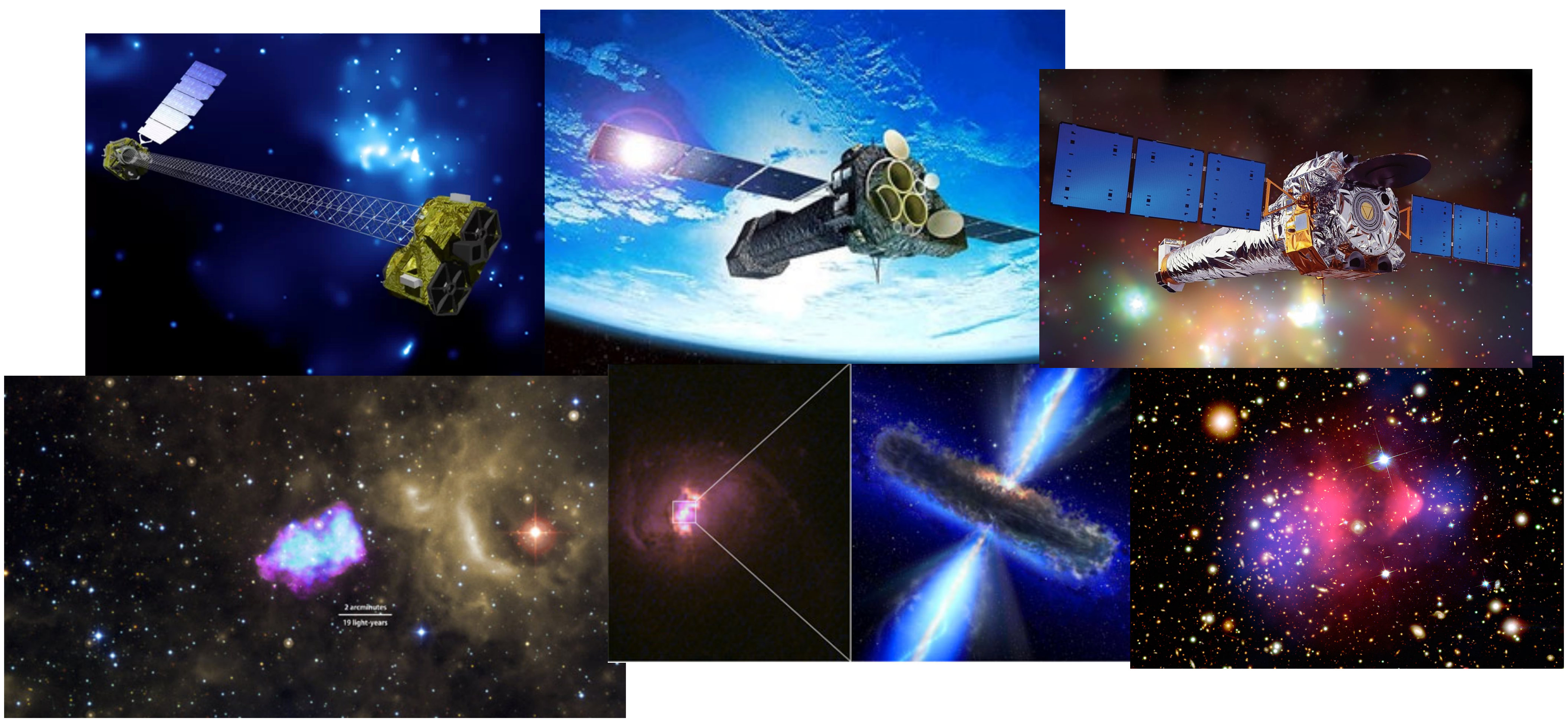
ISYA
INTERNATIONAL SCHOOL FOR YOUNG ASTRONOMERS

Lecture 4: Time Domain Astronomy

Rafael Martínez-Galarza
Harvard-Smithsonian Center for Astrophysics



Previously on… astrostats

- We have studied the basic rules of probability and learned how to estimate probabilities in terms of a frequency or repetition of an experiment

- We have learned what random variables are, and we have seen that we can characterize them with distributions.

- We have seen that PDFs are very common in astronomy, as they are related to the very problem of measurements that have uncertainties. But also they appear naturally in nature (e.g., the IMF).

This lecture:

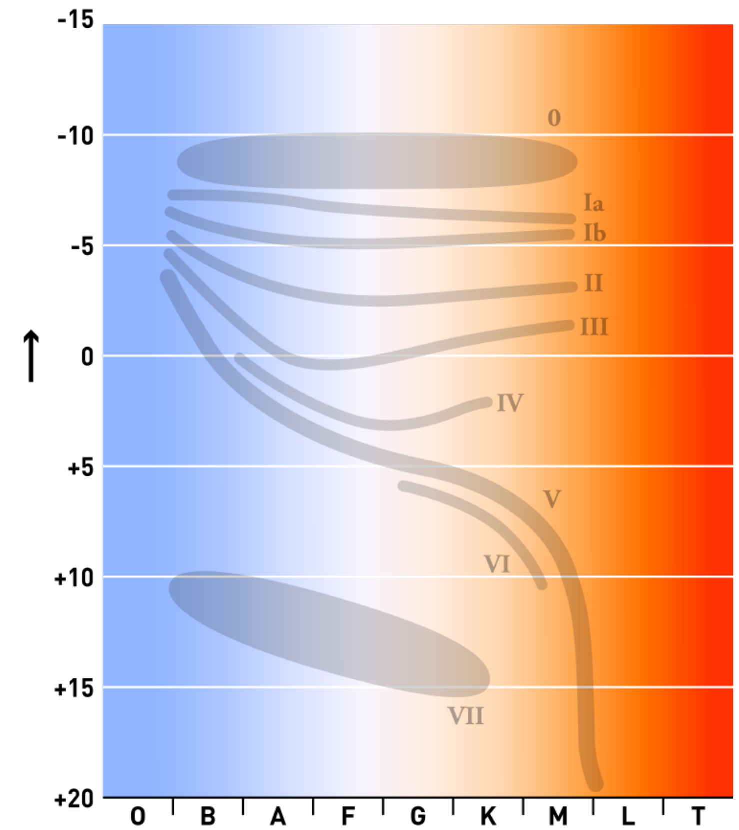
- We will learn how to use statistics in order to compare our models of Nature with data obtained with telescopes.
- We will learn how the problem of fitting a model to a set of data is philosophically different for frequentists and Bayesians.
- We will learn how we can evaluate how good our fit is to a particular set of data, given some assumptions.
- We will learn how to sample the full posterior distribution of model parameters, and fully characterize the uncertainties.
- We will apply this knowledge to the modeling of SNR spectra.
- Lots if credit to J. Vanderplas

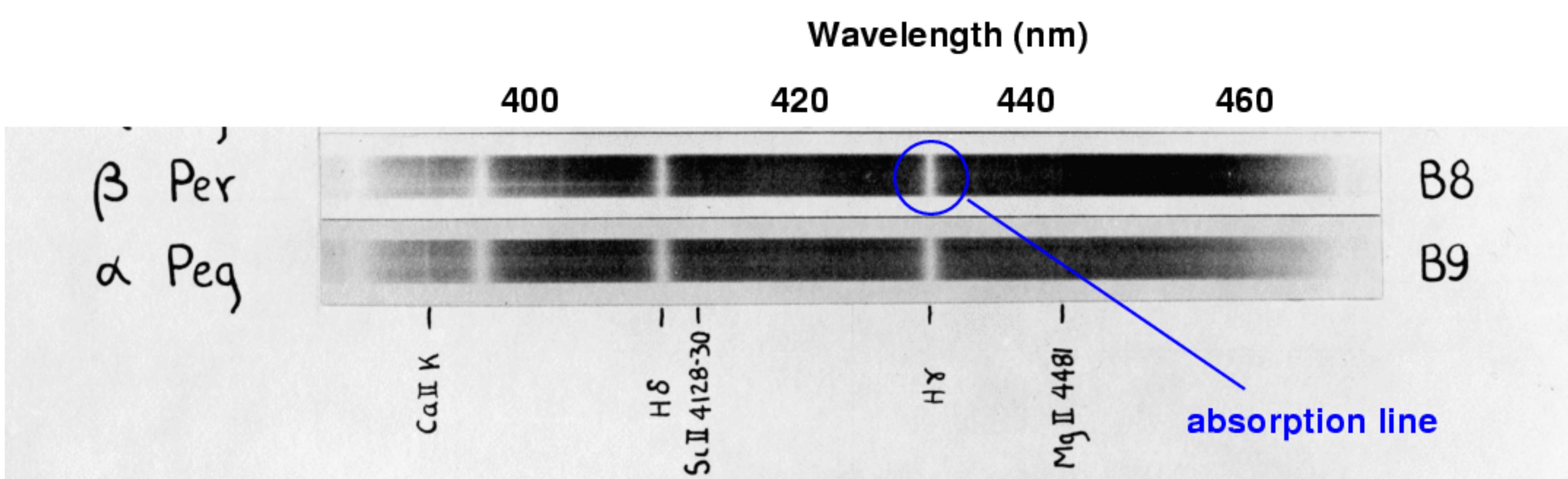
Big data in astronomy… then

Annie Jump Cannon at her Harvard desk

She analyzed over 300,000 stellar spectra during her lifetime…. by hand.

We owe her the stellar classification system we use today





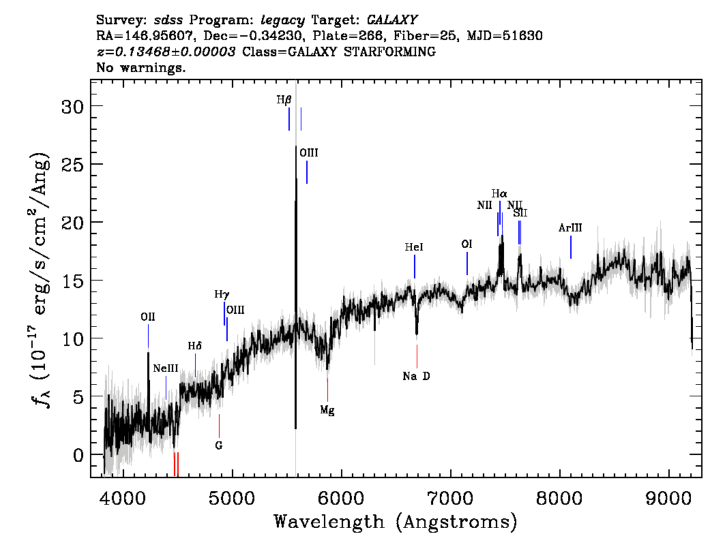
Big data in astronomy… now

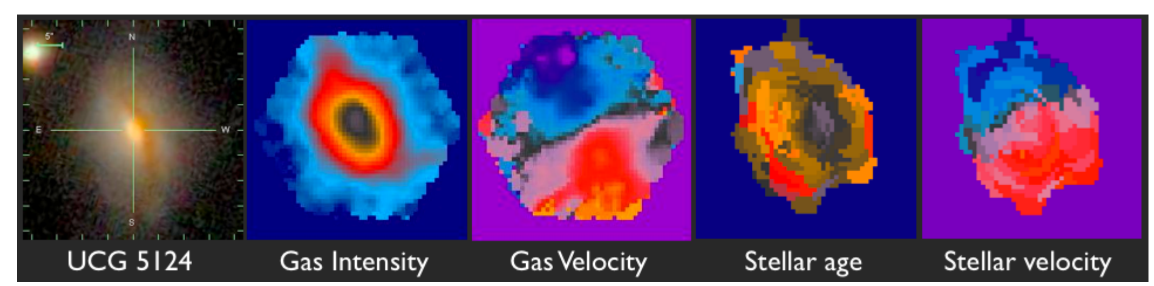
- Images. Limited by spatial resolution. Spatial resolution explosion with the next generation of telescopes.

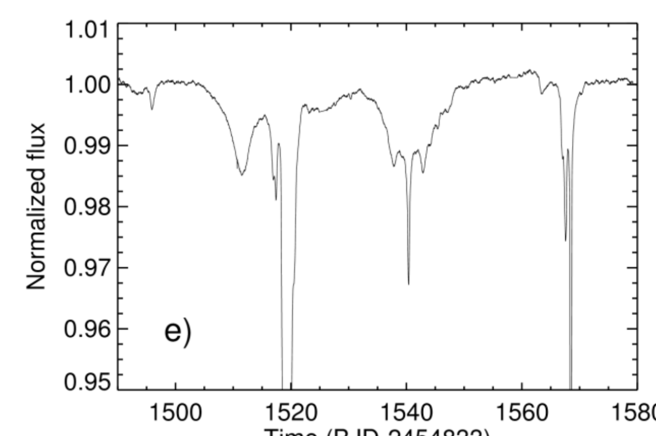
- Spectra. Limited by wavelength coverage and spectral resolution. Huge coverage of millions of spectra in the sky possible thanks to SDSS.

- IFU spectra. 2D spectra of astronomical objects. Spatial and spectral information together in the same dataset.

- Light curves. Variation of brightness as a function of time. Future synoptic surveys will imply an explosion of light curves.

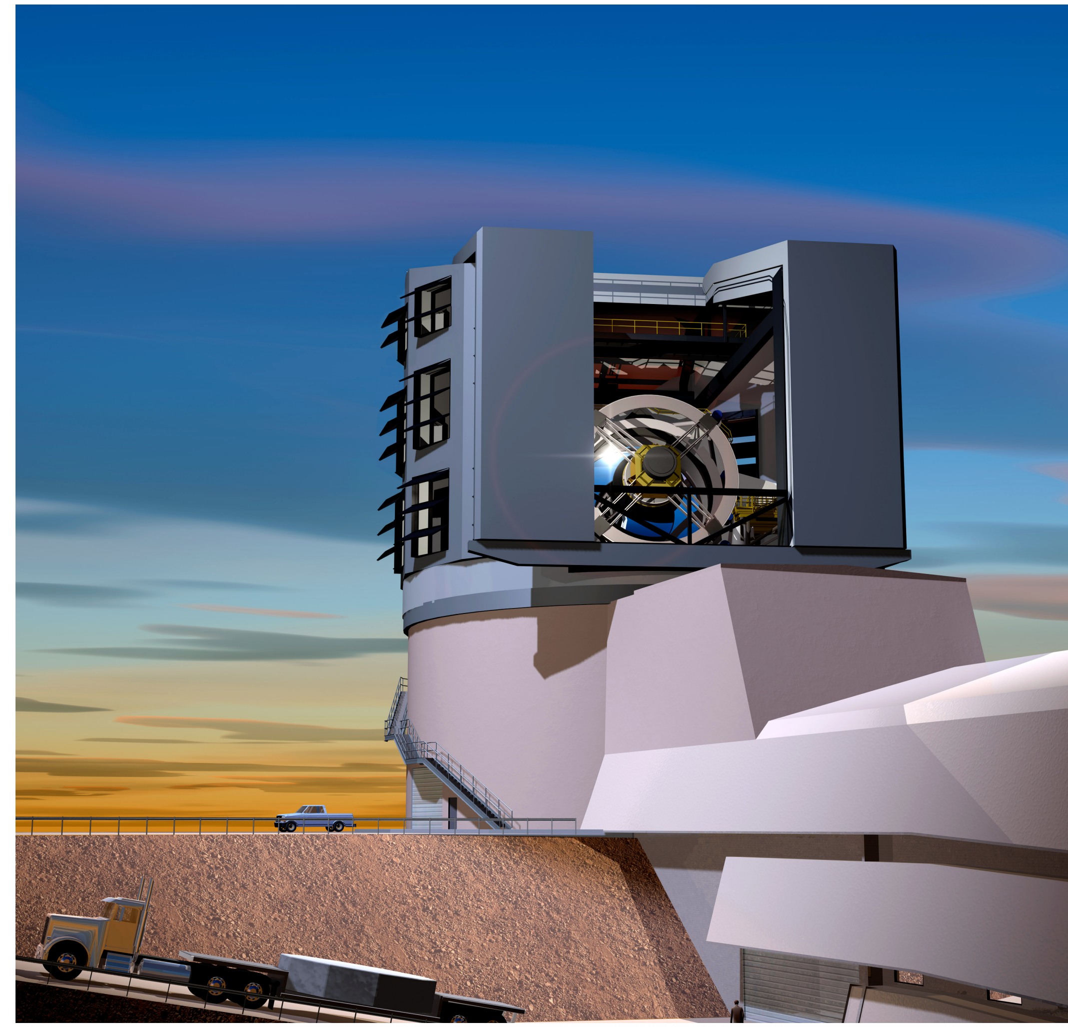


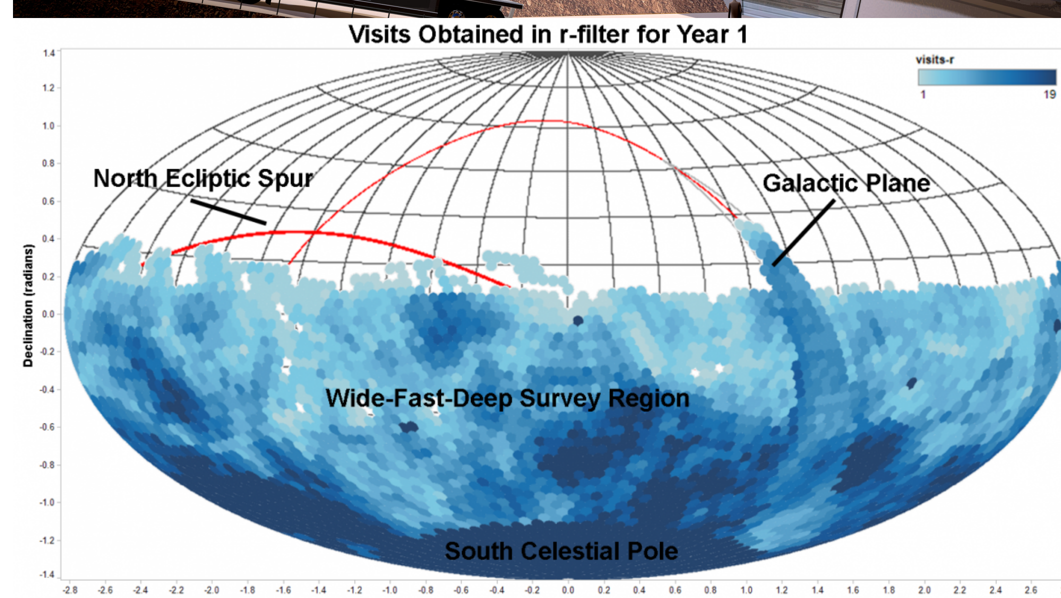




The motivation: LSST is coming

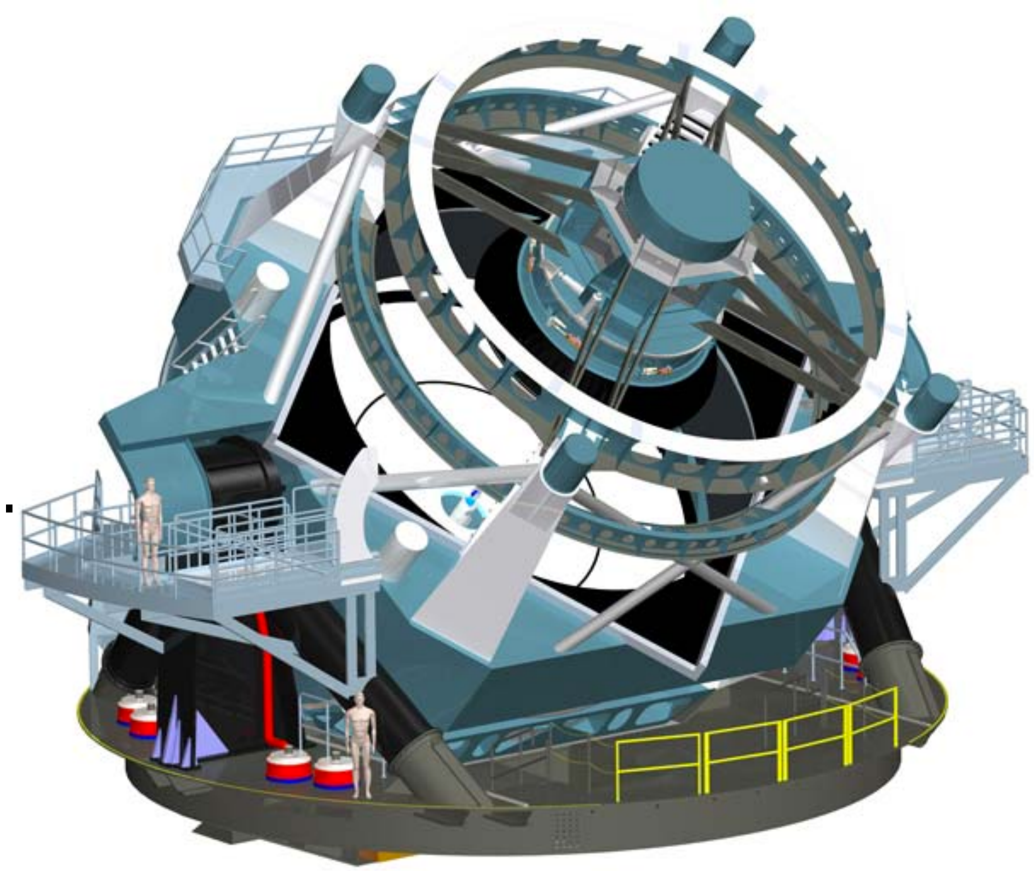
- The Large Synoptic Survey Telescope is a 8.4m reflector currently under construction in Chile (first light expected in 2021).
- Design concept: a survey that will take an image of every part of the entire visible sky every few nights, in six bands, for 10 years.
- Transients and variable stars: periodic and non-periodic variable sources will be studied in detail, and new types are expected at very short and very long timescales.





The Large Synoptic Survey Telescope

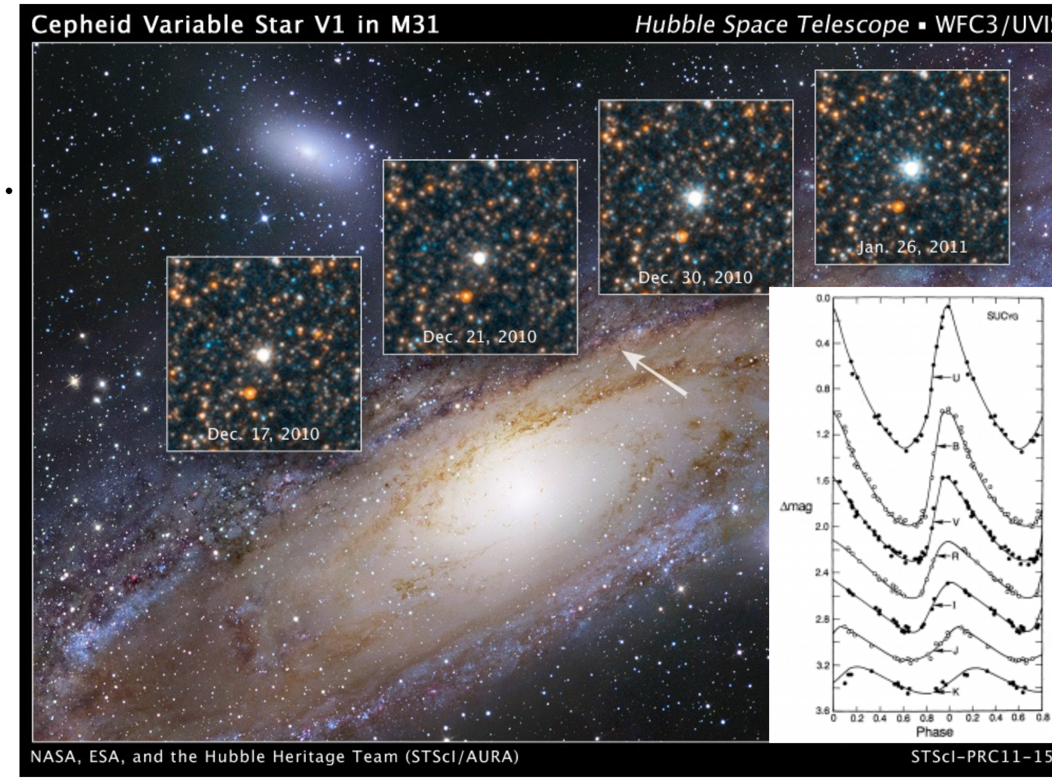
- LSST is an excellent example of what we mean by the new data-intensive astronomy
  - photometry of the entire southern sky every 3-4 nights for over 10 years.
  - ugrizy multiband data.
  - 30,000 GB per night.
  - Final catalog: 100s of petabytes.
  - ~1000 observations per field

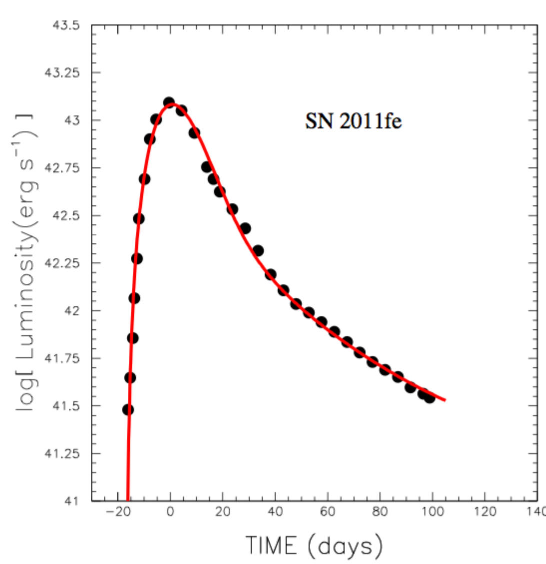


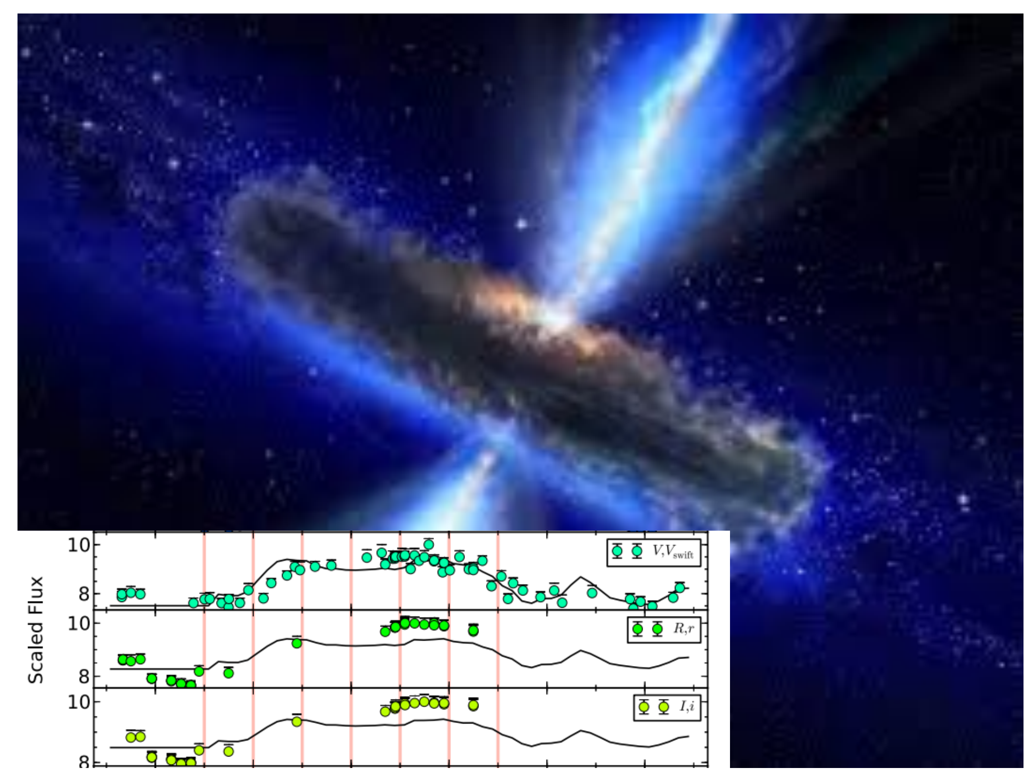

Slide by R. Hlozek
3.5°
slide: Andy Connolly for LSST Corporation

Challenge: Variability is Diverse

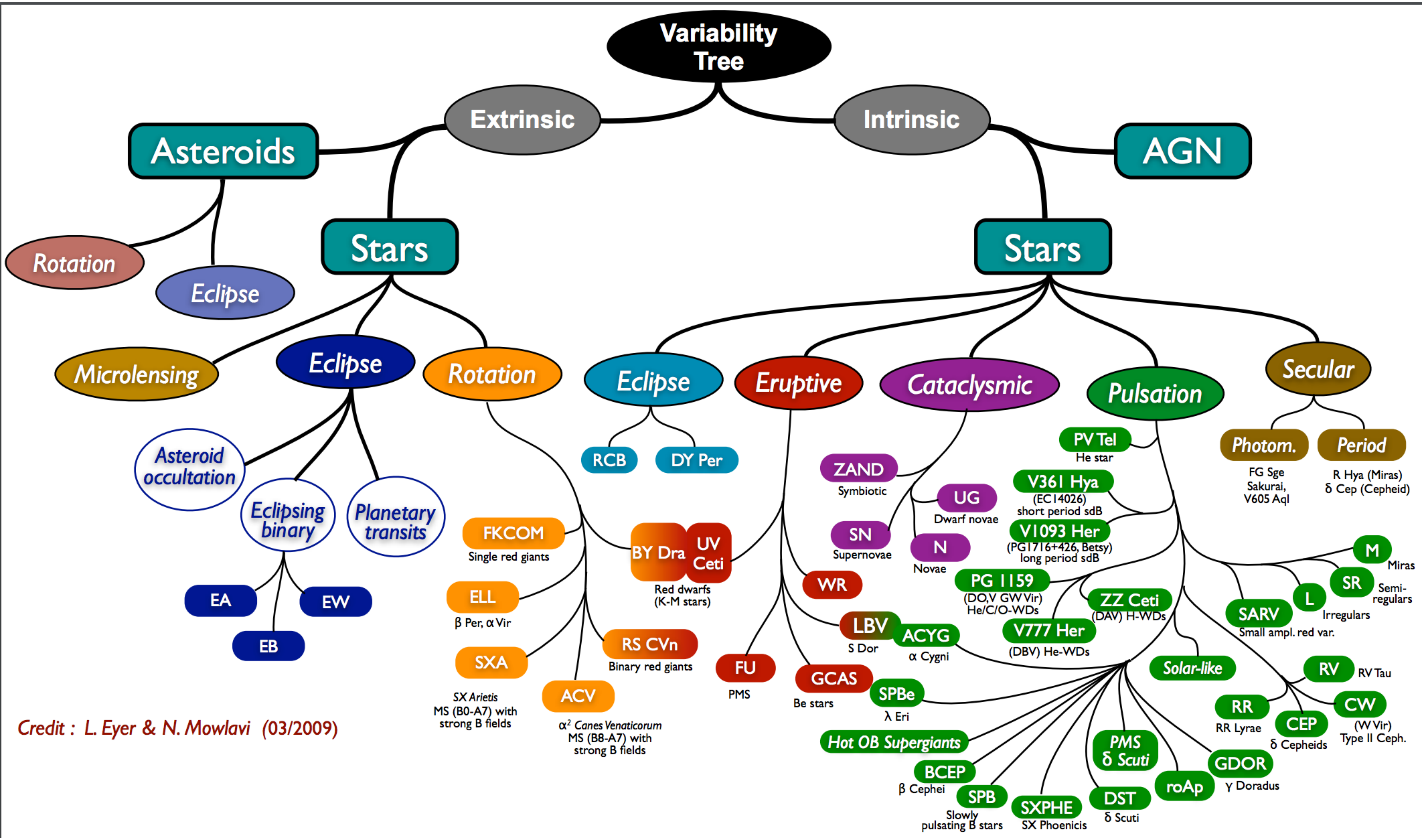
- Periodic (RR Lyrae stars, Cepheids)
    - Consistent in their periods and amplitudes.
- Quasi-periodic (Mira stars)
    - Dominating frequencies, but no consistency in phase or amplitude
- Stochastic (AGNs, QSOs)
    - Variability without any obvious patterns
- Transient (Supernovae, stellar flares, GRBs)
    - Short-time changes in flux, non periodic







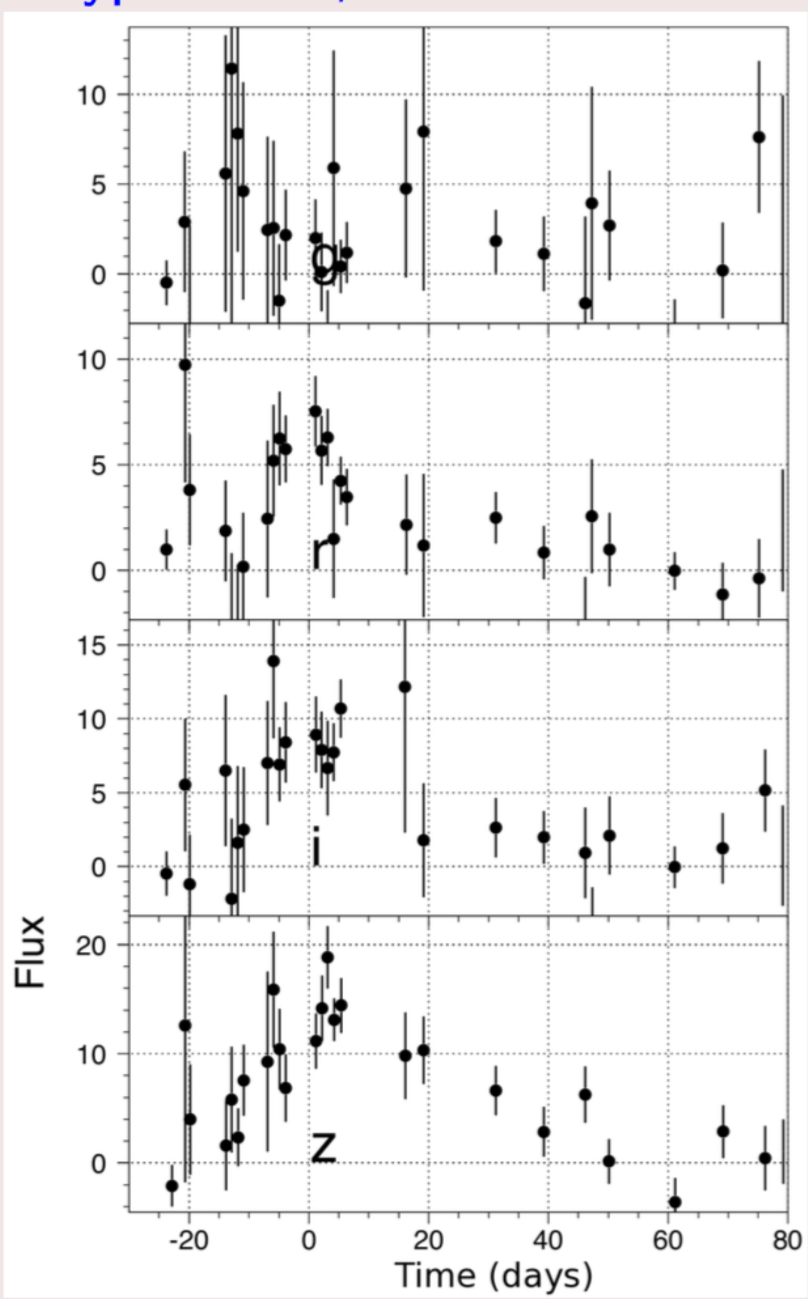
Variability Diagram



Light curves are sparse and non-uniform

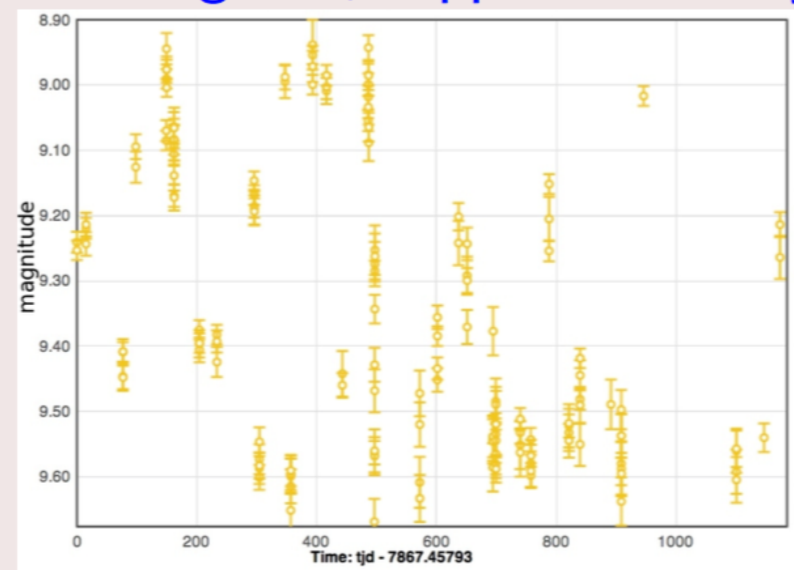
Supernovae

Type Ia SN, DES simulation

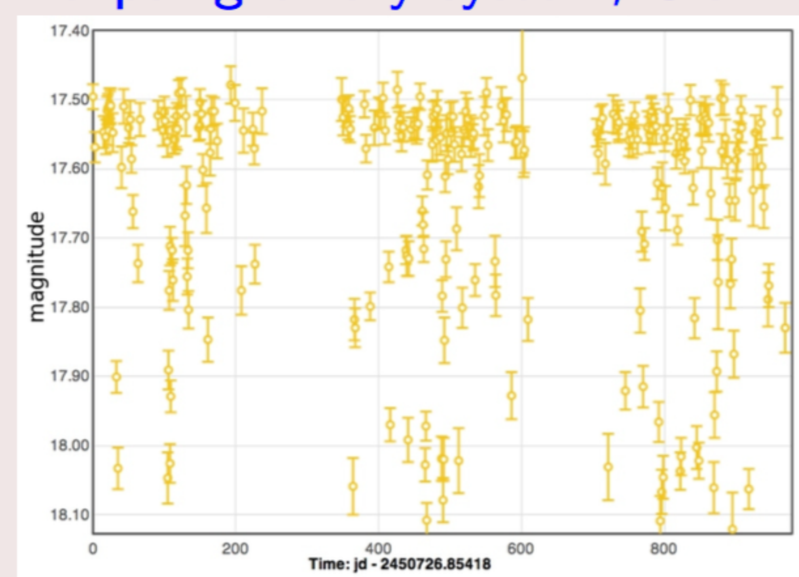


Variable Stars

Pulsating star, Hipparcos Survey



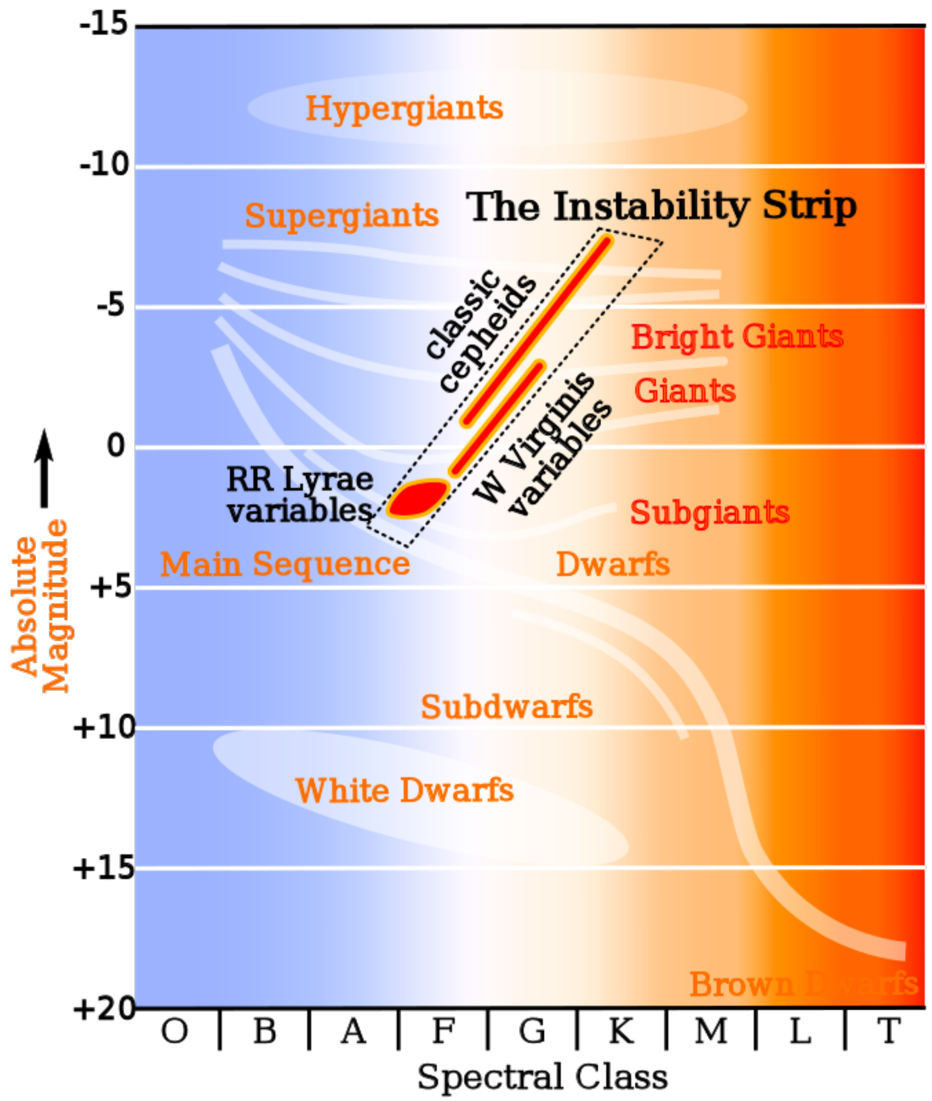
Eclipsing binary system, OGLE



http://dotastro.org

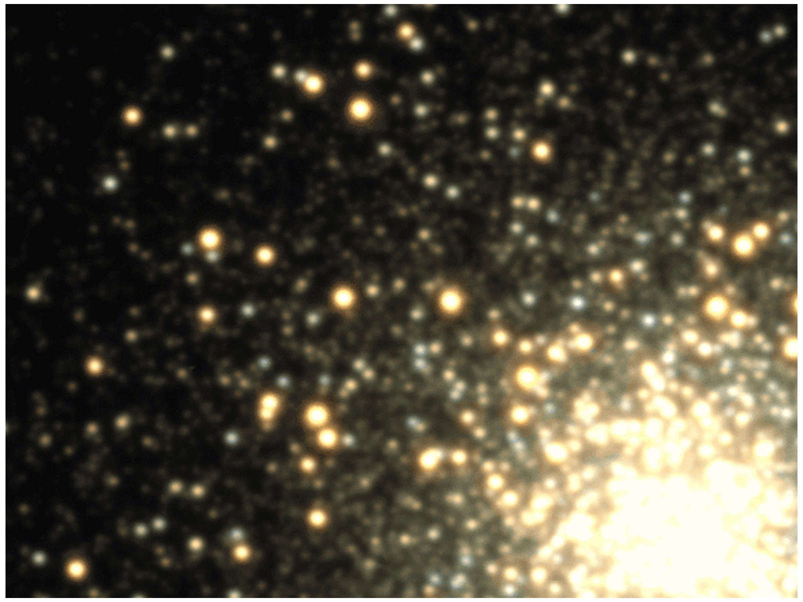
Slide: J. Richards.

RR Lyrae Stars: general properties

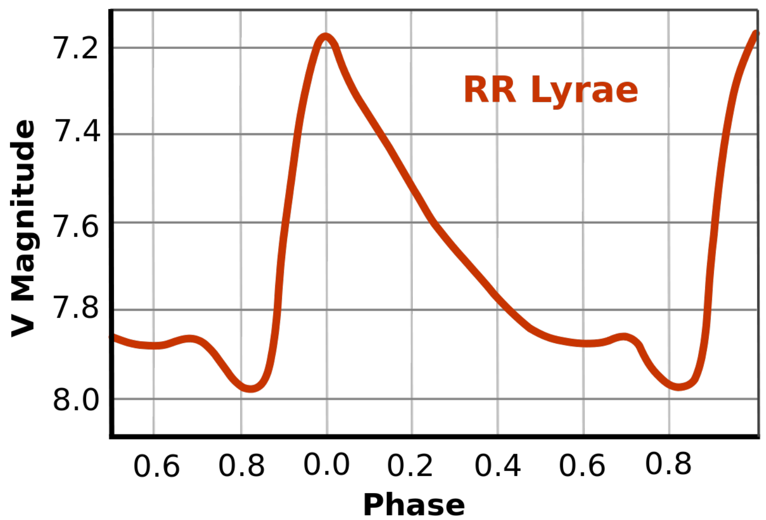


RR Lyrae are aging horizontal branch pulsating stars that have gone through a red giant phase, and are now in the instability strip. Pulsation is due to double ionization of He due to contraction and expansion of the star.

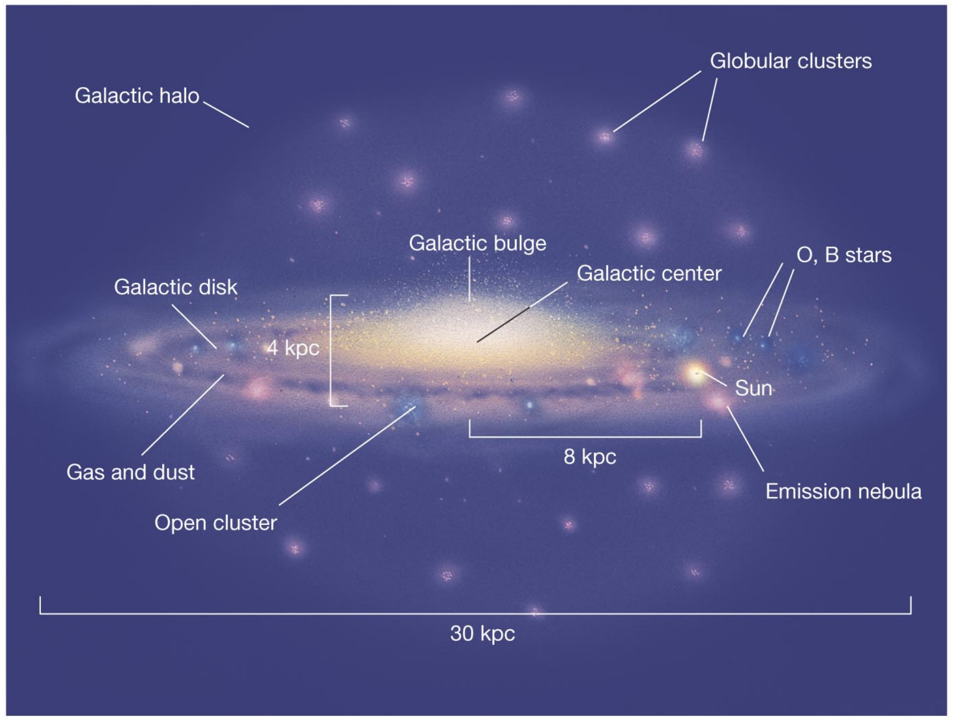
RR Lyrae stars in M3

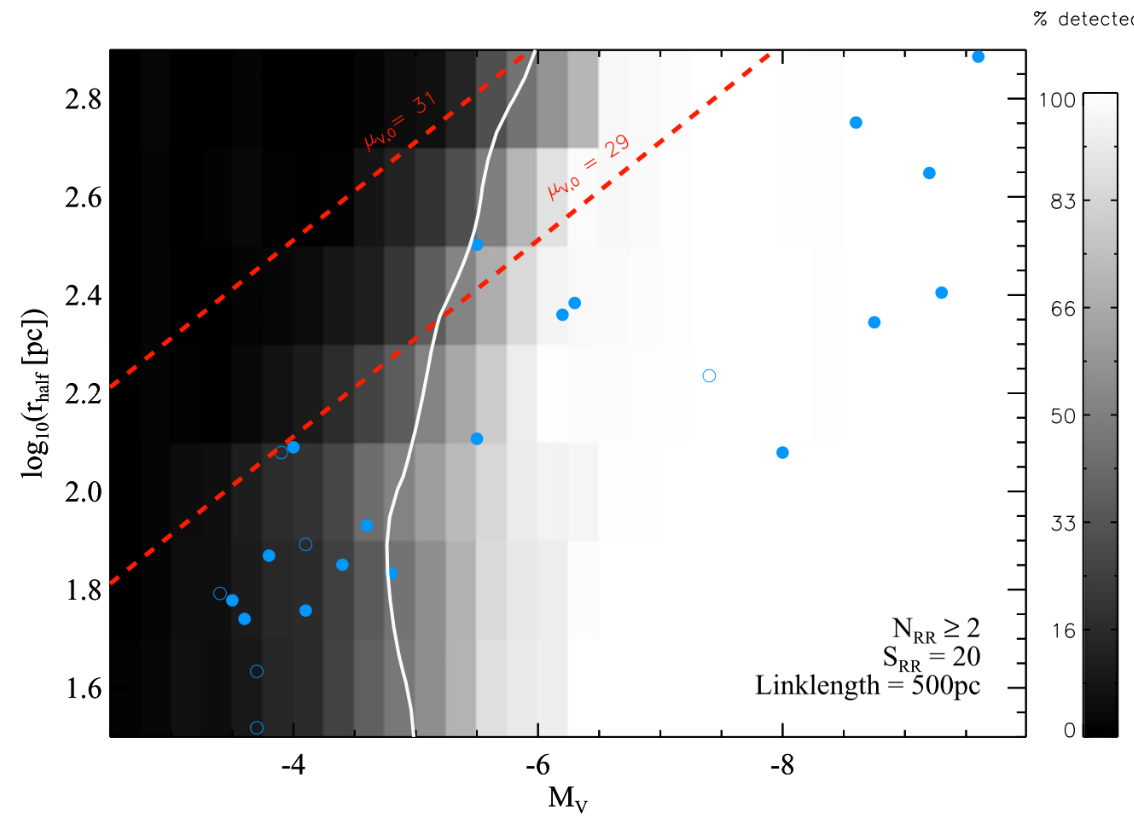


They have a very distinct light curve and a period-luminosity relationship that allows us to use them as standard candles



RR Lyrae stars: tracing structure in the galactic neighborhood.





The dynamical evolution of the galactic halo takes place at much slower timescales than the evolution of the galactic disk. Therefore, the halo keeps old secrets about the dynamical past of our galaxy: collisions, tidal tails and streams, etc.

But the galactic halo is mostly dark, except for the old globular clusters.

Every MW satellite has at least 1 RR Lyrae star detected. The above plot shows how likely it is to detect a satellite galaxy as a group of 2 or more RR Lyrae stars that are sufficiently close.

i.e. the LIKELIHOOD of two or more RR Lyrae stars to be very close above the galactic plane is very small.

How to find RR Lyrae stars: a recipe

1. Gather time-series data (with LSST, for example).

2. Detect objects with periodic variations (not so easy, as we have seen).

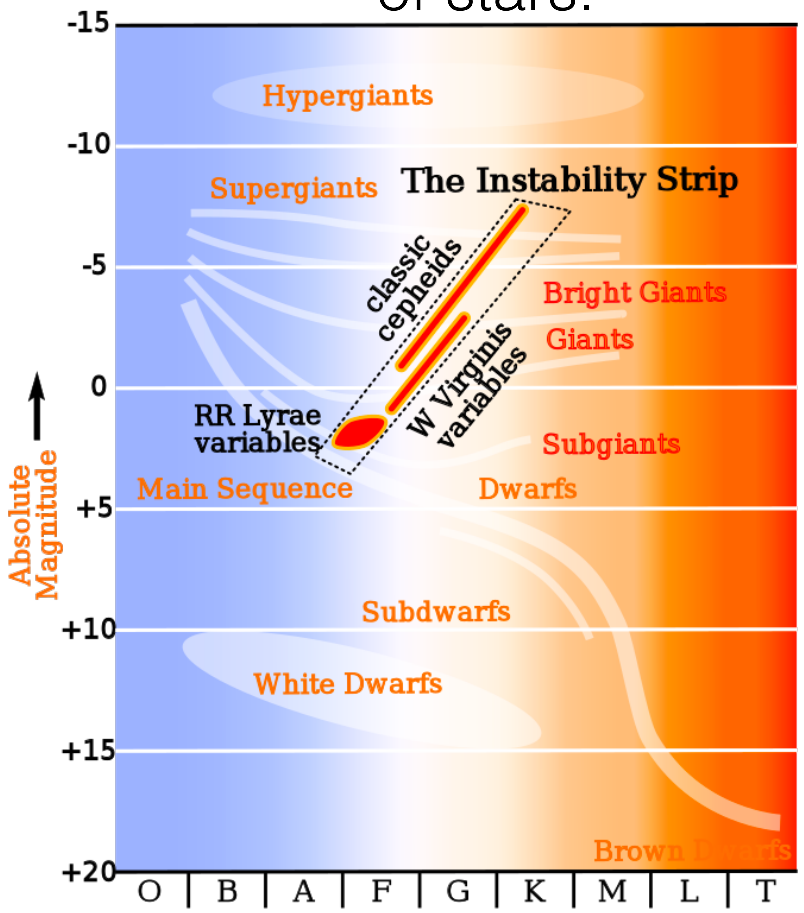
3. Fit RR Lyrae templates to light curves.

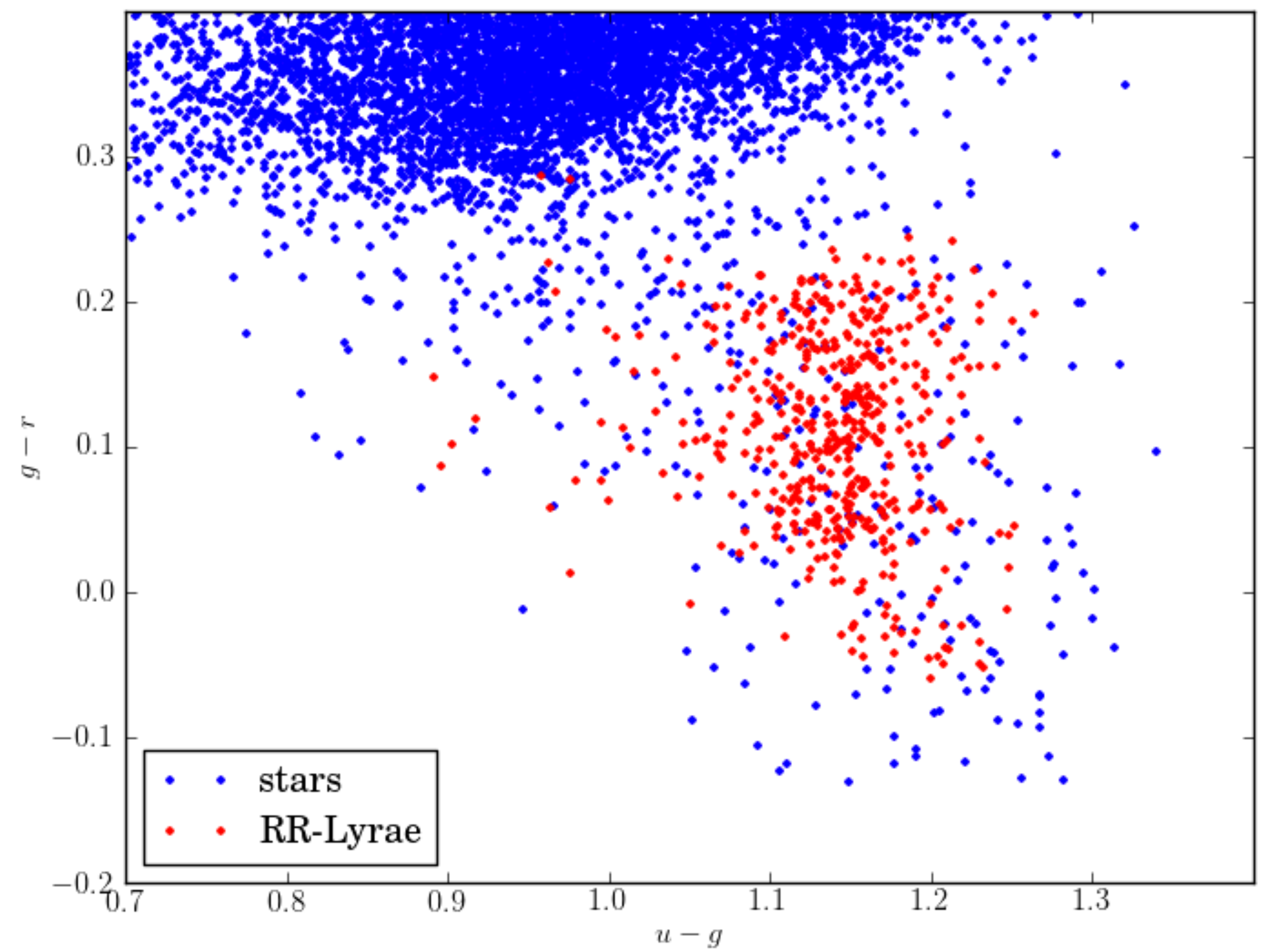
4. Use light curves to find distance, do the science!

   NOT SO STRAIGHTFORWARD!

What if I do not have light curve?

Using stellar models, we can predict colors based on the physical properties of stars.





- Sun

$$M_{\text{Sun}} = 2.0 \times 10^{33} \text{ g}$$

$$R_{\text{Sun}} = 7.0 \times 10^{10} \text{ cm}$$

$$\bar{\rho}_{\text{Sun}} = 1.4 \text{ g/cm}^3$$

$$\log g = GM/R^2 = 4.44 \text{ [cgs]}$$

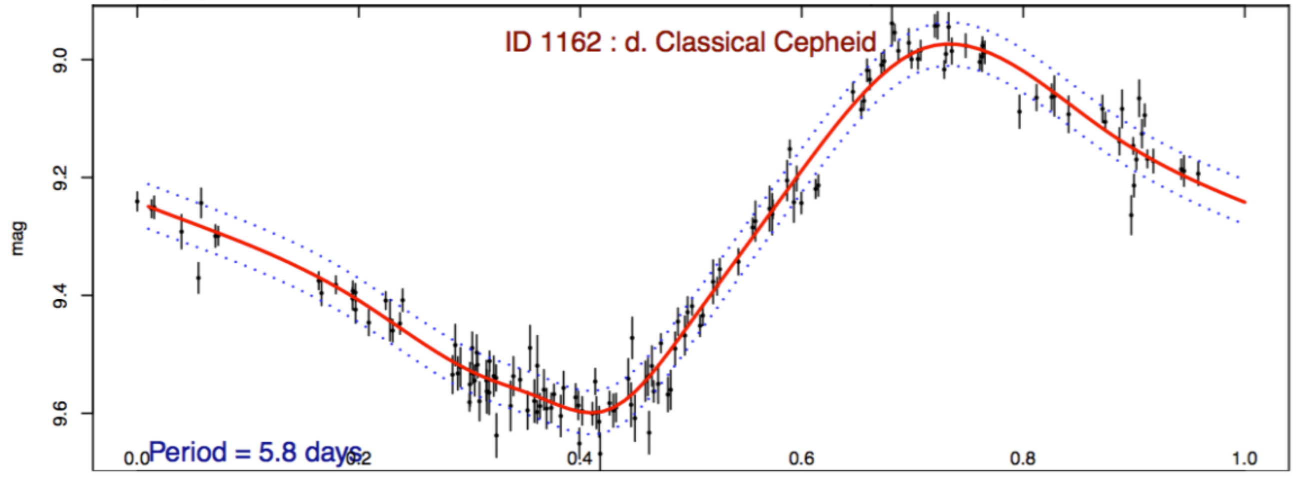
Identification of RR Lyrae stars is also possible based on their single-epoch colors only. However, contamination is an issue, as you can see above.

u-g color is sensitive to surface gravity, log g
g-r color is sensitive to T_{eff}.

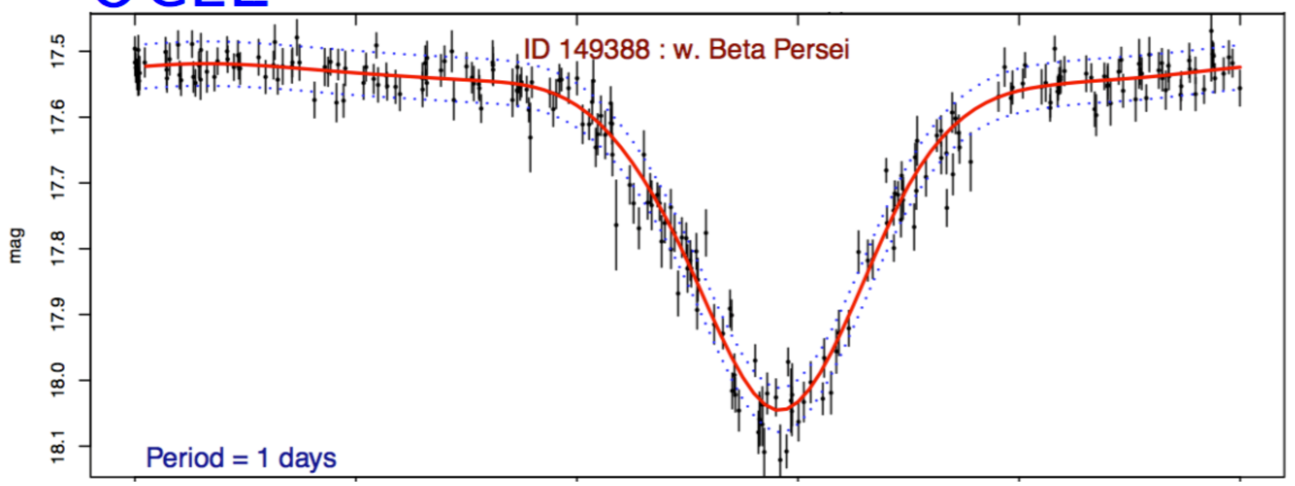
Characterizing light curves

Dense light curves would be more useful, but with limited telescope time we cannot equally sample light curves of different timescales.

Hipparcos



OGLE



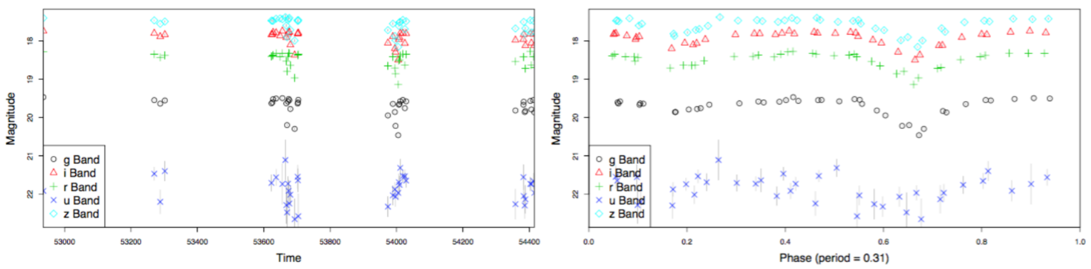
Traditionally, classification has been performed by human experts.

But the upcoming surveys with billions of sources will make this impossible,

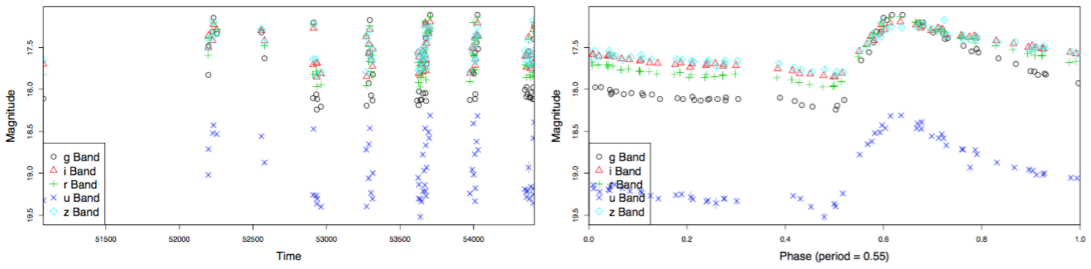
We need to characterize the variation of astronomical objects automatically.

First step of course is to find the periods, so we can 'fold' the light curve:

Eclipsing Binary (unfolded and folded)



RR Lyrae (unfolded and folded)

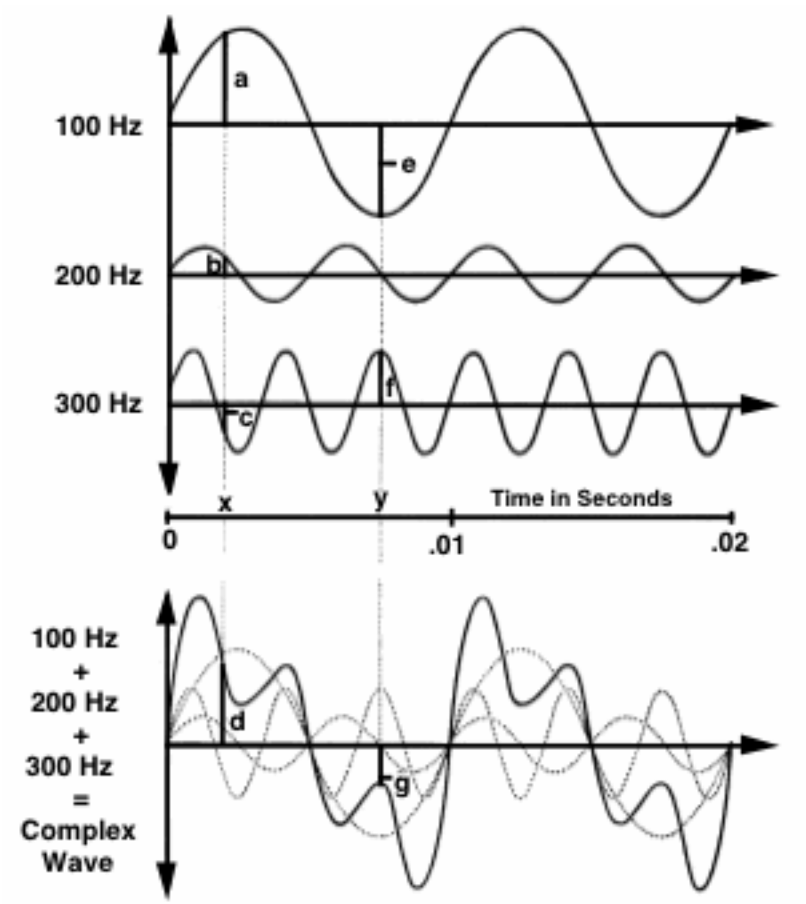


Fourier Analysis

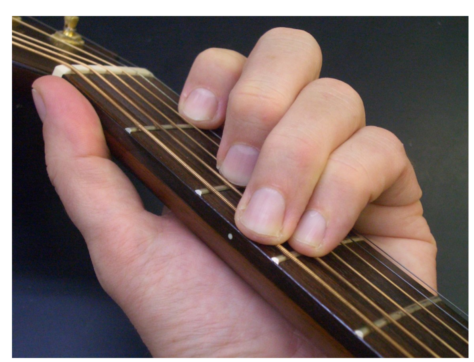
Any continuous signal can be represented as a series of harmonic components.

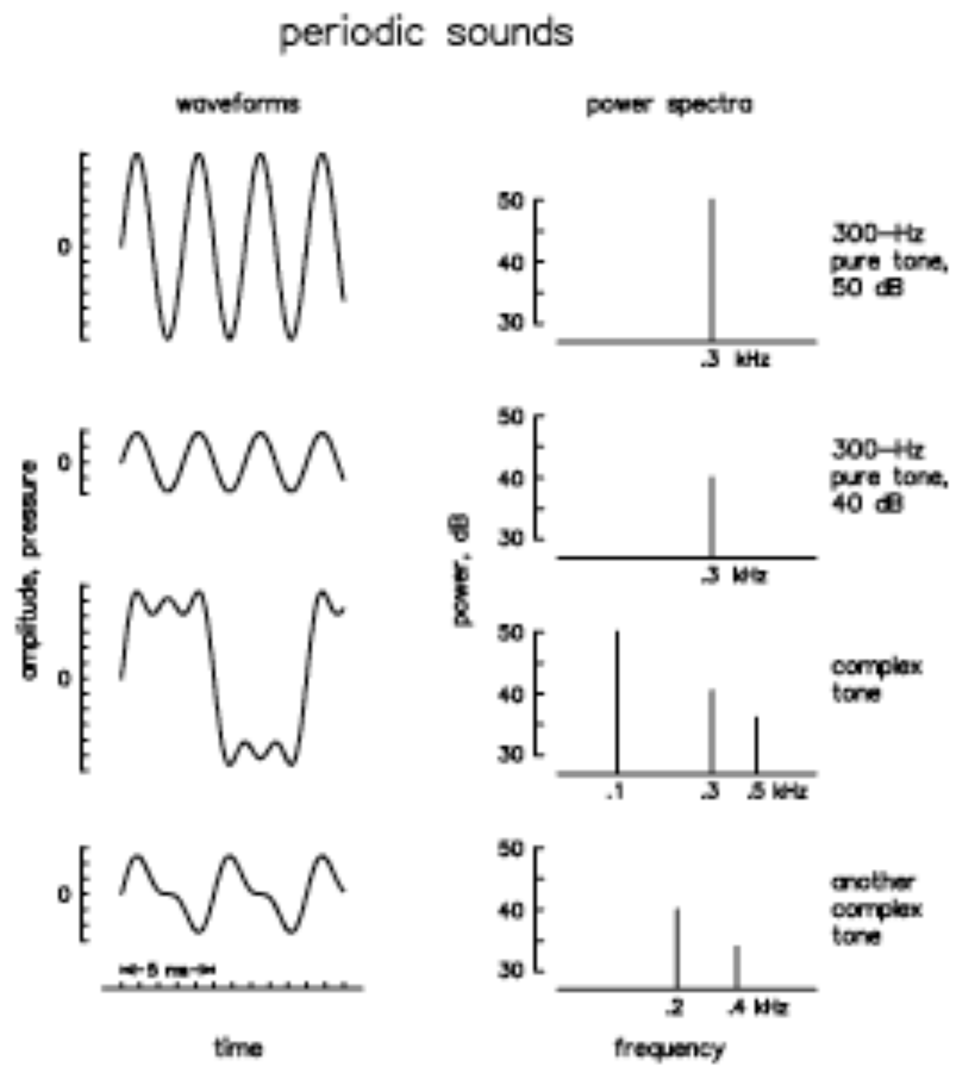
The square of the Fourier Transform gives the power spectrum, which indicates contributing frequency.

$$y(t|\omega,\theta) = \theta_0 + \sum_{n=1}^{N} \left[\theta_{2n-1}\sin(n\omega t) + \theta_{2n}\cos(n\omega t)\right]$$



$$C(\omega) = \frac{1}{N}\left|\sum_{k=1}^{N} y_k e^{i\omega t_k}\right|^2 ;$$

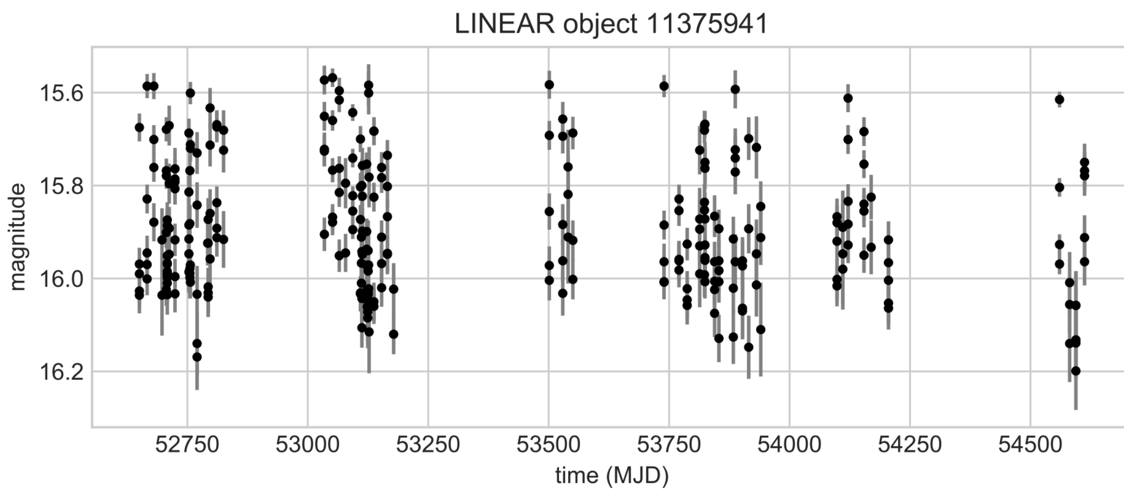




I could calculate the Fourier transform of a light curve to find which frequency has more power, but that is computational expensive. Instead…

How do we find periods?

This is a hard problem. One possibility is to use the Lomb-Scargle periodogram method.
Suppose you want to find the period of this light curve:



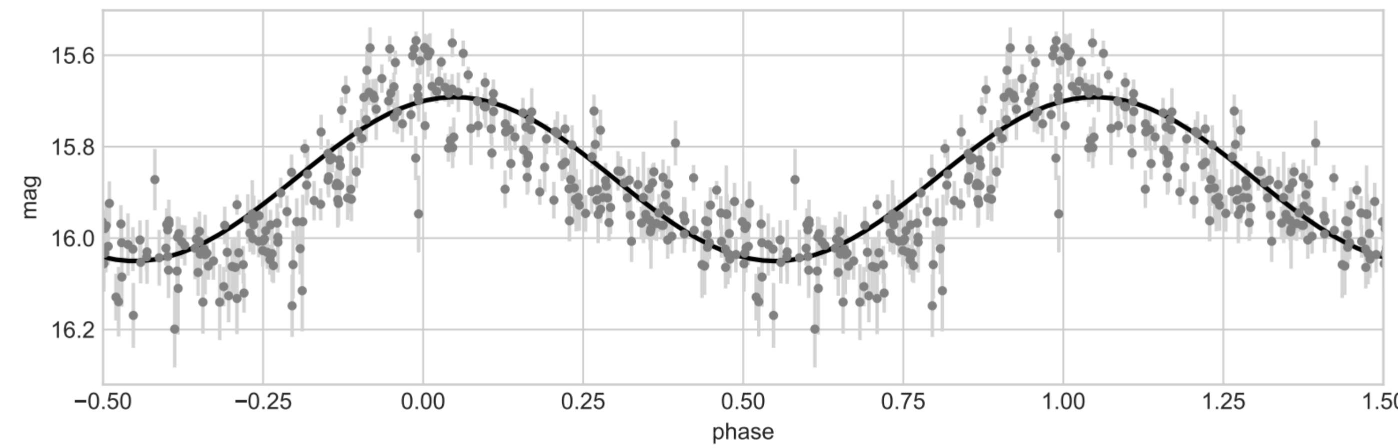
We propose a sinusoidal model for each frequency

$$y(t; f) = A_f \sin(2\pi f(t - \phi_f))$$

You can fit this model to the data by minimizing the X^2 statistic as a function of the parameters A_f and ϕ_f:

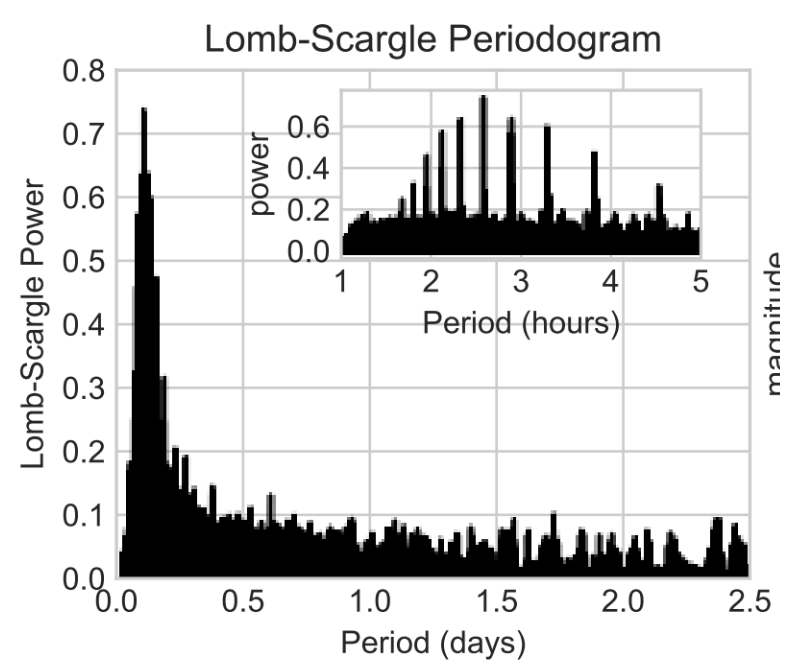
$$\chi^2(f) \equiv \sum_n \left(y_n - y(t_n; f)\right)^2$$

So you are assuming a sinusoidal model to that might not be the perfect description of the variability, but it contains the right frequency info.



You can build a periodogram (namely a plot of the strength of each frequency just like in Fourier analysis.) by defining:

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

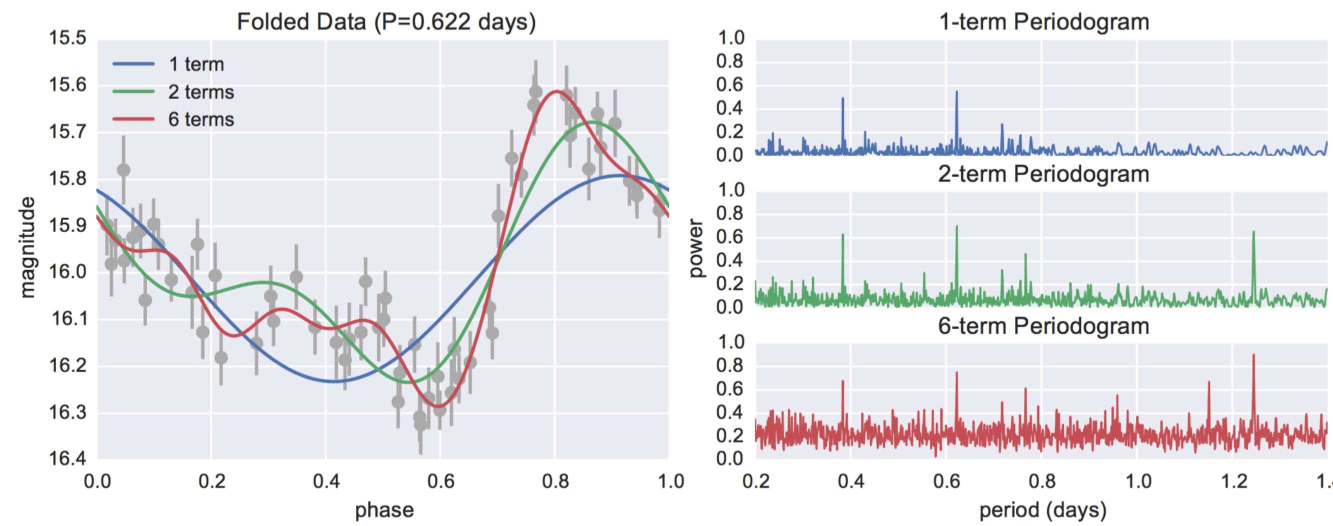


Multiband periodogram

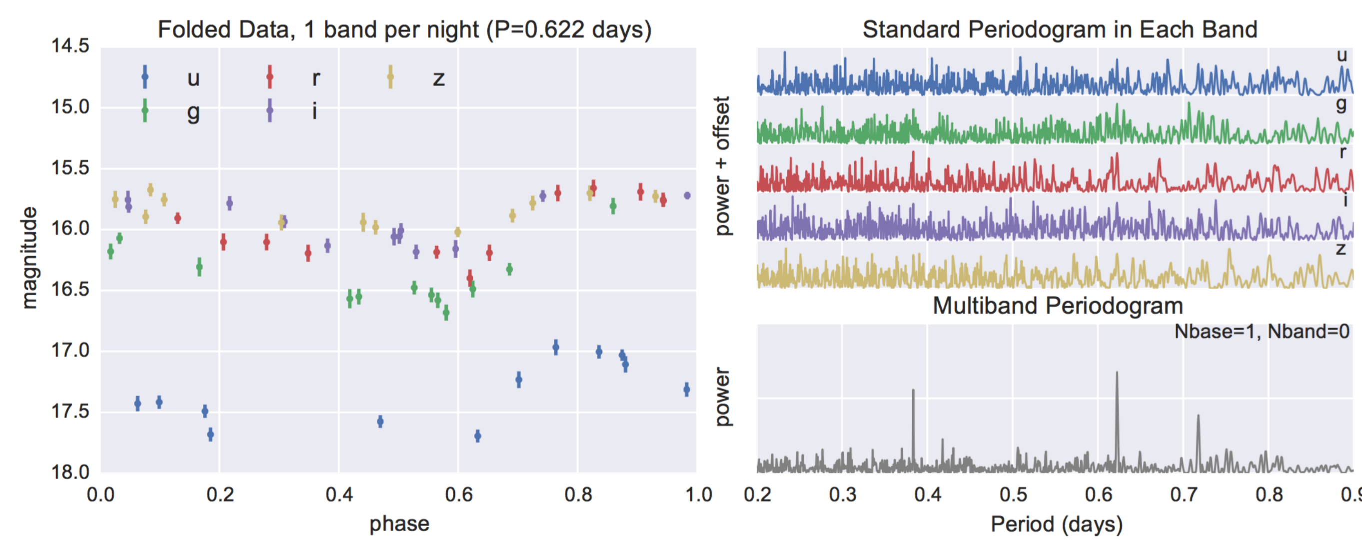
- We can also use the multi-band periodogram (van der Plas & Ivezic, 2015) to estimate periods.
- Outperforms existing methods, specially for non-simultaneous, sparsely sampled multi-band LCs.
- Method is linear on the θ parameters, and thus it is fast.
- Regularization is the key to allow multi-band analysis, and to avoid overfitting.

Single band (Lomb-Scargle):

$$y(t|\omega,\theta) = \theta_0 + \sum_{n=1}^{N} \left[\theta_{2n-1}\sin(n\omega t) + \theta_{2n}\cos(n\omega t)\right].$$



$$y_k(t|\omega,\theta) = \theta_0 + \sum_{n=1}^{M_{base}} \left[\theta_{2n-1}\sin(n\omega t) + \theta_{2n}\cos(n\omega t)\right] + \theta_0^{(k)} + \sum_{n=1}^{M_{band}} \left[\theta_{2n-1}^{(k)}\sin(n\omega t) + \theta_{2n}^{(k)}\cos(n\omega t)\right].$$



Estimating errors in distribution parameters

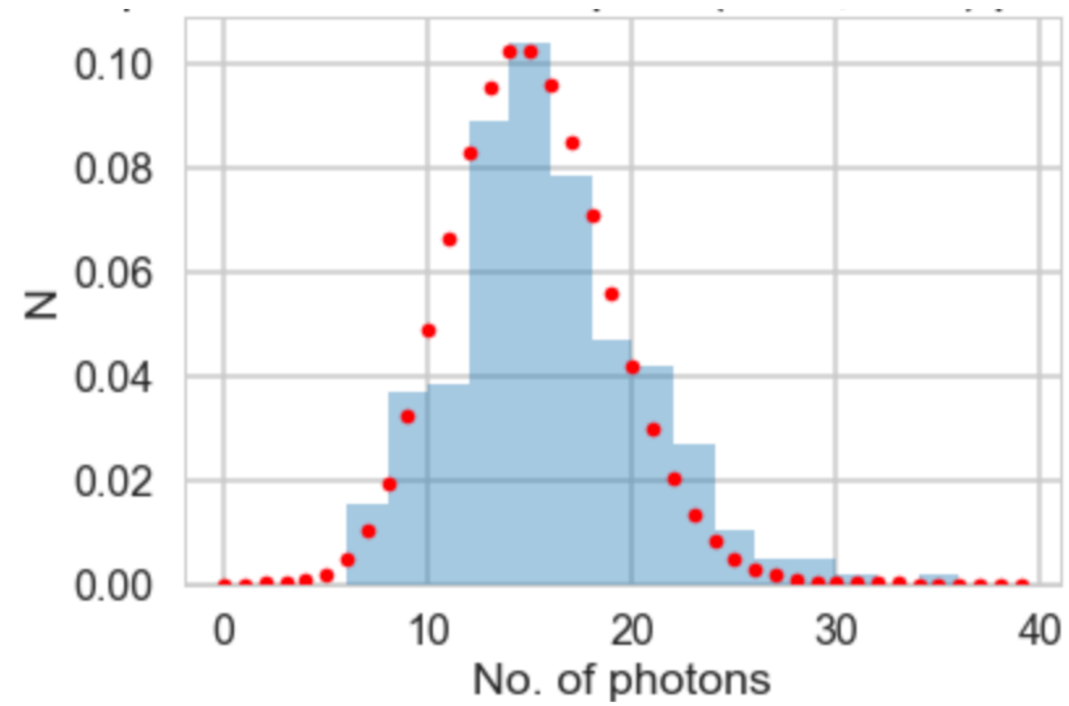
How do we estimate the significance of our period estimation? i.e., how do we estimate some kind of uncertainty?

Let's roll back to our discussion of photon arrivals on a Chandra ACIS detector.

As you remember, this had (almost) a poisson distribution.

Suppose you wanted to estimate the error in the mean or the σ distribution the results from sampling it.

One general method to estimate errors directly from data is bootstrapping.



Ideally, you would want to get many samples from the true population (not just one sample).

But you don't know the true population, so you use the empirical distribution of your sample.

You draw many 'samples' by resampling the empirical distribution with replacement.

What is Bootstrap

Bootstrap is a resampling procedure.

$\mathbf{X} = (X_1, \ldots, X_n)$ - a sample from F

$\mathbf{X}^* = (X_1^*, \ldots, X_n^*)$ - a simple random sample from the data.

$\hat{\theta}$ is an estimator of θ

θ^* is based on X_i^*

Examples:

$\hat{\theta} = \bar{X}_n,$ $\theta^* = \bar{X}_n^*$

$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2,$ $\theta^* = \frac{1}{n}\sum_{i=1}^{n}(X_i^* - \bar{X}_n^*)^2$

$\theta^* - \hat{\theta}$ behaves like $\hat{\theta} - \theta$

Nonparametric and Parametric Bootstrap

Simple random sampling from data is equivalent to drawing a set of i.i.d. random variables from the empirical distribution. This is Nonparametric Bootstrap.

Parametric Bootstrap if $X_1^*, \ldots, X_n^*$ are i.i.d. r.v. from $\hat{H}_n$, an estimator of F based on data $(X_1, \ldots, X_n)$.

Example of Parametric Bootstrap:

$X_1, \ldots, X_n \; i.i.d. \sim N(\mu, \sigma^2)$

$X_1^*, \ldots, X_n^* \; i.i.d. \sim N(\bar{X}_n, s_n^2); \quad s_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$

$N(\bar{X}_n, s_n^2)$ is a good estimator of the distribution $N(\mu, \sigma^2)$

Periodogram: peak significance

How significant is the highest peak in the Lomb-Scargle periodogram?

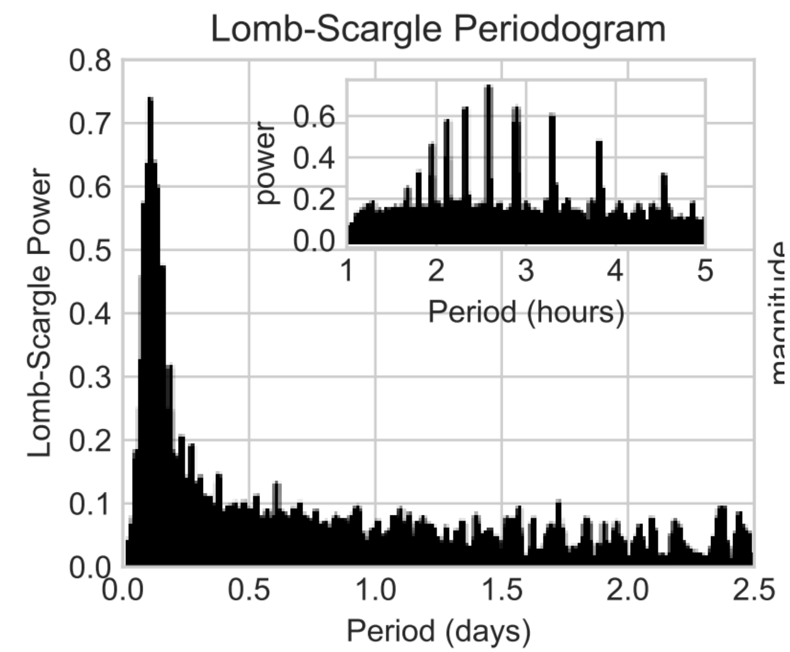
i.e., How likely is it to get a peak that high under the assumption that there is no periodic signal in the data?

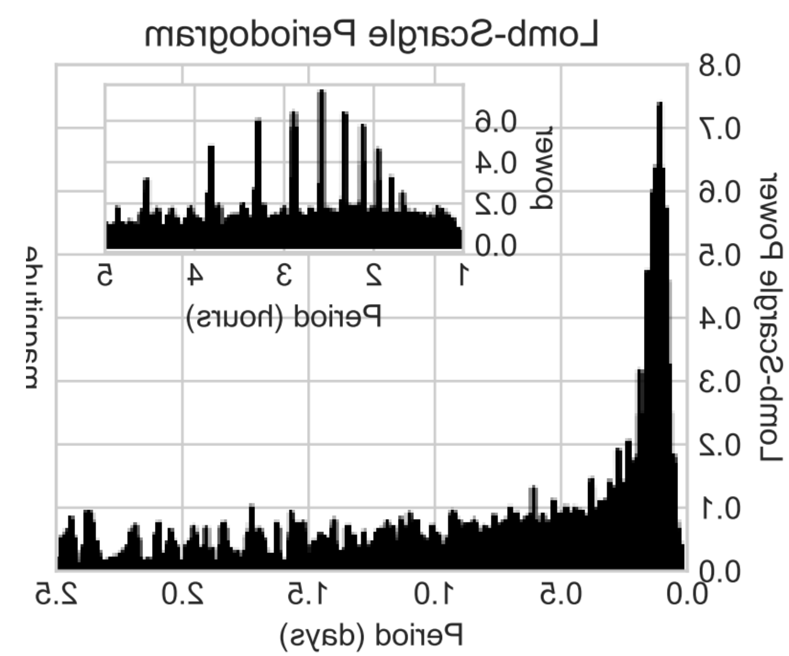
We can use bootstrap to answer this question:

1. Resample with replacement the periodogram keeping the temporal variable fixed.

2. In each case, record the resulting maximum of the periodogram.

For enough resamplings, the distribution of these maxima will approximate the true distribution for the case with no periodic signal present.





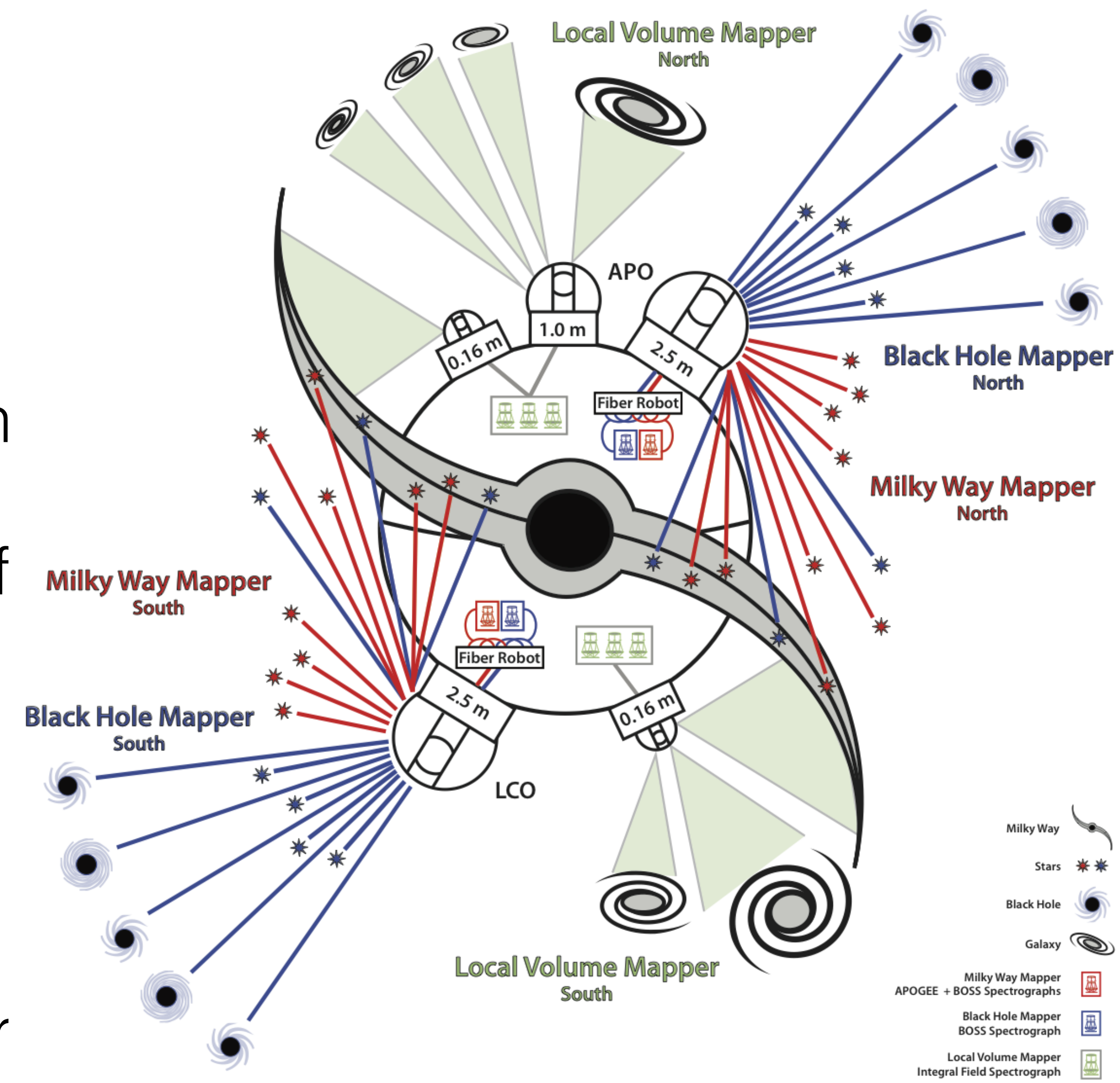
SDSS

- The Sloan Digital Sky Survey is one of the most successful surveys to day. It has obtained

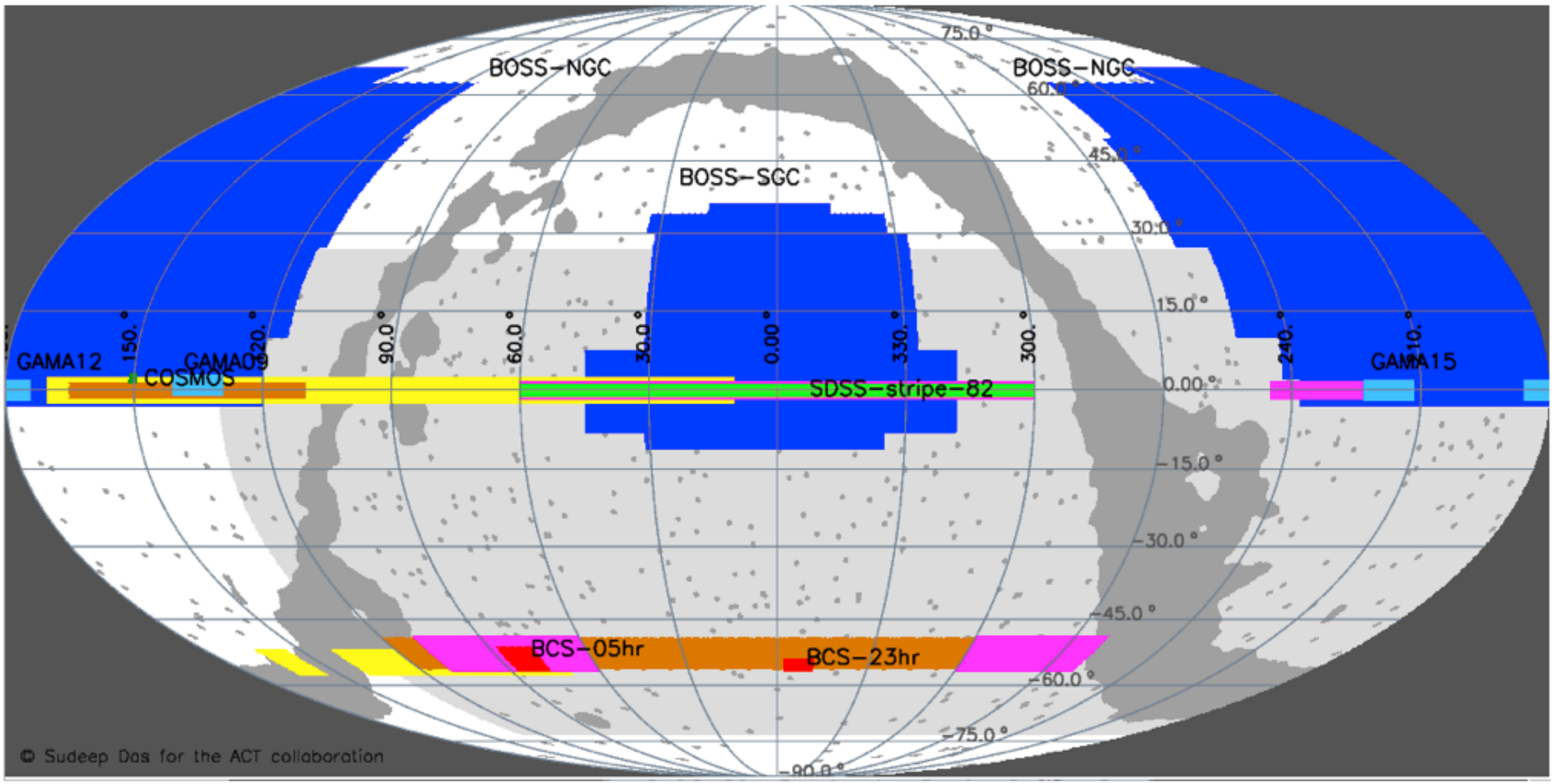
- It has collected over 3.5 million spectra and photometry for over 500 million objects in 15 years, and the current catalogue has a size of over 30,000 GB.

- Multi-epoch observations of millions of sources.

- SDSS is currently preparing for its phase V (all-sky, multi-epoch spectroscopy)



SDSS Stripe 82

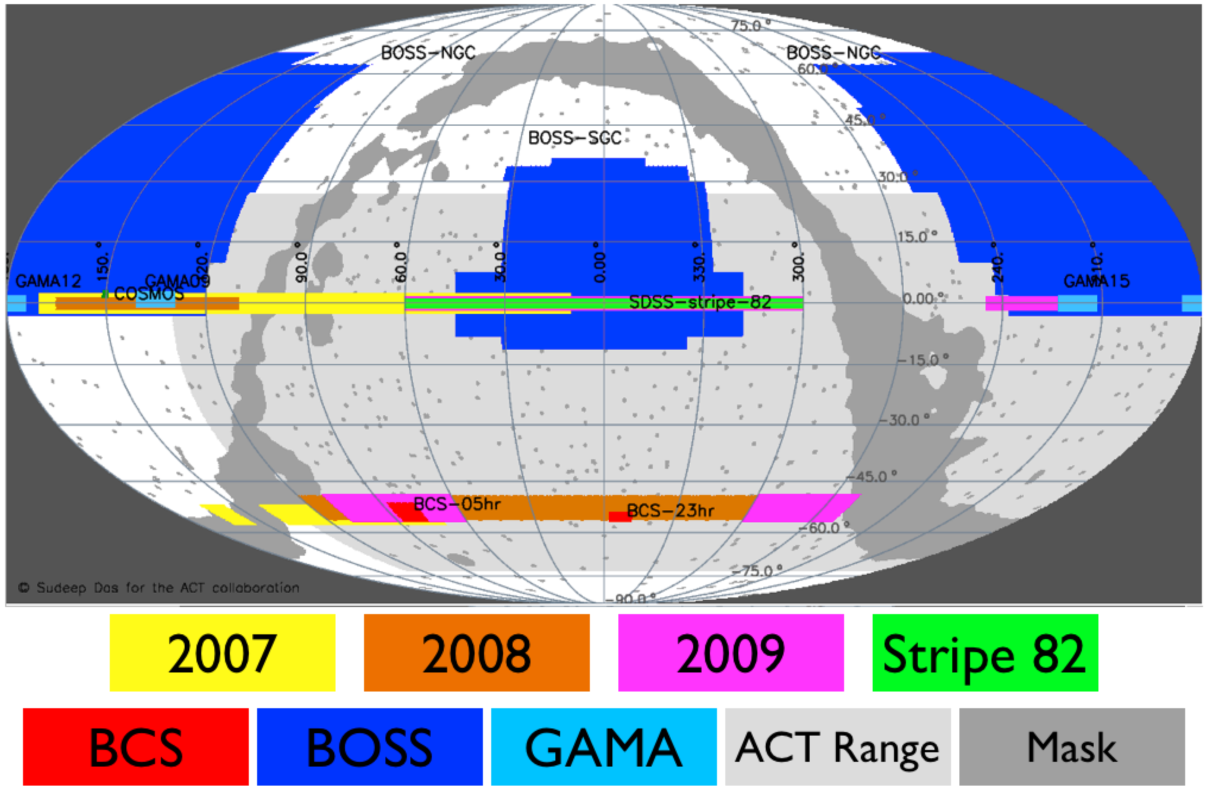


2007 | 2008 | 2009 | Stripe 82

BCS | BOSS | GAMA | ACT Range | Mask

SDSS Stripe 82

- 5 bands: ugriz
- ~60,000 variable sources
- ~50 epoch/object
- photometry roughly simultaneous across bands.



We will now look at a dataset from Stripe 82 and search for RR Lyrae stars

Exercise:
Finding periods of variable stars

- Load the Light Curves provided.

- Plot the unfolded light curves in the different SDSS bands. Can you say anything about cadence (How often are datapoint taken?)

- Use the Lomb-Scargle periodogram method to find the most likely period of the lightcurves. Is the period found similar in all bands?

- Plot the folded light curves? What kind of variability do you see?

- The use of bootstrapping

- See Jupyter notebook.