Understanding nonlinearities with Machine-Learning techniques

Javier Pascual-Granado[1] (javier@iaa.es); **J. R. Rodón Ortiz**[1]; **Cristina Roche Arroyos**[2]; **Mariel Lares-Martiz**[1]; **Giovanni Mirouh**[3]; **Miriam Rodríguez Sánchez**[4]

1 Departamento de Física Estelar. Instituto de Astrofísica de Andalucía, Granada, Spain
2 GMV Innovating solutions, Lisboa, Portugal
3 Departamiento de Física del Cosmos, Universidad de Granada, Granada, Spain
4 Departament d' Astronomia i Astrofísica, Universitat de Valencia, Burjassot, Spain

Abstract

Ultra-precise photometric data gathered by recent space missions allow to improve our knowledge of stars significantly but challenges to find models that fit the observations with the state-of-the-art theory are raised. In addition, the huge amount of data requires the application of innovative data analysis techniques to exploit the information.

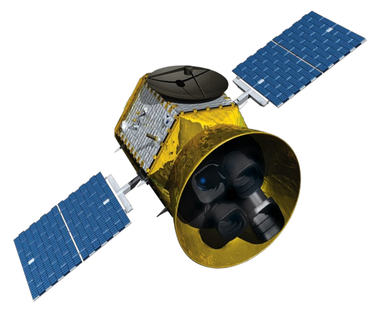
Machine learning could help us both to deal with the large amount of data and to explore new ways to process the lightcurves of pulsating stars. Of special interest for studies of intermediate mass pulsating stars is the decades-old unsolved non-linear phenomenon. **The aim of this project is to use machine learning tools in order to characterize a sample of pulsating stars through clustering analysis to shed some light into the nature of the nonlinear phenomenon.**

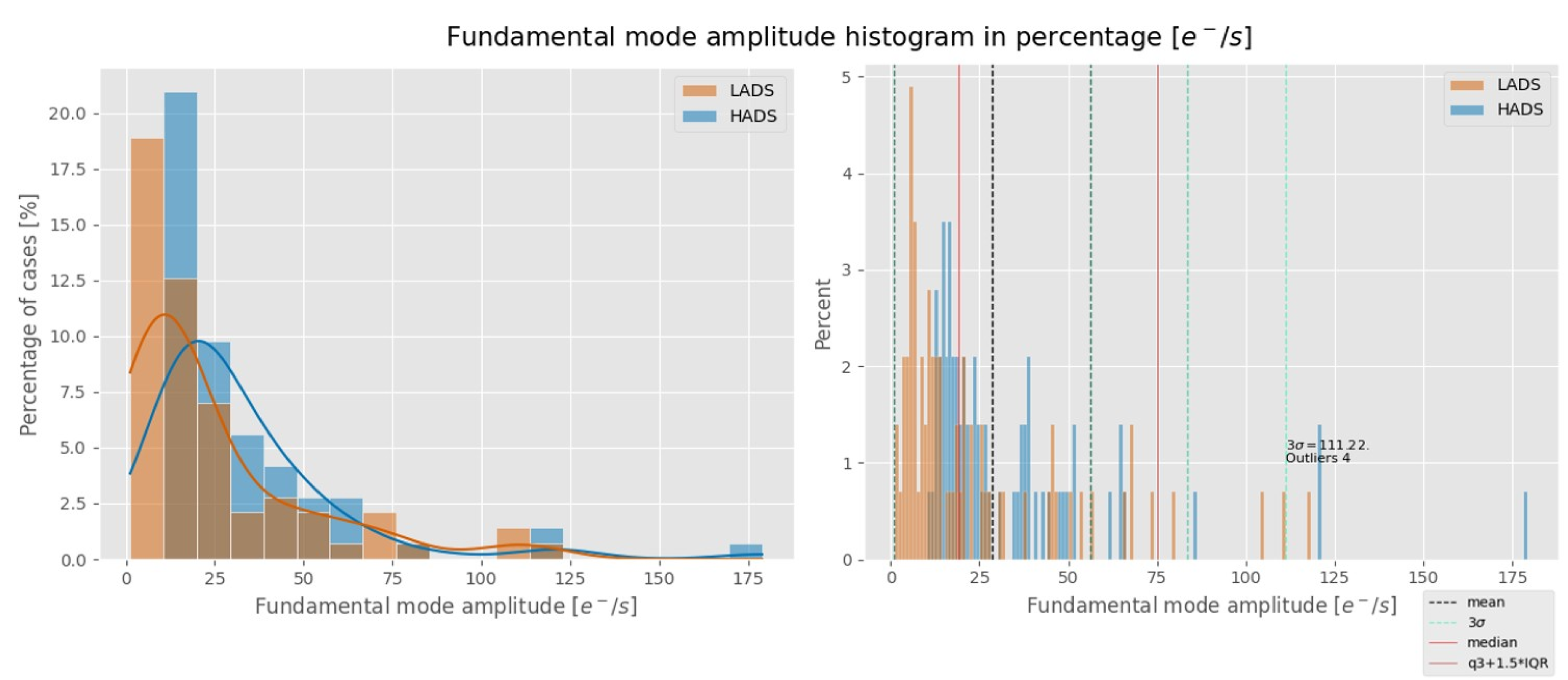
Specifically, we focus here on the **classification of Delta Scuti stars as High Amplitude (HADS) and Low Amplitude (LADS) through the analysis of the nonlinear parameters characterizing the lightcurves** (e.g. harmonics and combinations). The input parameters are extracted from the lightcurves of a **sample of stars observed by TESS space satellite**. The application of ML for understanding nonlinearities have a great potential use for the Complementary Science Program of PLATO

Overview

The current classification of δ Sct stars is based on an ad-hoc threshold on their light curve amplitude, yielding to High-Amplitude Delta Scuti and Low-Amplitude Delta Scuti **[1]**. In order to provide insight on the physical mechanisms that distinguish both classes of pulsating stars, here we have applied data analysis and different clustering techniques to frequency-domain features derived from the decomposition of the light curve of δ Sct stars to find new patterns in their classification.

The study is conducted on a sample of 142 δ Sct stars, balanced between high and low amplitudes from **[2]** where the parameters for the non-linear terms of the decomposition have been obtained through the Best Parent Method **[3]**.

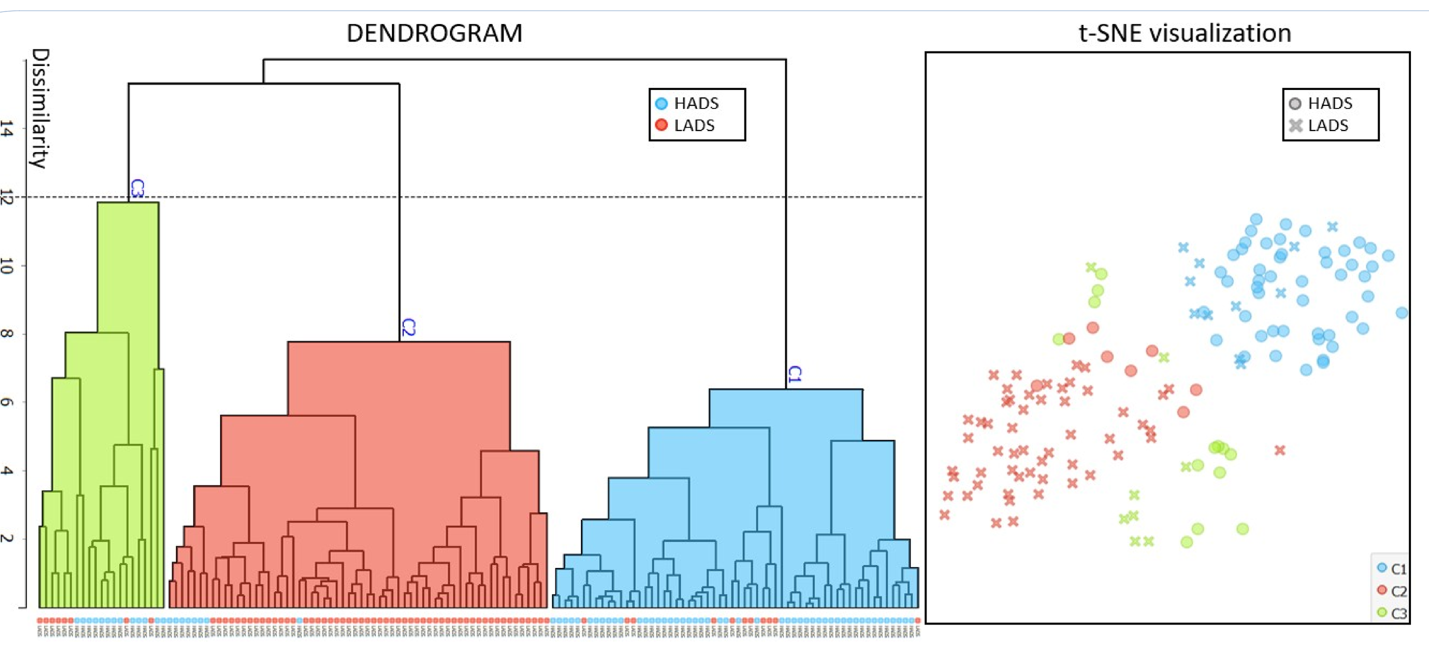




Clustering results based on fundamental amplitude

The insights we have obtained reveal that the amplitude of the fundamental radial mode is insufficient to obtain a separation between HADS and LADS.

The feature selection study proves that including other parameters related besides the fundamental amplitude, provide a better information coverage than with the amplitude alone, plus it provides higher dimensionality to the criterion, leading to extra patterns that can possibly enrich the study of δ Sct physics.



Hierarchical Clustering 9-frequency-domain feature space

$\nu_F, A_F, \phi_F, \nu_{1O}, A_{1O}, n_F, n_{1O}, n_+, n_-$

Dendrogram and t-SNE visualisation obtained by applying HC with Ward's method to the 142 δ Sct sample using 9-frequency-domain features. **Note that 3 clusters are clearly differentiated.**

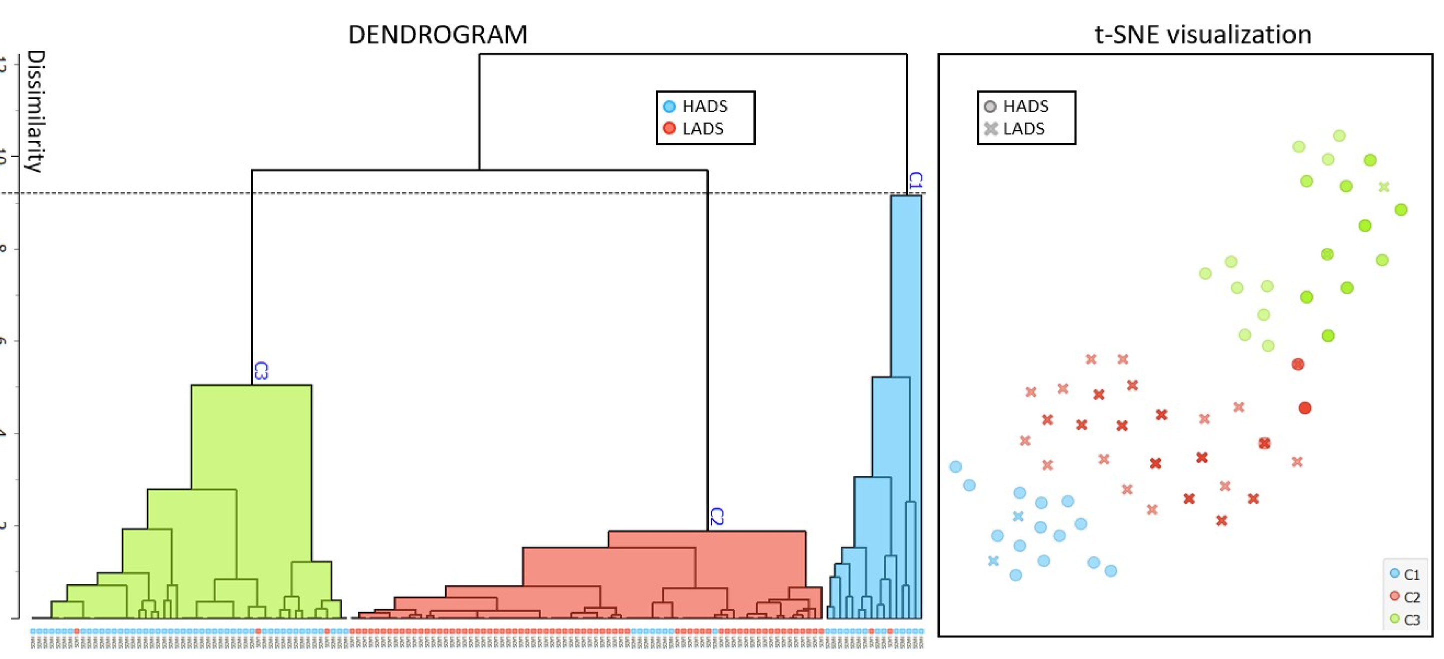
The hierarchical study led to the conclusion that the light curve amplitude may be masking the details of the stellar interior and different mechanisms within δ Sct observations.

A subcluster division within HADS was discovered, responding to features associated with non-linear mechanisms affecting pulsation.

Additionally, **difference combinations n_- were found to be a relevant non-linear feature for the clustering**[1]**. This might be pointing that subclusters are connected to different physical mechanisms causing the nonlinearities**.

The scripts developed for the features processing and clustering techniques application are based on Orange **[4]** and publicly available as a GitHub repository:

https://github.com/cristinaroche/DeltaScuti_Clustering.git



Hierarchical Clustering nonlinear frequency-domain feature space

n_F, n_+, n_-

Dendrogram and t-SNE visualisation obtained by applying HC with Ward's method to the 142 δ Sct sample using 3-non-linear frequency-domain features. **Note that here HADS are spread in 2 of the 3 different clusters.**

[1] see also the poster from M. Rodríguez et al. in this conference.

Conclusions

This study defines a methodology to find a potential new classification system and obtains a scheme of clusters and subclusters that represent a first step into it. It aims to support the development of δ Sct models and reduce the ambiguity in their classification. Due to the complexity found in the choice of adequate feature selection and clustering algorithms for the characteristics of the frequency-domain dataset, we have shown here only a summary of the methodology which provide a road map for future studies of clustering in the field. Further results will appear in Pascual-Granado et al. 2023 (in prep).

References

[1] McNamara, D.: 2000, in Delta Scuti and Related Stars, Vol. 210, p. 373

[2] Lares-Martiz 2021. "Non-linear terms in Delta Scuti stars power spectra", doctoral thesis, co-supervised by R. Garrido and J. Pascual-Granado

[3] Lares-Martiz, M., Garrido, R., Pascual-Granado, J., 2020, MNRAS, 498, 1194

[4] Demšar, J., et al (2013). Orange: Data Mining Toolbox in Python. Journal of Machine Learning Research, 14, 2349–2353