

MIARMA: a gap-filling method preserving the frequency content of the signal

J. Pascual-Granado¹, R. Garrido¹, and J. C. Suárez^{1,2}

¹ Instituto de Astrofísica de Andalucía (CSIC)

² University of Granada (UGR). Dept. Theoretical Physics and Cosmology, 18071, Granada, Spain

Abstract

Most of the techniques used in the analysis of time series from pulsating stars are based on the hypothesis of the analyticity of their underlying function. Thereby, gap-filling methods are also based on this hypothesis. Here we introduce MIARMA, a method for filling the gaps using Autoregressive Moving Average models. This method make use of a non-closed form expresion which can be analytic or not. We show that MIARMA can recover signal that no analytic interpolation method can. This method has been accepted by the CoRoT Scientific Committee for its implementation in the next correction of the data.

1 Introduction

Space missions like CoRoT and *Kepler* have observed a large sample of stars with two main objectives: transit detection and asteroseismology. For asteroseismology a long and uninterrupted observation is required to obtain reliable results. In order to detect transits more precise measurements are required. The requisites for both these objectives are fulfilled by these space missions but a photometric time series without gaps is an ideal case which is never reached in practice.

As pointed by [3], gaps change the original frequency content of any sampled signal, because the observed power spectrum is the result of the convolution of the original signal with the observational window. When the gaps are small as compared with the total time span, gap-filling techniques have been commonly used. For larger gaps, the prewhitening techniques become unavoidable.

Most of the popular gap-filling algorithms do not guarantee the preservation of the original frequency content, which is a requisite for asteroseismology to be reliable. A commonly used approach is a linear interpolation between valid data segments. This is the case for

CoRoT seismofield data. We show that in some cases this solution is insufficient to remove aliases.

We introduce a new gap-filling algorithm (MIARMA) which is based on a non-closed formula and can represent any kind of function, even non-analytic functions, thus preserving at best the original frequency content of the signal. This method make use of a forward-backward predictor based on autoregressive moving average (ARMA) models in order to fill the gaps. By applying this method to time series of stars with different pulsational characteristics observed by CoRoT satellite, new properties of the light curves are revealed.

2 MIARMA

As stated by Wold's theorem [8] any stationary signal can be represented as the sum of a deterministic (AR) plus a stochastic process (MA). In this way, these models are robust since an exact representation of the signal can be found independently of its properties: periodic or aperiodic, deterministic or stochastic, analytic or non-analytic, chaotic or not.

The algorithm MIARMA (Method of Interpolation by Autoregressive and Moving Average) is similar to the original algorithm introduced by [4] but instead of AR models ARMA models are used. Also, instead of a global model for the whole data set MIARMA implements local models.

A mixed autoregressive moving average process ARMA(p,q) can be expressed as

$$x_n = \sum_{k=1}^p a_k x_{n-k} + \sum_{k=1}^q b_k \epsilon_{n-k} + \epsilon_n, \quad (1)$$

where x_n is the time series, a_k are the p parameters of the autoregressive part of the model, ϵ_n is an independent white noise process and b_k are the q parameters of the moving average part of the model.

The order (p,q) of the model is selected by using a statistical criterion that guarantees minimal loss information, the Akaike Information Criterion (AIC) [1]. This is possible because ARMA models have a non-closed form expression, that is, no hypothesis about the form of the signal is introduced.

Once the order is selected using AIC, an ARMA model is locally fitted to the data segment located before and after every gap. In order to do this the coefficients a_k and b_k appearing in Eq.1 have to be determined for each data segment. MIARMA calculates the coefficients using an iterative algorithm that minimizes the sum of the quadratic errors checking the convergence after a number of iterations.

The algorithm MIARMA consists in three steps: first, a criterion based on physical principles for selecting the optimal order of the ARMA model; secondly, the data segments before and after a given gap are fitted using an ARMA model of the order selected in the first step; next, the gap is interpolated using a weighted function of a forward and a backward prediction based on the models of the selected data segments. The second and third steps are repeated for each gap contained in the entire time series.

This algorithm is particularly suited to process time series like the ones observed by CoRoT, with small gaps compared to the data segments between them.

3 Results: CoRoT data

The passing through the SAA introduces most of the invalid data (gaps) that are present in CoRoT observations. As described in [7], two sets are available in CoRoT level 2 data: gapped data, and a regularly sampled dataset obtained by interpolating linearly valid data segments.

We compare data interpolated using MIARMA with the original CoRoT linearly interpolated data in three cases of stars showing different pulsational characteristics: the δ Scuti HD 174966, showing periodic variations of the same order as the CoRoT observational window, the Be star HD 51193, showing longer time variations, and the solar-like HD 49933, with rapid time variations.

3.1 HD 174966

The case of A-F main sequence stars is particularly critical because they show pulsation frequencies close to the orbital frequency of the satellite. In particular the δ Scuti star HD 174966 could be considered as a paradigmatic case.

The power spectrum of the linearly interpolated time series is seriously affected by the spectral window of CoRoT (see Fig. 1) in contrast with the ARMA interpolation where the effects of the spectral window are eliminated. The impact of the linear interpolation in the power spectrum is so strong as gapped data.

Definitively, the linear interpolation does not improve the original power spectrum biased by orbital aliases in this kind of pulsating stars.

3.2 HD 51193

When the period of variation is much larger than the size of the gaps, which is the case of Be stars, it is commonly assumed that so small gaps have no effect on the power spectrum. Therefore, the filling provided by linear and ARMA interpolation is expected to be similar. However, we show here that for HD 51193 the power spectrum obtained from the linearly interpolated data is clearly biased by the spectral window of the CoRoT orbital frequency (13.97 cd^{-1}) in contrast to the ARMA interpolation, in which such an effect has been removed (see Fig. 2).

3.3 HD 49933

The solar-like star HD 49933 was chosen to test MIARMA with a light curve where CoRoT gaps contain several stellar pulsation cycles. Although, the most frequent gap length is 9 mins [2], some gaps last up to 0.2 days. Furthermore, this time series has a very high level of noise.

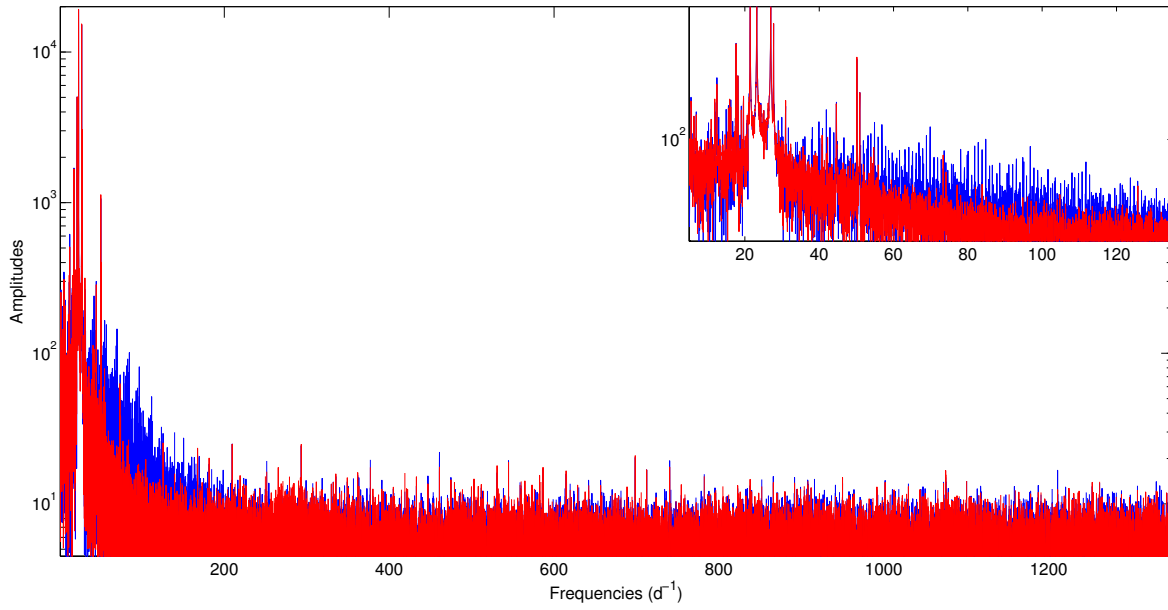


Figure 1: Periodogram of the light curve of HD 174966 observed by CoRoT after the ARMA interpolation (red), and after the linear interpolation (blue). In the inset the range where δ Scuti pulsations are excited is shown.

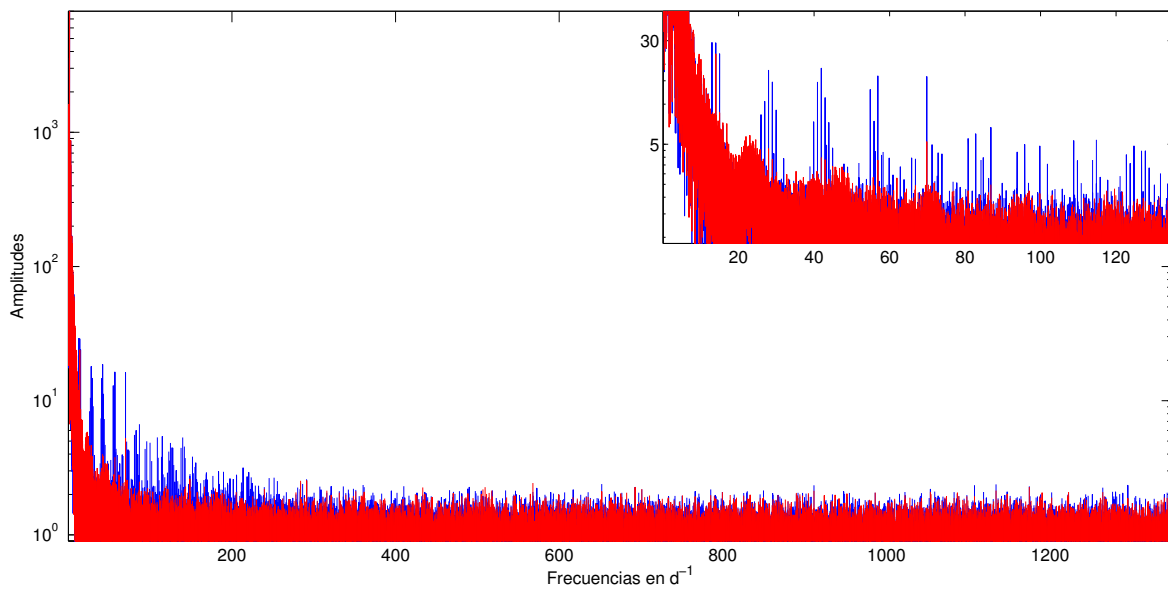


Figure 2: Periodogram of HD 51193 obtained after ARMA interpolation (red) and after linear interpolation (blue). In the inset the range from 0 to 130 d^{-1} is shown.

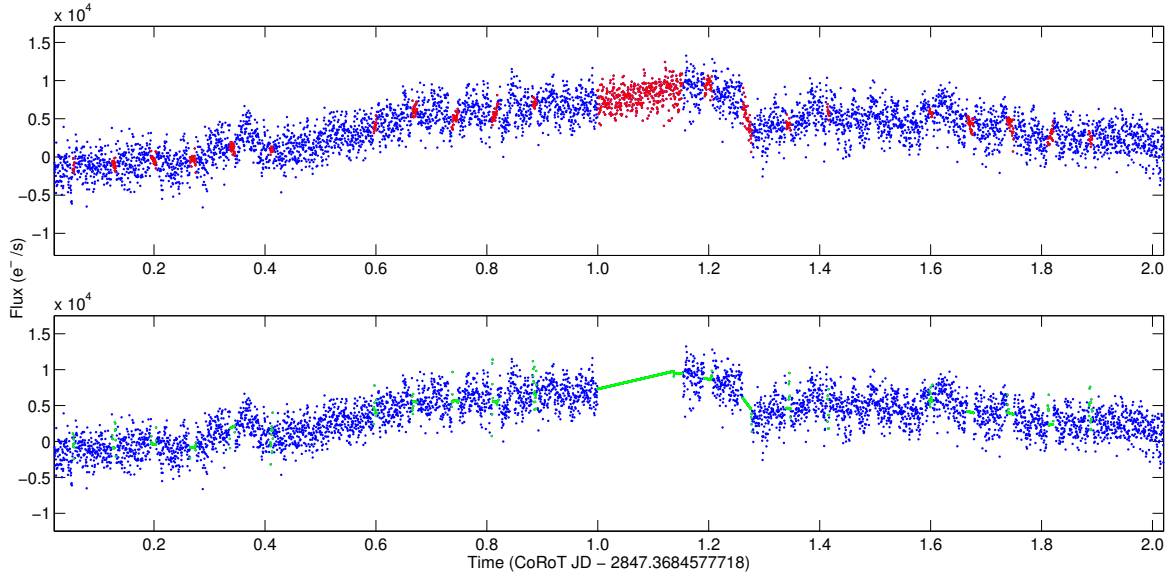


Figure 3: Comparison between ARMA interpolation (upper panel in red) and linear interpolation (lower panel in green) for the light curve of HD 49933 observed by the CoRoT satellite. Valid data is depicted in blue (both panels).

All these make the linear interpolation quite erratic (see Fig. 3). Also, in the gap between 1.0 and 1.2 days, ARMA predicts a signal with a kind of fine structure. We do not know the origin of such a structure, but we are confident that ARMA prediction does not correspond to a white noise process because the method is based on signal identification assuming white noise. The results obtained when studying the connectivities [5, 6] of this time series could be on the origin of this phenomenon. The comparison of power spectra obtained with the two interpolation methods (Fig. 4) looks quite similar with slightly higher amplitudes for the ARMA interpolation (see inset). However, a detailed study of their difference using the absolute value of the amplitudes ($\Delta A = |A(\text{ARMA}) - A(\text{LIN})|$) shows non-negligible and non-systematic differences (Fig. 4, lower panel).

4 Conclusions

In summary, we have shown here that MIARMA preserves at best the original information contained in the time series. This method proves to be very suitable for eliminating the aliases on CoRoT light curves whatever the type of variability considered.

We have found that the power spectrum of HD 174966 is clearly aliased when linear interpolation is used for filling the gaps; the light curve of HD 51193 presents a much more aliased spectrum than expected for a low frequency harmonic signal; and finally, although the linear interpolation does not affect noticeably the power spectrum of the CoRoT light curve of the solar-like star HD 49933, the ARMA interpolation shows rapid variations previously unidentified that MIARMA interprets as a signal.

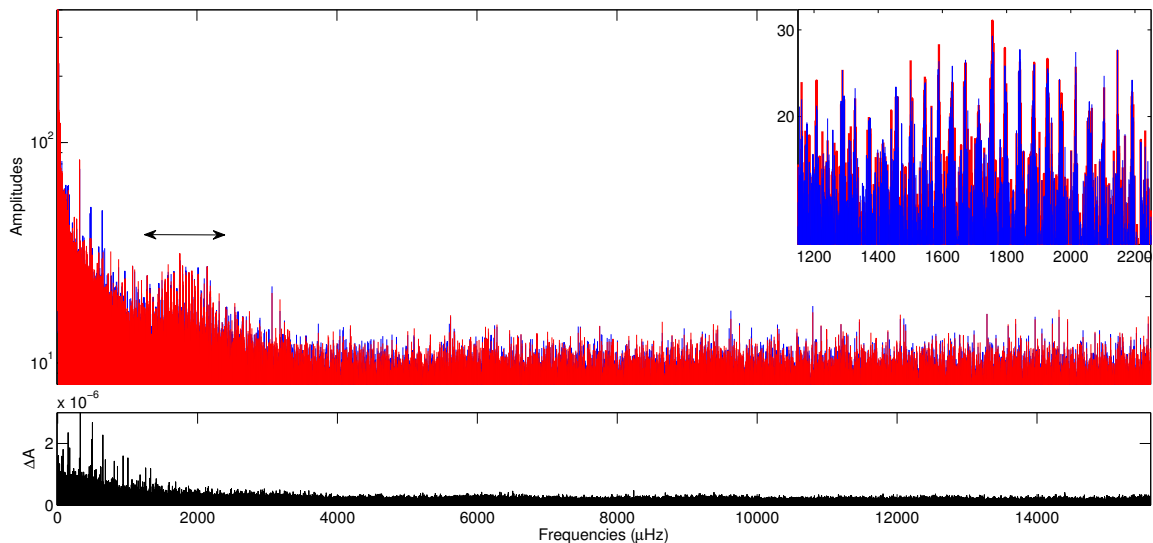


Figure 4: Periodogram of HD 49933 coming from data linearly interpolated (blue), and ARMA interpolation (red). Notice that the plotting order is inverted in the inset because the peaks in the p-mode region are higher for the periodogram of ARMA interpolated data. Lower panel shows the absolute differences in amplitude between both methods.

Acknowledgments

JPG and RG acknowledge support from the “Plan Nacional de Investigación” under project AYA2012-39346-C02-01, and from the “Junta de Andalucía” local government under project 2012-P12-TIC-2469. JCS acknowledges support from MINECO and FEDER funds through the Astronomy and Astrophysics National Plan under number AYA2012-39346-C02-01. Also acknowledges support from the European project SpaceInn with reference 312844 within the european SPACE program FP7-SPACE- 2011-1, and also acknowledges funding support from the Spanish Ministerio de Economía y Competitividad under Ramón y Cajal subprogram.

References

- [1] Akaike, H. 1974, IEEE Trans. Automatic Control, AC-19, 716
- [2] Appourchaux, T., Michel, E., Auvergne, M., et al. 2008, A&A, 488, 705
- [3] Deeming, T. J. 1975, ApSS, 36, 137
- [4] Fahlman, G. G. & Ulrych, T. J. 1982, MNRAS, 199, 53
- [5] Pascual-Granado, J., Garrido R., & Suárez, J. C., 2014. CUP, IAU Symposium 301, pp.85-88
- [6] Pascual-Granado, J., Garrido R., & Suárez, J. C., 2014. This volume
- [7] Samadi, R., Fialho, F., Costa, J. E. S., et al. 2007, arXiv, 3354
- [8] Wold, H. 1938, A study in the analysis of stationary time series 2nd ed. (Almqvist and Wiksell)