Application of the Ensemble Empirical Mode Decomposition (EEMD) method in astronomy

Alejandra Christen (1), Suryani Guha (2,3,) and Julieta P. Sanchez Arias (2)

(1) Institute of Statistics, University of Valparaiso, Chile; (2) Astronomical Institute, Czech Academy of Science, Ondrejov, Czech Republic; (3) Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic

While the Fourier method is widely known and used for frequency analysis, it has been found to be inadequate for signals with time-varying periods or with high amplitude noise. The empirical ensemble mode decomposition (EEMD) is a powerful tool that uses the Hilbert-Huang transform (HHT) to decompose non-stationary and nonlinear signals into finite and linear independent components called intrinsic modal functions (IMFs). To demonstrate its application, we choose a set of simple synthetic signals with single frequencies up to signals with multiple frequency components. In addition, we added random noise to the signal to test the sensitivity of this method in the presence of noise with different amplitudes. Furthermore, to investigate the effectiveness of this method, we apply EEMD for detecting frequencies in real light curves of selected known pulsating stars.

Empirical Mode Decomposition (EMD)

EMD is a signal analysis approach that breaks down multi-component signals Y(t) into a finite number of adaptive simple oscillatory modes, h_i known as intrinsic mode functions (IMFs) and a residual, r, the final residual or the data trend.

$$Y(t) = \sum_{i=1}^{n} h_i + r$$

Properties of EMD :

ADAPTIVE: EMD decomposes a signal based on its inherent characteristics without any predetermined basis functions. The basis functions (IMFs) are derived directly from the data.

BASIS ELEMENT: Each IMF has a zero mean and is symmetric with respect to local zero-crossings. IMFs also tend to be narrow-band signals, each representing a simple oscillatory mode.

VERSATILITY: This method is potentially viable for nonlinear and non-stationary data analysis.

ITERATIVE PROCESS: An iterative sifting process is performed, where local extrema are identified, and envelopes are created to isolate IMFs through the media.

TIME - FREQUENCY ANALYSIS: Hilbert–Huang transform (HHT) is the related spectral method for EEMD method. The HHT defines instantaneous frequency as the time derivative of phase, allowing to perform a time-frequency analysis.

EEMD or EMD ?

One challenge of EMD is mode mixing, where different signal components are not properly separated. This can occur when the signal contains closely spaced frequencies.

To overcome the mode-mixing problem a new approach consists of sifting an ensemble of white noise-added signal (data) and treats the mean as the final true result. This method is known as Ensambled Empirical Mode Decomposition (EEMD) method.

The proposed EEMD is developed as follows :

1. **Adds white noise series to the targeted data.**
2. **Decompose the data with added random white noise into IMFs.**
3. **Repeat step 1 and 2 several times.**
4. **Obtain the means (ensemble) of the corresponding IMFs as final result.**

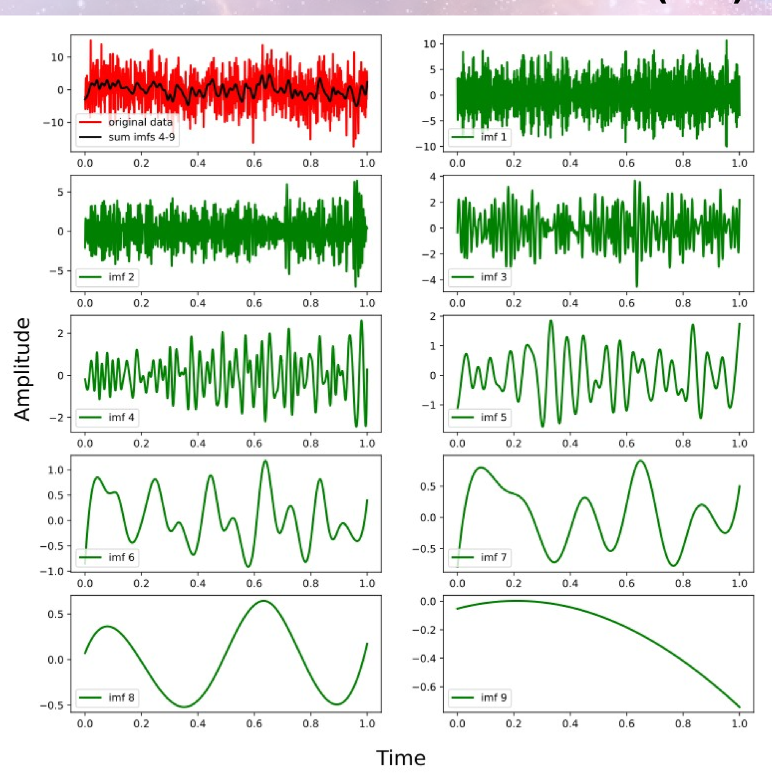
Comparison and strengths of (Fourier, WWZ and EEMD)

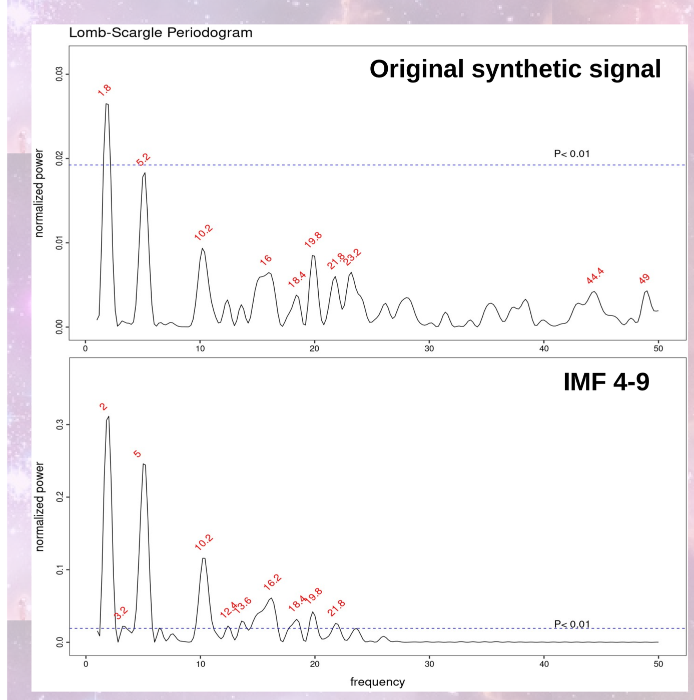
	Fourier Method	WWZ	EEMD
Basis	a priori	a priori	adaptive
Presentation	energy-frequency	energy-time-frequency	energy-time-frequency
Efficiency in Signal analysis	Linear and stationary	Non-stationary and linear	Non-sationary and nonlinear

Examples with synthetic signals

To test the efficiency of the EEMD method for detecting periods, we performed numerical simulations in R for 4 synthetic light curves with different types of noise and known periods.

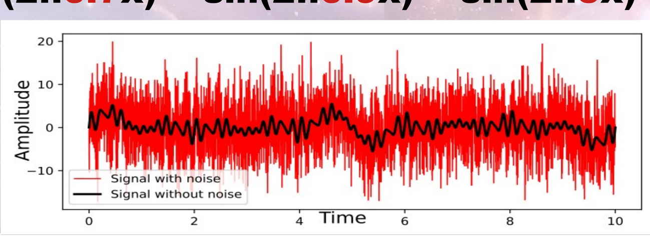
Case 1 : LARGE NOISE. White noise(σ=5) with Signal = sin(2π2x) + sin(2π5x) + sin(2π10x)

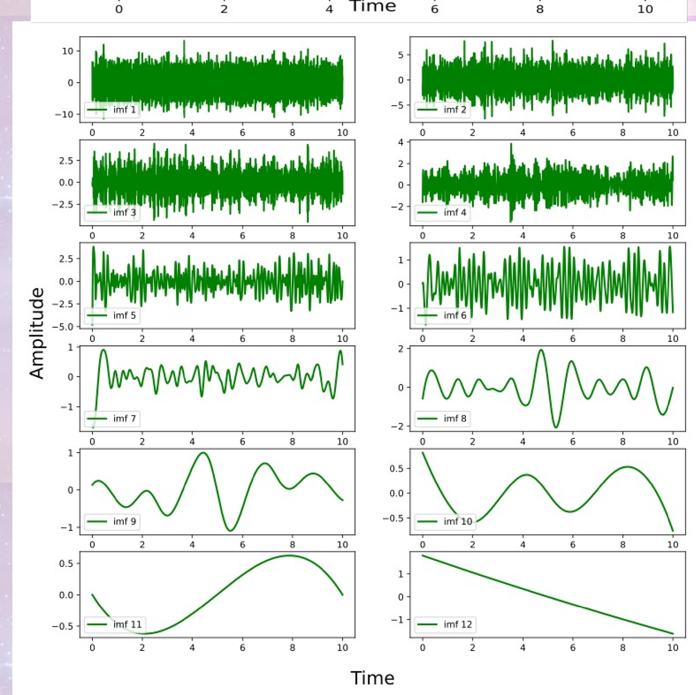


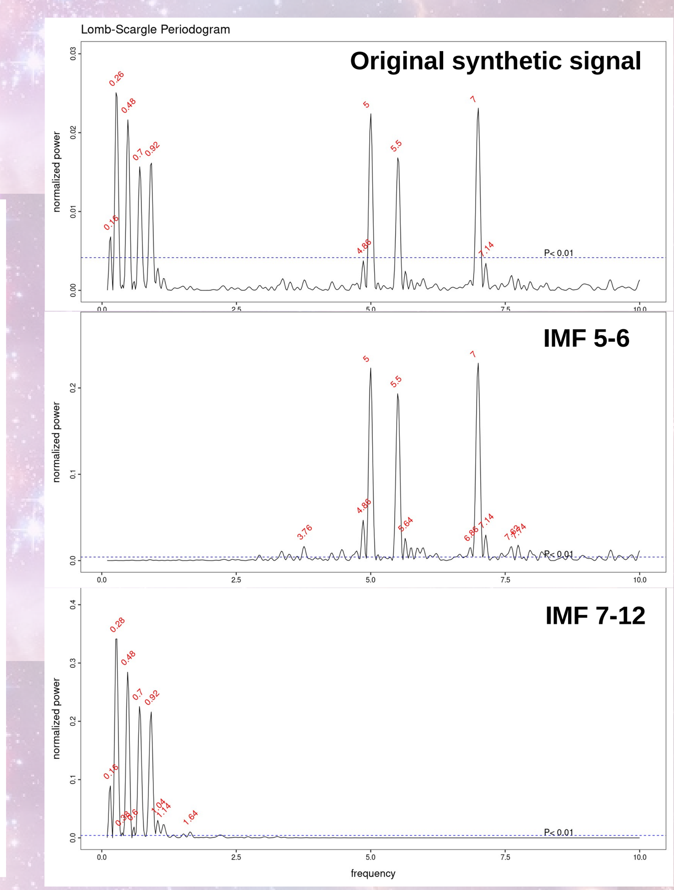


CASE 1	
LS-ORIGINAL	LS-IMF4-9
Recover true freq.	Recover true freq.
Bad SNR	Improved SNR
Low power for the freq. peaks	High power for the freq. peaks
Many false freq.	Fewer false freq.

Case 2 : LOW & HIGH FREQ. white noise(σ=5) with Signal = sin(2π0.3x) + sin(2π0.5x) + sin(2π0.7x) + sin(2π0.9x) + sin(2π5x) + sin(2π5.5x) + sin(2π7x)



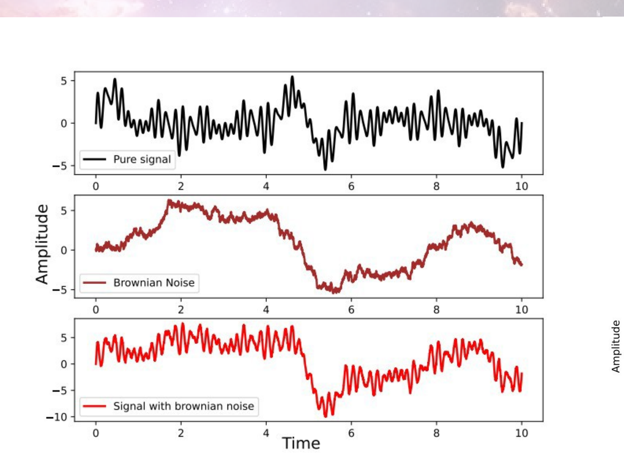


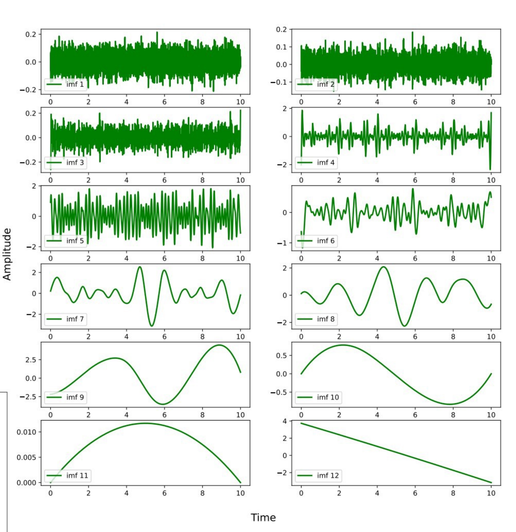


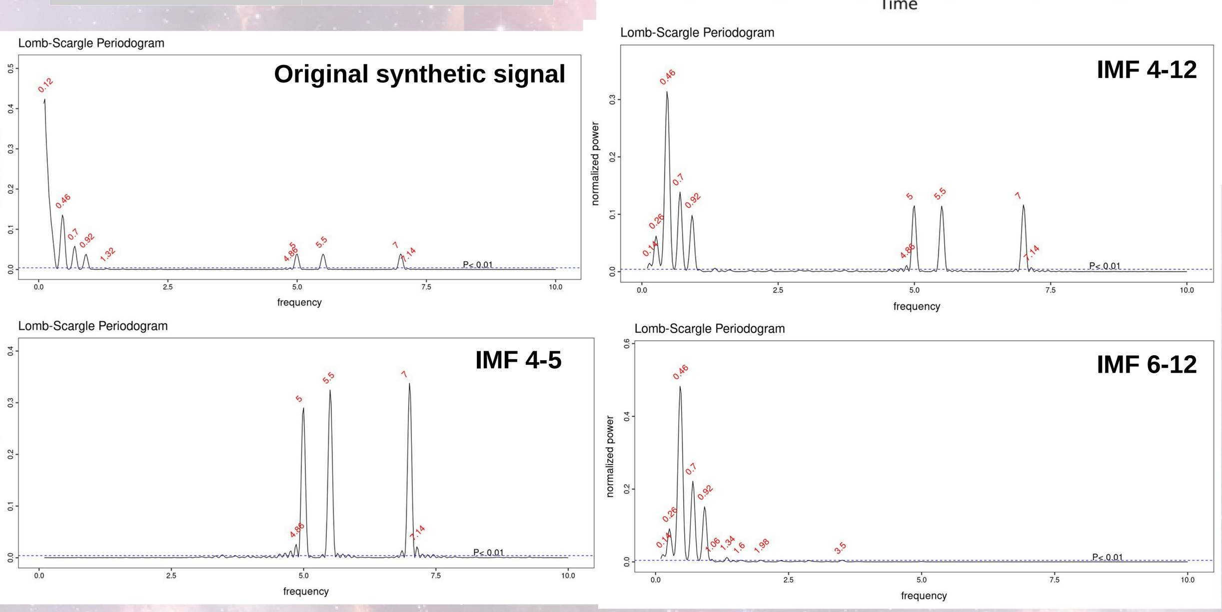
CASE 2	
LS-ORIGINAL	LS-IMF5-6/ LS-IMF7-12
It recovers the true high freq.	Recover the true high freq.
Recovers the low freq. but they are not sufficiently accurate.	Recover the low freq. but they are not sufficiently accurate.
Good SNR	Improved SNR
Low power for the freq. peaks	High power for the freq. peaks
Few false frequencies	Some false frequencies

Case 3 : LOW & HIGH FREQ. Brownian noise with Signal = sin(2π0.3x) + sin(2π0.5x) + sin(2π0.7x) + sin(2π0.9x) + sin(2π5x) + sin(2π5.5x) + sin(2π7x)

CASE 3	
LS-ORIGINAL	LS-IMF4-5/ LS-IMF6-12
Recover the true high freq. but the lowest.	Recover the true high freq.
Recover the low freq. but they are not sufficiently accurate.	Recover the low freq. but they are not sufficiently accurate.
Good SNR but red noise is present.	Improved SNR without red noise.
High power peaks for false freq.	High power peaks for true freq.
Some false low freq.	Some false low freq.





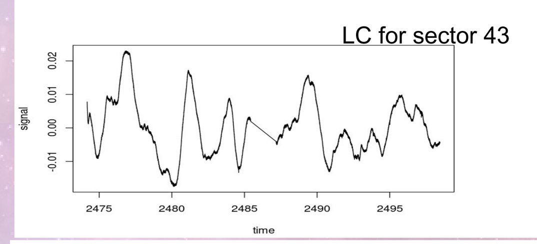


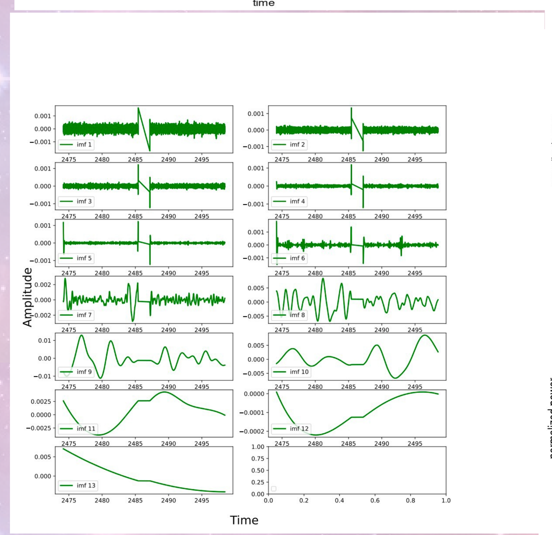
- Time computing depends strongly on the amount of sample data.
- Selection of the MFIs to be grouped depends on the signal.

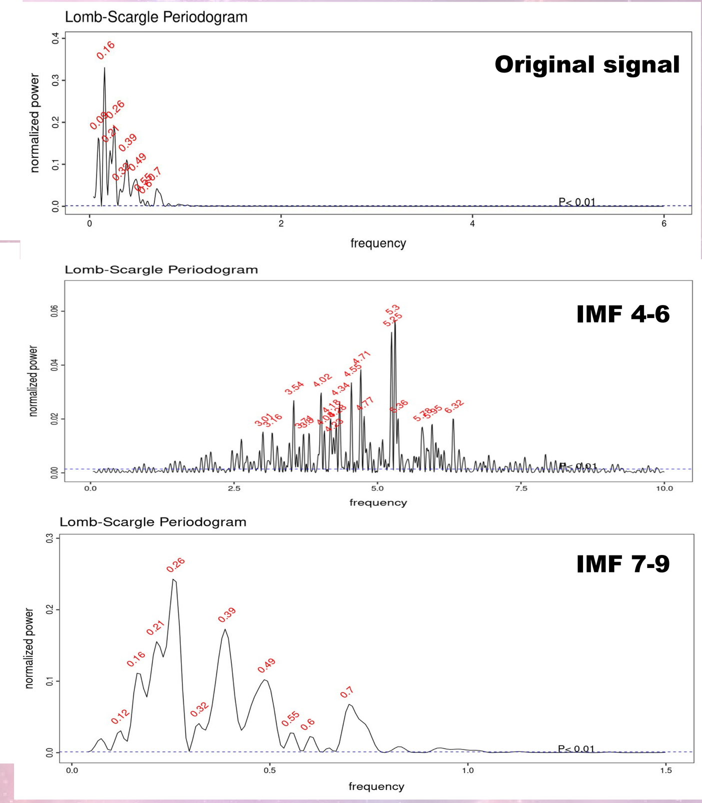
Applying EEMD to analysis the light curves of different stars :

Target 1 : HD 42087 – B supergiant star

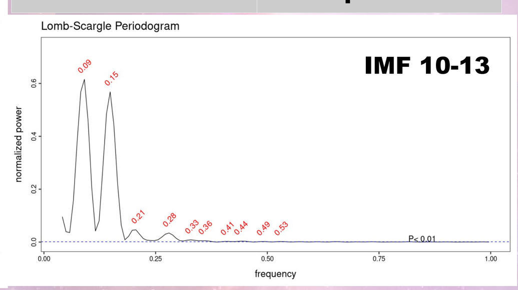
This object is a B supergiant star, proposed to be in the Terminal Age Main Sequence [4]. We used the 2 min TESS cadence from Sector 43. Our main interest in this the presence of red noise that usually affect the frequency spectra of B supergiants.





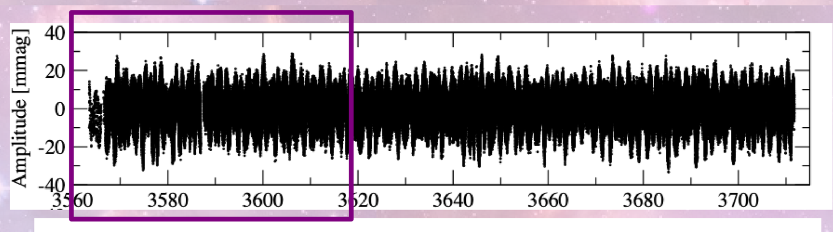


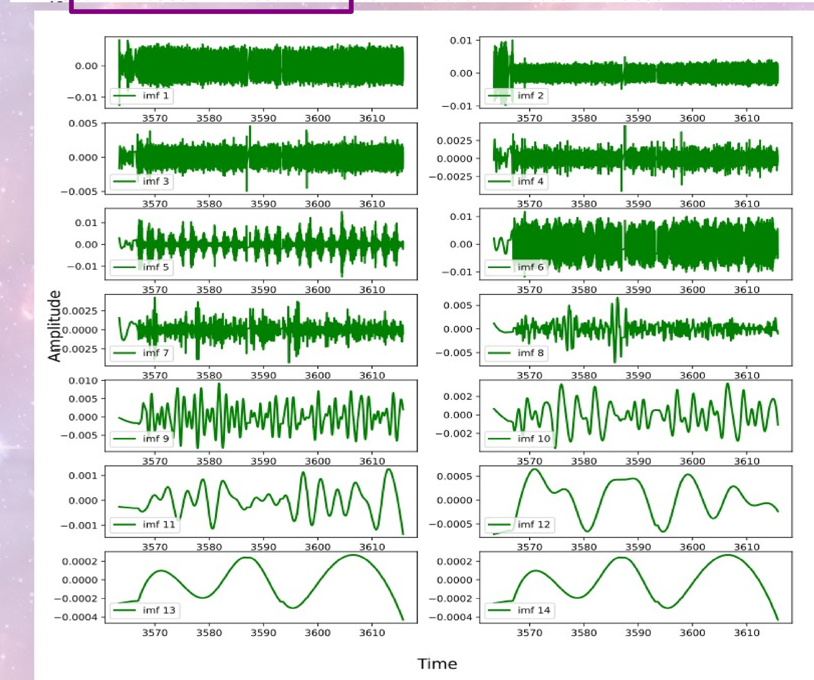
HD 42087 – B supergiant star	
LS-ORIGINAL	LS-IMF
Information displayed in on single periodogram.	Allows to retrieve frequencies in different ranges separately.
Exhibit gradual increase in power towards low frequencies	Don't exhibit increase in power towards low frequencies for middle IMFs sum (IMF 7-9).
Recover low freq.	Recover low freq.
Don't recover high freq.	Recover some high freq. with small power.

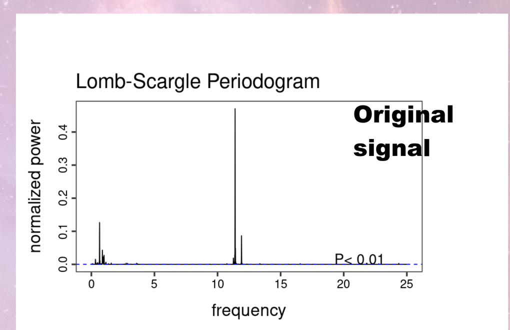


Target 2 : CoRoT 102314644 – Main sequence star

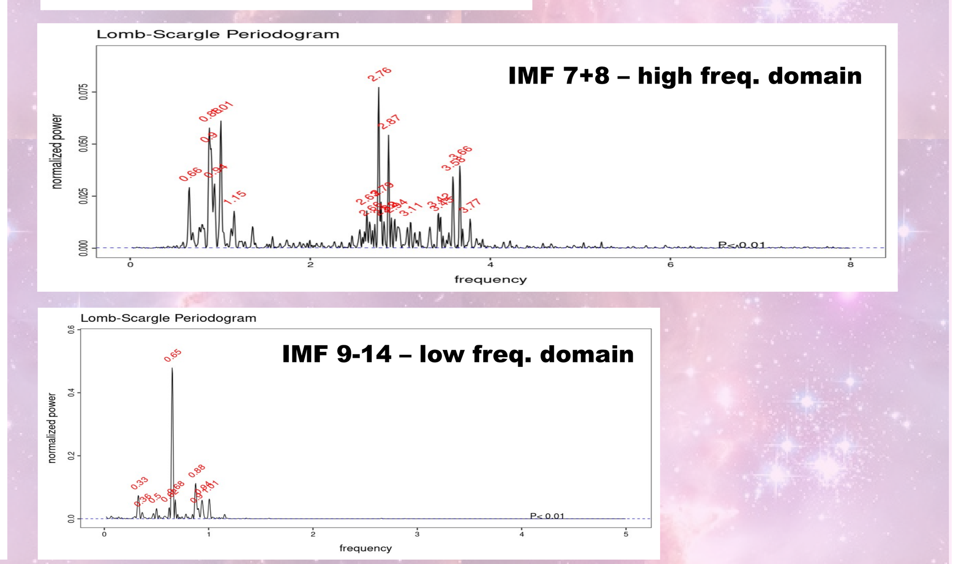
This object is an hybrid delta Sct-gamma Dor star. We employed the light curve from the third CoRoT long run LRa03 observations which lasted 148 d. Our main interest in this type of main sequence stars lie in their rich frequency spectrum having simultaneously p- and g-modes in different frequencies range.







RESULTS: We could efficiently detect separately the low and high frequency domain with the EEMD method.



Evolutionary and pulsational models

With the aim to compare the detected frequencies with theoretical predictions, we developed evolutionary and pulsational models with MESA [5] and GYRE [6] codes, respectively.

HD 42089 : We consider an effective temperature (Teff) of 18200 K as our first approximation and calculated evolutionary tracks with Teff=18200K and Mini=26, 28 and 30 Mo within the error box from [4] (see Fig.1). For each model we computed the adiabatic and non-adiabatic frequencies within the range [0.9, 6] c/d for l=0, 1 and 2. Fig 2 show the derived and observed frequencies. Our models have initial rotation O/Ocrit=0.4, Z=0.0142 and Y=0.2703. We have implemented Vink recipe for the mass loss.

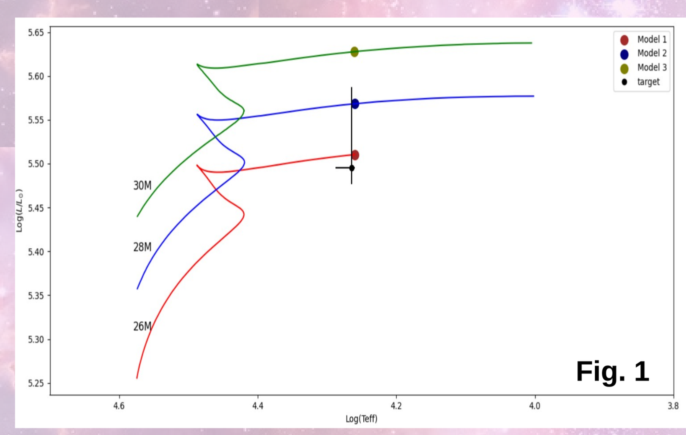


Fig. 1

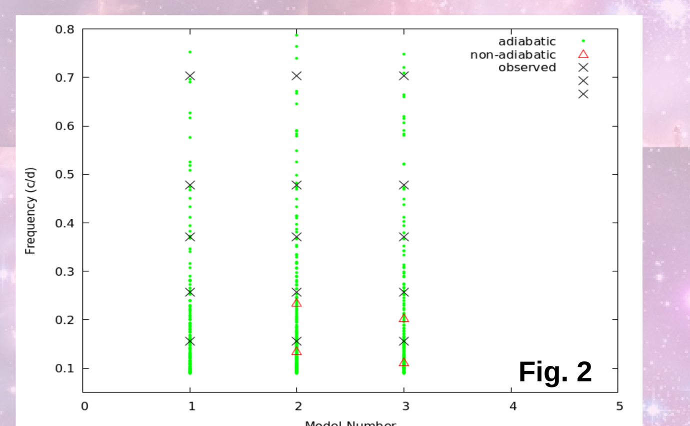


Fig. 2

RESULTS: Our non-adiabatic models predict only few excited frequencies below 0.25 c/d for . We could not predict the frequencies within [2.5, 6] c/d detected with the sum of IMF 7+8+9. The 28 Mo model fits better the observed frequencies.

CoRoT 102314644 : We consider Teff= 7244 K as our first approximation and calculated evolutionary tracks for Mini=1.75, 1.8 and 1.85 Mo within the derived error box in [7] (see Fig. 3). For each model we computed the adiabatic and non-adiabatic frequencies in the range [0.3, 24] c/d for l=0, 1 and 2. Fig. 4 and Fig 5 show the derived frequencies and those find in [7] for the p-modes and g modes domain, respectively. Our models have solar solar metalicity and nor rotation nor core overshooting included. We used the Henyey convection theory.

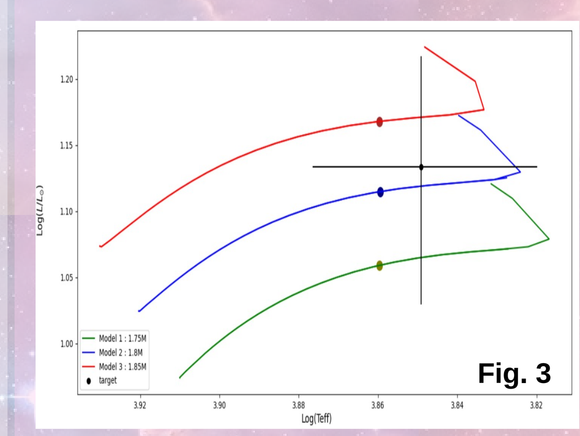


Fig. 3

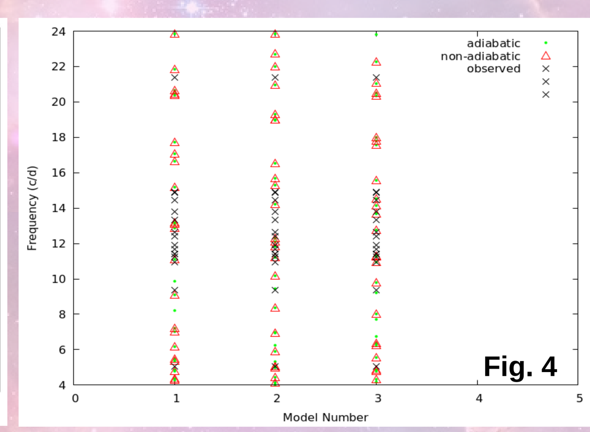


Fig. 4

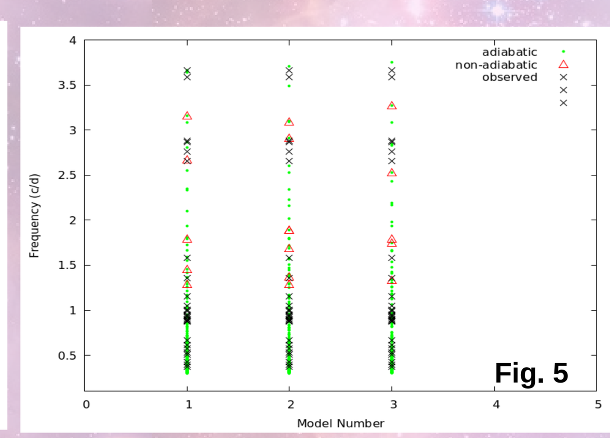


Fig. 5

RESULTS: Our non-adiabatic models do not reproduce the observed g-modes below 0.8 c/d, questioning the mode classification in [7]. We noticed prominent spikes in the Brunt Vaisala frequency at the end of the convective core indicating strong steep chemical transition between the H-core and the envelope when using the Henyey convection theory in contrast with the classical MLT, which affect the low g-modes frequency range. The excited low p–modes frequencies suggest a misclassification for modes in this range. However rotation should be included in our models to compare both results.

Conclusions and future work

We have demonstrated the potential of the EEMD method to retrieve frequencies in synthetic data. We could detected high frequencies in HD 42087 of uncertain origin. A full simulation design, including pulsations and exhibiting excess power at low frequencies (red noise phenomena) will be carried out to explore the performance of EEMD. Numerical improvements in pulsational codes are needed to retrieve non-adiabatic modes in the case of highly non-adiabatic envelopes. Different convection theories and rotation should be further explore for our COROT star.

Refences:

[1] Huang and Shen 2014. World Scientific Publishing Co. Pte. Ltd.
[2] Huang et al. 1998. Proceedings: Mathematical, Physical and Engineering Sciences, 454, 903-995.
[3] Gaci, 2016. Energy Procedia 97, 84 – 91.
[4] Sanchez Arias, J. P. et al. 2023. Galaxies, vol. 11, issue 5, p.93.
[5] Paxton B., et al., 2019, ApJS, 243, 1.
[6] Townsend R. H. D., Teitler S. A., 2013, MNRAS, 435, 3406.
[7] Sanchez Arias, J. P. et al.2023. A&A, 676, p.11.

Acknowledgement : This project has received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014-2020) under the Marie Skłodowska-Curie Grant Agreement No. 823734.