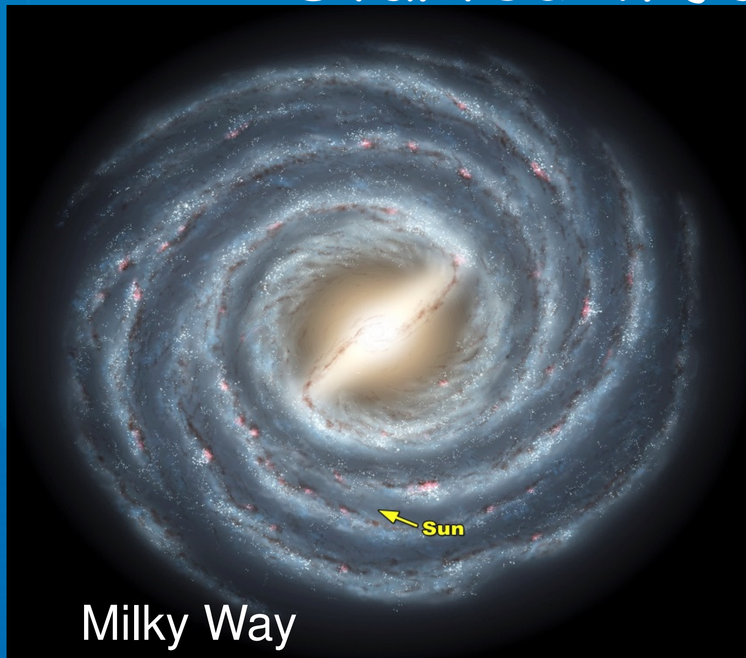
Pablo Estévez & Pablo Huijse
MAS & DEE, Universidad de Chile

Correntropy-based Spectral Methods for Detecting Periodic Variables

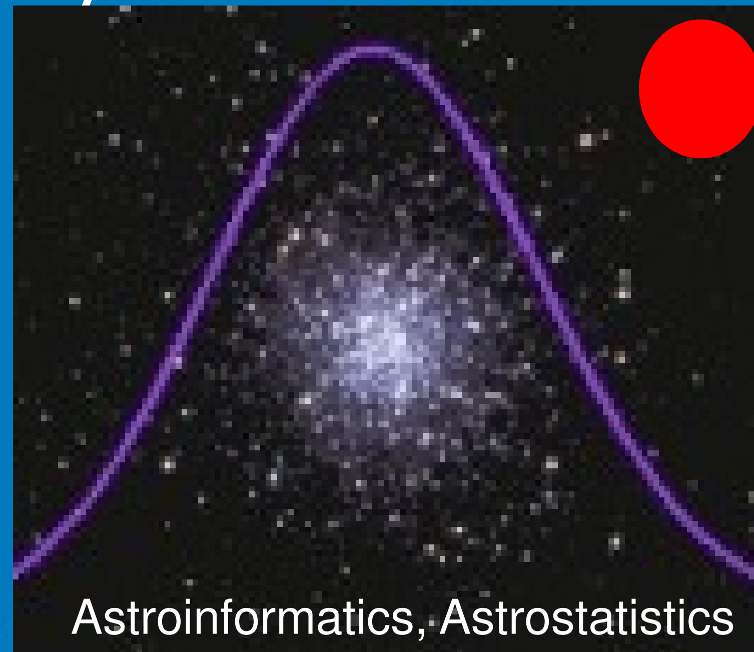
Astroinformatics 2014

Viña del Mar, Chile, August 26, 2014

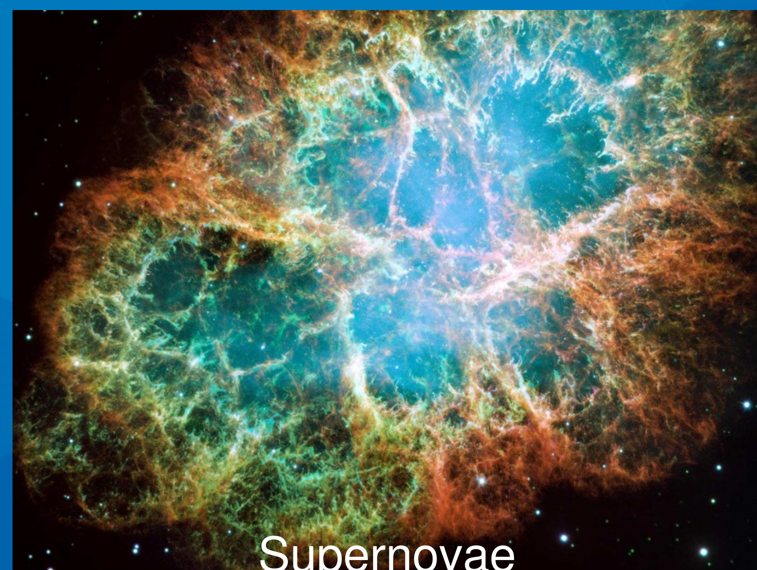
Millennium Institute of Astrophysics (MAS)
Started in January 2014



Milky Way

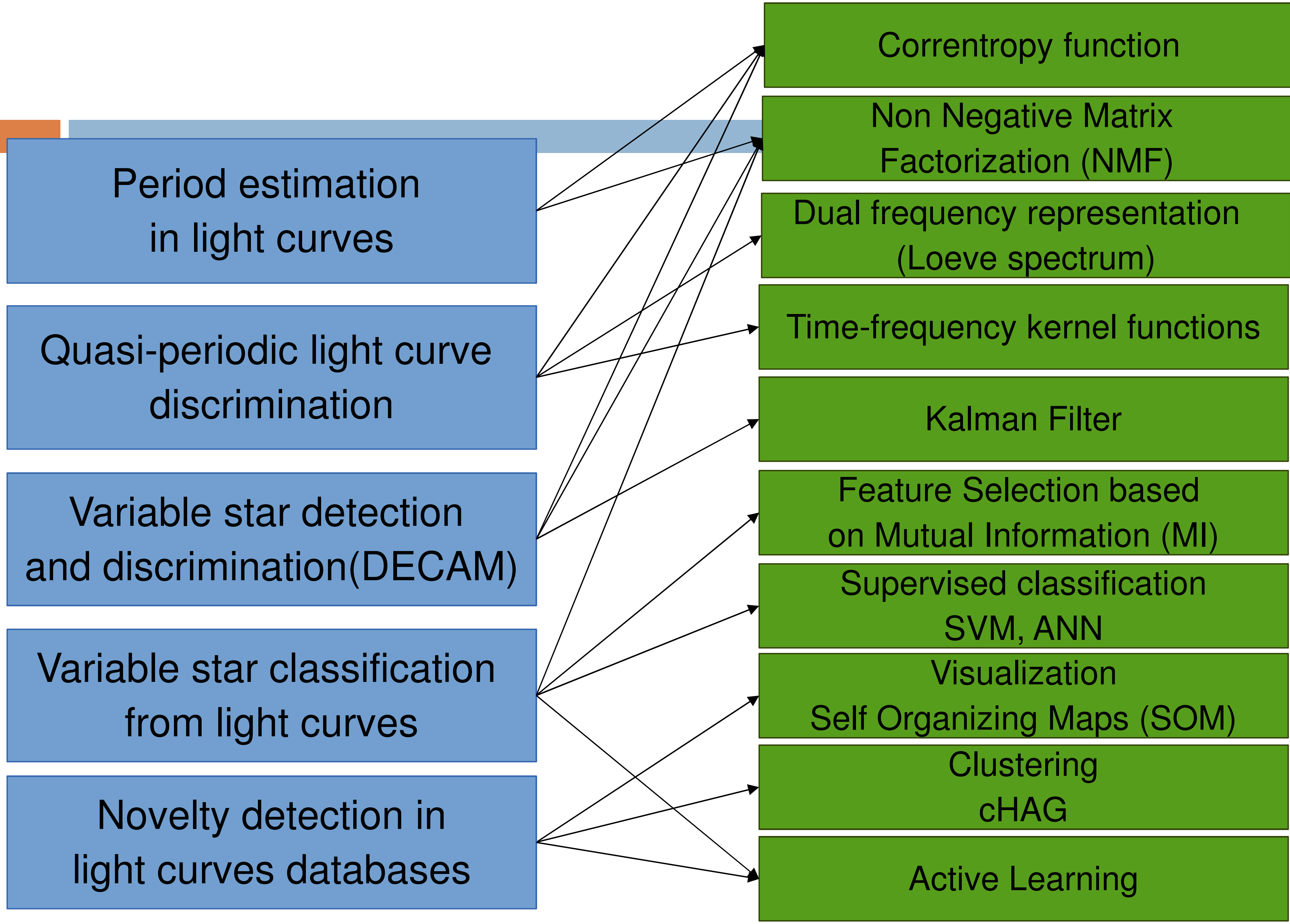

Astroinformatics, Astrostatistics


Exoplanets, Transients

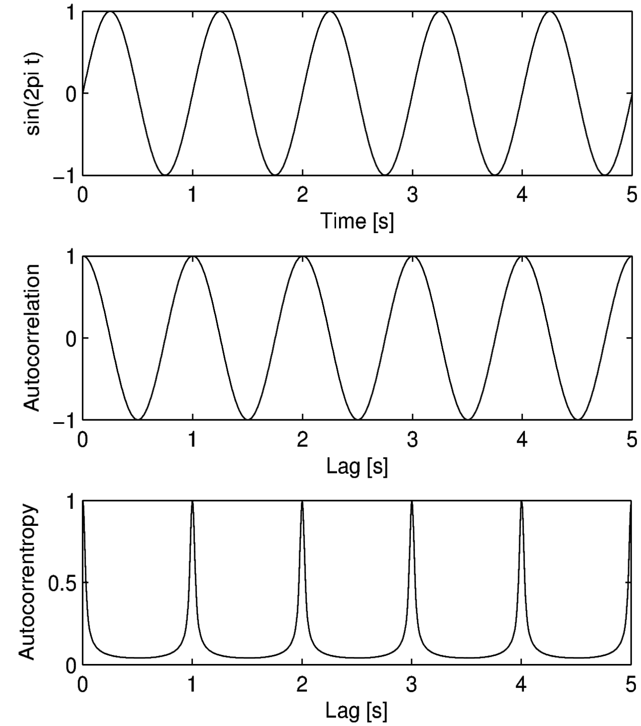

Supernovae

Astroinformatics/Astrostatistics Group

Problems Tools

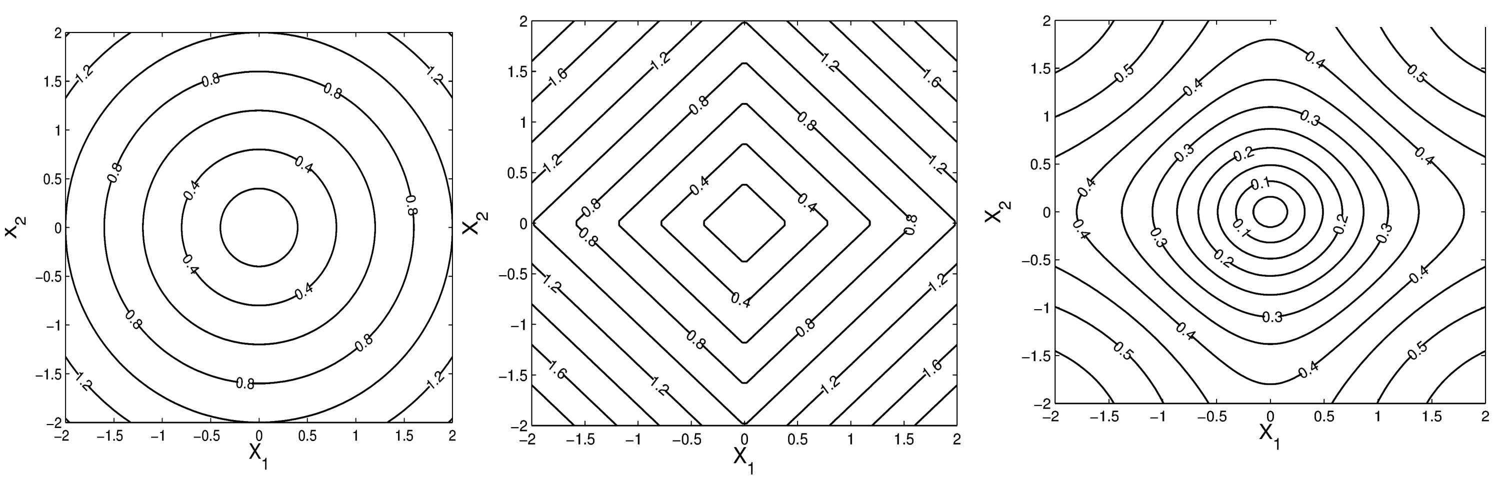


Correntropy



The correntropy is a localized similarity metric for time series. It generalizes the correlation function to Non-Gaussian and non-linear processes

$$v_\sigma(m) = \frac{1}{N-m+1} \sum_{n=m}^{N} \exp\left(\frac{-(x_n - x_{n-m})^2}{2\sigma^2}\right)$$

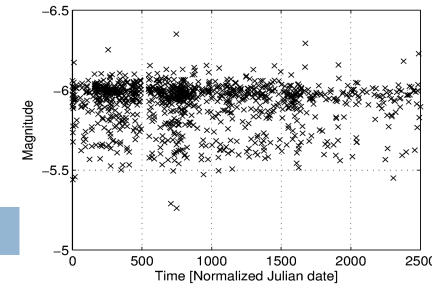


Automated period detection

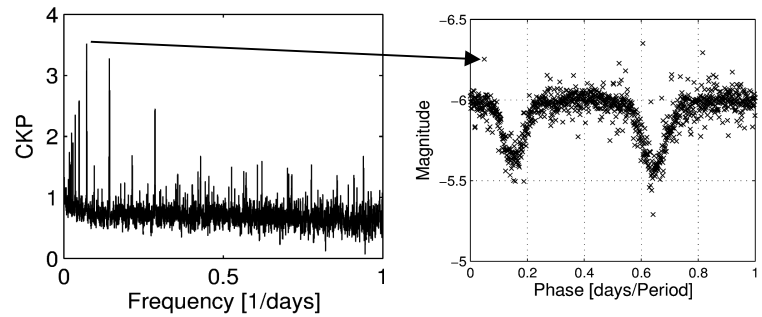
- Correntropy (generalized correlation) is used to compute similarities between samples
  - Go beyond second order statistics, taking into account higher order moments
  - Robustness to outliers and noise
- Spectral decomposition of correntropy using advanced signal processing techniques
  - Gaussian basis functions are used instead of sinusoids
  - Go beyond Fourier representation to get super-resolution, more localized and sparser spectra

Correntropy for periodicity detection in light curves

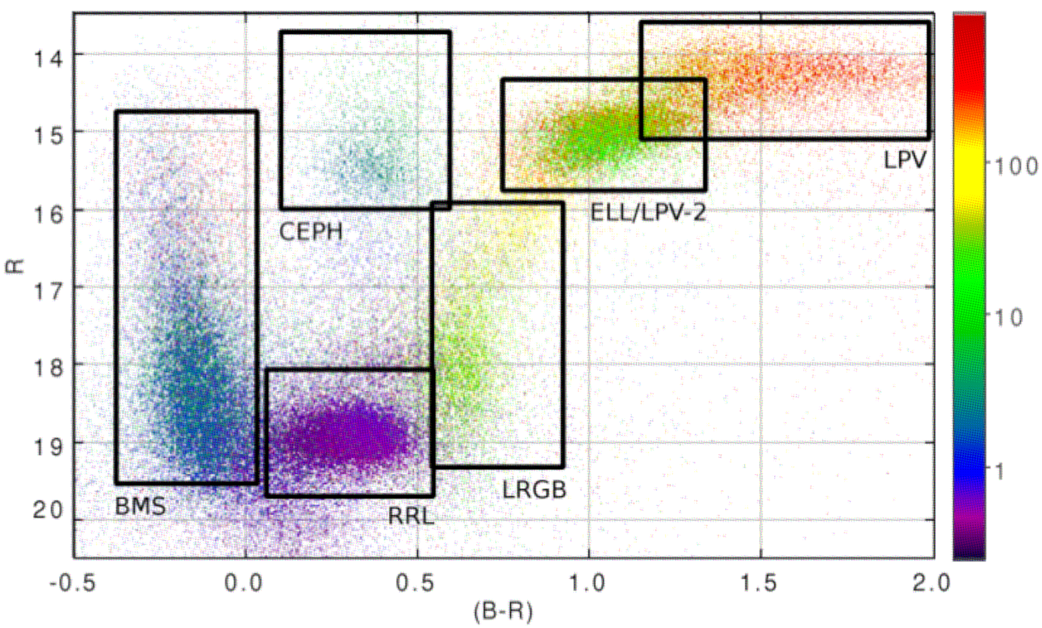
Dr. Pablo Huijse

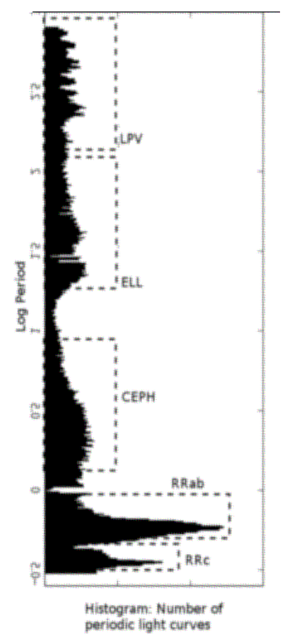


Correntropy kernelized periodogram (CKP), for unevenly sampled time series, robust against impulsive noise and outliers, can be tuned to different shapes.



$$CKP(f) = \sum_{i=1}^{N} \sum_{i=1}^{N} \left(G_\sigma(\Delta x_{ij}) - IP\right) \exp\left(-2\sin^2(\pi \Delta t_{ij} f)/\sigma_t^2\right)$$



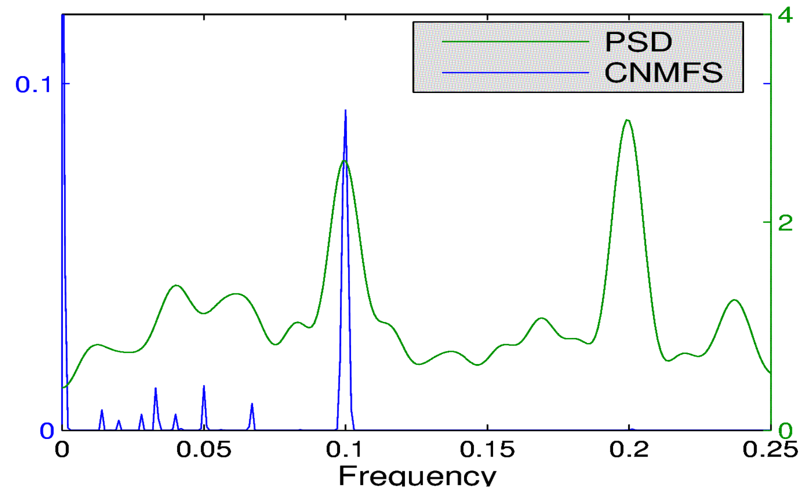


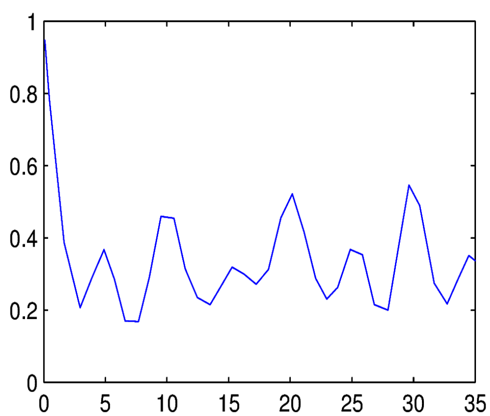
Implemented on GPGPU cluster from XSEDE, EROS survey (40M light curves) processed in 1 day, 140,000 periodic light curves: LMC and SMC found.

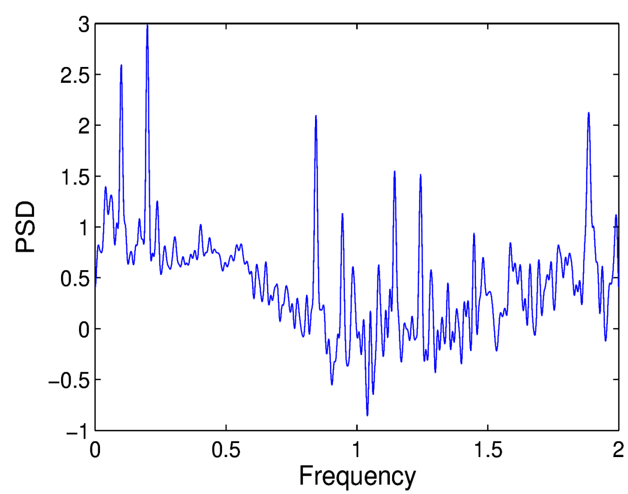
Spectral Decomposition based on Correntropy and Non-negative Matrix Factorization

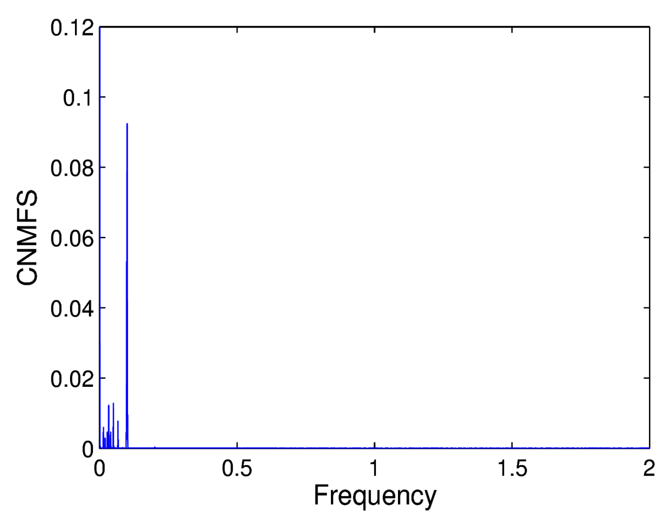
Sparse and superesolved frequency representation for correntropy using NMF (similar to LASSO with adaptive dictionary and NN constraint)

<u>Advantages:</u> **Robust to noise/outliers**, better frequency resolution, localized, cleaner, sparser

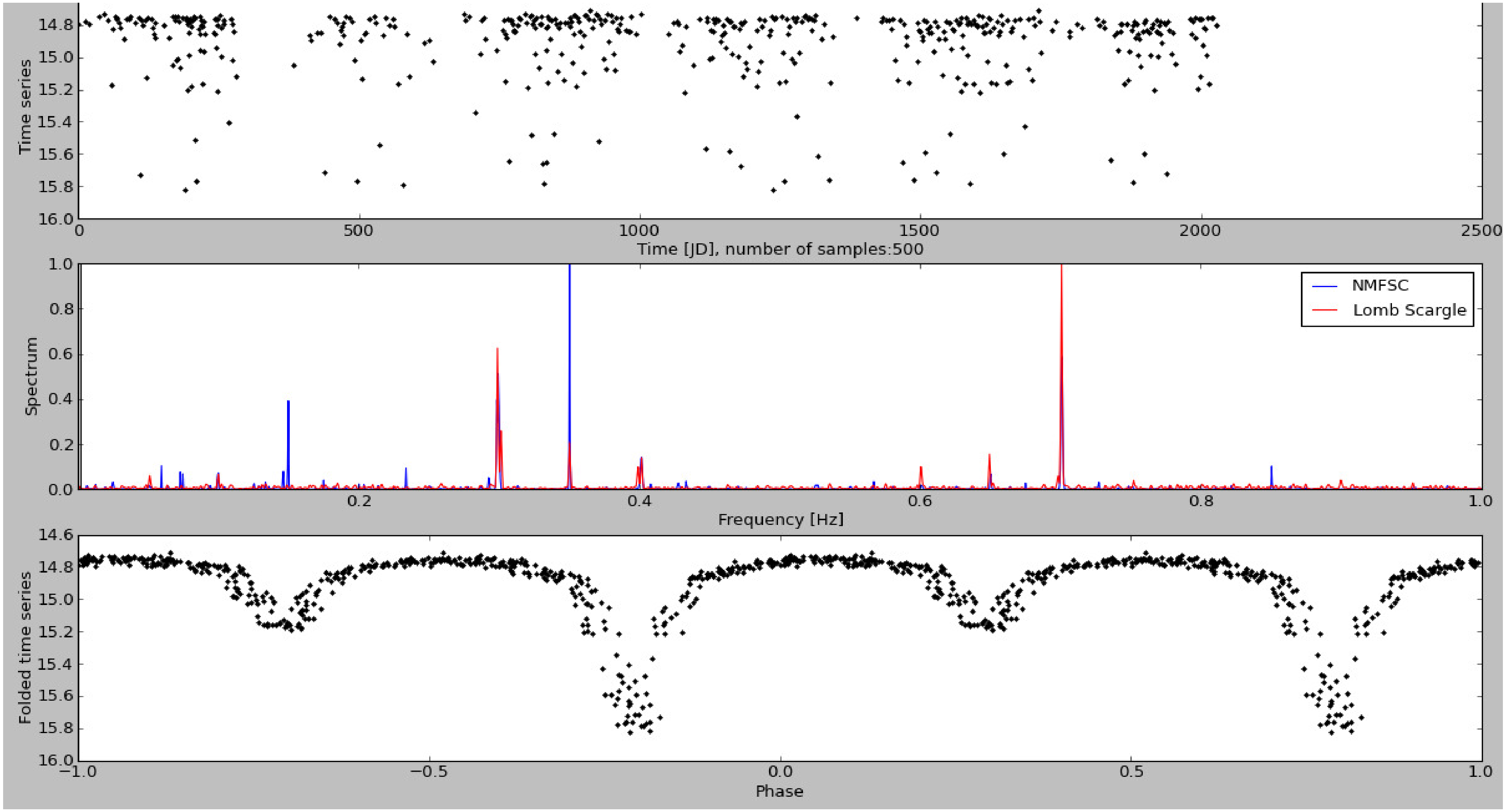








Example

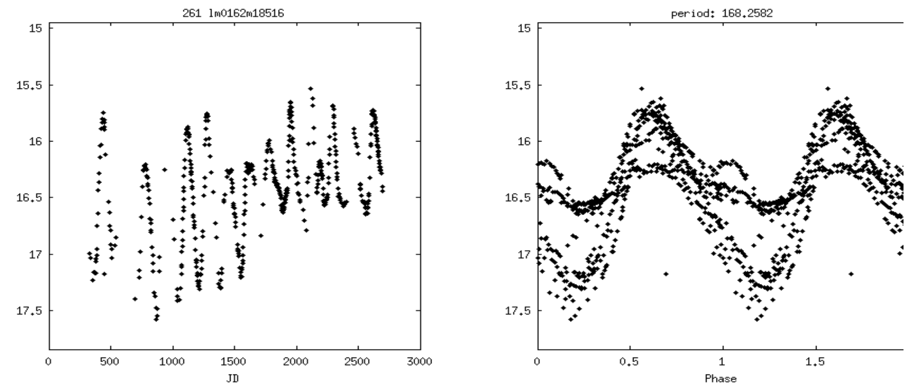


Time series
14.8
15.0
15.2
15.4
15.6
15.8
16.0
0
500
1000
1500
2000
2500
Time [JD], number of samples:500
Spectrum
1.0
0.8
0.6
0.4
0.2
0.0
0.2
0.4
0.6
0.8
1.0
Frequency [Hz]
NMFSC
Lomb Scargle
Folded time series
14.6
14.8
15.0
15.2
15.4
15.6
15.8
16.0
−1.0
−0.5
0.0
0.5
1.0
Phase

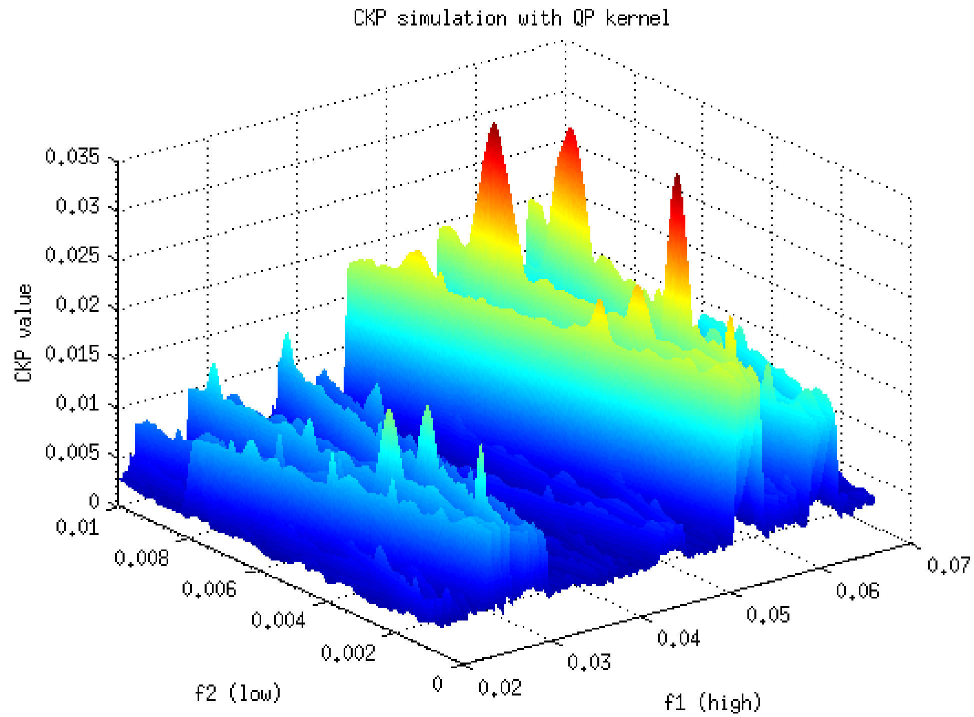
Non-stationary kernels for quasi-periodicity detection

MSc student: Catalina Elzo

Extend current methodologies to recognize time-frequency features



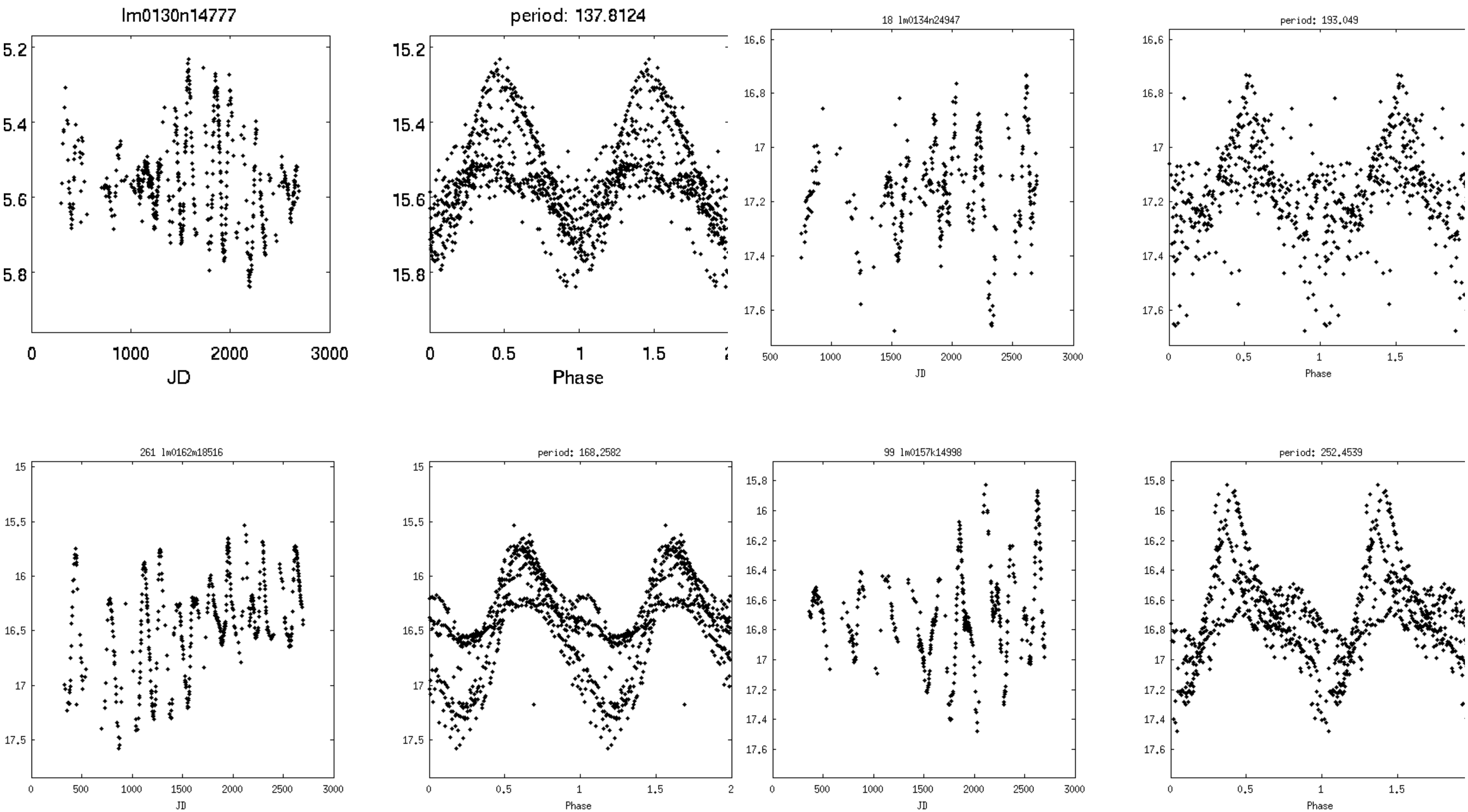
Design quasi-periodic kernels (generative model), plug them into the CKP (detection)



$$G_{AM}(\Delta t, f_1, f_2) = \exp(-2\sin^2(\pi \Delta t f_1)/\sigma_t^2) + \exp(-2\sin^2(\pi \Delta t (f_1 + f_2))/\sigma_t^2) + \exp(-2\sin^2(\pi \Delta t (f_1 - f_2))/\sigma_t^2)$$

Example: AM kernel

Time to go fishing (for quasiperiodics)



Detection of groups of complementary features

PhD student: Jorge Vergara

Selecting relevant features improves the performance of ML classifiers (better generalization) and reduce computational time.

Some features might be relevant only when they work in a group.

Mutual information criterion to sort relevant features and detect complementary groups. Sequential strategy to detect interactions between features.

$$r_j^{MAXGI} = \max_i \left(I(f_i;C \mid S), I(f_i;C \mid \neg f_i) \right)$$

$$r_1 \rightarrow r_2 \rightarrow r_3 \rightarrow r_4 \rightarrow r_5 \rightarrow r_6 \rightarrow \cdots \rightarrow r_d$$

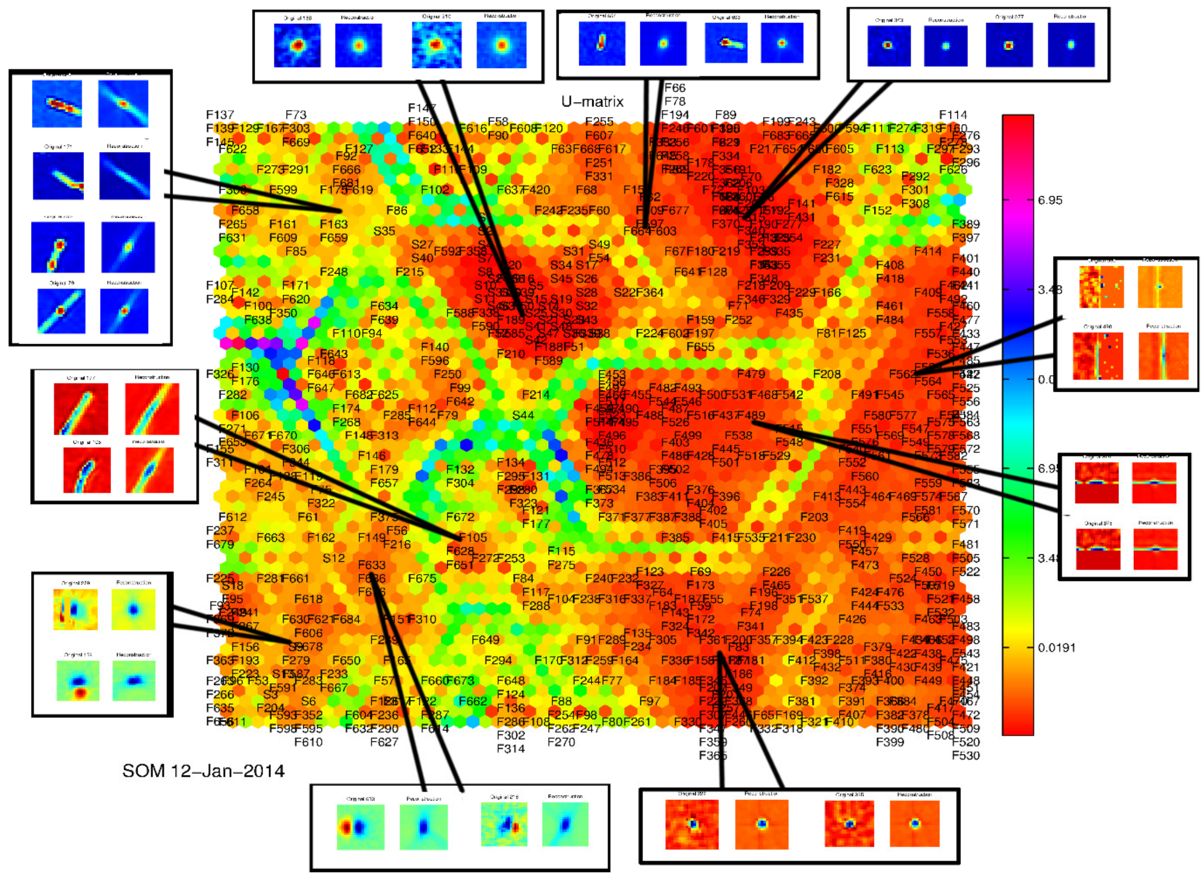
$$Int_{i,\dots,k} = \max_{k=i+1}^{j} \left(I(r_i;C \mid \{r_{i+1},\dots,r_k\}) \right) \geq \Theta$$

$$\{r_1,r_2\} \rightarrow r_3 \rightarrow \{r_4,r_5,r_6\} \rightarrow \cdots \rightarrow r_d$$

NMF, SOM and SVM: a semi-supervised approach to variable star detection for DECAM

Undergrad Student: Emanuel Berrocal

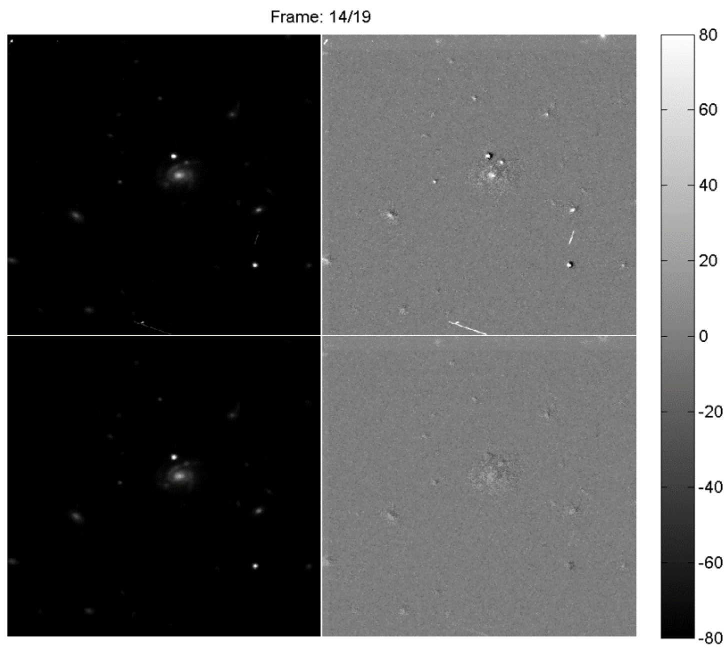
- 21x21 stamps of candidates from the DECAM
- Visualize and define clusters using SOM
- Label the data and classify using supervised methods



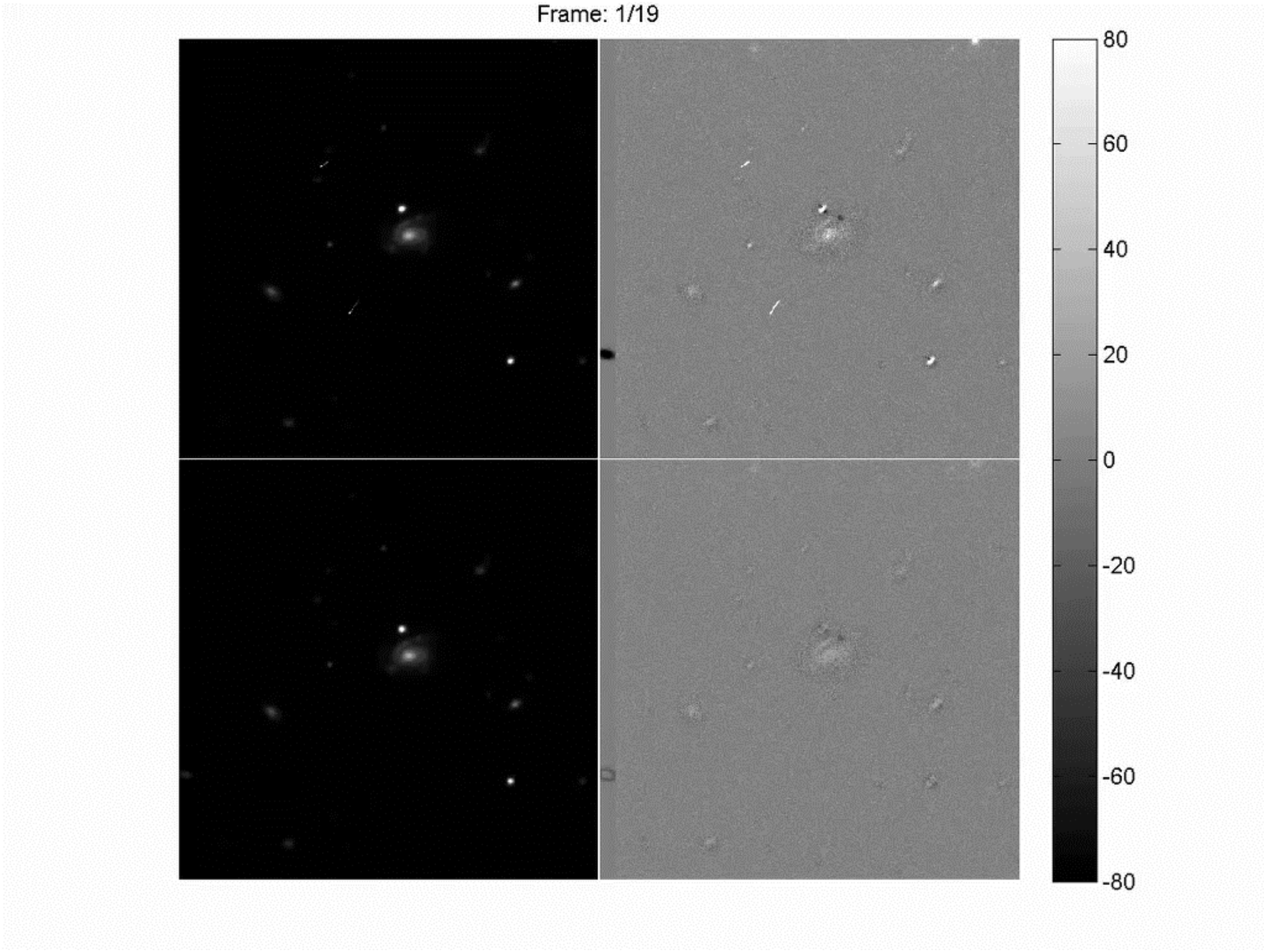
Correntropy Kalman filter for variability detection on images

MSc student: Pablo Huentelemu

- A Kalman filter is used to discriminate the dynamic background on a set of astronomical images
- The correntropy gives the Kalman filter robustness to noise and outliers, and enhanced performance on non-gaussian settings.
- Smooth increments in brightness are saved



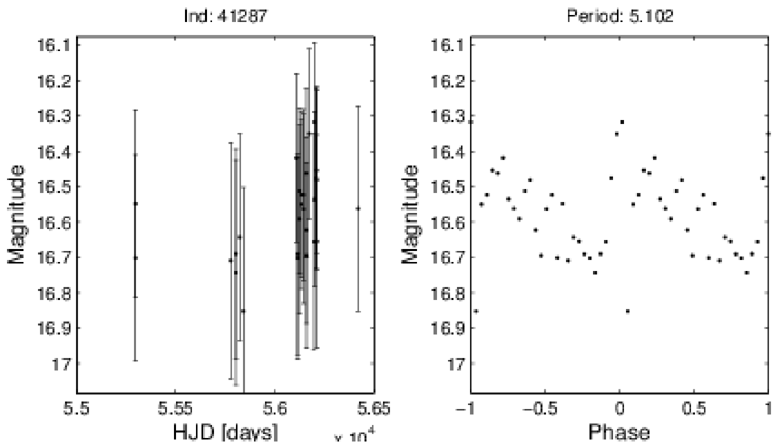
Frame 100/500

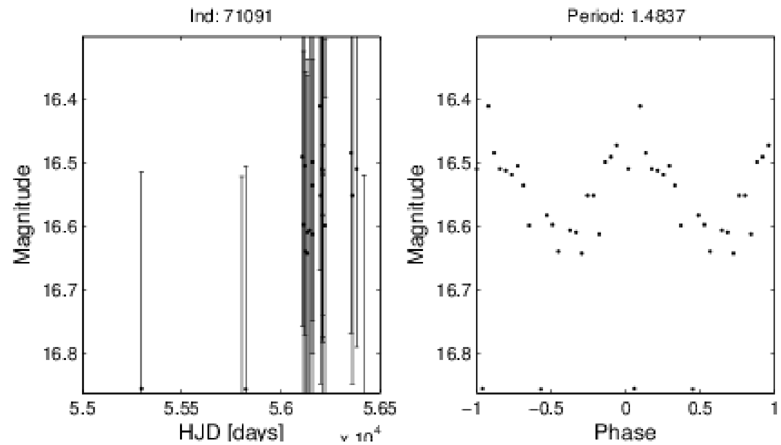


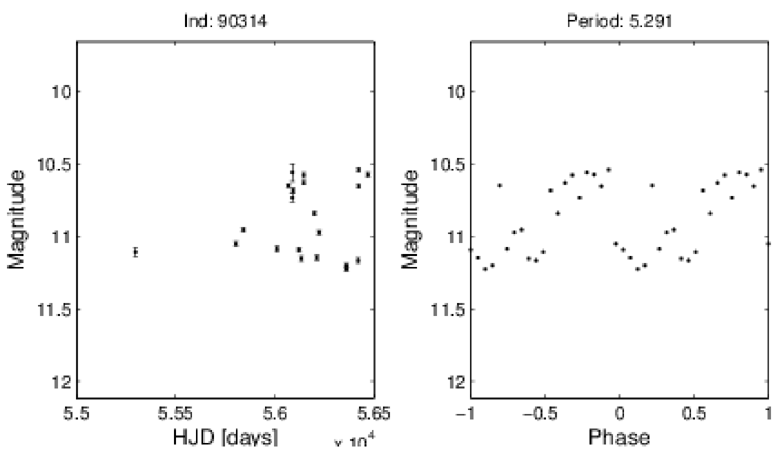
Frame: 1/19
80
60
40
20
0
-20
-40
-60
-80

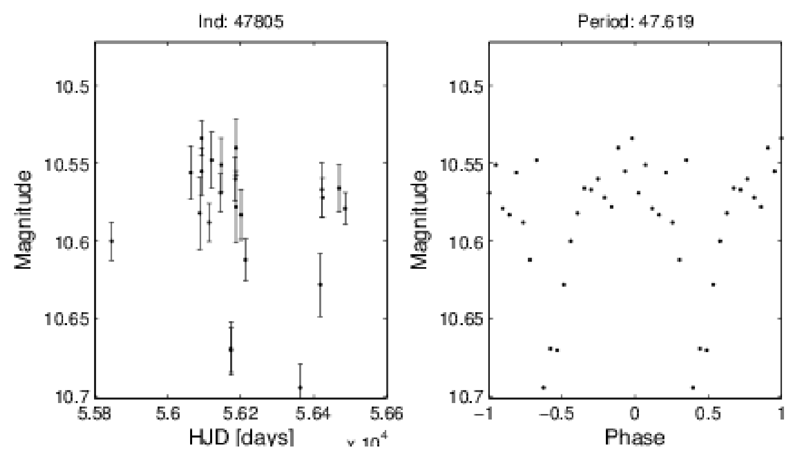
Preliminary results on periodicity detection for VVV

Results with N<20









Computational Intelligence Challenges and Applications on Large-Scale Astronomical Time Series Databases

IMAGE LICENSED BY INGRAM PUBLISHING

Pablo Huijse
Millennium Institute of Astrophysics, CHILE

Pablo A. Estévez
Department of Electrical Engineering,
Universidad de Chile
and the Millennium Institute of Astrophysics, CHILE

Pavlos Protopapas
School of Engineering and Applied Sciences,
Harvard, University, USA

José C. Príncipe
Computational NeuroEngineering Laboratory,
University of Florida, USA

Pablo Zegers
Facultad de Ingeniería y Ciencias Aplicadas,
Universidad de los Andes, Chile

Abstract—Time-domain astronomy (TDA) is facing a paradigm shift caused by the exponential growth of the sample size, data complexity and data generation rates of new astronomical sky surveys. For example, the Large Synoptic Survey Telescope (LSST), which will begin operations in northern Chile in 2022, will generate a nearly 150 Petabyte imaging dataset of the southern hemisphere sky. The LSST will stream data at rates of 2 Terabytes per hour, effectively capturing an unprecedented movie of the sky. The LSST is expected not only to improve our understanding of time-varying astrophysical objects, but also to reveal a plethora of yet unknown faint and fast-varying phenomena. To cope with a change of paradigm to data-driven astronomy, the fields of astroinformatics and astrostatistics have been created recently. The new data-oriented paradigms for astronomy combine statistics, data mining, knowledge discovery, machine learning and computational intelligence, in order to provide the automated and robust methods needed for the rapid detection and classification of known astrophysical objects as well as the unsupervised characterization of novel phenomena. In this article we present an overview of machine learning and computational intelligence applications to TDA. Future big data challenges and new lines of research in TDA, focusing on the LSST, are identified and discussed from the viewpoint of computational intelligence/machine learning. Interdisciplinary collaboration will be

Digital Object Identifier 10.1109/MCI.2014.2326100
Date of publication:

1556-603X/14©2014IEEE