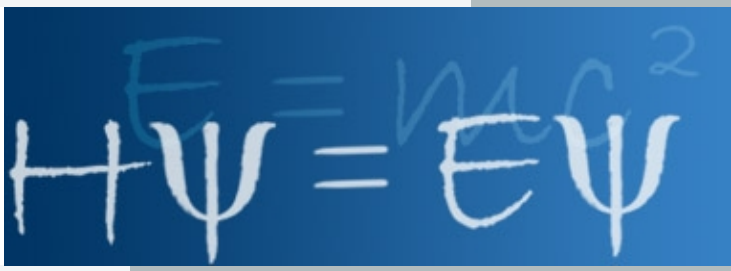
Philipps Universität Marburg

Development of an Open Source light curve classificator

Christian Dersch



Philipps-Universität Marburg, Fachbereich Physik, AG Astronomie

15. September 2015

Christian Dersch

University part

- Master student (physics) @University of Marburg
- Bachelor in 2013: "Photometry at Wolf-Rayet-Stars"
- Currently working on master thesis "Knowledge Discovery in the OGLE-III database of variable stars"

Community part

- Developer @Fedora Linux
- Working in groups for Astronomy, Science & Technology
- Including packages like astroML or Astromatic Tools in Fedora

Astronomy Group at University of Marburg

Gerling Observatory

Research group

- History of Astronomy and Observational Astronomy
- Prof. Dr. Andreas Schrimpf

Research topics

- Analysis of Sonneberg Plate Archive
  - Poster "First Steps Towards a Photometric Analysis of the Sonneberg Sky Patrol Plates" by M. Spasovic
- **Data Mining in context of variable stars**
- History of Astronomy (Christian Ludwig Gerling)

In cooperation with

- Sonneberg Observatory (P. Kroll)
- Department of Physics & Astrophysics, University of Delhi, India (H. P. Singh)

About this Talk

Concept:

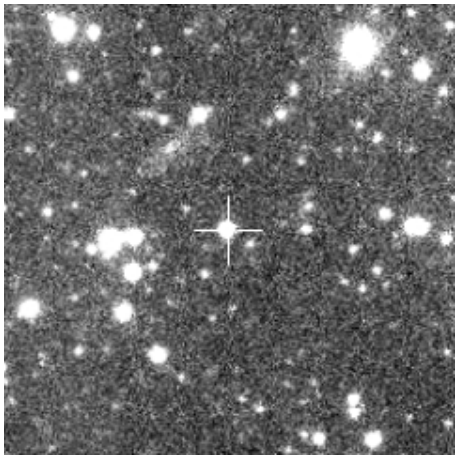
- Presentation of a concept for development of a classificator
- Based on experiences from work @master thesis
- Work in progress $\Rightarrow$ Focussed on preparation of data

What we want to have

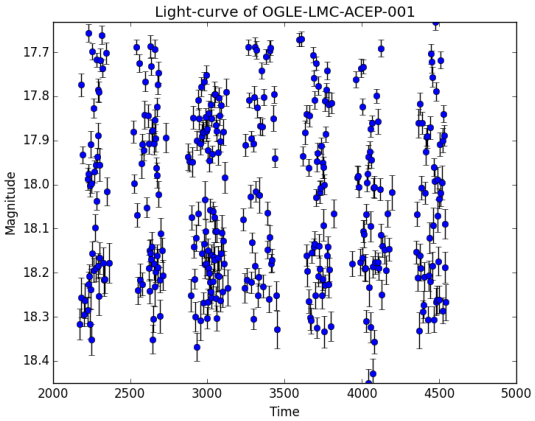
- Reliable classification of light curves
- Different methods depending on ambition
- Here: Machine Learning (ML)
  - Supervised: Filter for known classes
  - Unsupervised: Search for new (sub-) classes
- Reproducible results

State of the Art - Light curve analysis

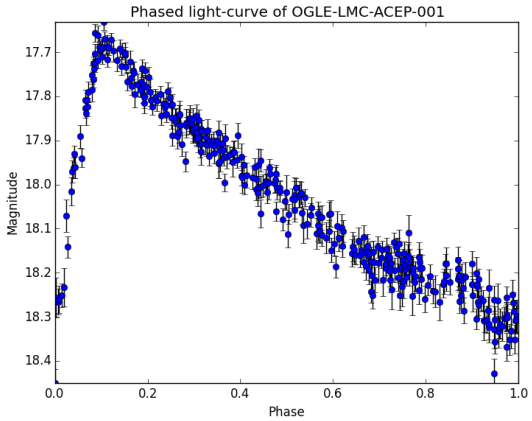
From measurement to final results



(a) Measurement



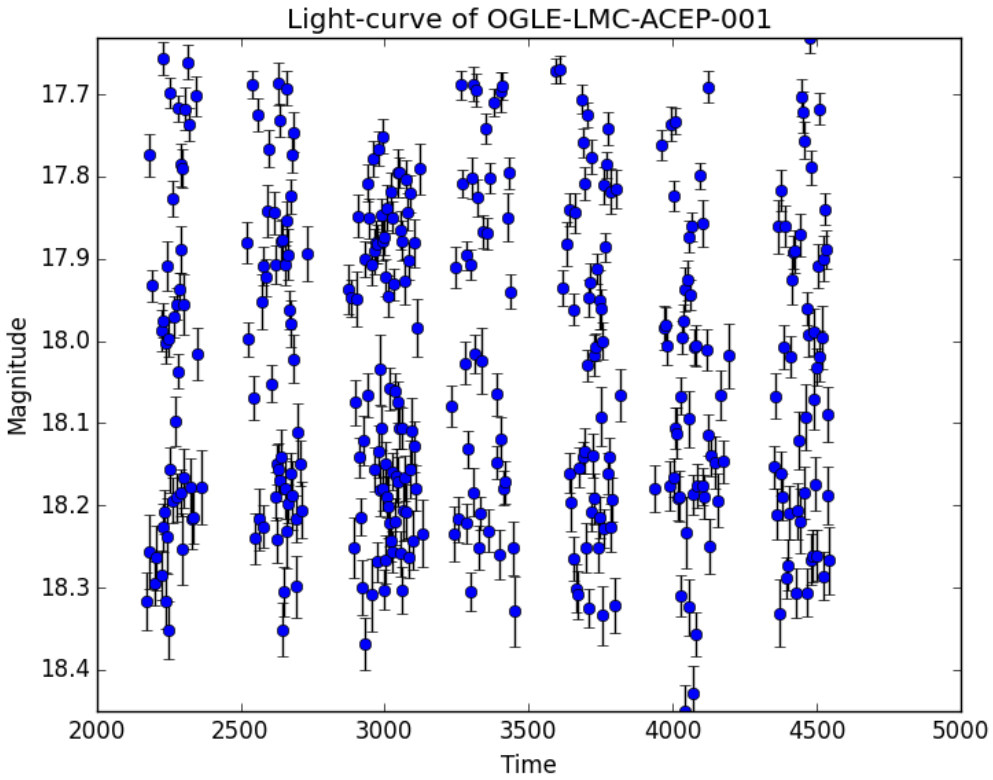
(b) Light curve



(c) Phased light curve

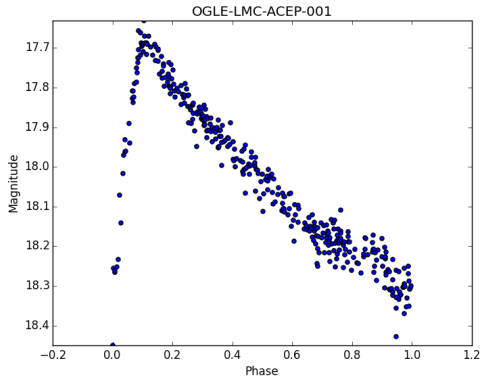
- Look at (c): "Oh, looks like an anomalous Cepheid???"
- Output of step (c) $\Rightarrow$ Machine Learning
- Interpret output of ML-based analysis

The input data

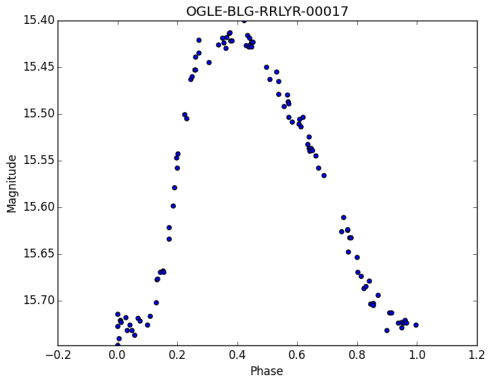


- Photometrical data reductions done
- Problems
  - Data unequeally spaced at time axis
  - $\Rightarrow$ Not a direct input for ML
- We have to calculate classifiable data from input

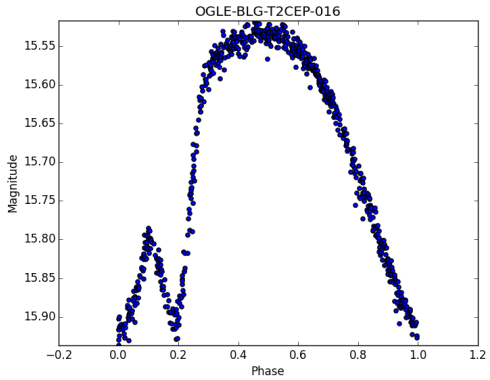
How to solve?



(d) ACEP



(e) RRLYR



(f) T2CEP

Assumption (for master thesis): periodic variable stars

- Idea: Fold all measured data in one "period" and analyze $mag(\phi(t))$ instead of $mag(t)$

$$\phi(t) = \frac{t - t_0}{P} - \text{Int}\left(\frac{t - t_0}{P}\right)$$

- Result: We loose dependency on date of measurement and get characteristic light curves
- Standard technique in field of variable stars

Generating classifiable data

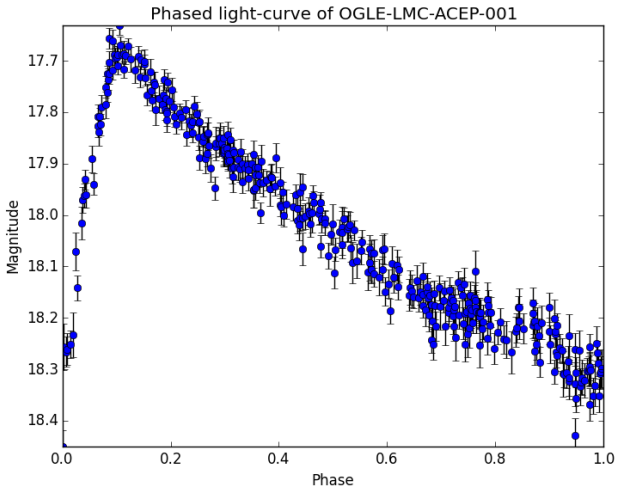
- Problem: Data still unequally spaced and not normalized
- Solutions:
  - Normalize mag into interval $[0, 1]$
  - Fit a polynomial to the normalized phased light curve
    - Use fit parameters for classification
    - Generate a synthetic equispaced light curve using fitted function
  - Fit Fourier parameters R_{ij}, ψ_{ij} to light curve and use them for classification

$$m(t) = A_0 + \sum_{n=1}^{N} A_n \cos\left(2\pi n \phi\left(t\right) + \psi_n\right)$$

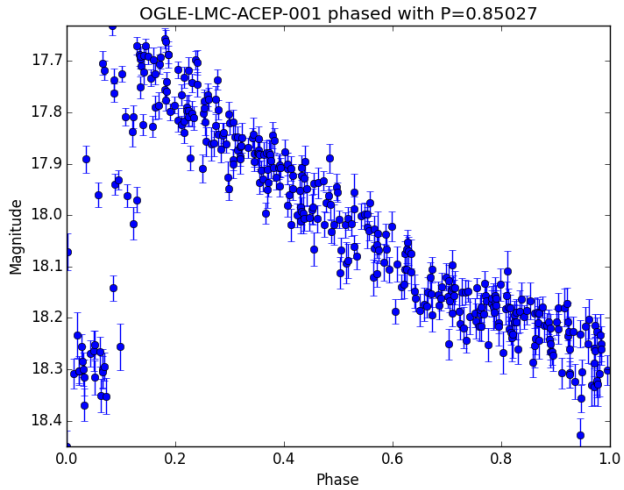
$$R_{ij} = \frac{A_i}{A_j} \qquad \psi_{ij} = \psi_i - i\psi_j$$

- Fourier-based classification (without ML) established method

Importance of period search implementation



(g) OGLE



(h) FNPEAKS

- Slightly different period: 3.5 seconds, P about 20.4 hours
- Calculated period with FNPEAKS[1]
- Calculated Fourier parameters using OGLE and FNPEAKS periods:

	R_{21}	R_{31}
OGLE	0.460 ± 0.014	0.364 ± 0.013
FNPEAKS	0.379 ± 0.039	0.219 ± 0.038

[1] http://helas.astro.uni.wroc.pl/deliverables.php?lang=en&active=fnpeaks

Measure quality

- Difference in Fourier parameters show: We need some kind of quality measurement
- Idea: Compare standard deviation of phased light curve with error of measurements
- Classification of generated equispaced light curve more reliable

After data preprocessing and optional quality measurement: Ready for classification!

Classification

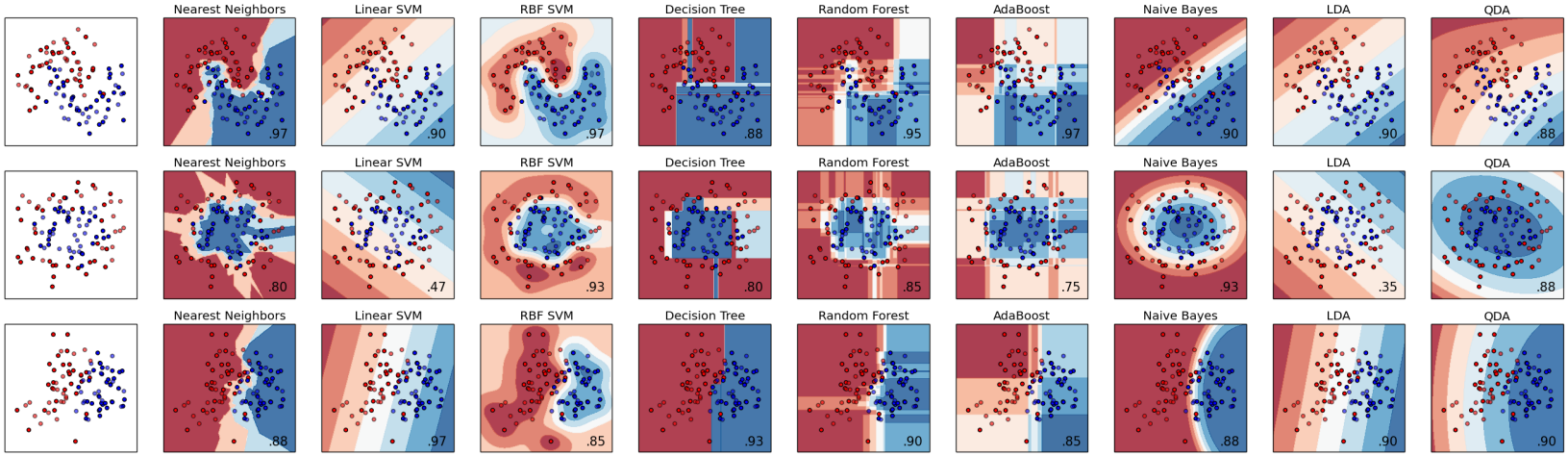
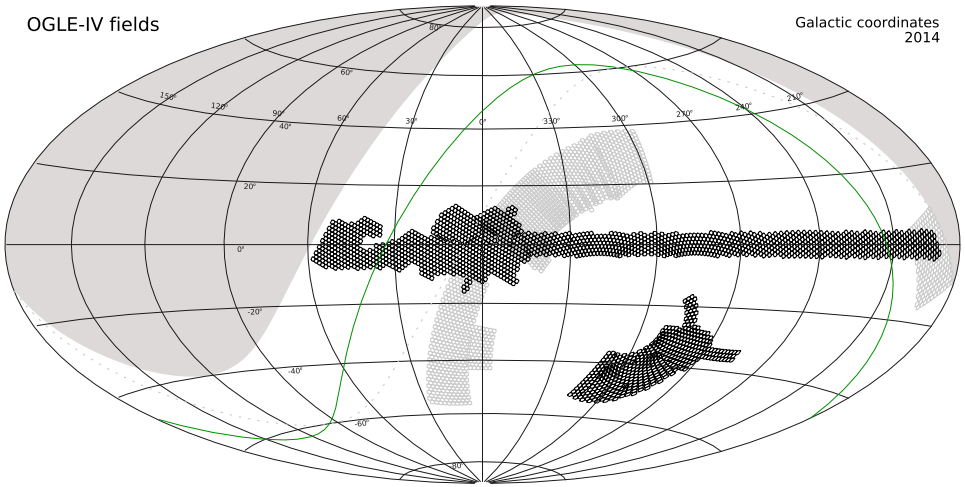


Figure: Different classifiers implemented by scikit-learn[2]

- Data now viable for analysis by common ML software
- ML toolboxes typically provide implementations for
  - Supervised Learning
  - Unsupervised Learning
  - **Principal Component Analysis (PCA)**
  - **Clustering**

[2] `http://scikit-learn.org/stable/index.html`

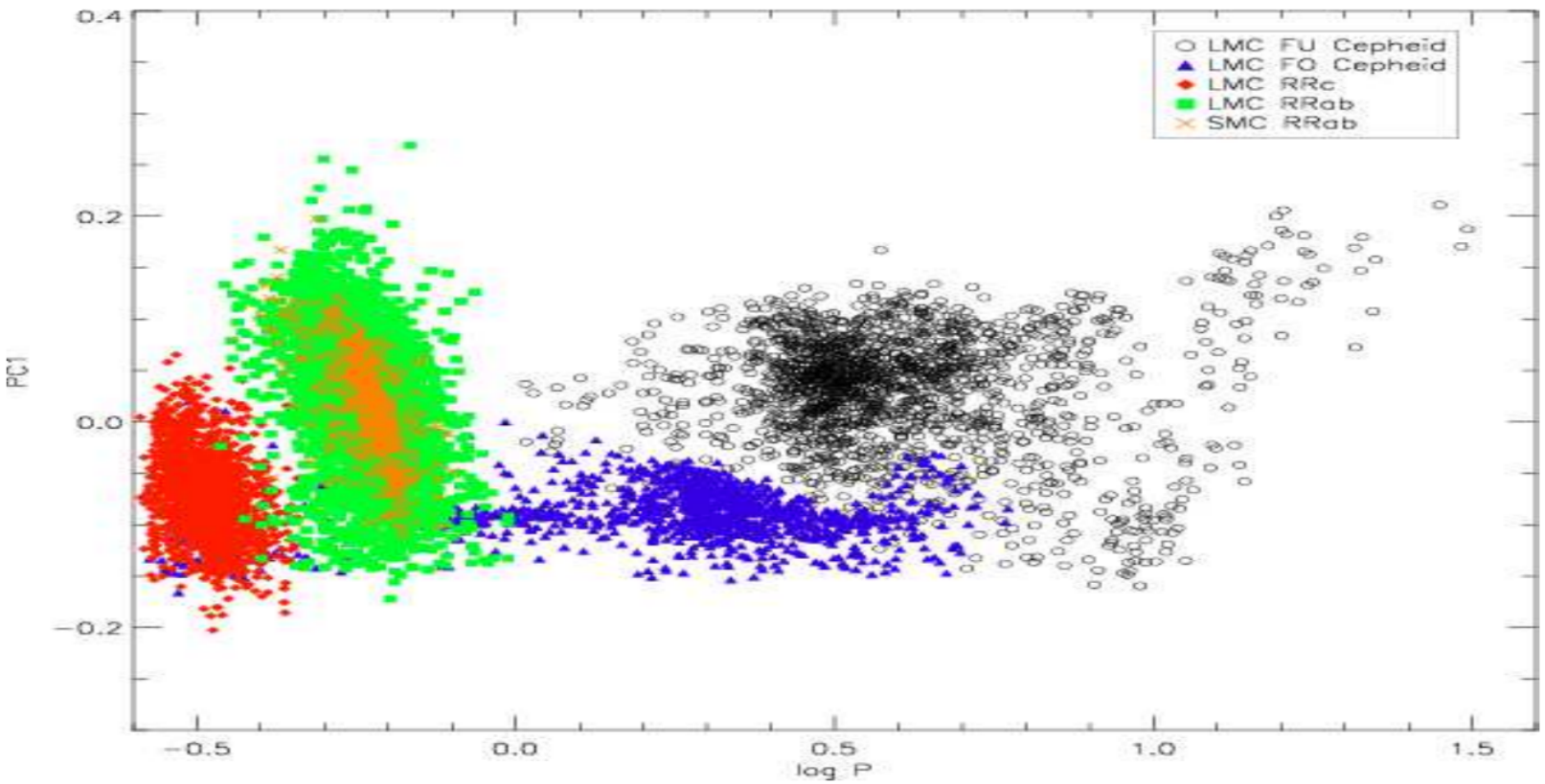
Why OGLE?



- Optical Gravitational Lensing Experiment
- Photometrical measurements
- OGLE-III catalog of variable stars (approx 400 000 stars)
- Provides "ready to use" light curves
- **Well analyzed, also ML-based**

Nice base for development of a light curve classificator

ML-based classification of OGLE-data in LMC/SMC



Deb, Singh (2009): "Light curve analysis of Variable stars using Fourier coefficients and Principal Component Analysis"

Implementation

Develop as free software

- Better reproducibility by third parties
- Reasonalbe algorithms
- Available for whole community
- Many basic packages (Python + modules) are free software too
- Avoids the "reinventation of the wheel"

How to implement? Use Python!

Which language shall we use?

- In Astronomy community: Python is a common programming language
- astropy, numpy, scipy, ...
- $\Rightarrow$ New software should at least have an interface to Python
- Guru: C/C++ is much faster...

Already done (just to be rewritten in a distributable way...)

- Data preprocessing using numpy/scipy

Available

- scikit-learn (machine learning package for Python): PCA, Clustering etc. implemented
- Python wrappers for shogun (fast C++ machine learning toolbox)

Solution: Create modular software, write critical components using C++ and generate wrappers for them (swig)

astroML

- Machine Learning and Data Mining for Astronomy
- Many tasks already implemented (periodogram for example)
- Based on
  - Python
  - scikit-learn, numpy, matplotlib
- Astropy Affiliated Package
- astroML-addons: example for integration of critical components implemented in C++

Goal: Implement light curve analysis in a way, that we can contribute it to astroML instead of initiating a new project

Conclusion

- Implement data preprocessing (draft already done)
- Quality measurement
- Test classification using scikit-learn
- Use OGLE as test data, compare with previous analysis
- Document well and publish as free Software

Thank you for your attention!

Any questions?

References

Sukanta Deb and Harinder P Singh. "Light curve analysis of variable stars using Fourier decomposition and principal component analysis". In: *Astronomy & Astrophysics* 507.3 (2009), pp. 1729–1737.

Ž. Ivezić et al. *Statistics, Data Mining and Machine Learning in Astronomy*. Princeton, NJ: Princeton University Press, 2014.

J. P. Long et al. "Optimizing Automated Classification of Variable Stars in New Synoptic Surveys". In: PASP 124 (Mar. 2012), pp. 280–295. DOI: 10.1086/664960. arXiv: 1201.4863 [astro-ph.IM].

F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

I. Soszyński. "The OGLE-III catalog of variable stars: First results". In: *IAU Symposium*. Ed. by J. T. Van Loon and J. M. Oliveira. Vol. 256. IAU Symposium. Mar. 2009, pp. 30–35. DOI: 10.1017/S1743921308028214.

J.T. Vanderplas et al. "Introduction to astroML: Machine learning for astrophysics". In: *Conference on Intelligent Data Understanding (CIDU)*. Oct. 2012, pp. 47–54. DOI: 10.1109/CIDU.2012.6382200.