Lecciones en Astrofisica Avanzada (Semester 1 2024)

# Mapping the Universe with Variable Stars (III)

**Nina Hernitschek**
Centro de Astronomía CITEVA
Universidad de Antofagasta

May 17, 2024

# Student Presentation



10 min presentation

Q&A

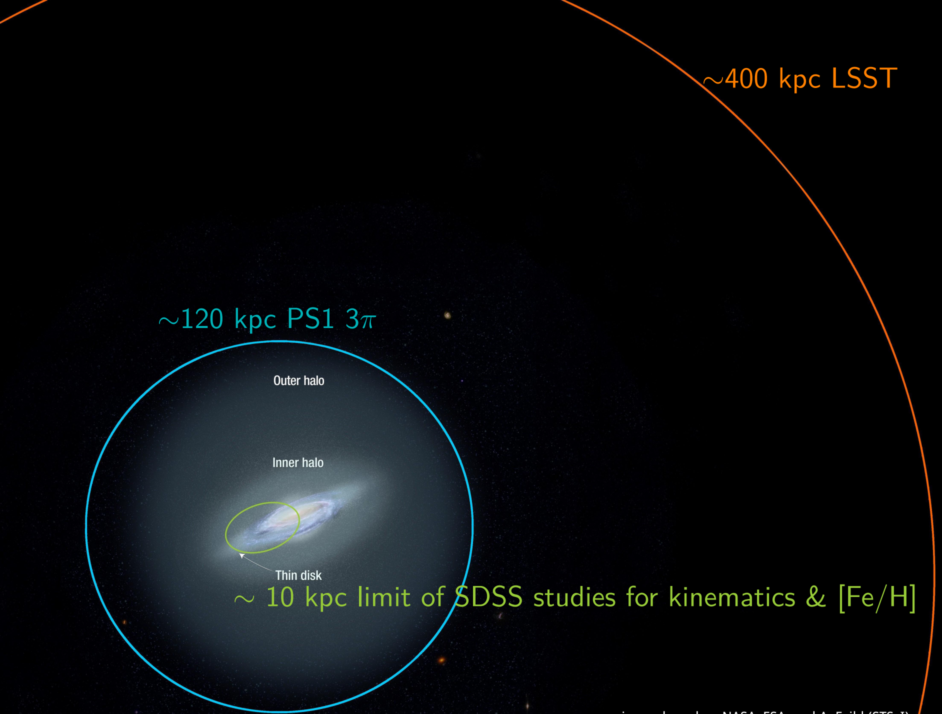


image based on NASA, ESA, and A. Feild (STScI)

# Motivation



**catalogs of variable sources from deep and wide all-sky surveys**

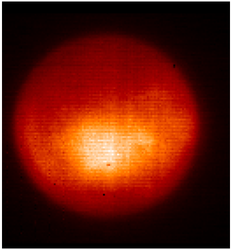
to **model a survey**, tools are needed for

- describing data quality $\rightarrow$ outlier might fake or hide true variability
- describing light curve characteristics $\rightarrow$ "features" with scientific relevance
- classifying sources $\rightarrow$ catalogs others can use
- finding substructure $\rightarrow$ clumps, overdensities, ... the science we want to do

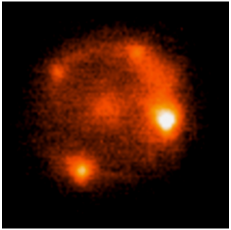
different telescopes study different things

some telescopes let scientists look at specific parts of the sky in high resolution to study the fine details of objects, or for a long time to collect more light to see fainter objects: **targeted observations**

example: Keck, Magellan



moon Europa from the Keck Observatory, credit: Mike Brown



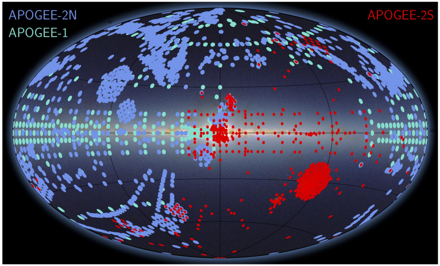
near-IR image of gravitationally lensed Type Ia SN, Keck Observatory

# Astronomical Surveys

different telescopes study different things

other telescopes let scientists study lots of objects in wide areas of the sky, but at lower resolution: **survey telescopes**

example: SDSS, Pan-STARRS, Rubin Observatory




APOGEE-2: A stellar spectroscopic survey of the Milky Way, composed of a northern survey with Apache Point Observatory (APOGEE-2N), and a southern survey with the 2.5m du Pont Telescope at Las Campanas (APOGEE-2S).

# LSST/ Rubin Observatory



**Survey telescopes**, like at Rubin Observatory, map the night sky by scanning and taking pictures of all parts of the sky instead of taking pictures of one specific object or set of objects.

# LSST/ Rubin Observatory



**Survey telescopes**, like at Rubin Observatory, map the night sky by scanning and taking pictures of all parts of the sky instead of taking pictures of one specific object or set of objects.

**Rubin Observatory** is a unique survey telescope

# LSST/ Rubin Observatory



**Survey telescopes**, like at Rubin Observatory, map the night sky by scanning and taking pictures of all parts of the sky instead of taking pictures of one specific object or set of objects.

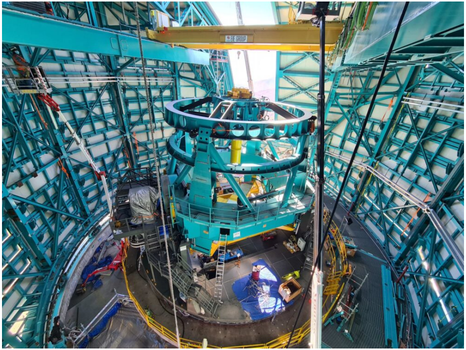
**Rubin Observatory** is a unique survey telescope

it is specially designed to:

- quickly take huge pictures of the entire Southern hemisphere sky
- repeat those pictures every few nights for ten years
- take those pictures in super-high detail while also being able to see very faint objects

# LSST/ Rubin Observatory

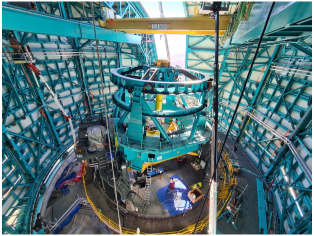
Rubin Observatory[1] will conduct the Legacy Survey of Space and Time (LSST)



credit: www.rubinobservatory.org

# LSST/ Rubin Observatory

Rubin Observatory[1] will conduct the Legacy Survey of Space and Time (LSST)



credit: www.rubinobservatory.org

HOW?

# LSST/ Rubin Observatory

? HOW?

Every night for ten years, Rubin Observatory will take hundreds of images of the Southern Hemisphere sky producing about 20 terabytes of data every night.
By the end of the survey, the resulting data set will be enormous: about 60 petabytes!

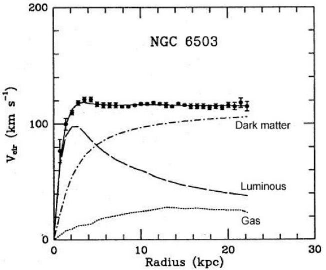
# LSST/ Rubin Observatory



the namesake:

Dr. Vera C. Rubin was an American astronomer who made essential contributions to the study of dark matter by recognizing that galaxy rotation curves show some "missing matter":

credit: Vassar College Library



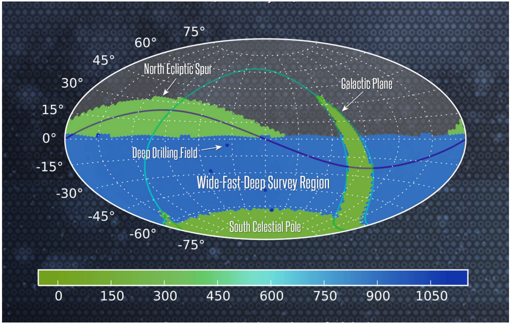

K.G. Begeman, A.H. Broels, R.H. Sanders. 1991. Mon.Not.RAS 249, 523.

stars at the outer edges move just as fast as those towards the center - high velocities caused by some invisible mass holding the galaxies together

# The LSST Survey



- 10-year photometric *ugrizy* survey (near-UV, optical, near-IR)
- depth of r$\sim$27.5 mag
- 1000 images/night = 15 TB/night, 10 million transients/night
- start of operations: 2024




LSST survey strategy, number of visits incl. sub-surveys. (credit: www.lsst.org)

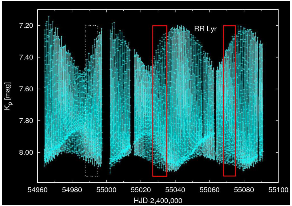
# Research Questions



large data volume of LSST (and other all-sky surveys) enables for

population studies
e.g.: larger samples of RR Lyrae stars to understand the Blazhko effect

finding rare 'one-in-a-million', 'one-in-a-billion' events, often called *anomalies*

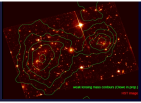



e.g.: extremely low mass (ELM) white dwarf
(El-Badry et al. 2021)

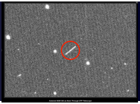
# Science with the LSST Survey

LSST is designed to address four science areas:

Probing Dark Energy and Dark Matter



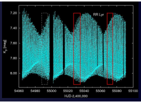
Cataloging the Solar System



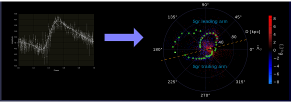
Exploring the Variable/ Transient Optical Sky

TVSSC

Transients and Variable Stars Science Collaboration



Mapping the Milky Way



# LSST/ Rubin Observatory

Rubin Observatory can accomplish all of these things because:


Installation of fiber optic cables on the telescope mount.
credit: www.rubinobservatory.org

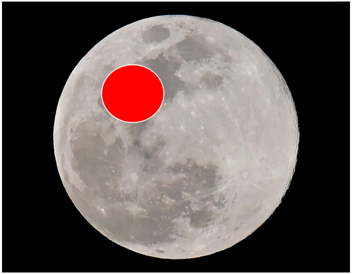
- Rubin Observatory's telescope can move much faster than other telescopes its size - it can take pictures faster and create a more detailed map of the night sky
- Rubin Observatory also has a big field of view - one picture covers the same area as 40 full moons
- Rubin Observatory's camera is the highest resolution camera ever created for astronomy and astrophysics
- Rubin Observatory has a big, 8.4 m main mirror that lets it collect a lot of light and see faint objects

# LSST/ Rubin Observatory



big telescopes have (usually) small Fields of View

most telescopes in the 8 - 10 m class image the red circle:



sky: 40,000 deg$^2$
moon: 0.2 deg$^2$
large telescopes: $\sim$0.01 deg$^2$

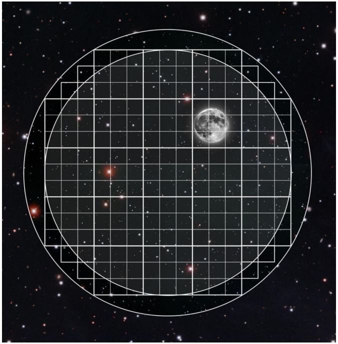
$\Rightarrow$ $\sim$4 million images to cover the sky

# LSST/ Rubin Observatory



big telescopes have (usually) small Fields of View

Rubin Observatory (VRO) FOV:



VRO: 9.6 $\text{deg}^2$

$\Rightarrow$ $\sim$4,300 images to cover the sky

$\Rightarrow$ image the whole sky once every 5 nights

# LSST/ Rubin Observatory



## Rubin Observatory Key Numbers[2] (excerpt)

**Telescope System:**

- FOV: 3.5 deg (9.6 deg$^2$)
- Primary mirror diameter: 8.4 m
- Mean effective aperture: 6.423 m
- Final f-ratio: f/1.234
- Etendue ($A\Omega$): 319 m$^2$deg$^2$
- Camera weight: 3060 kg

Etendue is a measure of the flux gathering capability of an optical system.
etendue = aperture [m$^2$] $\times$ FOV [deg$^2$]

[2] https://www.lsst.org/scientists/keynumbers

# LSST/ Rubin Observatory

## Rubin Observatory Key Numbers[2] (excerpt)

**Telescope System:**

- FOV: 3.5 deg (9.6 $\text{deg}^2$)
- Primary mirror diameter: 8.4 m
- Mean effective aperture: 6.423 m
- Final f-ratio: f/1.234
- Etendue ($A\Omega$): 319 $\text{m}^2\text{deg}^2$
- Camera weight: 3060 kg

Etendue is a measure of the flux gathering capability of an optical system.

etendue = aperture [$\text{m}^2$] $\times$ FOV [$\text{deg}^2$]

**Dataset:**

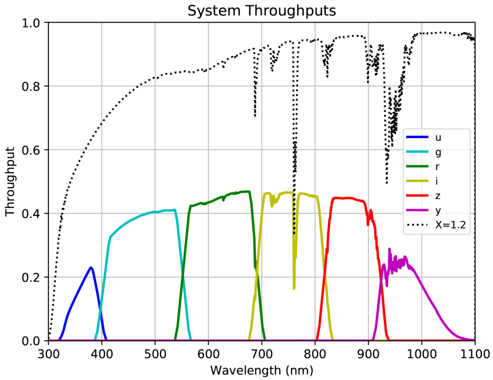
- Nightly data size: 20TB/night
- Final database size (DR11): 15 PB
- Real-time alert latency: 60 seconds

[2] https://www.lsst.org/scientists/keynumbers

# LSST/ Rubin Observatory

## Rubin Observatory Key Numbers[2] (excerpt)

**Spectral response/throughputs:**




[2]https://www.lsst.org/scientists/keynumbers

# LSST/ Rubin Observatory



## survey design

science cases lead to **competing constraints** on the LSST Survey Strategy:

e.g.:

**Cosmological parameter estimation** requires uniform coverage of 18,000 $\deg^2$. Obtaining accurate photometric redshifts requires a specified number of visits in each filter.

**Weak lensing shear measurements** benefit from allocating times of best seeing to observations in the $r$ and $i$ bands. Maximizing S/N requires choosing the next filter based upon the current sky background.

**Supernova** cosmology requires frequent, deep photometry in all bands.

Detecting the motion of **solar system objects** and transients, characterizing **stellar variability** on various timescales, and acquiring the best proper motions and parallaxes place further demands upon the distribution of revisit intervals and observation geometries to each point on the sky.

# LSST/ Rubin Observatory

## survey design

science cases lead to **competing constraints** on the LSST Survey Strategy:

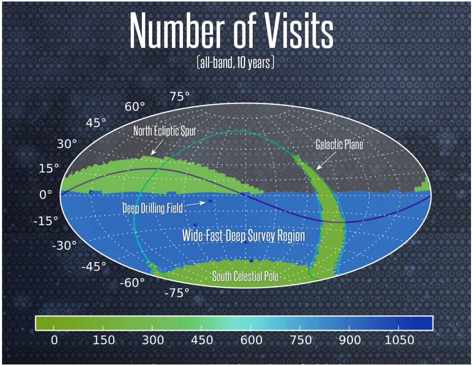
Synthesizing the requirements to accomplish the four primary science objectives of Rubin Observatory,

- Probing dark energy and dark matter
- Taking an inventory of the Solar System
- Exploring the transient optical sky
- Mapping the Milky Way

# LSST/ Rubin Observatory

## survey design

90% of time* will be spent on a uniform survey: every 3 - 4 nights, the whole observable sky will be scanned twice per night

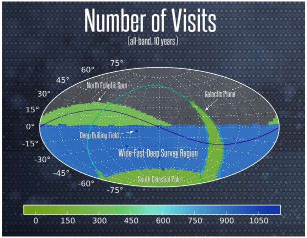



LSST survey strategy, number of visits incl. sub-surveys. (credit: www.lsst.org)

# LSST/ Rubin Observatory

## survey design

90% of time* will be spent on a uniform survey: every 3 - 4 nights, the whole observable sky will be scanned twice per night




The survey area and cadence have been (and will be) fine-tuned to support all four science themes and enable the discover and characterization of transient objects.

*A small (<10% of time) set of "special survey programs" is designed to explore extreme corners of discovery space: deep drilling fields, mini-surveys

# LSST/ Rubin Observatory



## survey design

These **fiducial survey plans** can be optimized for science output:

- Operations Simulations (OpSim)
- Image Simulations (ImSim)
- Base catalogs of stars and galaxies in LSST filters (CatSim)
- Key Project Documents (Science Requirements Document, Data Products Definition Document)

The **Operations Simulator (OpSim)** is an application that simulates the field selection and image acquisition process of the LSST over the 10-year life of the planned survey.

It has a sophisticated model of the telescope and dome to properly constrain potential observing cadences.

LSST operations can be simulated using realistic seeing distributions, historical weather data, scheduled engineering downtime and current telescope and camera parameters.

# LSST/ Rubin Observatory

## survey design

The community is participating in this process by evaluating the simulated surveys:

**Cadence Note: Cadence impacts on reliable classification of standard-candle variable stars, including detection of amplitude period, phase modulation effects (e.g., Blazhko effect)**

NINA HERNITSCHEK[1,*] AND KEIVAN G. STASSUN[1]

[1] *Vanderbilt University*

1. INTRODUCTION

The Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST) will carry out its science goal of "Mapping the Milky Way" through both astrometry and photometry, with a single-exposure depth of $r{\sim}24.7$ and an anticipated baseline of 10 years. This will enable LSST to access the Milky Way's old halo not only deeper, but also with a longer baseline and better cadence than e.g. PS1 $3\pi$ (Chambers et al. 2016), making this survey ideal to study populations of variable stars such as RR Lyrae (Hernitschek et al. 2016; Sesar et al. 2017a).

As members of the Transients and Variable Stars (TVS) Classification group, we focus on the specific science case of detecting period/ phase shift effects, so-called Blazhko effect (Blazhko 1907), of RR Lyrae stars. So far, due to depth and cadence of typical all-sky surveys, it was nearly impossible to study this effect on a larger sample. Surveys such as PS1 $3\pi$ with relatively few observations over a moderately long baseline allowed only for fitting the period and phase of RR Lyrae stars while integrating over the complete survey length, thus not giving any information regarding whether the period and/or phase of the light curve might have changed during the survey. On the other hand, surveys specialized for detecting slightly changing light curves due to very finely sampled cadence (such as TESS, see Ricker et al. 2015) usually have a relatively small footprint. LSST's cadence and depth, however, will allow for studying variable stars in the Milky Way's old halo in a way that makes population studies possible.

2. SCIENCE CASES

# LSST Data Products



readout — **Raw Data**
30 s image, 20 TB/night

⇩

37 s — **Prompt Data Products**
difference image analysis (DIA)
alerts: up to 10 million/night

⇩

24 h — **Prompt Products Database**
images, object & source catalog
solar system bodies orbit catalog

⇩

yearly — **Annual Data Release**
via LSST Science Platform
& LSST Data Access Centers

⇩

10 years — **Final 10 Year Data Release**
images: 5.5 million $\times$ 3.2 Gpx
catalog: 15 PB, 37 billion objects

⇨

**Alert Brokers**

ALeRCE

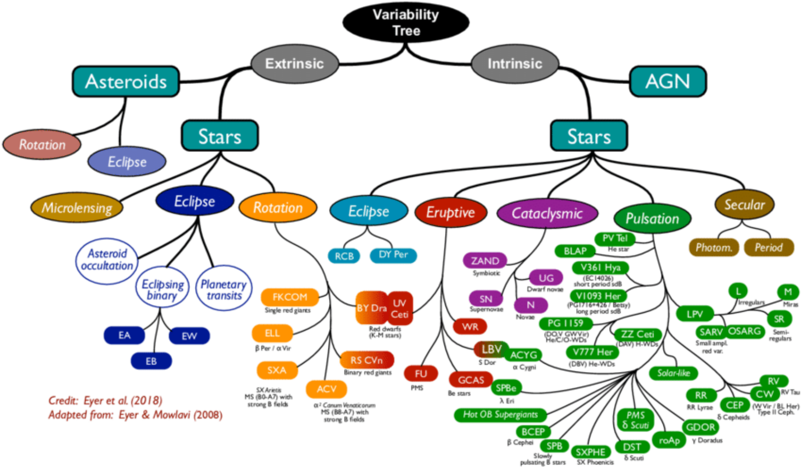
FINK

ANTARES

Lasair

$\Rightarrow$ science as soon as LSST begins

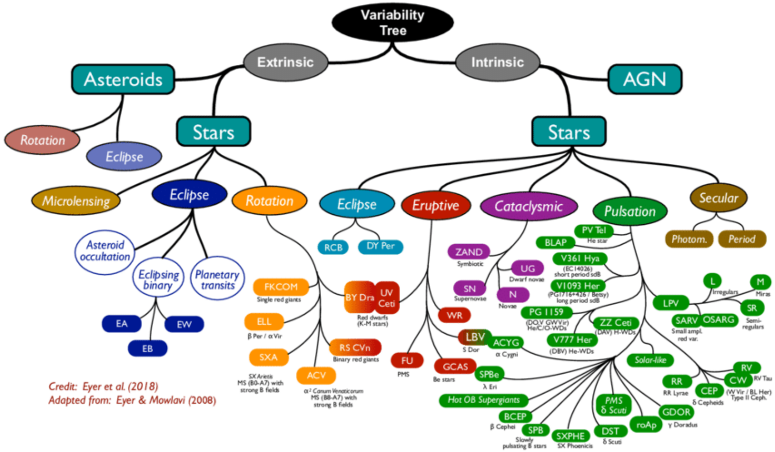
# Variable & Transient Sources with LSST




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

# Variable & Transient Sources with LSST






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

# Variable & Transient Sources with LSST



## Rubin LSST Transients and Variable Stars Science Collaboration

https://lsst-tvssc.github.io/

one of the LSST Science Collaborations

- Dark Energy
- Solar System
- Transients and Variable Stars
- Stars, Milky Way, and the Local Volume
- Galaxies
- Active Galactic Nuclei
- Strong lensing
- Large-scale Structure
- Informatics and Statistics

# Community

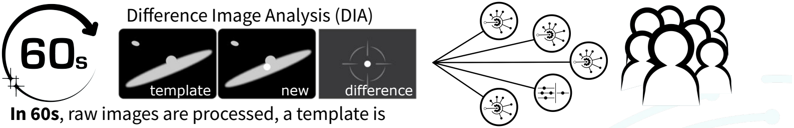


why **citizen science**?

- citizen science is vital for astronomy
- industry drives rapid advances in machine learning
- LSST data rate demands machine learning for identifying time-domain events
- citizen scientists now include thousands of machine learning experts

# Community

within 60 s after readout:
Stream of Alerts is released to Alert Brokers and to the LSST Alert Filtering Service




**In 60s**, raw images are processed, a template is subtracted, and difference-image sources are detected, associated, characterized, and…

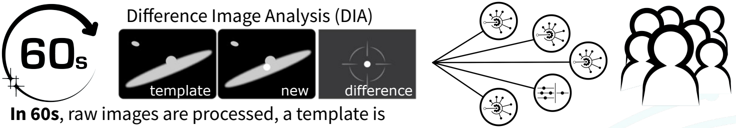
…distributed as alerts to brokers, where they can be rapidly analyzed by users.

Alerts: packets of LSST data for a difference image
Brokers: receive & process Alerts (external to LSST)

# Community



within 60 s after readout:
Stream of Alerts is released to Alert Brokers and to the LSST Alert Filtering Service




**In 60s**, raw images are processed, a template is subtracted, and difference-image sources are detected, associated, characterized, and…

…distributed as alerts to brokers, where they can be rapidly analyzed by users.

brokers will deliver scientific classification & interpretation to filter sources
Example uses include:

- collections of transient discoveries
- (pre-)classification using features & machine learning
- forwarding to *downstream brokers*
- alerting users
- alert distributions as ways to learn more about object types

# Community



within 60 s after readout:
Stream of Alerts is released to Alert Brokers and to the LSST Alert Filtering Service

ALeRCE
Automatic Learning for the Rapid Classification of Events

FINK

ANTARES

Lasair

brokers currently process a stream from Zwicky Transient Facility (ZTF)

# Community



Cross-Collaboration Work: LSST Data Challenge

Data: Simulated LSST light curves of $\sim$3.5 million objects, including full range of astronomical phenomena

Challenge: Accurately classify the objects based on the available photometry

for the simulation, TVS members contributed models of galactic variability

# Community



## Data challenges: PLAsTiCC /ELAsTiCC

*Photometric LSST Astronomical Time-series Classification Challenge*
and its extension
*The Extended LSST Astronomical Time-Series Classification Challenge*

ELAsTiCC uses simulated alerts, delivered to the alert brokers, to mimic the future rate, volume, and complexity of the LSST prompt data products. Realistic contextual information is incorporated into synthetic alerts.

The Challenge:

- Types are unbalanced
- Small number in the training set
- The training set is not representative of the test data
- Seasonal gaps
- Non-uniform cadence

# Classification



?

How does classification of variable sources happen?

Classification relies on detected **features**.

# Goals of Time Series Analysis

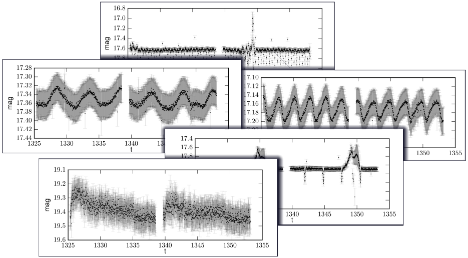


Time series analysis extracts meaningful statistics and other characteristics of the dataset in order to understand it.

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 $\chi^2$ 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 $\chi^2$ 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):

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

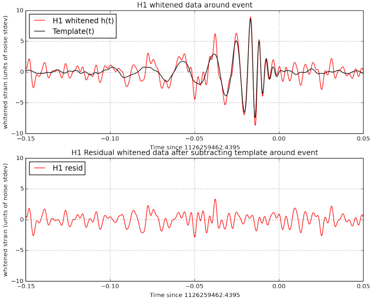
1. For $N = 100$ data points (not 100 objects), the minimum detectable amplitude is $A_{\rm min} = 0.92\sigma$
2. For $N = 1000$, $A_{\rm min} = 0.52\sigma$

# 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

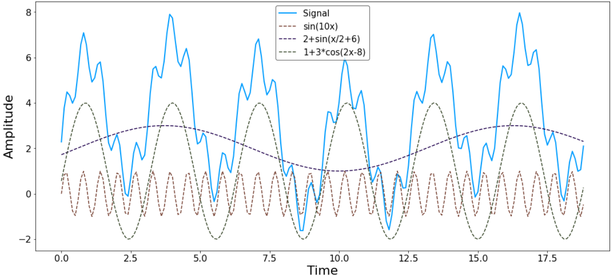
**example:** gravitational wave event GW150914




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

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

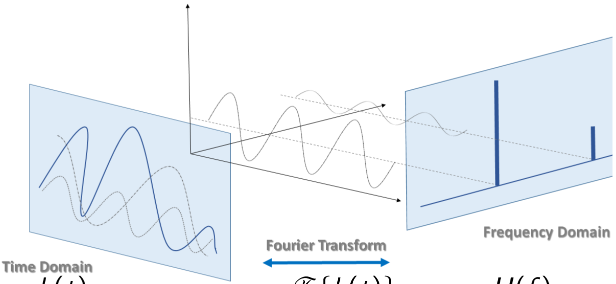
with inverse transformation

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

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**:




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

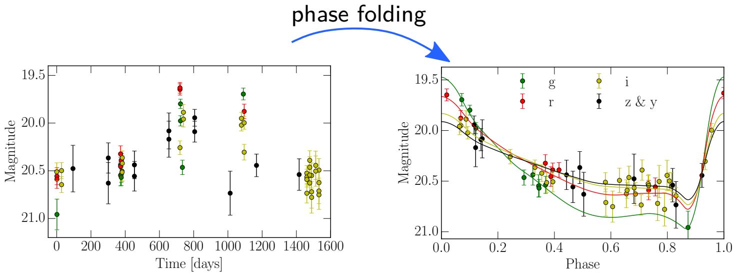
# Detecting Periodic Signals



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

For a periodic signal, if the period is known

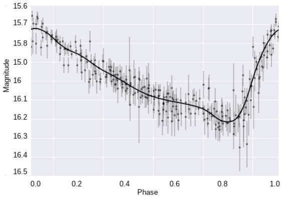
- we can write $y(t+P) = y(t)$, where $P$ is the period.
- we can create a **phased light curve** that plots the data as function of phase: $\phi = \frac{t}{P} - \mathrm{int}\left(\frac{t}{P}\right)$ with $\mathrm{int}(x)$ being the integer part of $x$.



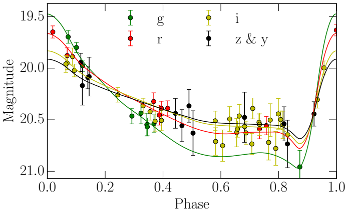
# Detecting Periodic Signals

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

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



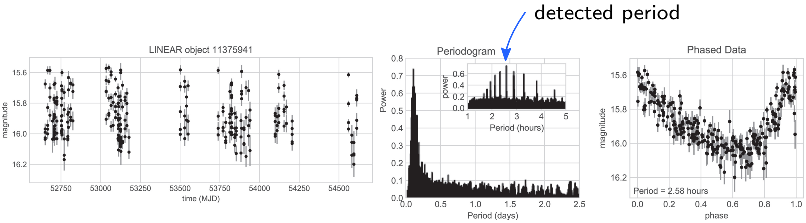

vs.




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** $\omega$ 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 $\chi_0^2$ is the $\chi^2$ 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 $\omega$.

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

# Machine Learning

… is the sub-field of computer science that gives computers the ability to learn without being explicitly programmed
(Arthur Samuel, 1959)

# Machine Learning

... is the sub-field of computer science that gives computers the ability to learn without being explicitly programmed (Arthur Samuel, 1959)

$\Rightarrow$ allows to **uncover hidden correlation patterns** through iterative learning by sample data

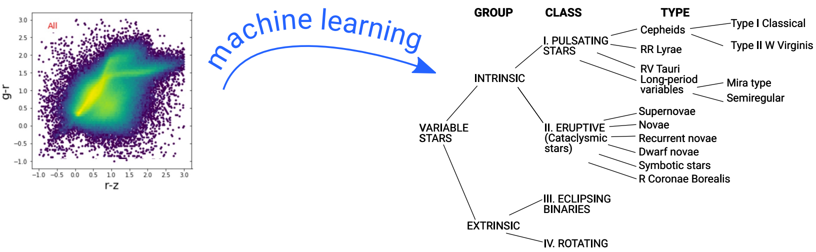
# Machine Learning

… is the sub-field of computer science that gives computers the ability to learn without being explicitly programmed (Arthur Samuel, 1959)

⇒ allows to **uncover hidden correlation patterns** through iterative learning by sample data

parameter space of measurements

parameter space of astrophysical objects



# Machine Learning

... is the sub-field of computer science that gives computers the ability to learn without being explicitly programmed (Arthur Samuel, 1959)

$\Rightarrow$ allows to **uncover hidden correlation patterns** through iterative learning by sample data

$\Rightarrow$ allow "to model a survey":

- describing data quality $\rightarrow$ outlier
- describing light curve characteristics $\rightarrow$ "features"
- classifying sources $\rightarrow$ catalogs
- finding substructure $\rightarrow$ clumps, overdensities, ...

# Summary

Variable stars have enabled us to dramatically change our view of the universe during the last $\sim$100 years:

- causes of variability
- distances
- the shape of our Milky Way
- the composition of our Universe
- the expanding Universe