Cepheid pulsation mode identification via unsupervised machine learning

Santiago Henao Castellanos, Alejandro García Varela

s.henao10@uniandes.edu.co, josegarc@uniandes.edu.co

Physics department, Universidad de los Andes, Bogotá, Colombia

Introduction

The Period-Luminosity (PL) relation of Cepheid variable stars is an essential tool for measuring distances in our Universe, and it has a direct impact on the model-free value of the Hubble parameter (Riess et al. 2023). This relation is affected by the presence of pulsation overtones in the stars, which displace the tendencies within the PL diagram.

In the literature, the mode is determined using relations between the fit parameters of a Fourier series (Pietrukowicz et al. 2021; Simon & Lee 1981), but this approach hides the connection between pulsation modes and light curve shapes: the fundamental (F) mode Cepheids have an asymmetrical light curve, with a faster rising branch, whereas the first and second overtones (1O, 2O) are progressively more sinusoidal.

From this intuition we present a set of features that quantify the shape of the light curve, and we use them to perform clustering of the OGLE Cepheids to find their pulsation modes.

Data

The data to test our features was taken from the OGLE Catalog of Variable Stars (OCVS), specifically the anomalous, classical, and type II Cepheids from the LMC (Soszyński et al. 2015a,b, 2018) and from the Galactic bulge (Soszyński et al. 2017). I band photometry was taken from both OGLE III and IV phases, as the observation times do not present any overlap. The result was a very inhomogeneous dataset in terms of class representation:

label	Classical			Anomalous		Type II		
	F	1O	2O	aF	a1O	BLHer	WVir	RVTau
size	5367	4246	40	242	101	978	1026	482

Nevertheless, this was accepted as a property of the dataset, as some star types are rarer than others.

Light curve modeling

The features to be proposed will work best with a smooth model of the phased light curve. We will not use the Fourier coefficients for the clustering, but our model comes from a truncated series:

$$f_n(x) = A_0 + \sum_{k=1}^{n} A_k \cos(2\pi(x + \phi_k))$$

which is fitted increasing the order n until $\frac{\text{RSS}}{N-(2n+1)^{3/2}}$ reaches a minimum, with N being the length of the data and RSS the residual sum of squares.
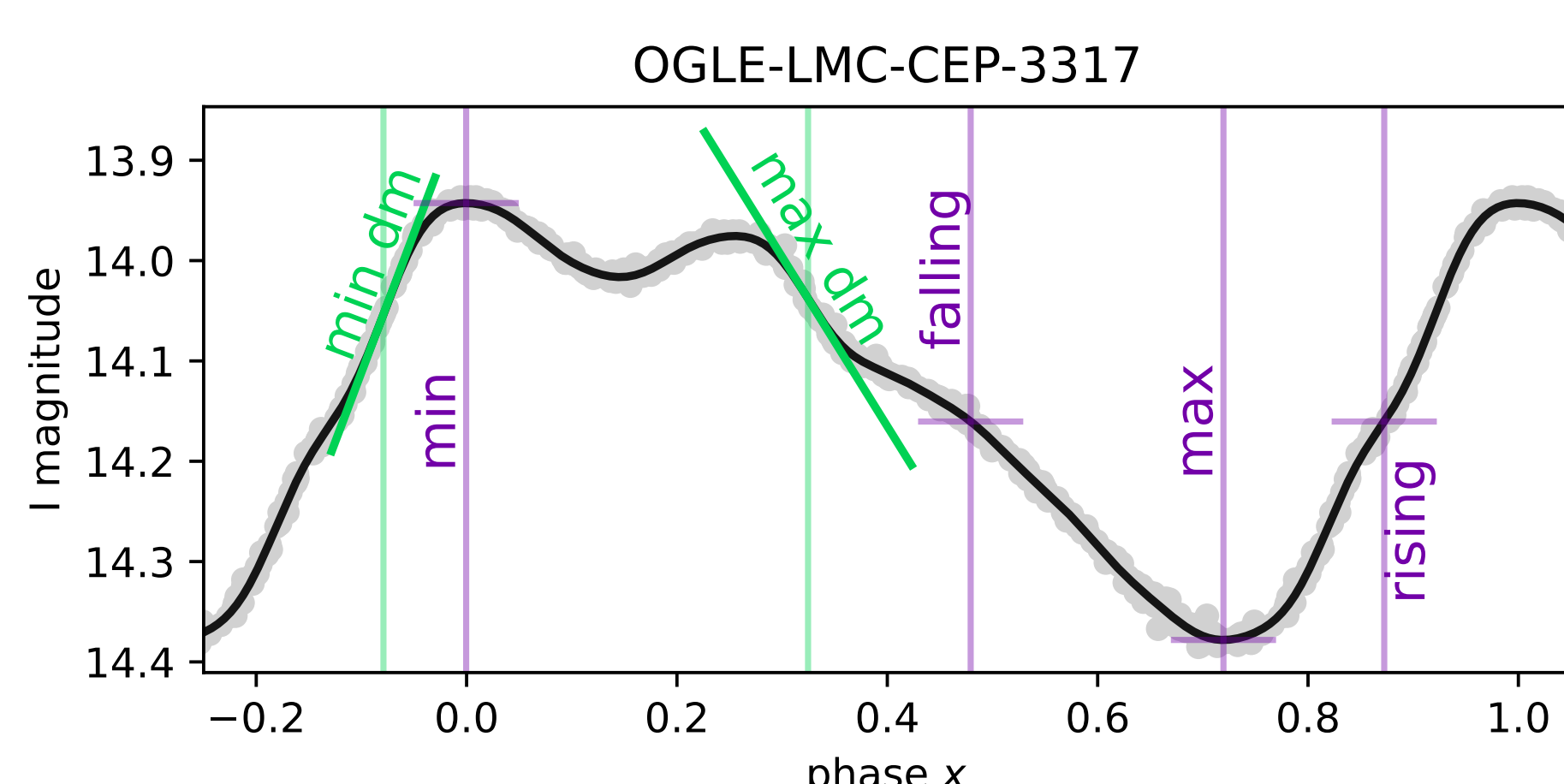

Figure 1: Fundamental mode Cepheid fitted to 11th order, with features marked.

Features

In order to describe the general form of the light curve $\{x_i, m_i\}$, we propose as features:

`x max`, `x min`: phases on the extrema.
`K x max`, `K x min`: curvatures on the extrema.
`dm max`, `dm min`: extrema of the derivative,
  `x dm max`, `dm x min`: and their phases.
`x falling`, `x rising`: midpoint crosses.
`P`: pulsation period, in days.
`A`: pulsation amplitude, in magnitudes.

The fit ephemeris (`x min`) was subtracted from all the other phases, and `K x min` was rejected after examination. This left us with 10 features, which are exemplified in Figure 1. After that, the phases were scaled with $\text{arctanh}(2x - 1)$, and the other features with $\log_{10}$.

Identification on the PL diagram

We examined the resulting clusters in the PL diagram (Figure 2). The largest three seemed to cover the most part of the F, 1O and WVir classes. There were many small clusters on the upper and lower parts of the linear tendencies on the PL diagrams, which were aggregated with the bigger classes.

As can be seen on Figure 3, all of the OCVS classes were present, but due to their size differences, the classical 2O and the a1O are only marginally identified.
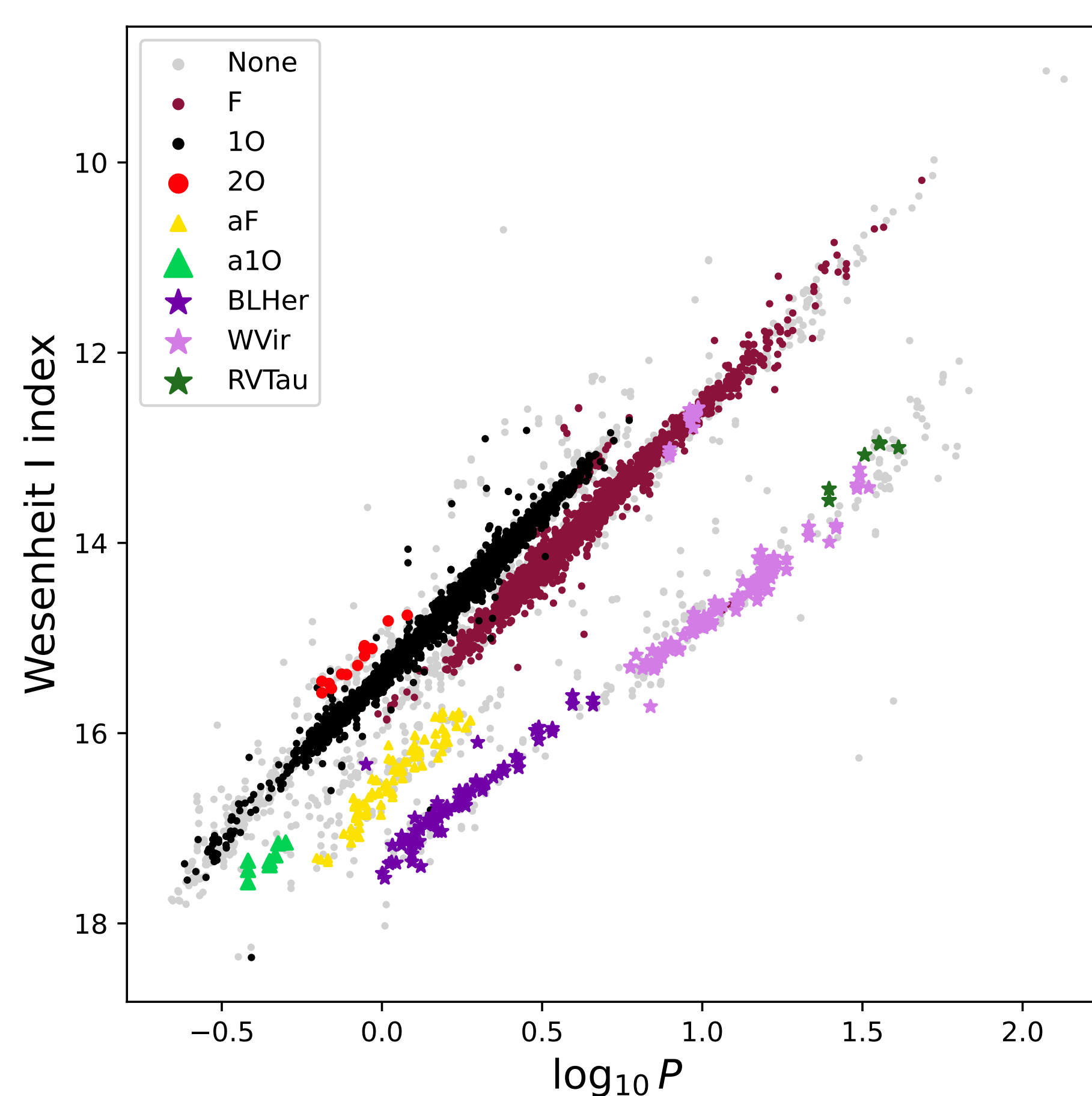

Figure 2: PL relation of the LMC Cepheids with the aggregated cluster labels. Here the Wesenheit index is defined as $W = I - 1.55(V - I)$.
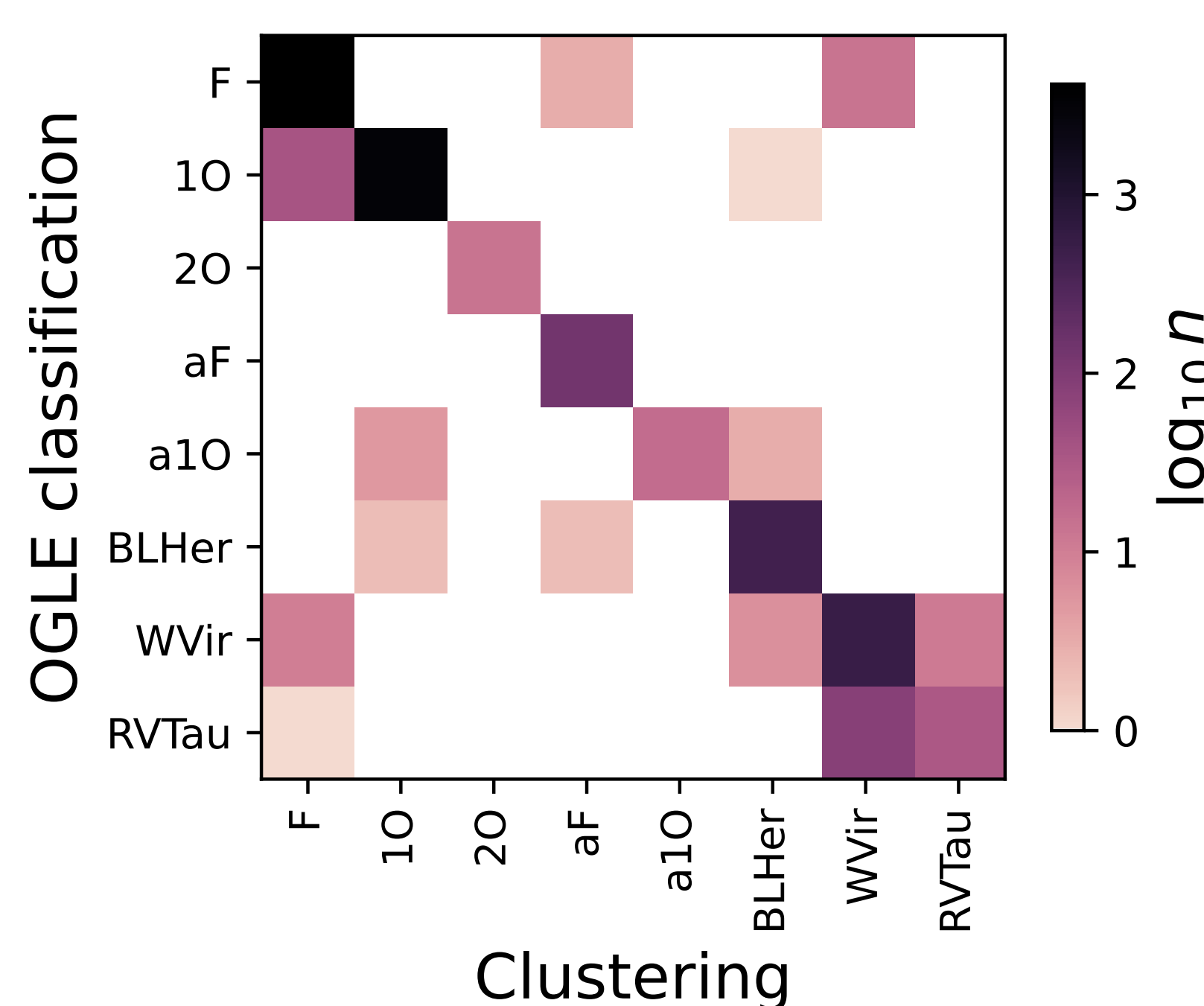

Figure 3: Confusion matrix between cluster identification and OCVS labels. n is the amount of data in each cell.

Clustering

The HDBSCAN* algorithm (Campello et al. 2013) was used to cluster the data, using the Python implementation by McInnes et al. (2017).

We used the Chebyshev metric, defined on the feature space as $d_\infty(\vec{x}, \vec{y}) = \max_k(|x_k - y_k|)$, because it gave us 67% of selection ratio and captured the desired detail on the clusters.

The selected values for the hyperparameters were `min_cluster_size=5` and `min_samples=4`. This generated more clusters than expected, but with the hybrid algorithm of Malzer & Baum (2020), using $\epsilon = 0.047$, the nearest clusters were merged, leaving us with 63 classes and 29% of rejected data.

Hidden morphology of Cepheids

Additional to the PL relations, some structure was found on the F, 1O and aF classes as can be seen in Figure 4.
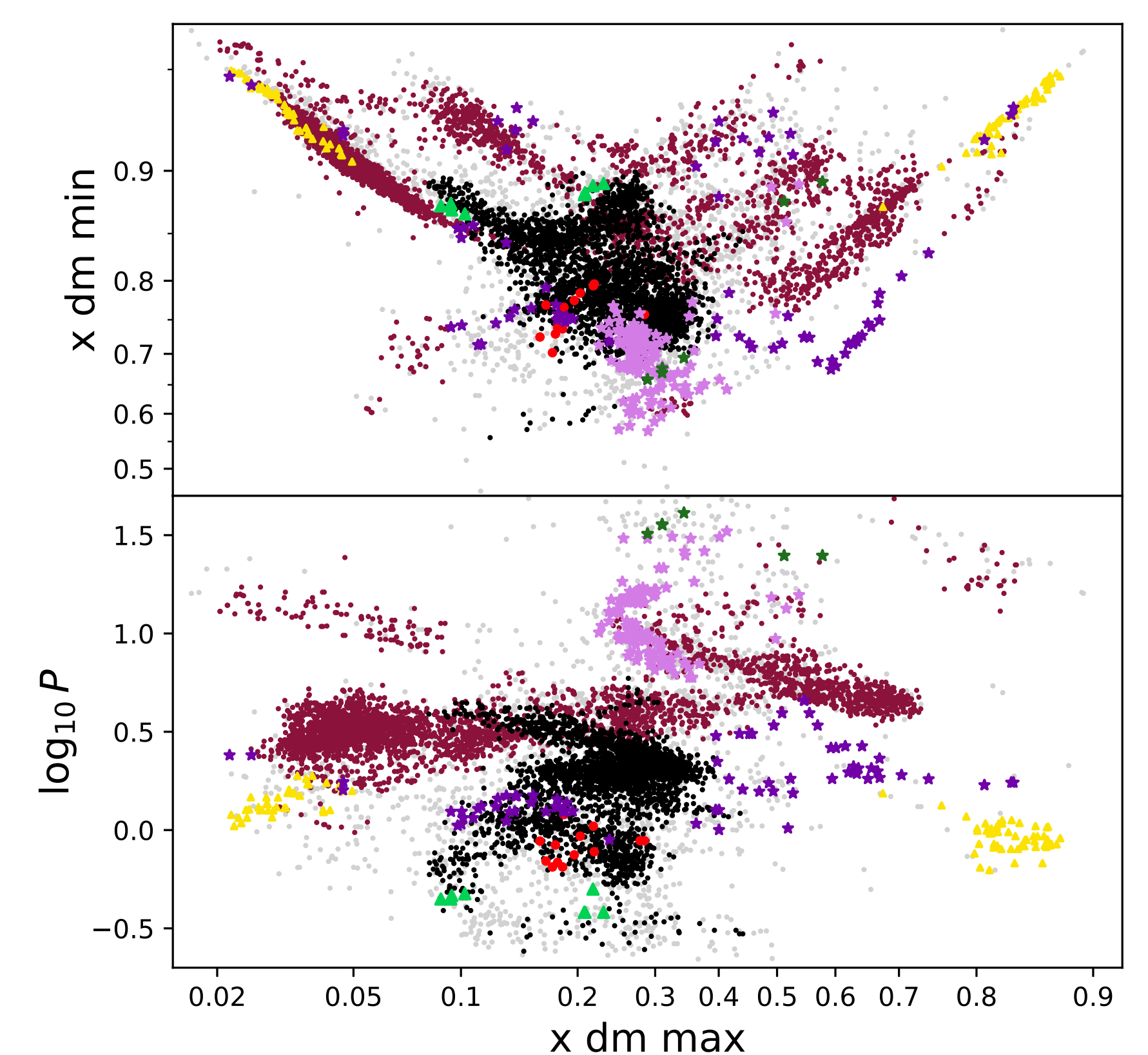

Figure 4: Selection of pair plots from the clustering features. Same legend as in Figure 2.

It remains to be seen if this structure is related to the Hertzprung progression, as the presence of a bump in the curve could affect the derivative values.

Conclusions and acknowledgments

The proposed set of light curve features proved capable of determining the pulsation mode of classical Cepheids, and are able to discern them from the anomalous and type II Cepheids. The presence of Cepheids with a secondary bump reaching beyond the real amplitude compromises some of the features, and it will be explored in the future.

The authors acknowledge Universidad de los Andes for the grants and computational resources used in this project.

References

Campello, R. J. G. B., Moulavi, D., & Sander, J. 2013, Density-Based Clustering Based on Hierarchical Density Estimates (Springer Berlin Heidelberg), 160–172
Malzer, C. & Baum, M. 2020, in 2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI) (IEEE)
McInnes, L., Healy, J., & Astels, S. 2017, The Journal of Open Source Software, 2
Pietrukowicz, P., Soszyński, I., & Udalski, A. 2021, Acta Astron., 71, 205–222
Riess, A. G., Anand, G. S., Yuan, W., et al. 2023, ApJ, 956, L18
Simon, N. R. & Lee, A. S. 1981, ApJ, 248, 291
Soszyński, I, Udalski, A., Szymański, M. K., et al. 2015a, Acta Astron., 65, 233
Soszyński, I, Udalski, A., Szymański, M. K., et al. 2015b, Acta Astron., 65, 297
Soszyński, I, Udalski, A., Szymański, M. K., et al. 2018, Acta Astron., 68, 89
Soszyński, I, Udalski, A., Szymański, M. K., et al. 2017, Acta Astron., 67, 297