Time Series Applications in Astronomy

Pavlos Protopapas

Institute for Applied Computational Science (IACS)
Harvard

Collaborators: Isadora Nun, K.Pichara, D-W Kim, P. Huijse, P. Estevez, J. Principe, P. Zegers, etc

Traditional Astronomy

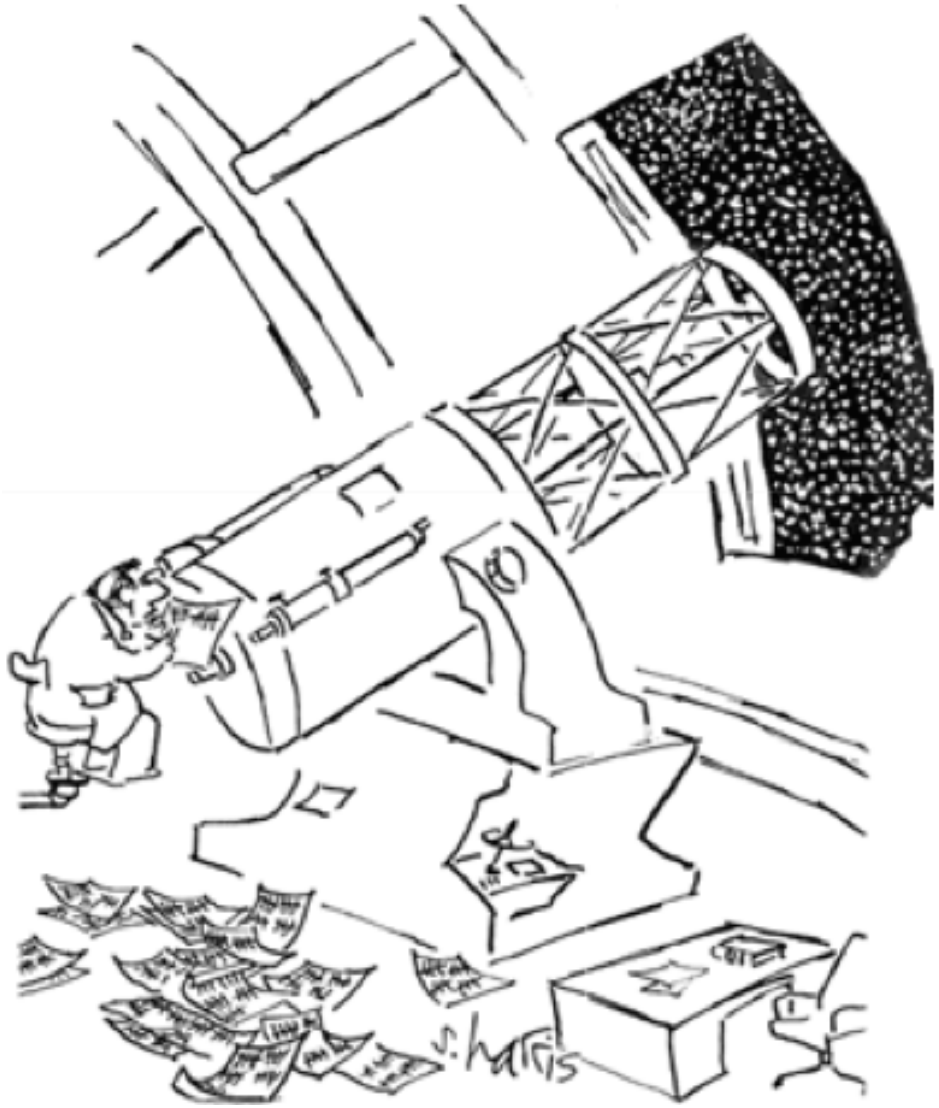


image taken from K. Borne

Today

image taken from K. Borne

Good news/bad news

- Bad news: We are facing a tremendous challenge
  - Storage
  - Transfer
  - Analysis
- Good new: Data science is SEXY

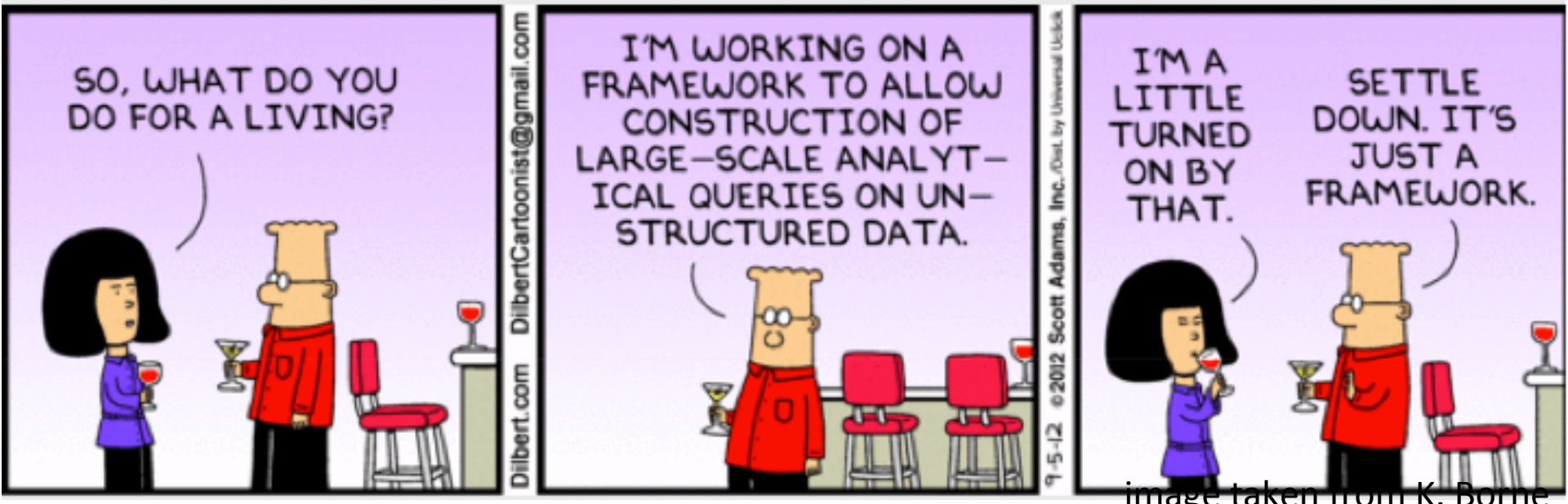


image taken from K. Borne

Unanswered Questions/Wish List

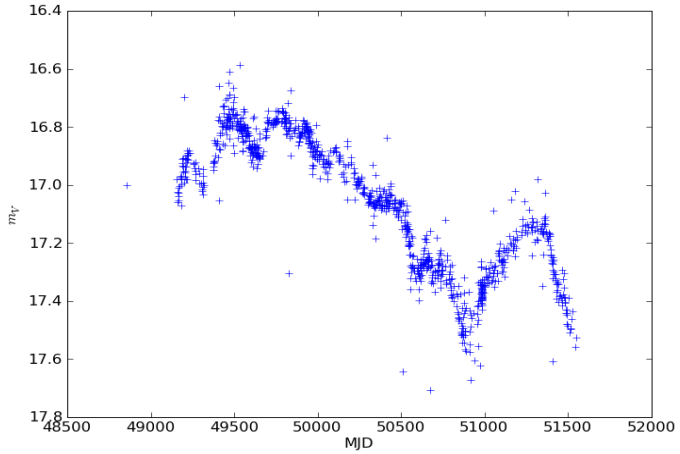
- Classification

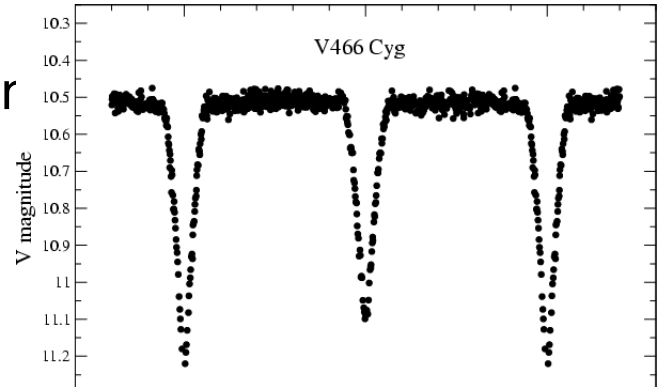
  Be able to classify objects based on their variability characteristics: quasars, variable stars, supernovae, etc

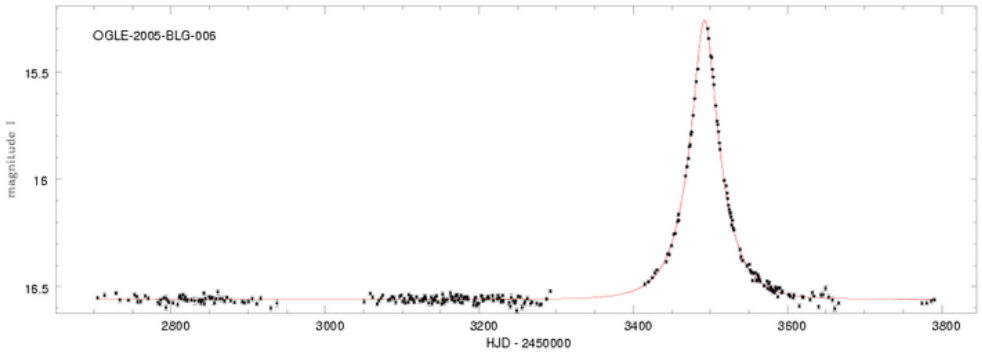
- Event detection of rare, low signal-to-noise ever

  Occultation, Microlensing, Stellar Flares

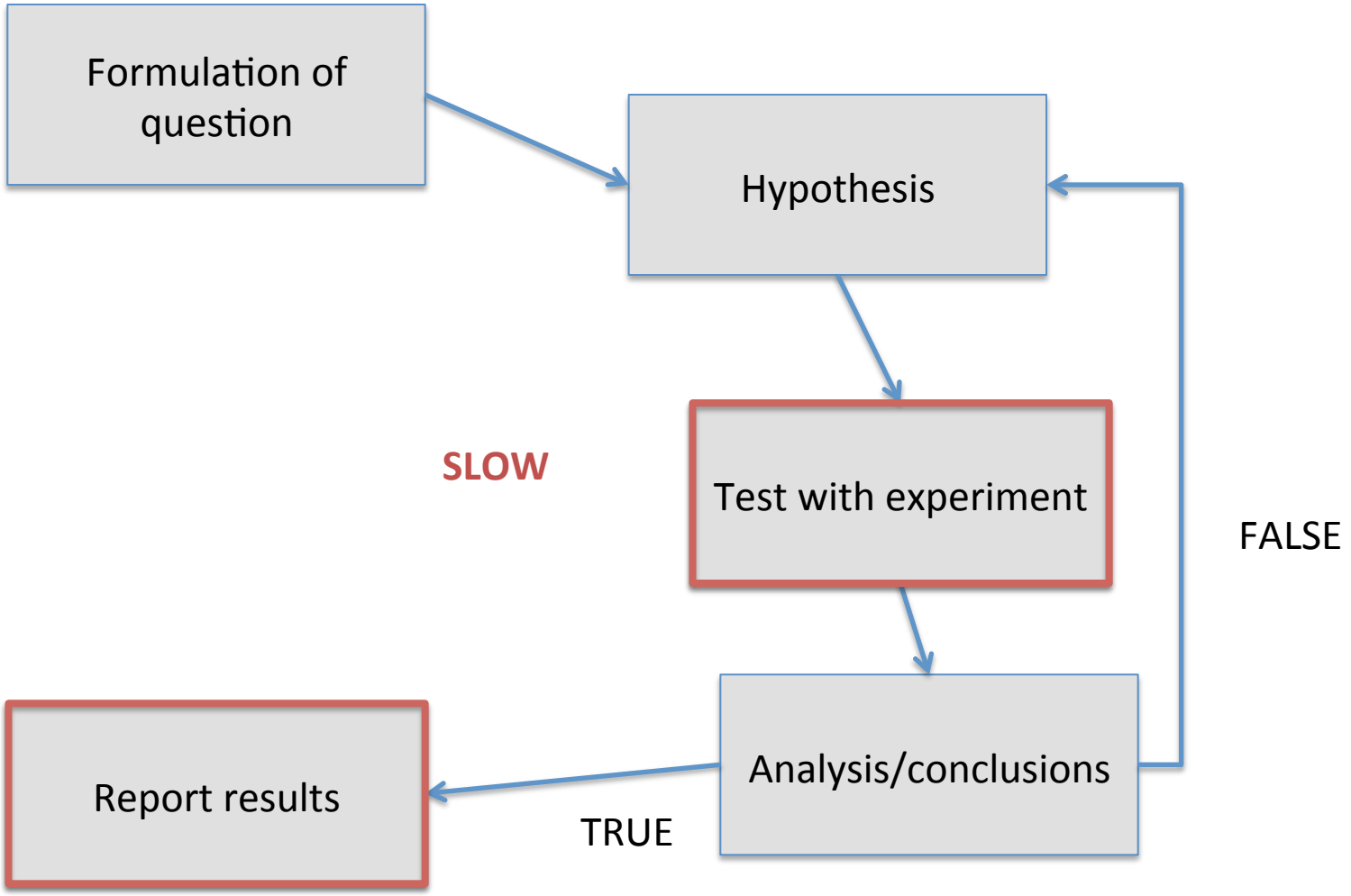
- Time Series modeling
- Designing observations



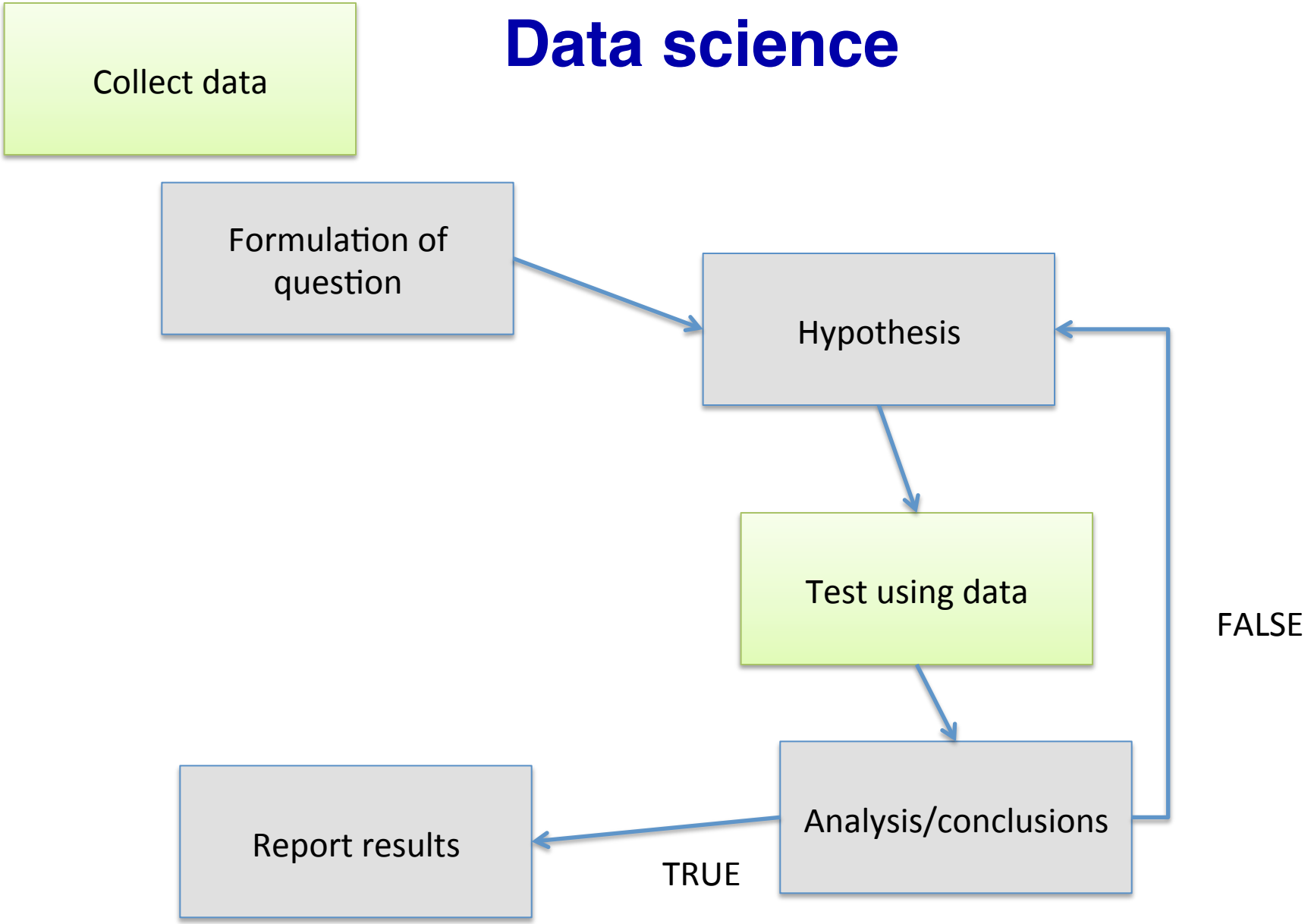




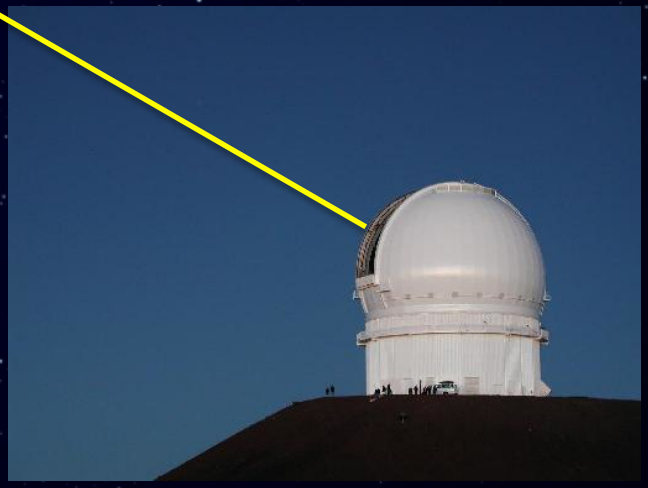
Traditional science



Data science



Time t0

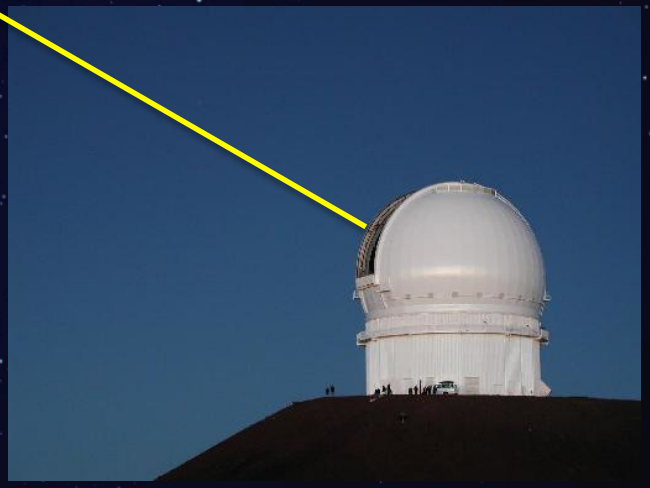


brightness

time

t0

Time t1



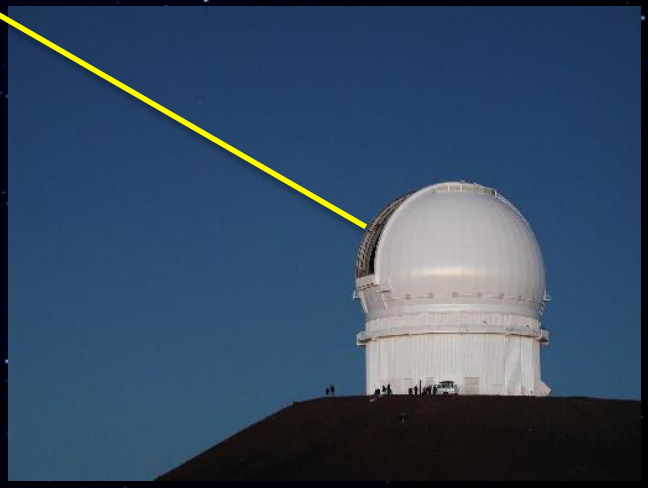
brightness

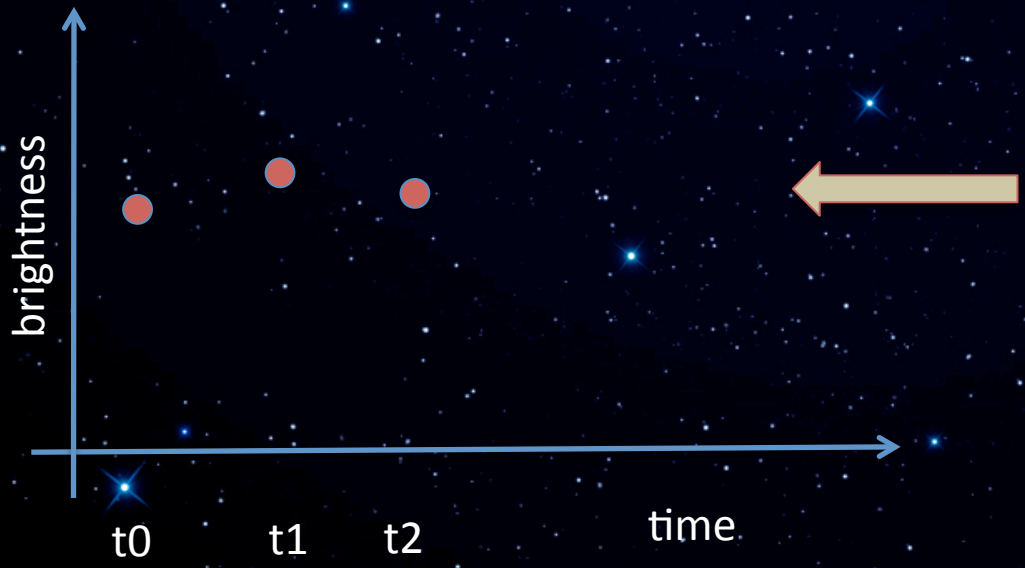
time

t0

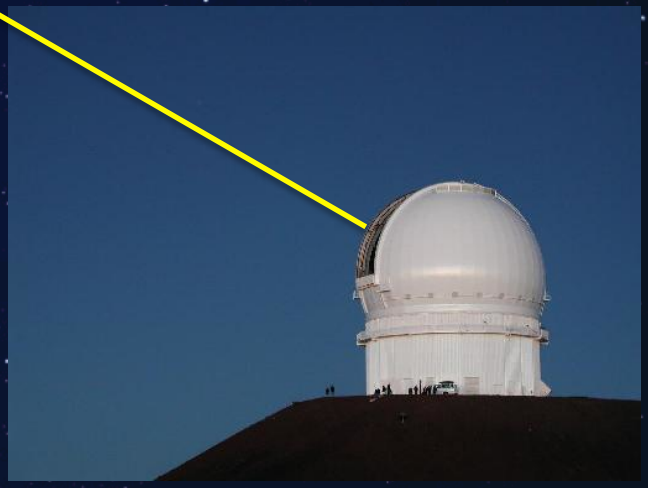
t1

Time t2





Time t2



brightness

t0 t1 t2 t3 time

Brief history of data collection in astronomy

Humans have collected data over the long history of astronomy,

- Cuneiform tablets of ancient Babylon (700 BC) about Venus
- The Greeks, the Chinese, the Indians, [add your favor civilization]
- Tycho Brahe
- …

CCD + modern computers

- MACHO and related surveys for dark matter objects (1990-2000): ~ 10 Terabytes
- Digitized Palomar Sky Survey: 3 Terabytes
- 2MASS (2-Micron All-Sky Survey): 10 Terabytes
- GALEX (ultraviolet all-sky survey): 30 Terabytes
- Sloan Digital Sky Survey (1/4 of the sky): 40 Terabytes
- Pan-STARRS (2013): 40 Petabytes
- LSST (2021): 100 Petabytes

The next year that number is going to double, and the year after that it will double again, and so on and so on.

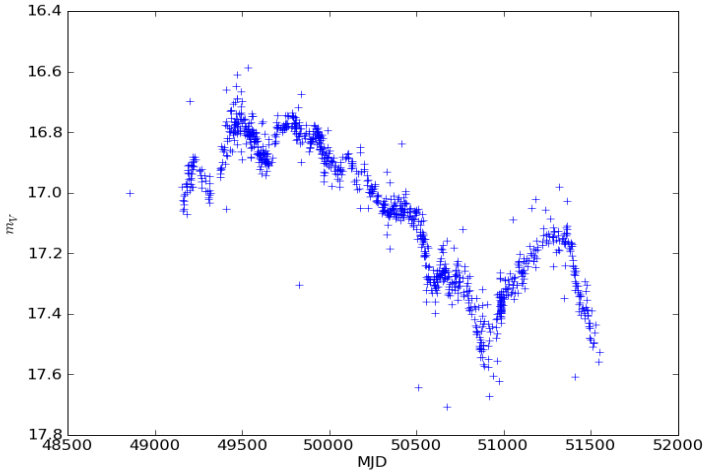
Questions/Wish List

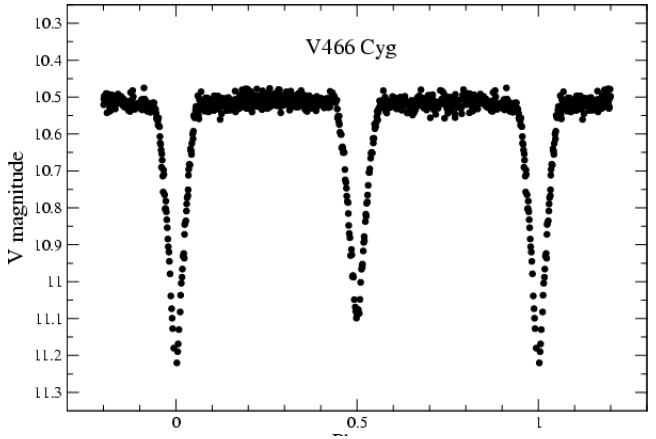
- Classification

  Be able to classify objects based on their variability characteristics: quasars, variable stars, supernovae, etc

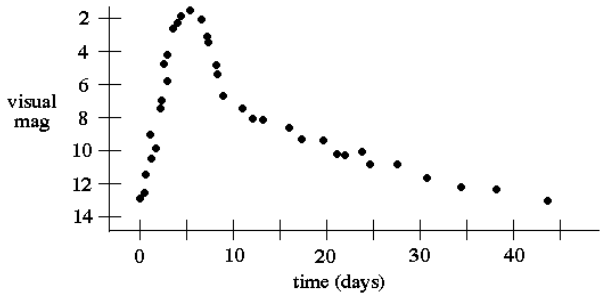
- Period finding

  For sparse and noisy data, period determination is not easy.



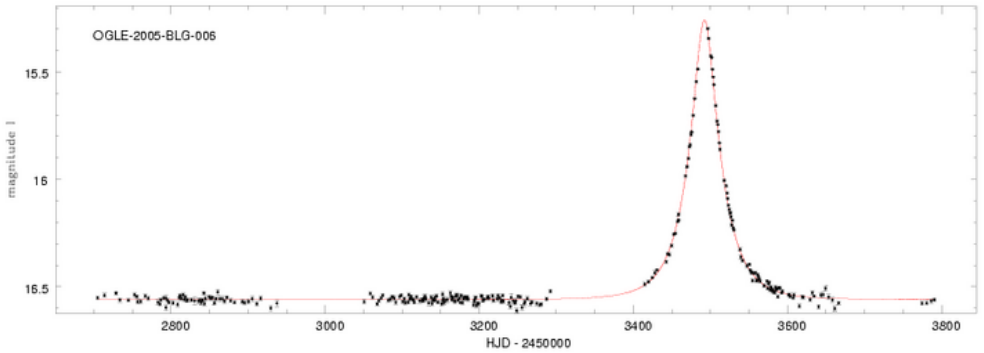


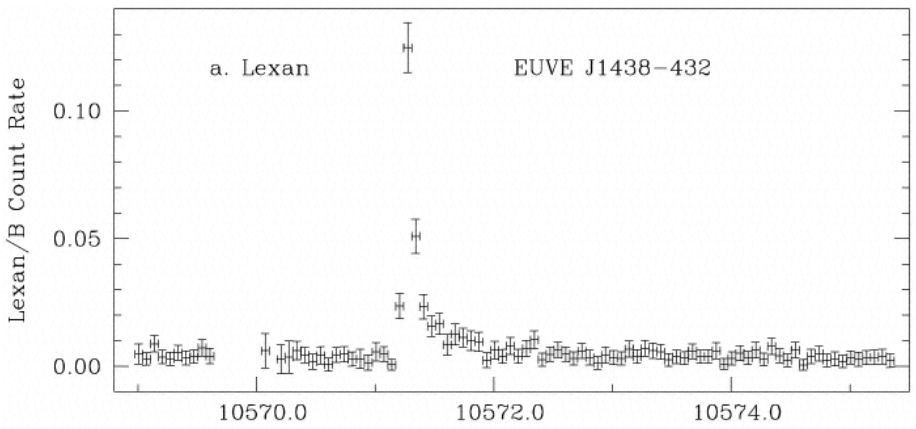
Nova Light Curve

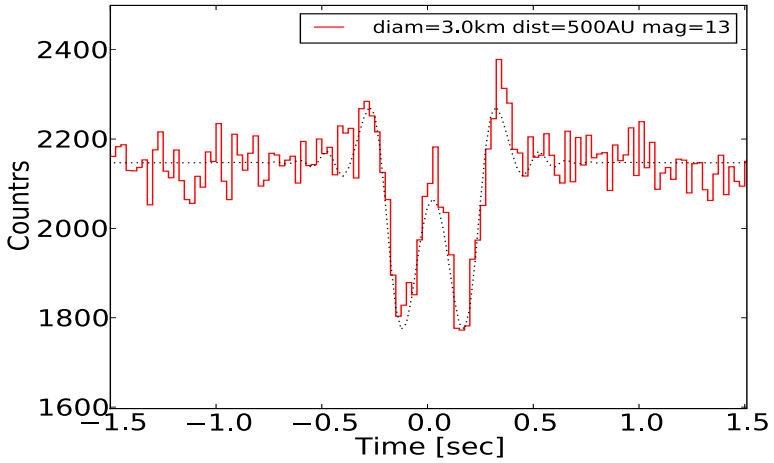


Questions/Wish List

- Event detection of rare, low signal-to-noise events
  - Occultation
  - Microlensing
  - Stellar Flares







Questions/Wish List

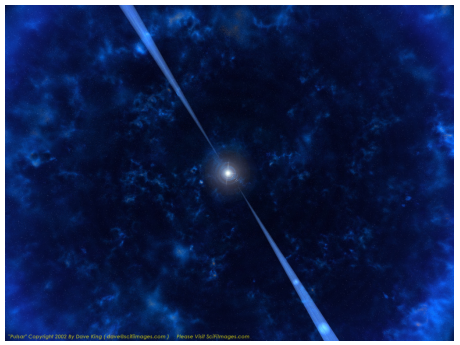
- Time Series modeling
  - Autoregressive model
  - Gaussian processes
  - …

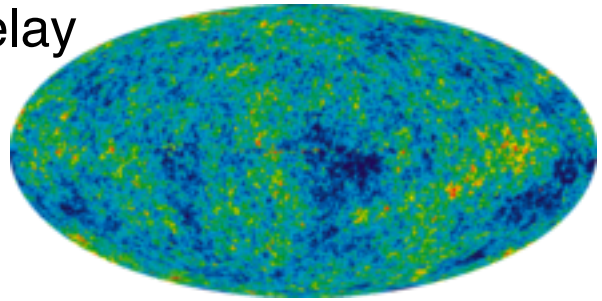
$$Y_t = \sum_{i=1}^{p} \alpha_i Y_{t-i} + \epsilon_t$$

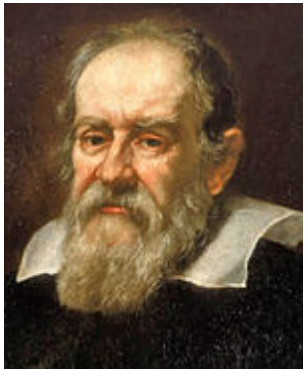
- Designing observations
  - Use model and observations to design future observations

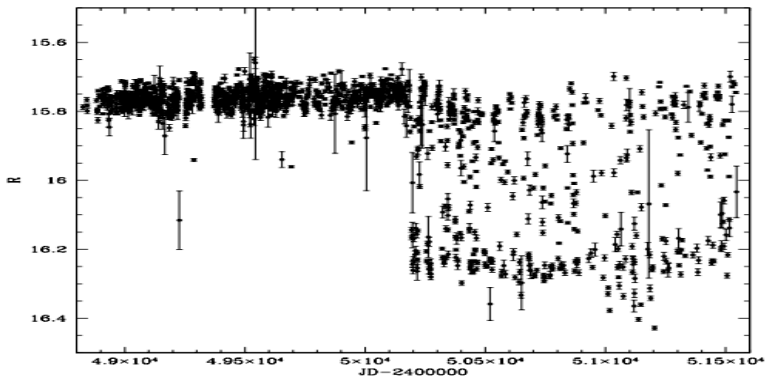
Questions/Wish List

- Novelty detection
  - Classify something as novel serendipitously
    - Pulsars (Jocelyn Bell Burnell and Antony Hewish 1967 while looking for Quasars).
    - CMB (Arno Penzias and Robert Wilson 1967 using a horn antenna designed to relay telephone calls via satellite)
  - Four Jovian moons (Galileo 1609)



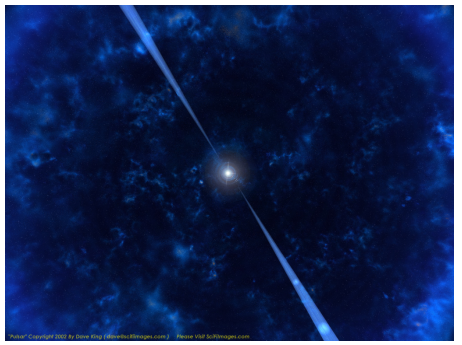


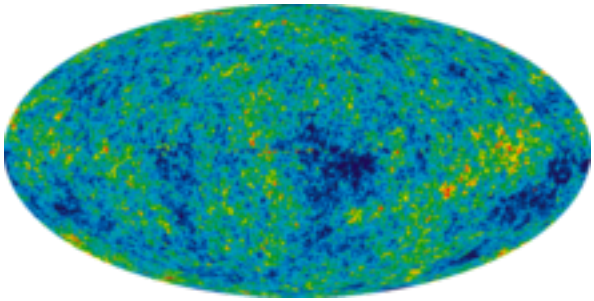


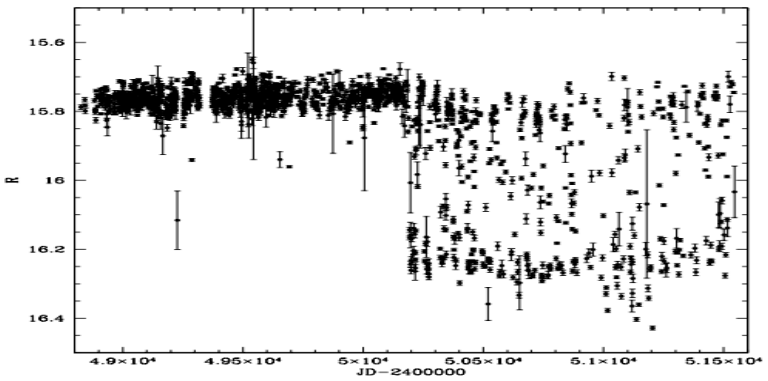


Questions/Wish List

- Novelty detection
  - Classify something as novel serendipitously
    - Pulsars (Jocelyn Bell Burnell and Antony Hewish 1967 while looking for Quasars).
    - CMB (Arno Penzias and Robert Wilson 1967 using a horn antenna designed to relay telephone calls via satellite)
  - Four Jovian moons (Galileo 1609)







Outlier detection (event detection)

- Event detection is to task of finding something we know about, but it is rare or difficult to find

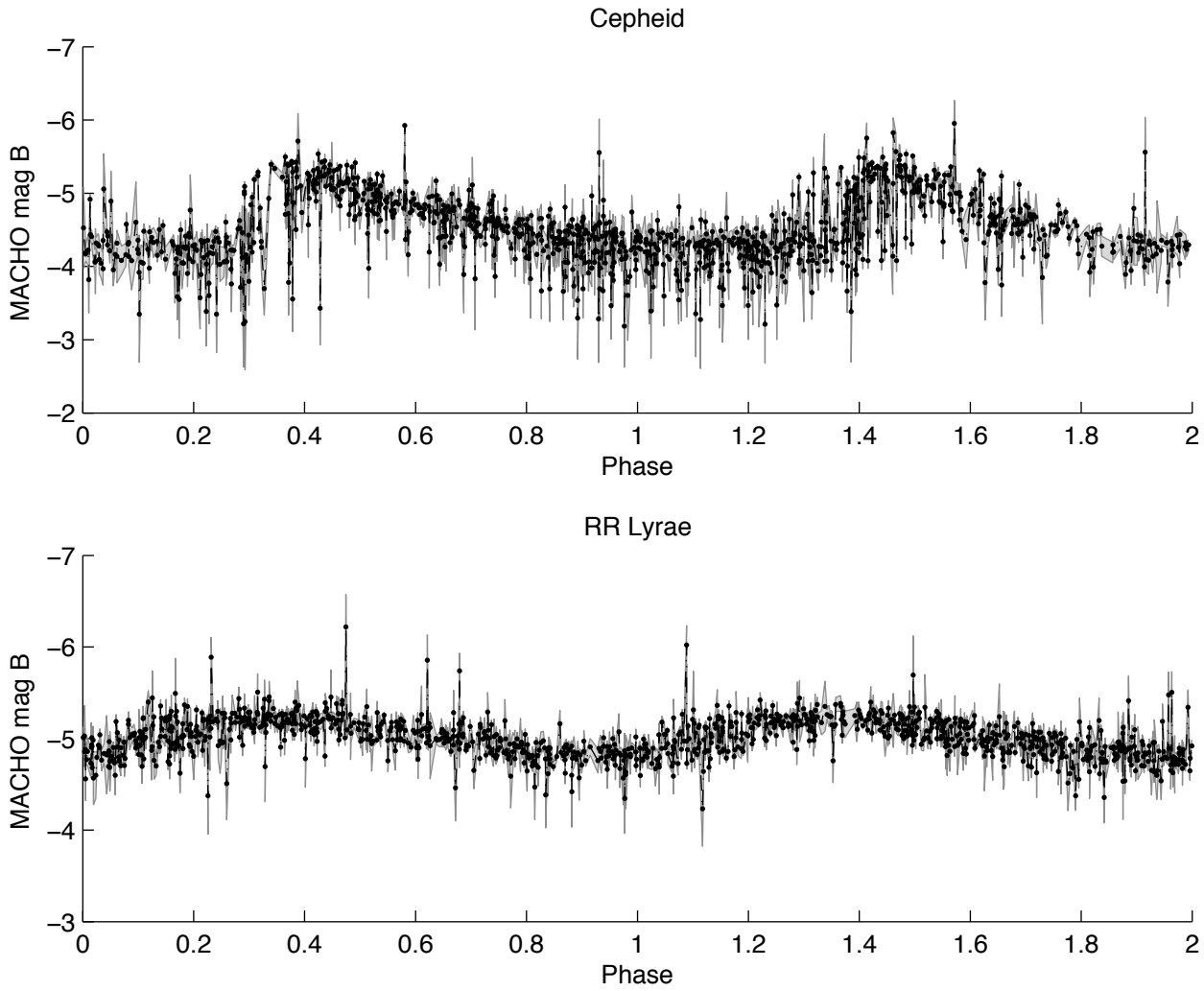
  Preston et. al. 2008, Blocker et. al 2011

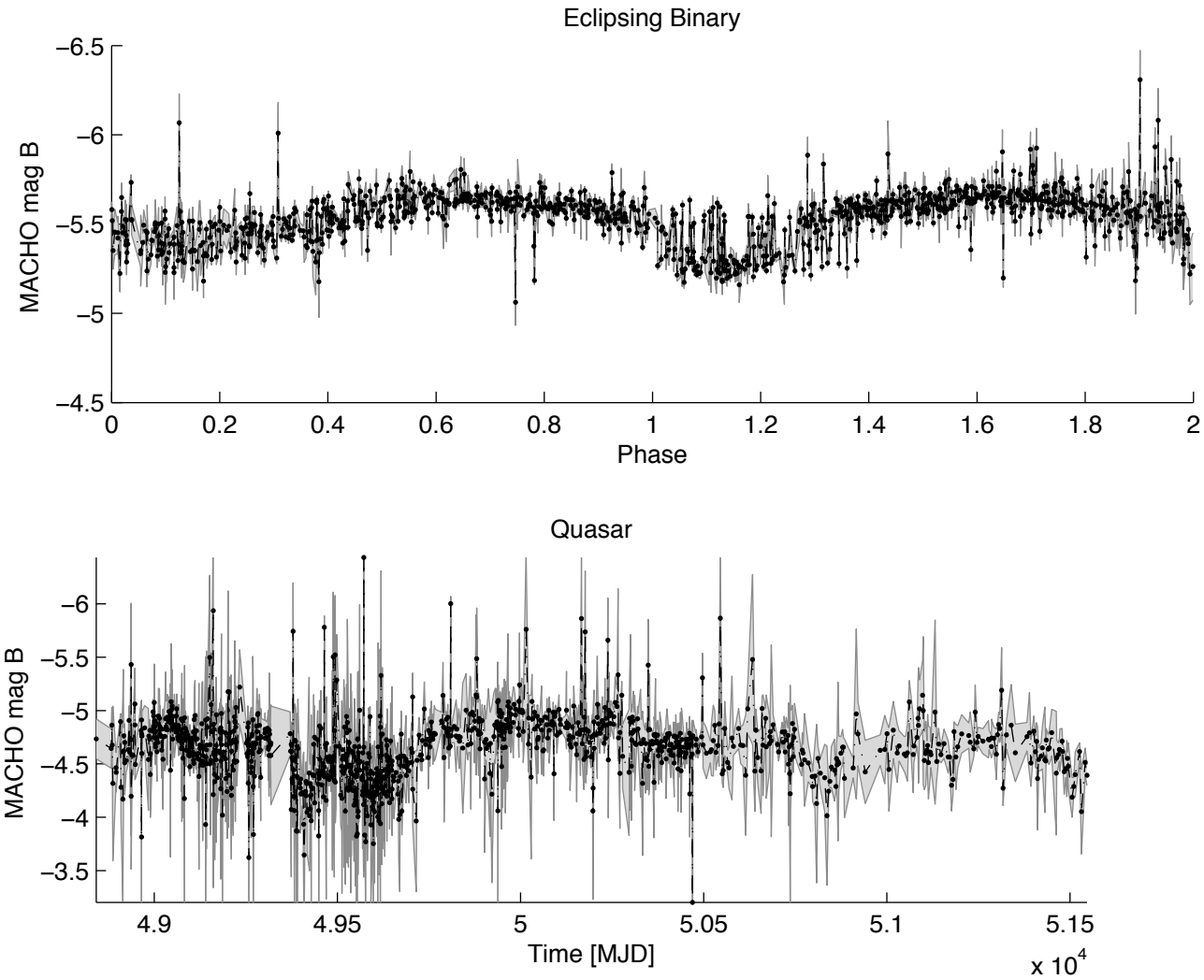
  - Scan statistics (sequential scan through the data)
  - Rank statistics
  - Found all microlensing in EROS/OGLE and some new

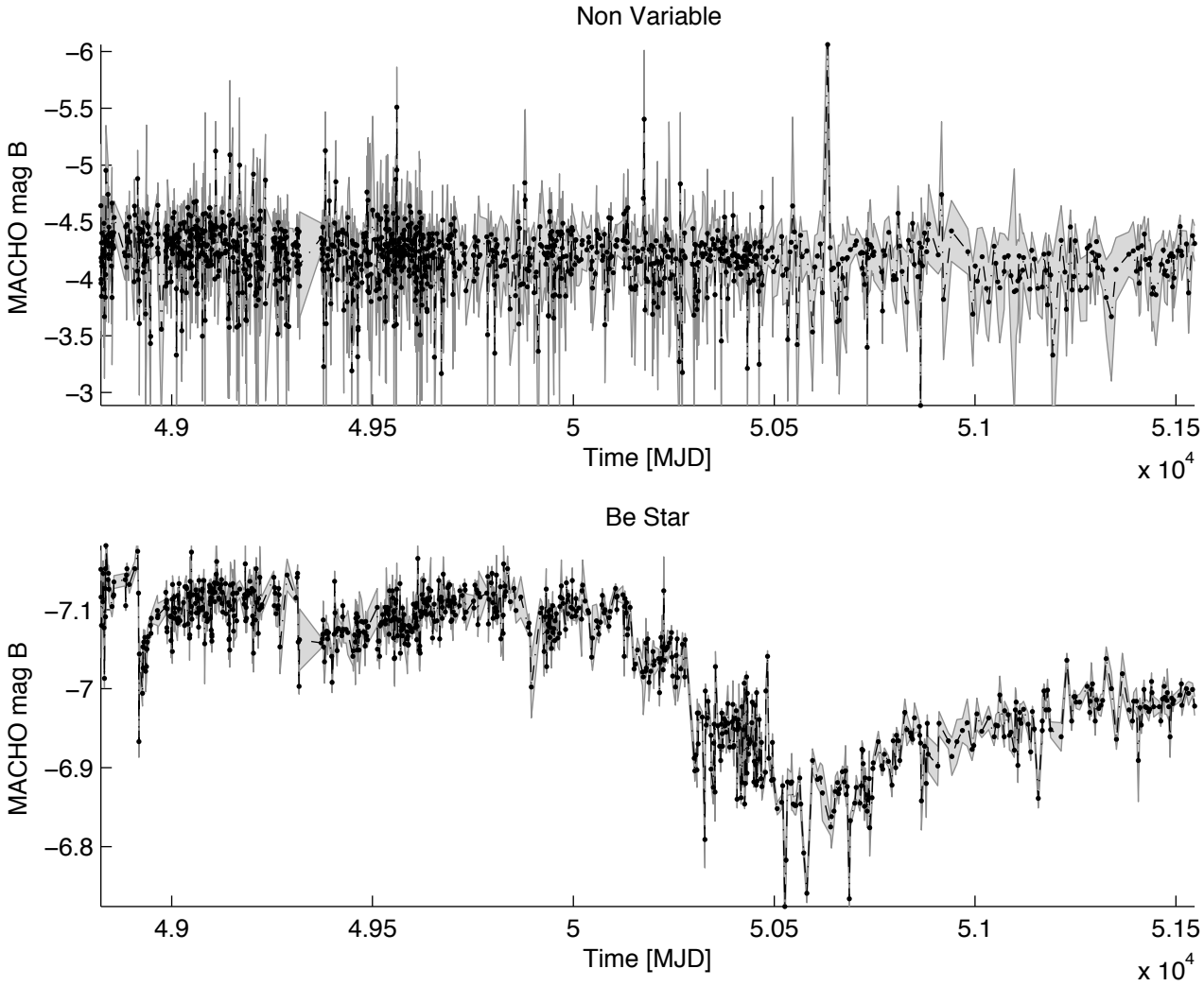
- **Outlier detection** is the task of finding something we did not know about before

  Rebragaata et. al. 2007, Keogh et. al. 2008, Richards et al 2012 etc

What we know


Cepheid
MACHO mag B
−7
−6
−5
−4
−3
−2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Phase
RR Lyrae
MACHO mag B
−7
−6
−5
−4
−3
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Phase

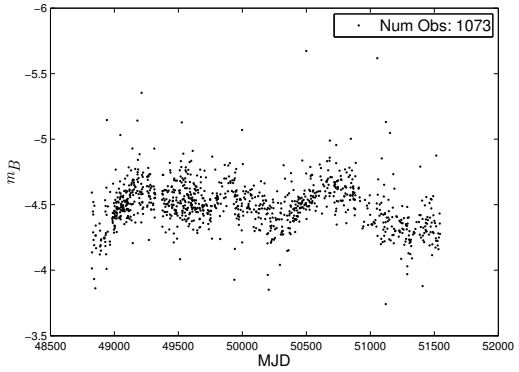

Eclipsing Binary
MACHO mag B
−6.5
−6
−5.5
−5
−4.5
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
Phase
Quasar
MACHO mag B
−6
−5.5
−5
−4.5
−4
−3.5
4.9
4.95
5
5.05
5.1
5.15
Time [MJD]
x 10^4

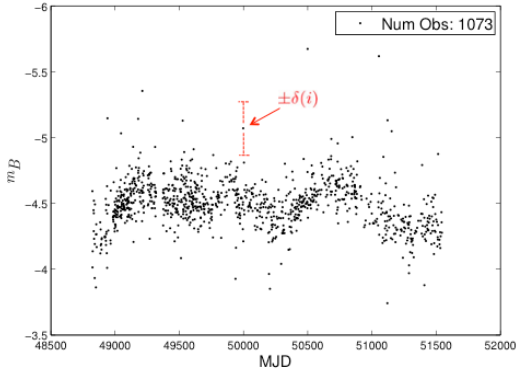

Non Variable
MACHO mag B
−6
−5.5
−5
−4.5
−4
−3.5
−3
4.9
4.95
5
5.05
5.1
5.15
Time [MJD]
x 10^4
Be Star
MACHO mag B
−7.1
−7
−6.9
−6.8
4.9
4.95
5
5.05
5.1
5.15
Time [MJD]
x 10^4

Lightcurve Representation, FEATURES

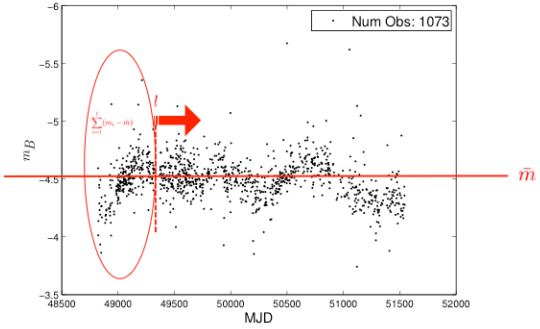
Numerical descriptors of Lightcurves

Lightcurve





$$K = \frac{1}{\sqrt{N}} \frac{\sum_{i=1}^{N} |\delta(i)|}{\sqrt{\sum_{i=1}^{N} \delta(i)^2}}$$



$$S_l = \frac{1}{N\,\sigma} \sum_{i=1}^{l} (m_i - \bar{m})$$

$$R_{cs} = \max(S) - \min(S)$$