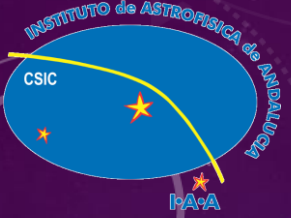




# MIARMA: AN INFORMATION PRESERVING GAP-FILLING ALGORITHM

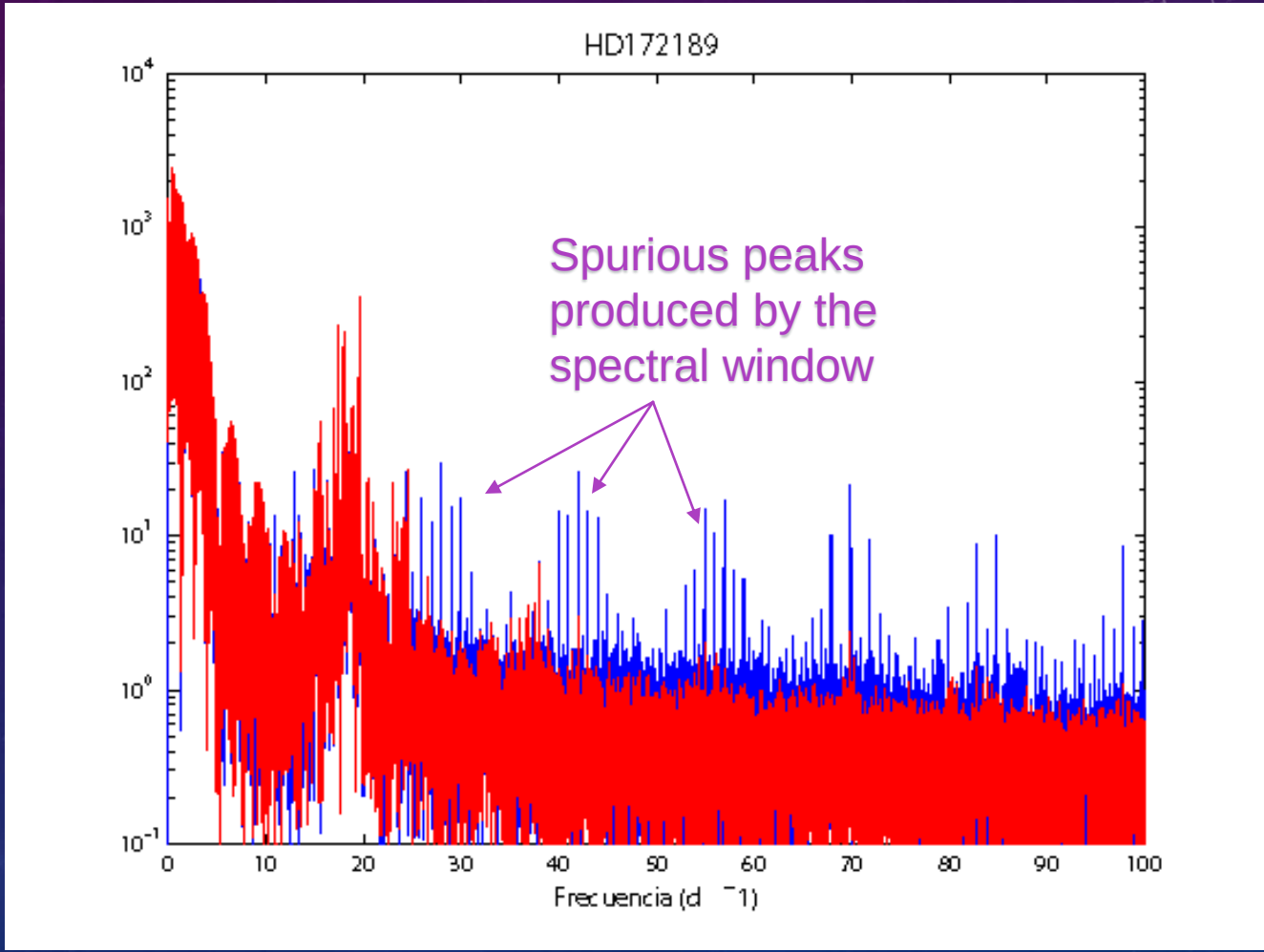
**Javier Pascual Granado**

**Rafael Garrido Haba**

Missing data in physics,

11-12 May 2015 Nice

# CoRoT crossing through the SAA introduce spurious peaks




Pascual-Granado et al., CUP 2012, IAUS285, pp.392-393

A gap-filling

preserving information

Unbiased

Non-closed form expression, fitting functions that can be analytic or not.

# GAP-FILLING TECHNIQUES

| Hypothesis | Specific | Approximation | Order |
|---|---|---|---|
| Analyticity<br>Stationarity<br>Linearity<br>Sparse signal<br>Signal properties<br>Noise properties<br>…. | Solar observations<br>Satellites | Limited by the number of terms in the model | No clear criterion for selection of the order |
| **Polinomial**<br>**CLEAN**<br>**Inpainting** | **Brown & JCD**<br>**Repetitive music** | **Fahlman & Ulrych** | **Roques** |

No hypotheses, general method, signal representation, clearly defined criterion for the selection of the order?

**ARMA**

# THE MODEL:

**AR** $$x_t = \sum_{k=1}^{p} \alpha_k x_{t-k} + a_t$$ Purely Autoregressive

**MA** $$x_t = -\sum_{k=1}^{q} b_k n_{t-k}$$ Moving Average

**ARMA** $$x_t = \sum_{k=1}^{p} \alpha_k x_{t-k} - \sum_{k=1}^{q} b_k n_{t-k} + a_t$$ Mixed AR + MA

**Wold's theorem**: any stationary signal can be represented by as an ARMA process.

# CRITERION FOR SELECTION OF THE ORDER (P,Q)

- An ungapped data segment is modelled. Iteration through p, q

- Given the k model, its Akaike coefficient is obtained ($AIC_k$)

$$AIC_k = N.\log(V) + 2(p+q)$$

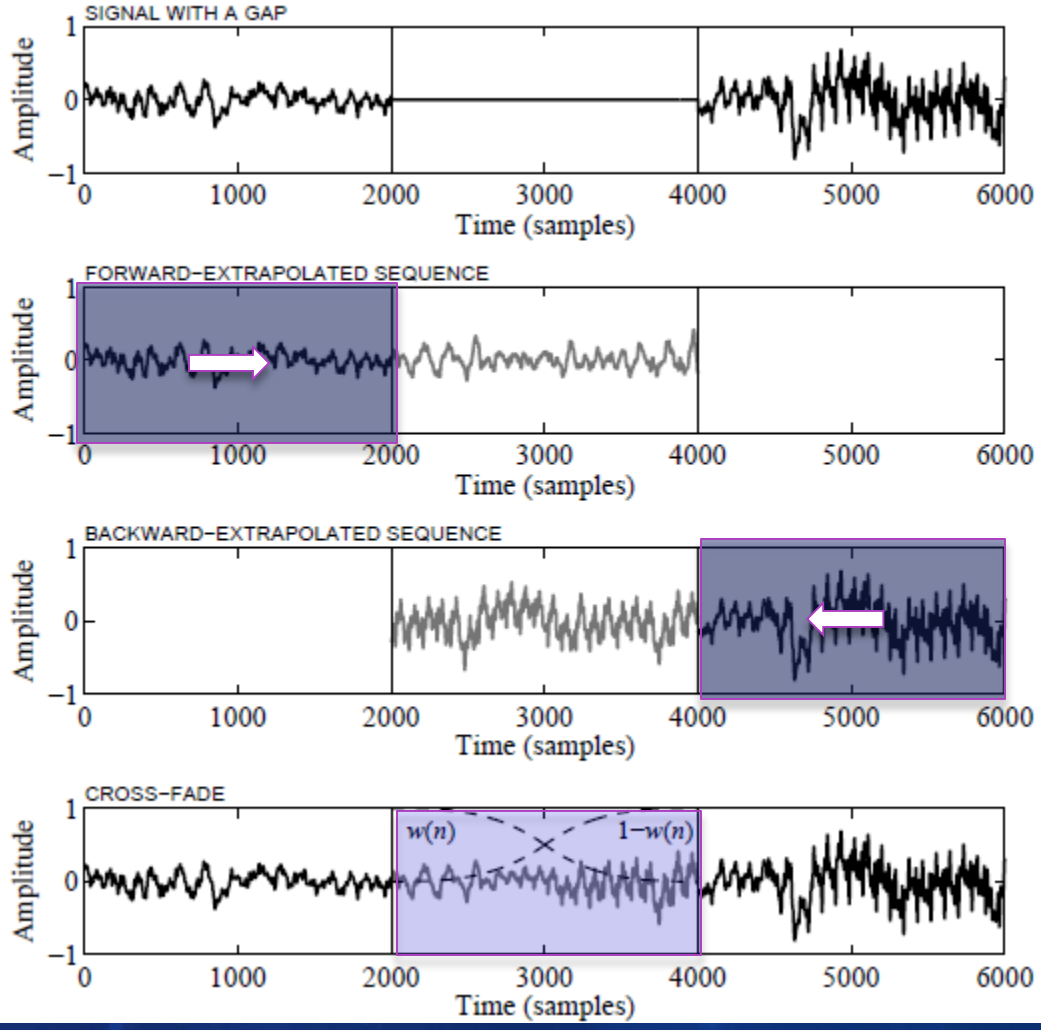
N = length of the data segment,

V = mean quadratic error of prediction

- Akaike criterion: the optimal model has min $AIC_k$

- Maximum Entropy Principle: guarantees that it is the best model that we can find with the information available.

# (MIARMA)


SIGNAL WITH A GAP
FORWARD–EXTRAPOLATED SEQUENCE
BACKWARD–EXTRAPOLATED SEQUENCE
CROSS–FADE
Amplitude
Time (samples)
w(n)
1−w(n)

# RESULTS

- High freqs.: HD49933. Solar-like

  P ~ min- hours

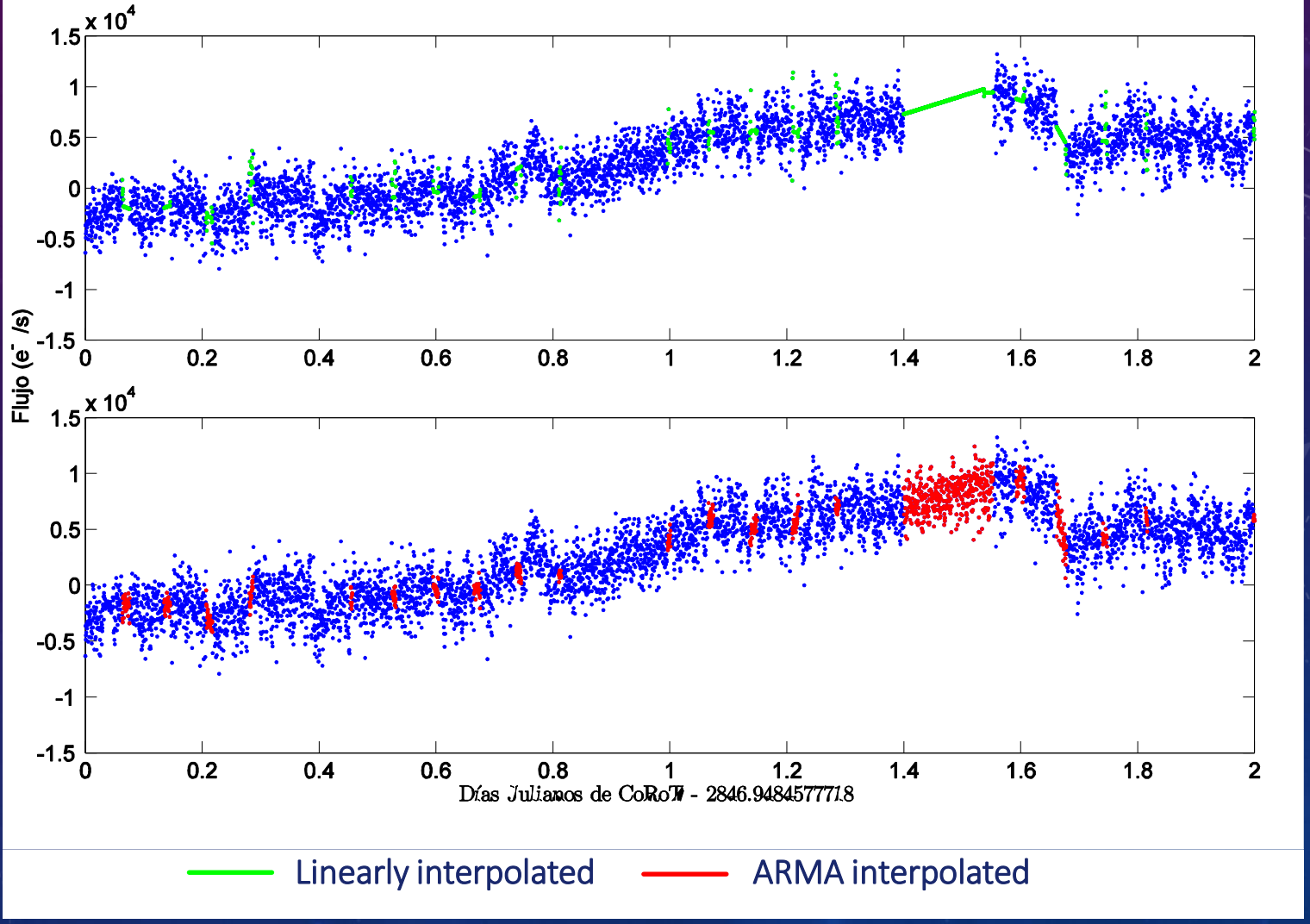
- Intermediate freqs.: HD174966. δ Scuti

  P ~ 0.3 – 3 hours

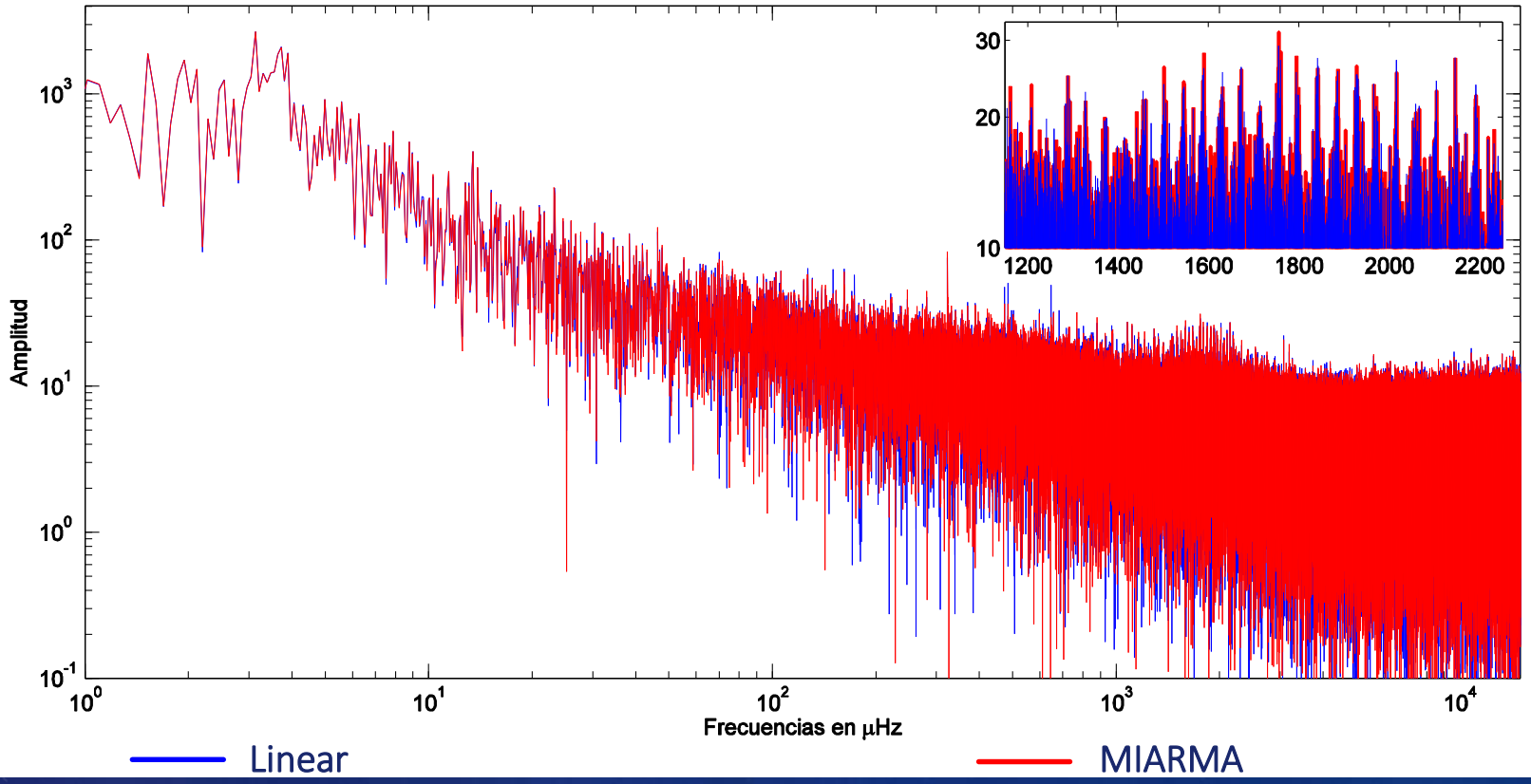
- Low freqs: HD51193. Be pulsating star

  P ~ days

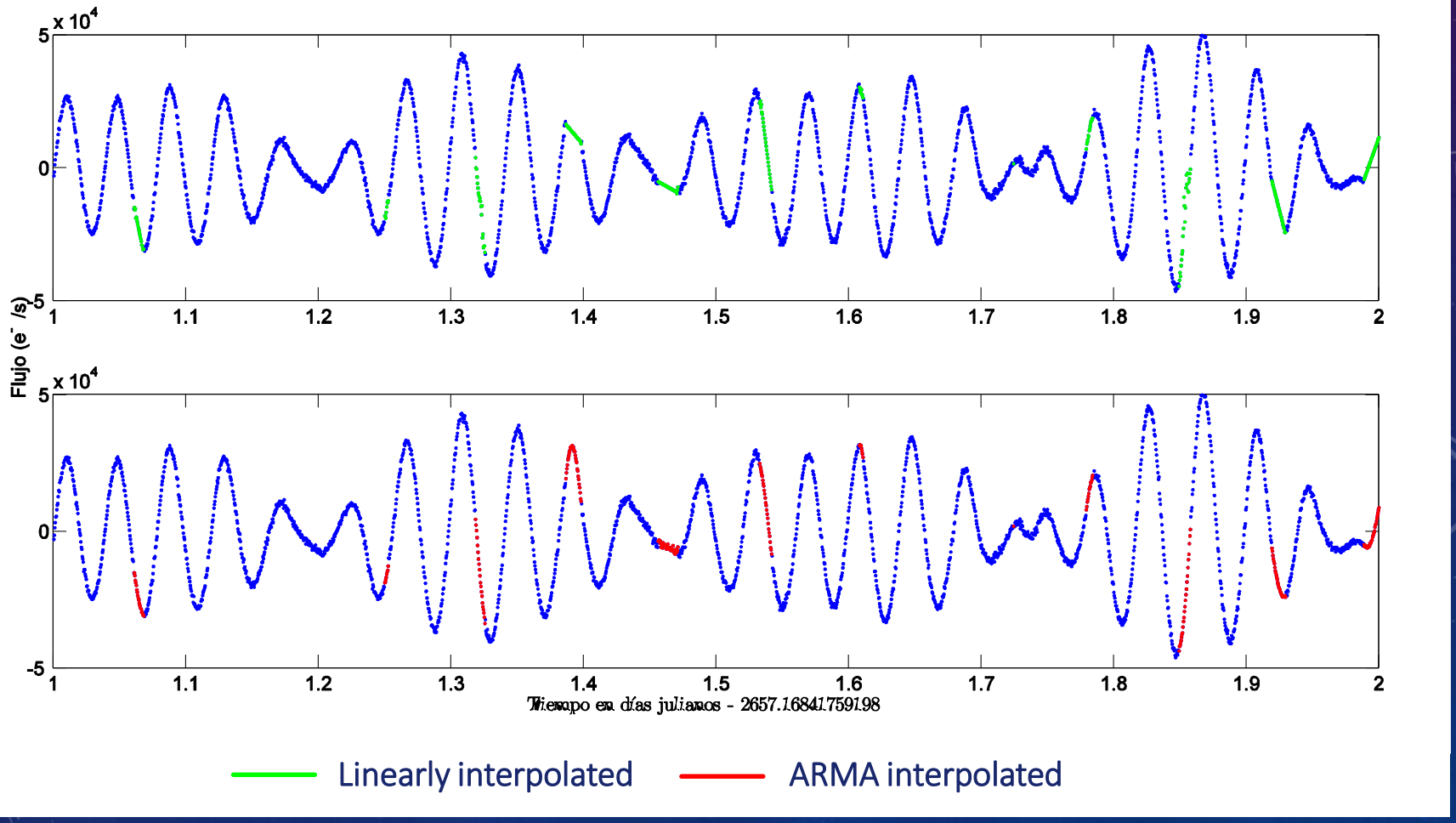
# RESULTADOS: HD 49933



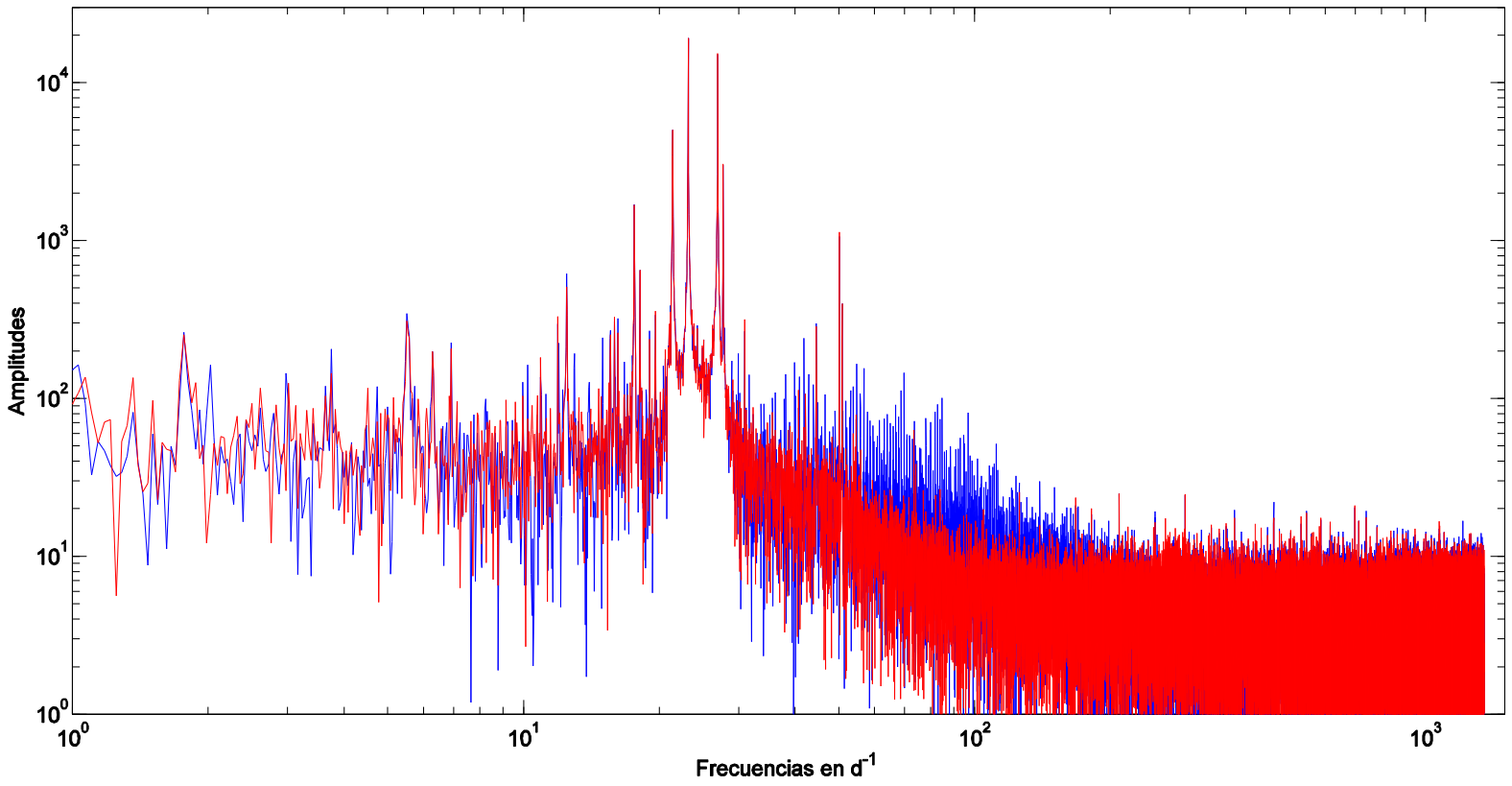
# RESULTS: HD 49933



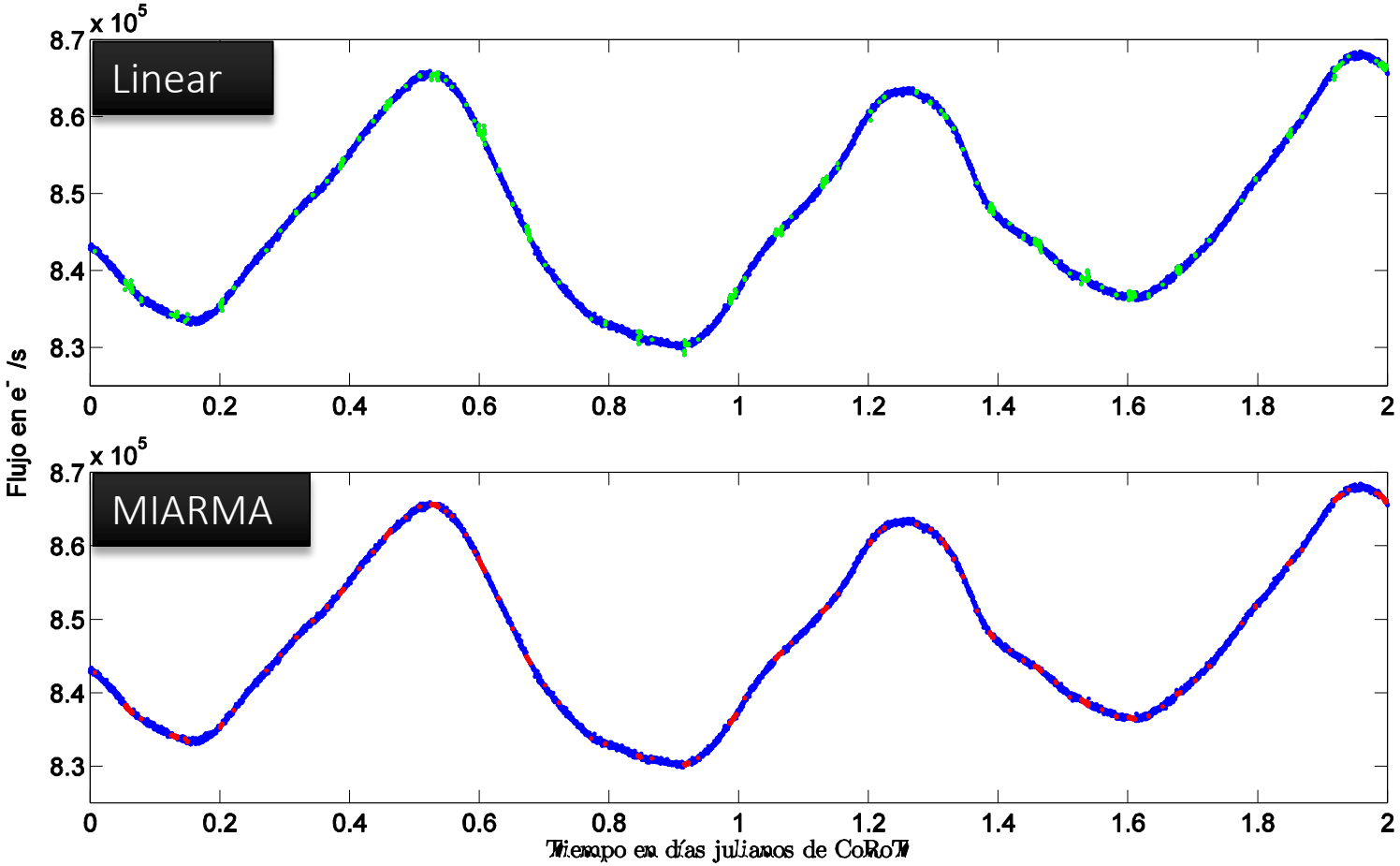
# RESULTS: HD 174966



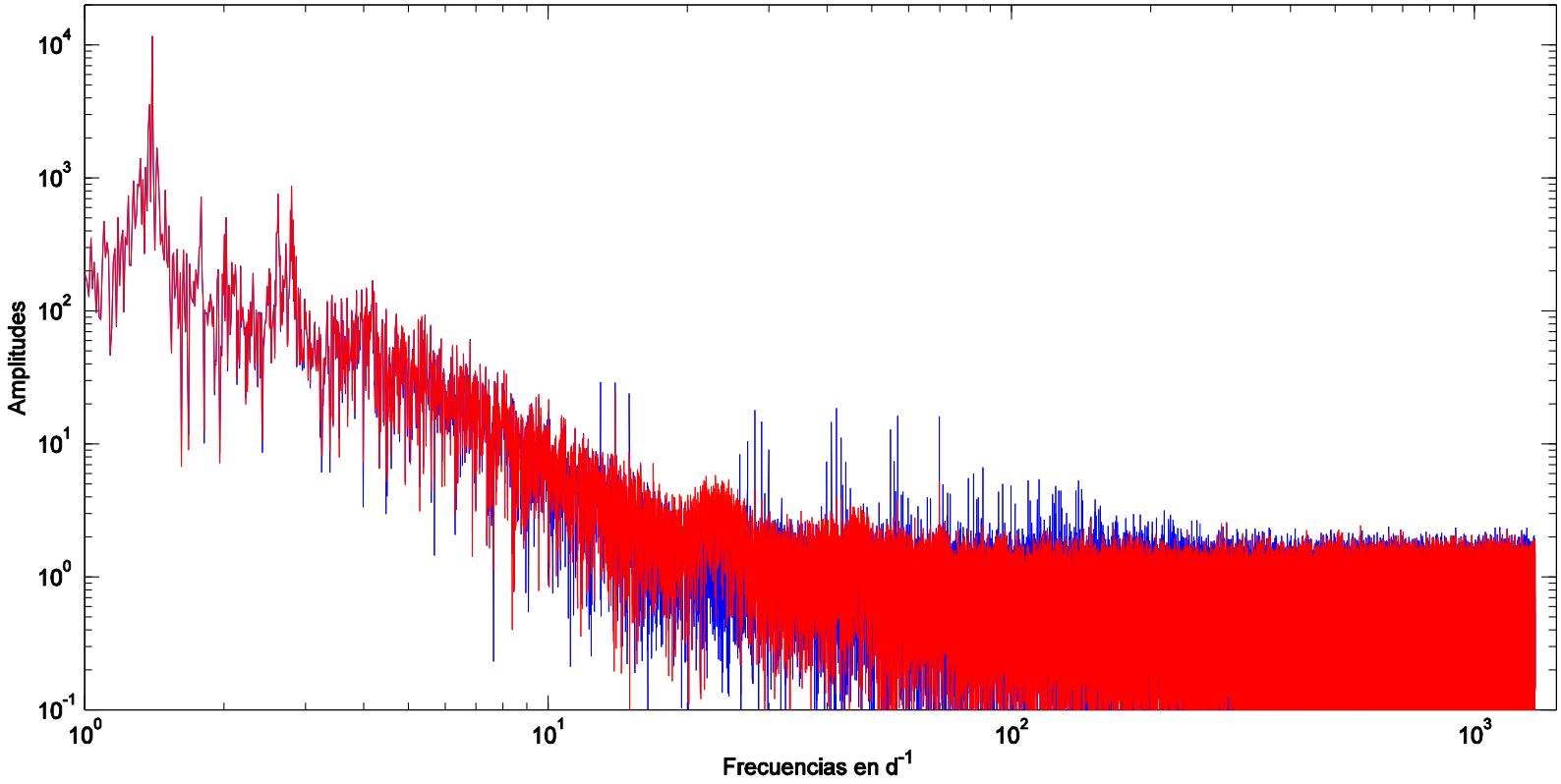
# RESULTS: HD 174966



# RESULTS: HD 51193

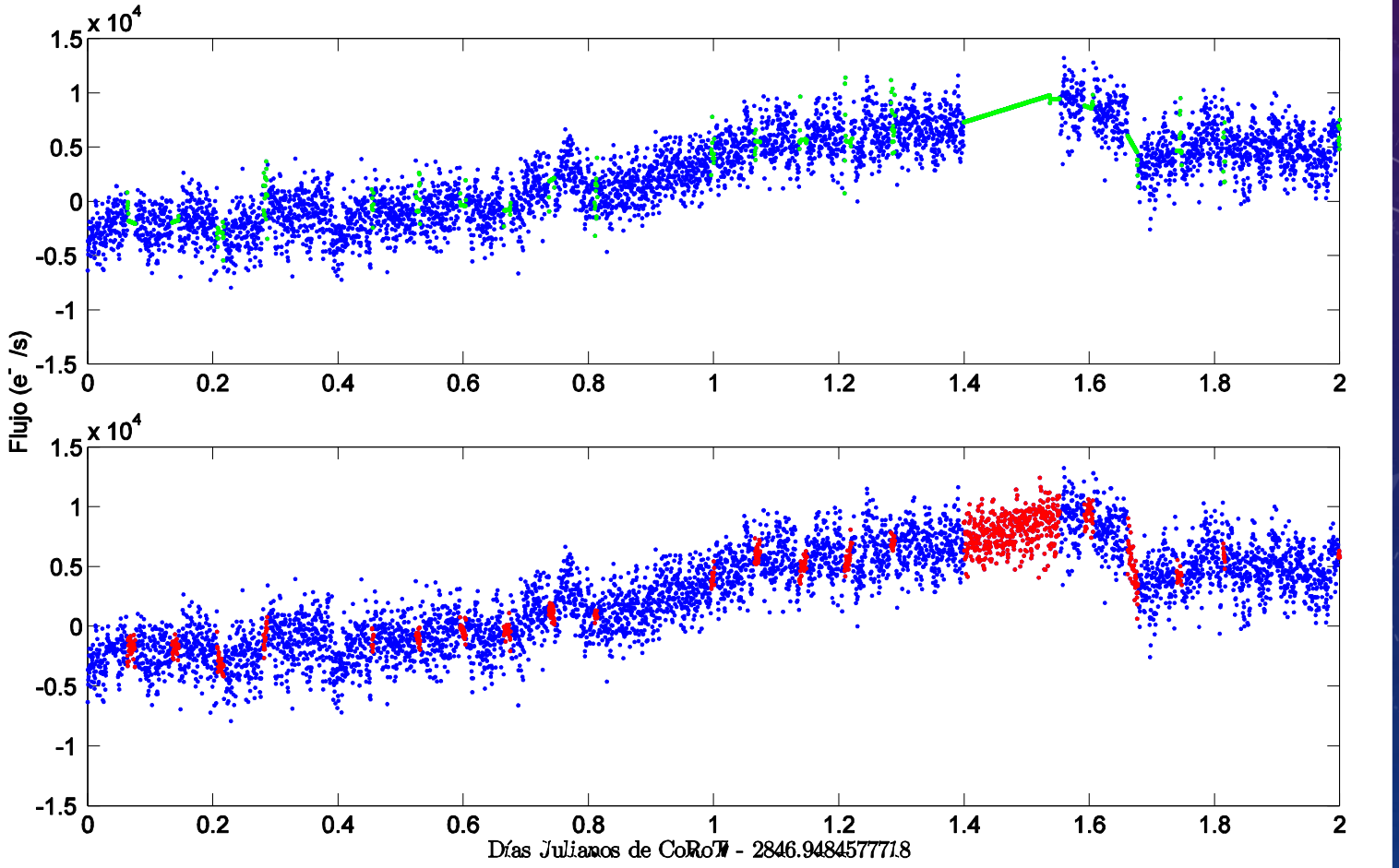


# RESULTS: HD 51193

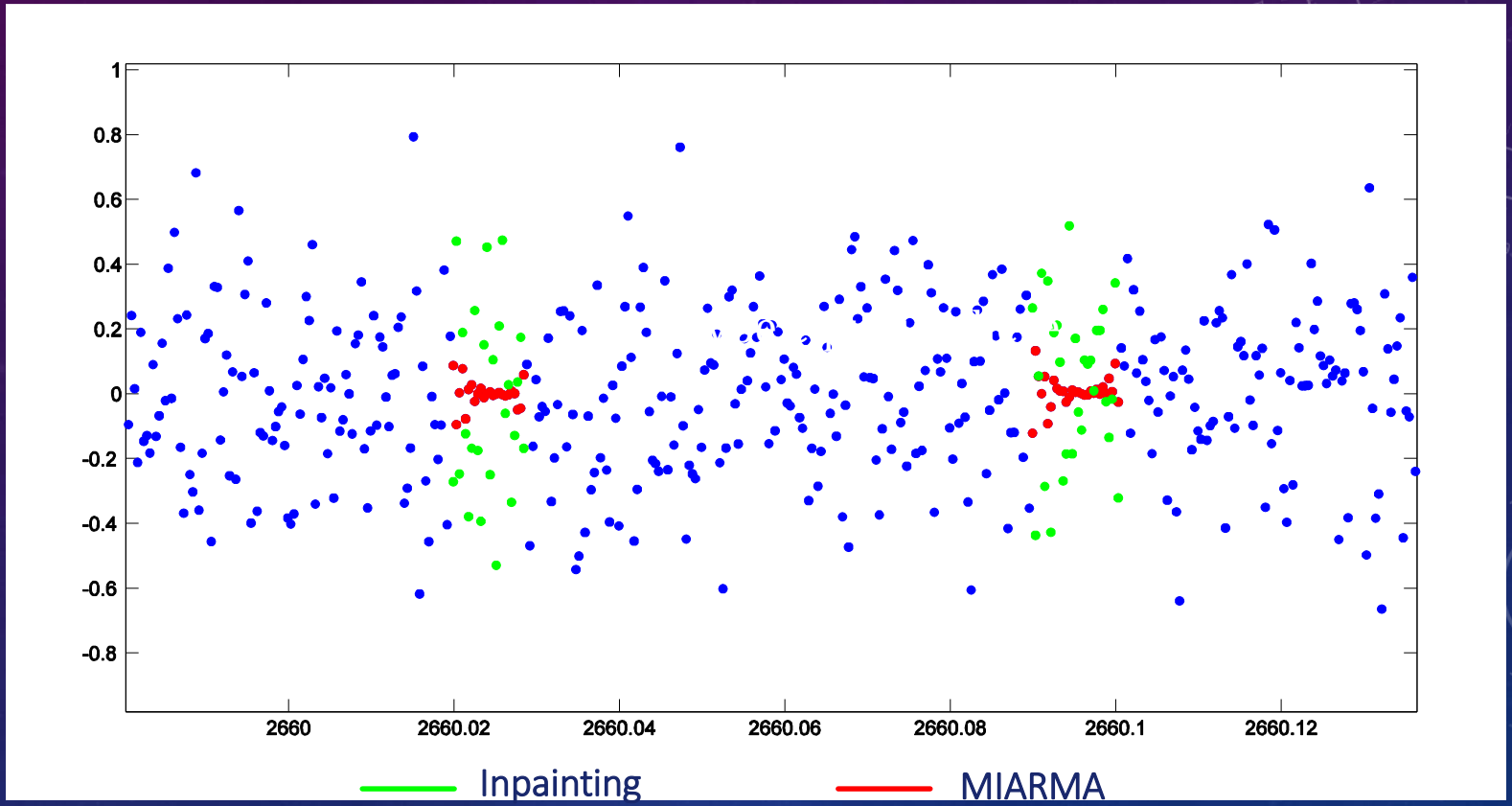



Linearly interpolated

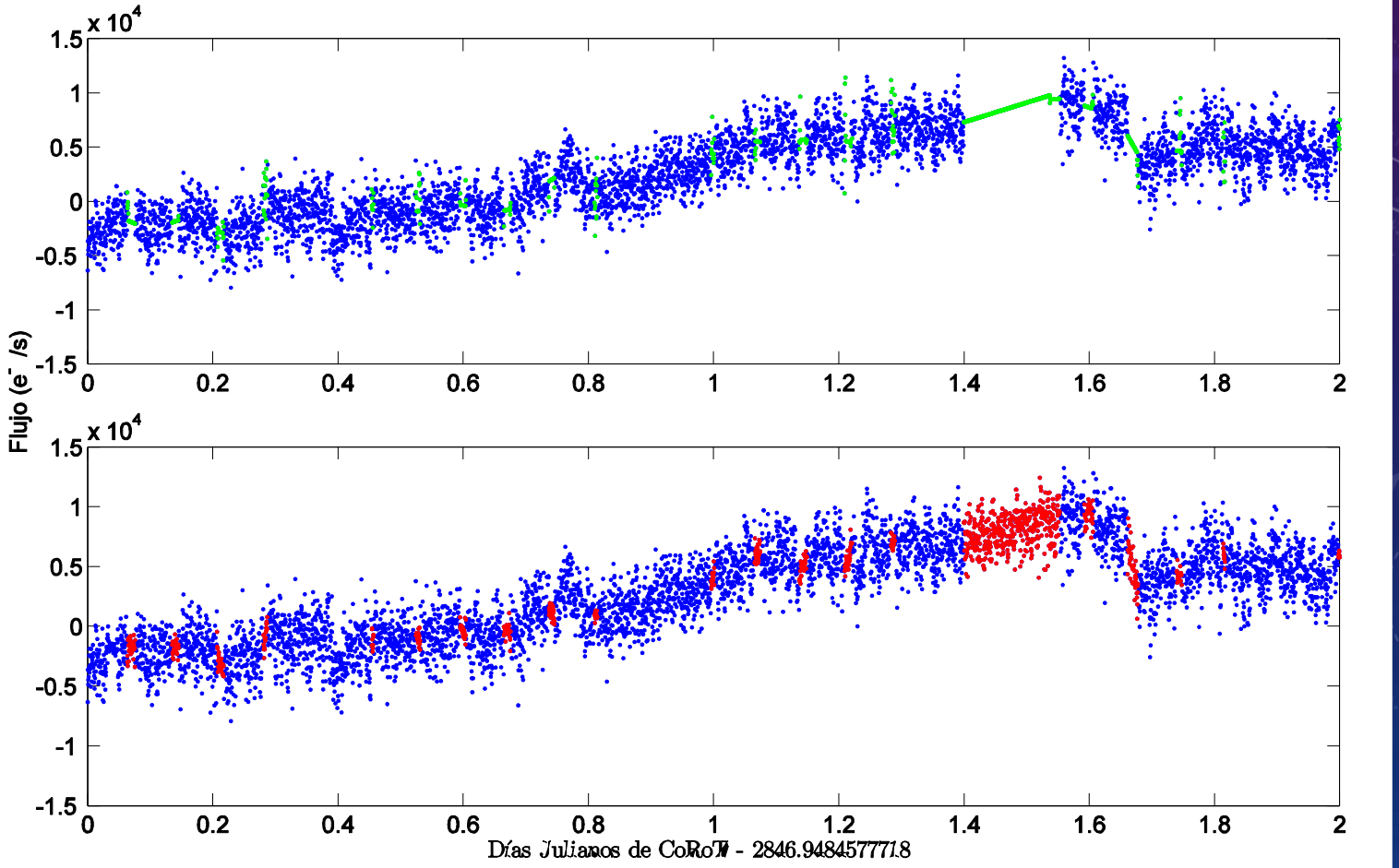
ARMA interpolated


x 10^4
1.5
1
0.5
0
-0.5
-1
-1.5
Flujo (e⁻ /s)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
x 10^4
Días Julianos de CoRoT - 2846.948457771.8
Linearly interpolated
ARMA interpolated

# WHITE NOISE




$\sigma_n = 0.2502$ $\sigma_{inpainting}$= 0.2730 $\sigma_{miarma} = 0.0753$


x 10^4
1.5
1
0.5
0
-0.5
-1
-1.5
Flujo (e⁻ /s)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
x 10^4
Días Julianos de CoRoT - 2846.948457771.8
Linearly interpolated
ARMA interpolated

Are the continuous functions underlying the light curves of pulsating stars analytic functions?

# CONCLUSIONS

- We have shown that linear interpolation is not reliable for asteroseismology whatever the case of study.

- On the other side, with a reliable gap-filling method there's no need to interpret an aliased periodogram. Prewhitening techniques (CLEAN) are neither necessary.

- We have introduced a gap-filling method based on ARMA models which is information preserving.

# CONCLUSIONS

- The method works for all kind of signals and has been tested in three different frequency ranges showing a reduction of the aliases in each case.
- Contrary to the expectations, the power spectrum of the Be star (low freq. pulsations) is strongly affected by the aliases. This could point to the non-analyticity of the underlying function of this light curve.
- The same could be happening in the case of the solar-like star HD49933 where a fine structure has been found of unknown origin.

# CONCLUSIONS

- If this hypothesis about the underlying function is confirmed, no gap-filling using a base of analytic functions would preserve the original information.

- If we want to solve interpretation problems in asteroseismology we have to use information preserving methods.