Searching for extragalactic variable stars using Machine Learning algorithms

Javier Alejandro Acevedo Barroso

Advisor: Alejandro García

Departamento de Física
Universidad de los Andes

May 10, 2020

The farther, the dimmer

Light is conserved over the surface of a sphere

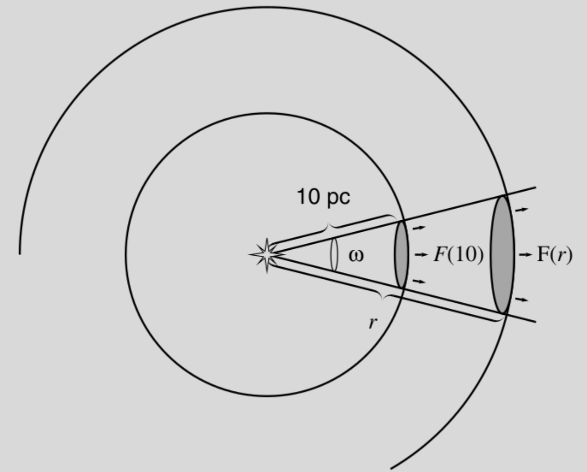
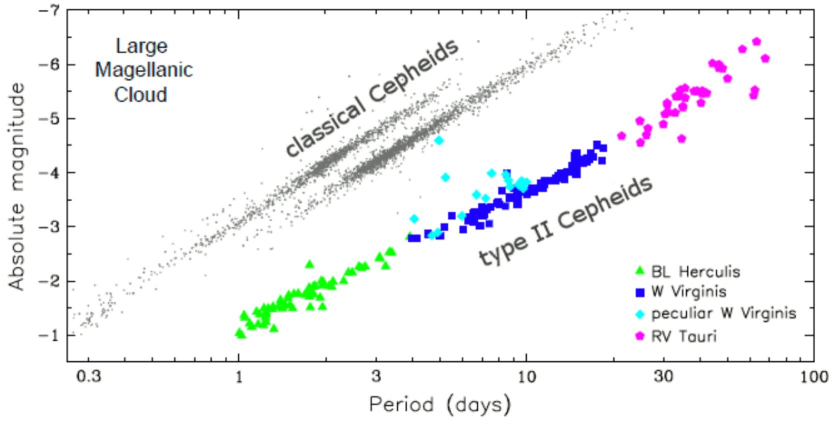


Image taken from Karttunen *et al.* 2017

Cepheid variable stars

The absolute magnitude of a Cepheid depends on its periodicity.



The OGLE project - Igor Soszyński.

Main objective

- To create catalogues of variable for the galaxies NGC55, NGC247 and NGC7793; using The Araucaria Project data and Machine Learning algorithms to automatise the search and classification.

Photometry related objectives

- ✓ To download, preprocess (bias, flat and stacks) and organise the old Araucaria data for NGC55, NGC247 and NGC7793.
- ☐ To perform PSF photometry over the stacked images and generate photometric catalogues for every night.
- ☐ Crossmatch the sources from the different catalogues to create light curves for each resolvable star.

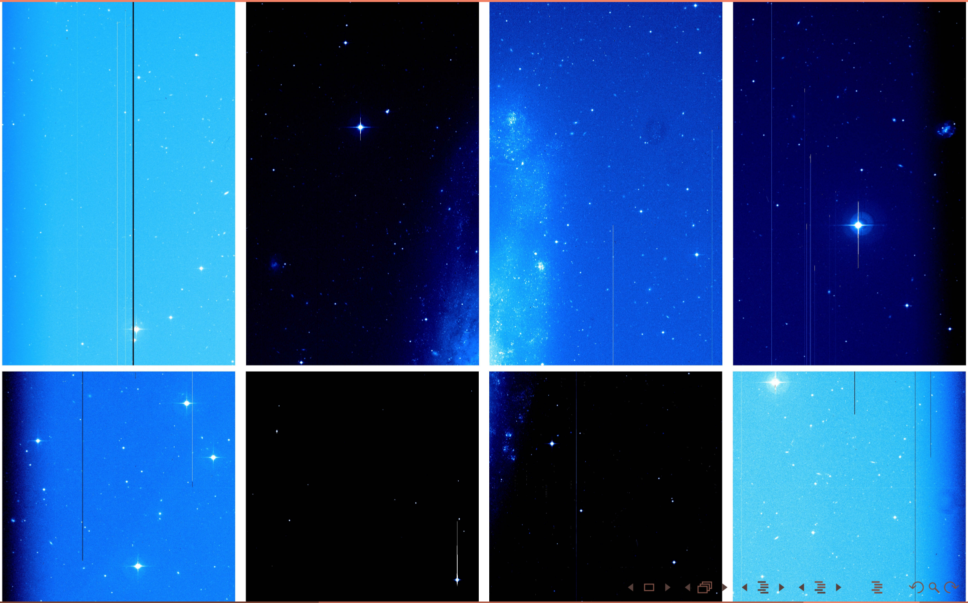
Machine Learning related objectives.

- Define a space of representative features, and project the light curves into a features vector.
- Design and train a variable star classifier using the "OGLE catalogue of variable stars"; and algorithms such as Random Forest, Support Vector Machine, or different kinds of Neural Networks.
- Use the classifier to create a variable star catalogue from the light curves generated before.

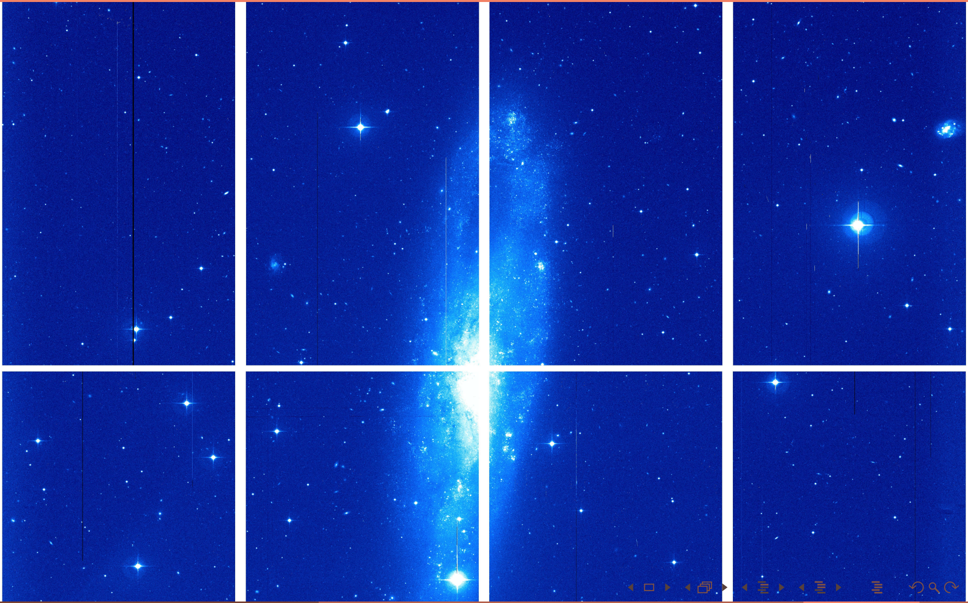
CCD corrections and stacks

- ✓ Download bias frames for the night, and combine them using median to create a “master bias” frame.
- ✓ Download flat frames for each filter, correct for bias, and combine them into a “master flat” using a median stack. (So far, we are only concern with images in V).
- ✓ Download the science images and correct them for bias and flat (using the aforementioned master frames).
- ⊡ Stack the dithered science images to fill the gaps between the chips, and create a final science frame per night.

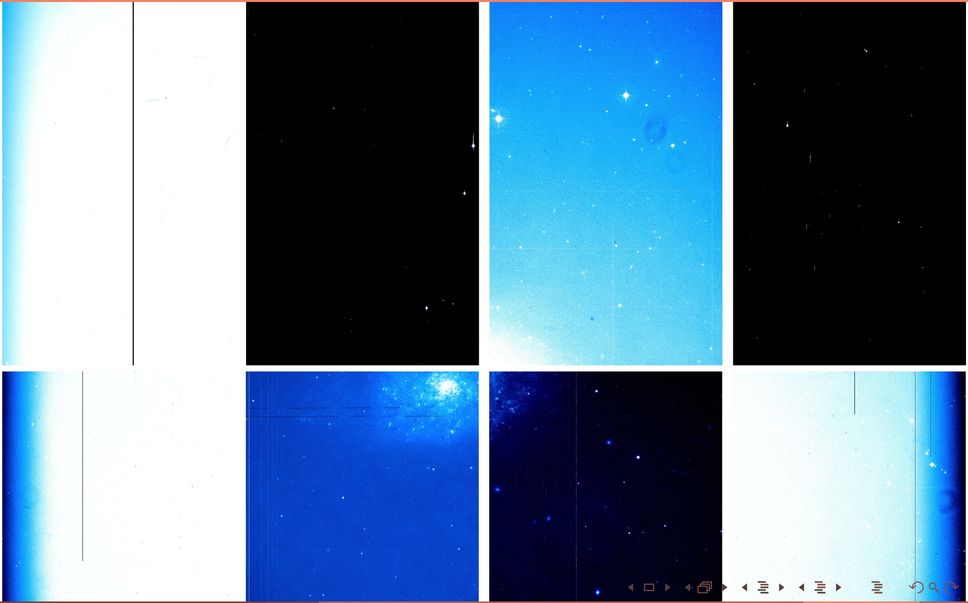
NGC247 raw science image



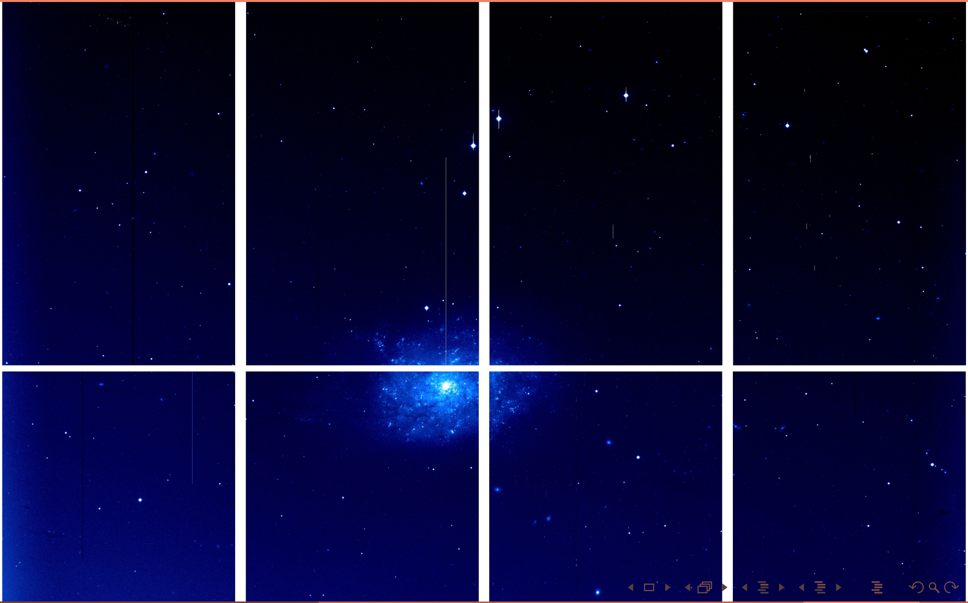
NGC247 processed science image



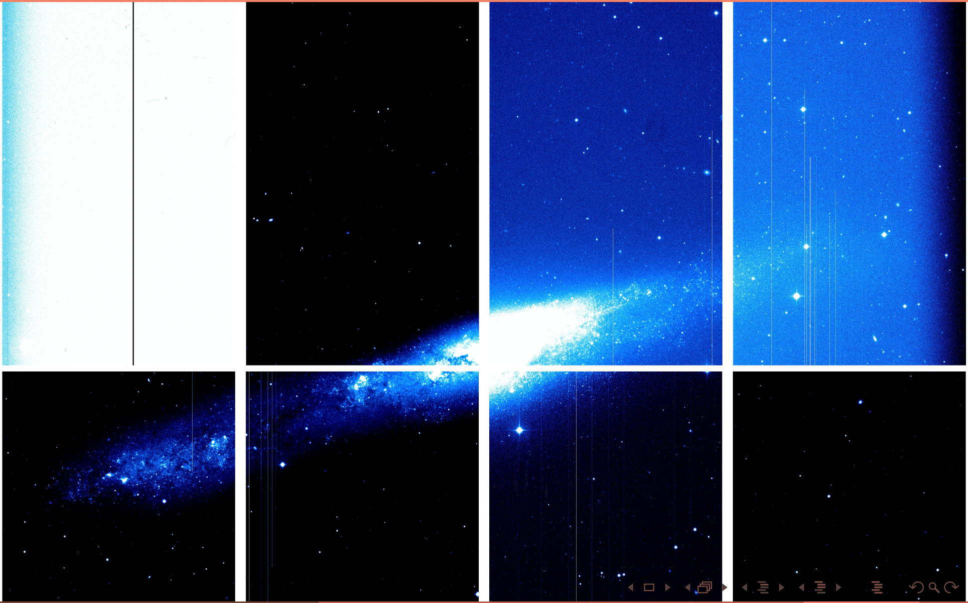
NGC7793 raw science image



NGC7793 processed science image



NGC55 raw science image



NGC247 raw science image

NGC55 final stacked frame of a night

Processing pipeline

- To massively apply the CCD corrections, we developed a pipeline that takes retrieved files from the ESO public catalogue, processes them using IRAF, and returns CCD corrected images.
- The stacking is only possible on nights where five or more dithered images were taken.
- To stack the images, we are using modified versions of the IRAF packages ESOWFI and MSCRED (Valdes 1998), along some adapted scripts from Phil Massey written originally for the “Local Group Survey” (Massey, Olsen, *et al.* 2006) and available at their website (Massey & Holmes 2003).

Stacking dithered frames

1. Update the World Coordinated System (WCS) of the frames, that is the transformation between (x, y) CCD coordinates to (RA,DEC) in the sky.
2. Resample the images in the same tangent plane. (Align their (x,y) coordinates).
3. Correct for the different background levels and intensities. This is done using Massey scripts (The most crucial step in order to get a good stack).
4. Combine the images using a median stack.

Theses steps were taken from different manuals and guides. (Valdes 1993, 1997, 1998; Jones & Valdes 2000; Massey & Holmes 2003; Jannuzi *et al.* 2004; Schweiker 2016)

Comparision between stacks

DS9

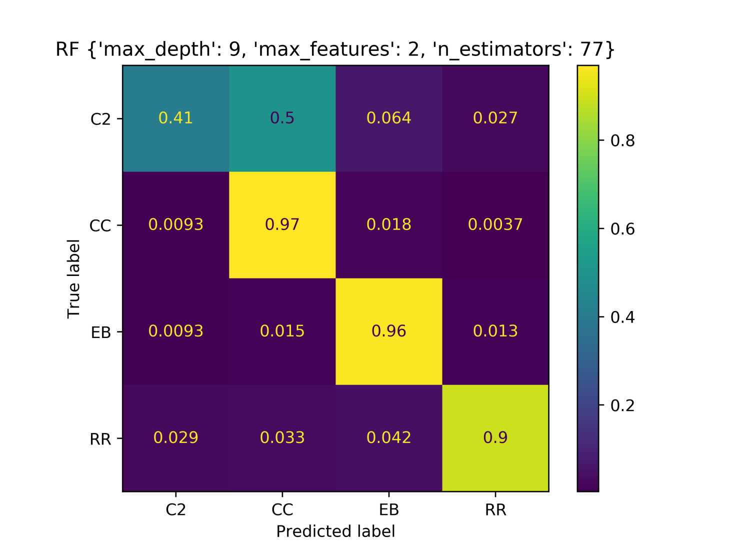
Tasks on the Machine Learning front while the light curves are generated

- ✓ Run and adapt the R scripts inherited from Uniandes students Muriel (2017) and Bayron (2020).
- ⊡ Project the light curves into a representative feature vector. (So far we are using the features from Pérez-Ortiz *et al.* 2017: median, MAD, octile skewness, left octile weight, right octile weight and robust Abbe value).
- ✓ Reproduce Muriel's work on stellar classification (Pérez-Ortiz *et al.* 2017), but using V instead of I.
- ☐ Design and implement binary classifiers based on the best multiclass classifiers per class.
- ☐ Explore options to weed out non variable stars on the Araucaria light curves.

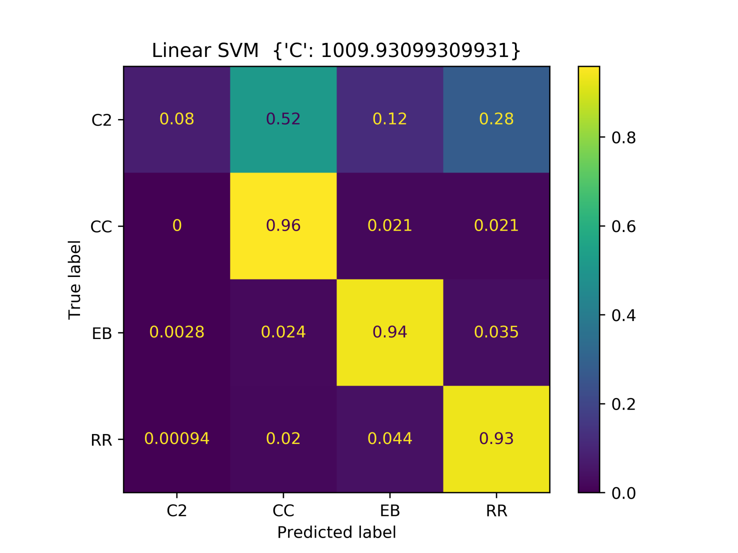
Some details about the implemented models

- All the models used OGLE 3 light curve catalogue as training data, and OGLE 4 light curve catalogue as test data.
- We used two algorithms: Random Forest and Support Vector Machine.
- Random Forest had the maximum depth of each tree, the maximum number of features and the number of estimators as hyperparameters.
- Support Vector Machine had the C value as hyperparameter.
- We explored the hyperparameter space several times optimising different metrics each time: accuracy, weighted accuracy (to try account for the class imbalance), weighted recall, and weighted F1 score.

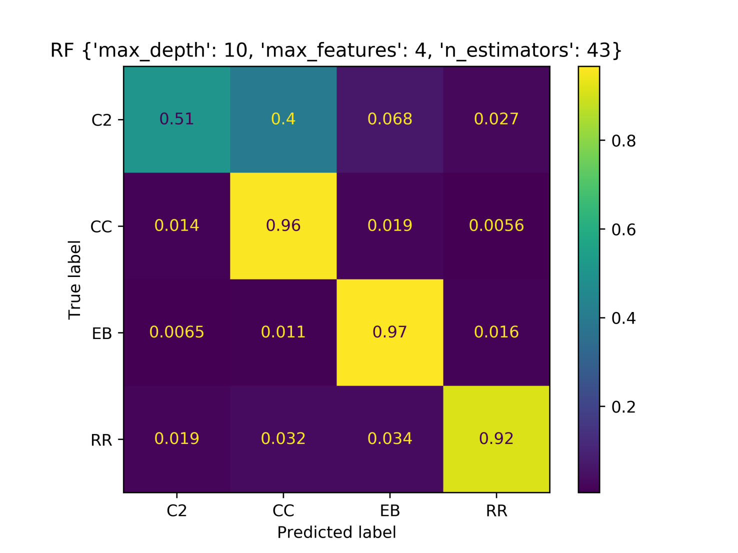
Optimising accuracy - Random Forest



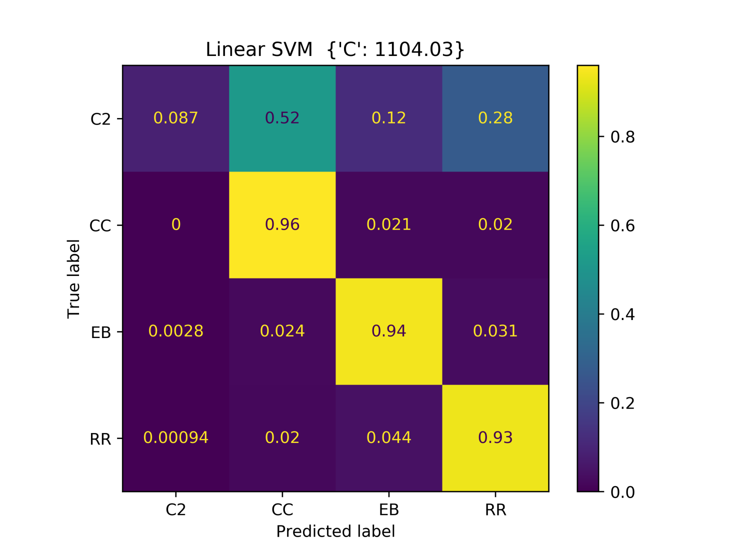
Optimising accuracy - SVM



Optimising weighted accuracy - Random Forest



Optimising weighted recall - SVM



References I

Valdes, F. *The IRAF/NOAO Spectral World Coordinate Systems.* in *Astronomical Data Analysis Software and Systems II* (eds Hanisch, R. J., Brissenden, R. J. V. & Barnes, J.) **52** (Jan. 1993), 467.

Valdes, F. *The Reduction of Mosaic Images.* https://www.noao.edu/noao/noaonews/dec97/node23.html. [Online; accessed 10-May-2020]. 1997.

Valdes, F. *The IRAF Mosaic Data Reduction Package.* in *Astronomical Data Analysis Software and Systems VII* (eds Albrecht, R., Hook, R. N. & Bushouse, H. A.) **145** (Jan. 1998), 53.

Jones, H. & Valdes, F. *Handling ESO WFI Data with IRAF.* http://www.ls.eso.org/lasilla/Telescopes/2p2T/E2p2M/WFI/docs/reduction/WFIinIRAF/WFIinIRAF.ps. 2000.

References II

- Massey, P. & Holmes, S. *Local Group Survey: Mosaic Reduction Notes.* http://www2.lowell.edu/users/massey/lgsurvey/splog2.html. [Online; accessed 10-May-2020]. 2003.
- Jannuzi, B., Claver, J. & Valdes, F. *The NOAO Deep Wide Field Survey MOSAIC Data Reductions.* https://www.noao.edu/noao/noaodeep/ReductionOpt/frames.html. [Online; accessed 10-May-2020]. 2004.
- Massey, P., Olsen, K. A. G., *et al.* A Survey of Local Group Galaxies Currently Forming Stars. I. UBVRI Photometry of Stars in M31 and M33. AJ **131**, 2478–2496. arXiv: astro-ph/0602128 [astro-ph] (May 2006).

References III

- Schweiker, H. *KPNO MOSAIC-3 IMAGER USER MANUAL.* https://www.noao.edu/kpno/mosaic/manual/. [Online; accessed 10-May-2020]. 2016.
- Karttunen, H., Kröger, P., Oja, H., Poutanen, M. & Donner, K. J. *Fundamental Astronomy.* (2017).
- Pérez-Ortiz, M. F., García-Varela, A., Quiroz, A. J., Sabogal, B. E. & Hernández, J. Machine learning techniques to select Be star candidates. An application in the OGLE-IV Gaia south ecliptic pole field. A&A **605,** A123. arXiv: 1707.04560 [astro-ph.IM] (Sept. 2017).

The end. Thanks for your attention.