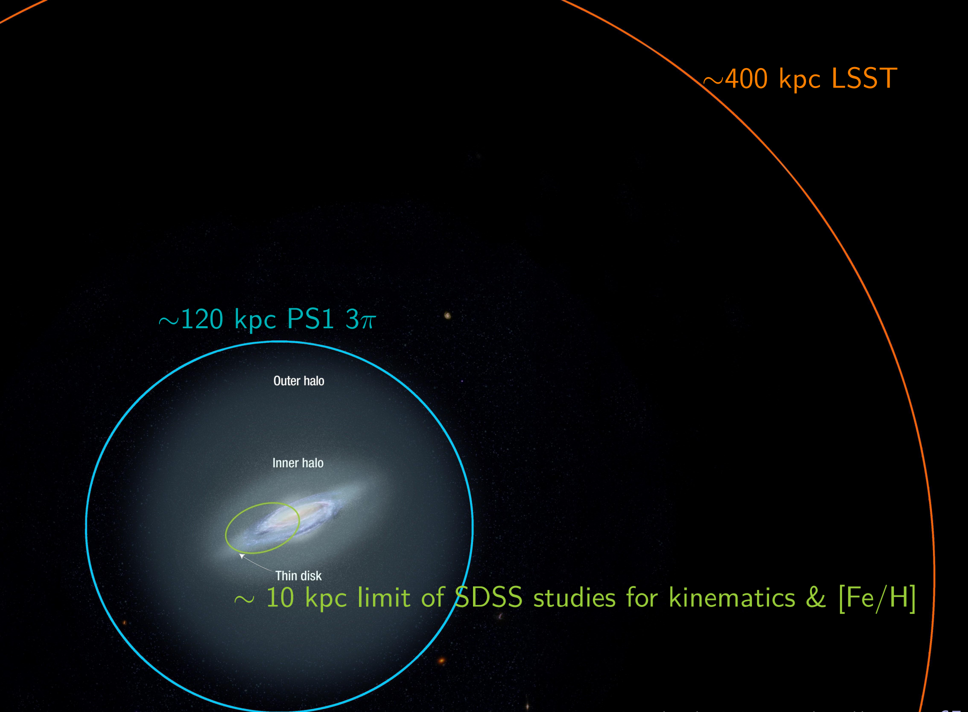


image based on NASA, ESA, and A. Feild (STScI)