



UNIVERSITÀ  
DI TORINO

# DIPARTIMENTO DI MATEMATICA GIUSEPPE PEANO

Corso di Laurea Magistrale in Matematica

## *Time Series analysis of PRISMA all-sky camera data*

Candidato:

Alessandro Daghero

Relatrice UNITO:

Cristina Zucca

Correlatori INAF:

Donata Bonino, Dario Barghini, Daniele Gardiol

# Sommario

- Analisi temporale di serie storica
  - Ricerca di un trend pluriennale
    - lineare
    - completo di stagionalità
  - Analisi dei residui
- Analisi in frequenza di serie storica
  - Ricerca di frequenze significative
  - Simulazioni per robustezza livello di significatività

# Framework matematico: analisi di serie storiche

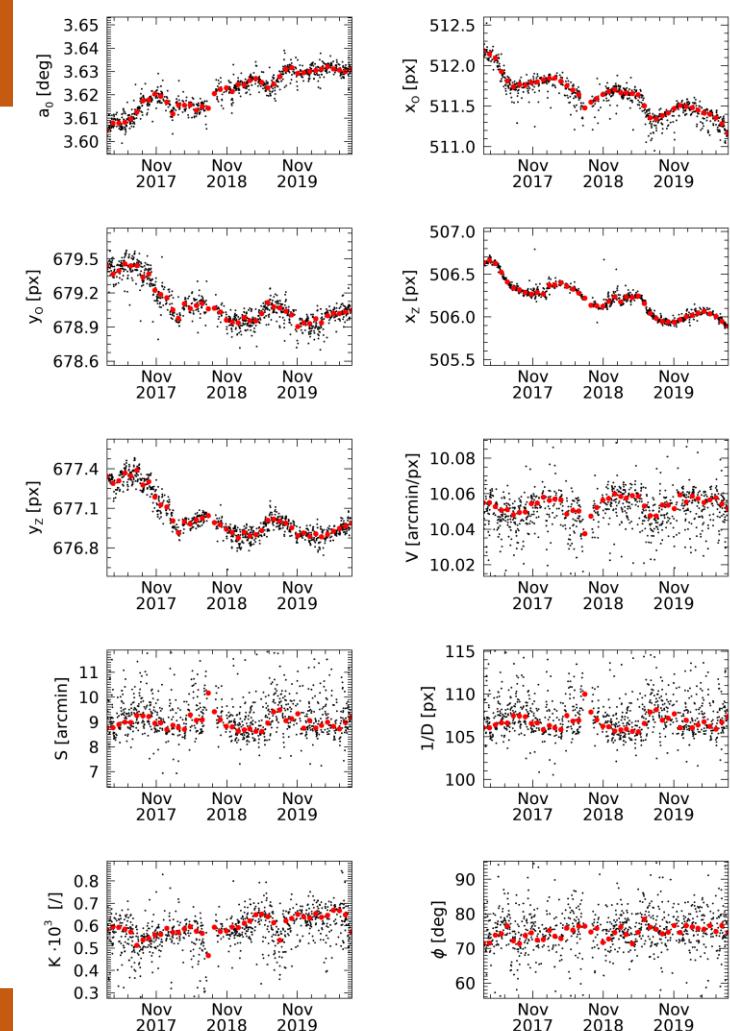
**Serie storica:** una sequenza ordinata di variabili aleatorie  $\{X_t\}$ ,  $t$  discreto o continuo.

- **Caso più semplice:** tutte le v.a.  $\{X_t\}$  seguono la stessa distribuzione di probabilità e sono indipendenti  
→ campione casuale
- **In generale:** la distribuzione statistica di  $\{X_t\}$  può cambiare per ogni  $t$ , e le  $X_t$  possono essere correlate. I momenti statistici possono essere stazionari (invarianti nel tempo) o non-stazionari.

# Parametri dalla calibrazione

I modelli astrometrici si basano su 10 parametri:

- $(x_0, y_0)$  e  $(x_z, y_z)$ : centro ottico e proiezione dello zenith, in pixel
- $a_0$ ,  $V$ ,  $S$ ,  $D$ : legati alla rotazione e alla distorsione dell'immagine
- $K$ ,  $\Phi$  : legati alla distorsione, secondo modello astrometrico



## Dataset: ITVA01



All-sky camera di Pino Torinese

- Dati provenienti dalla camera dell’Osservatorio Astronomico della Regione Autonoma della Valle d’Aosta, la seconda camera entrata in funzione il 16/03/2017
- Contiene dati raccolti dal 16/03/2017 al 30/11/2021, per un totale di 1720 giorni di osservazione.
- Presenta 181 buchi temporali, di cui uno di 45 giorni e uno di 26 giorni.
- Il numero effettivo di giorni è 1397.
- Per ogni giorno osservato, sono registrati:
  - il numero di stelle osservate.
  - 12 parametri con rispettivi errori, utili per la calibrazione e il calcolo delle orbite di eventuali meteore.

## Data cleaning

Per garantire dati affidabili e coerenti, escludiamo:

- Le osservazioni con meno di 3000 stelle identificate,
- Nei casi in cui due osservazioni abbiano la stessa data giuliana, quella con meno stelle osservate. Questo accade perché il dataset è organizzato per mesi e date giuliane, e l'ultimo giorno di un mese può coincidere con il primo del mese successivo.

Il dataset finale presenta:

- 1019 giorni di osservazioni.
- 700 giorni mancanti, distribuiti su 269 buchi temporali.

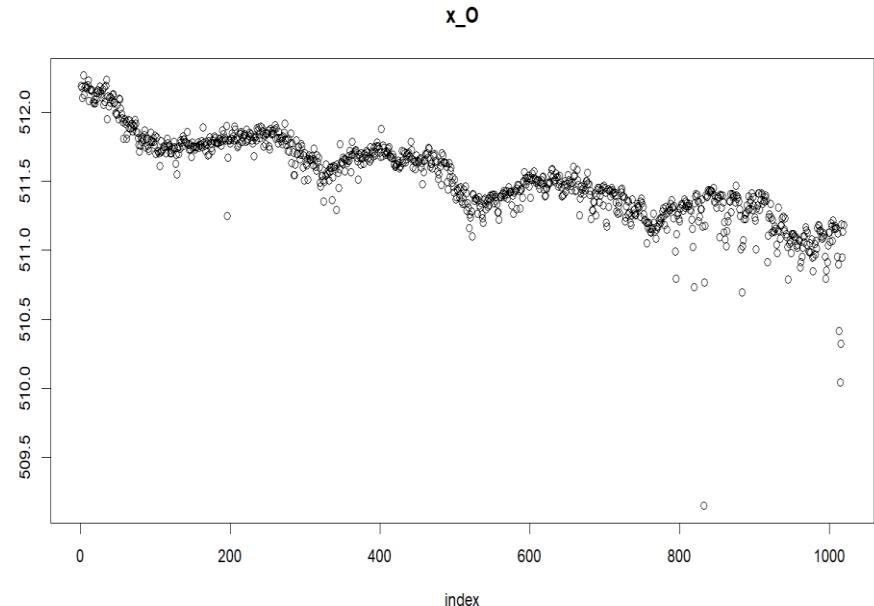
# Analisi temporale

- Analisi temporale

- Ricerca di un trend pluriennale
  - lineare
  - completo di stagionalità
- Analisi dei residui

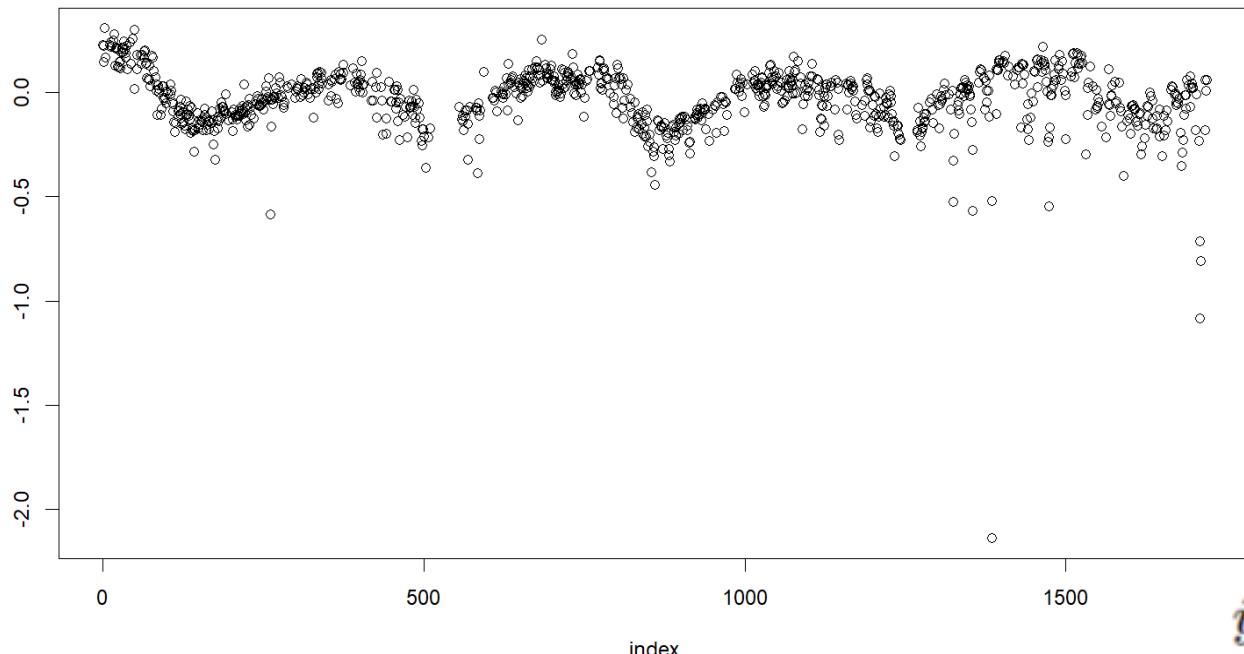
- Analisi in frequenza

- Ricerca di frequenze significative
- Simulazioni per robustezza livello di significatività



# Trend lineare

Linear detrended  $x_O$



Weighted Least Square:

$$\min \sum_{t=1}^{1019} w_t [x_O(t) - (a + b \cdot t)]^2$$

$$w_t = \frac{1}{s_{x_O}^2}$$

$$\hat{y} = 512 - 0.000486 * t$$

R<sup>2</sup>: 0.8315

adj R<sup>2</sup>: 0.8314

# Modellizzazione con spline

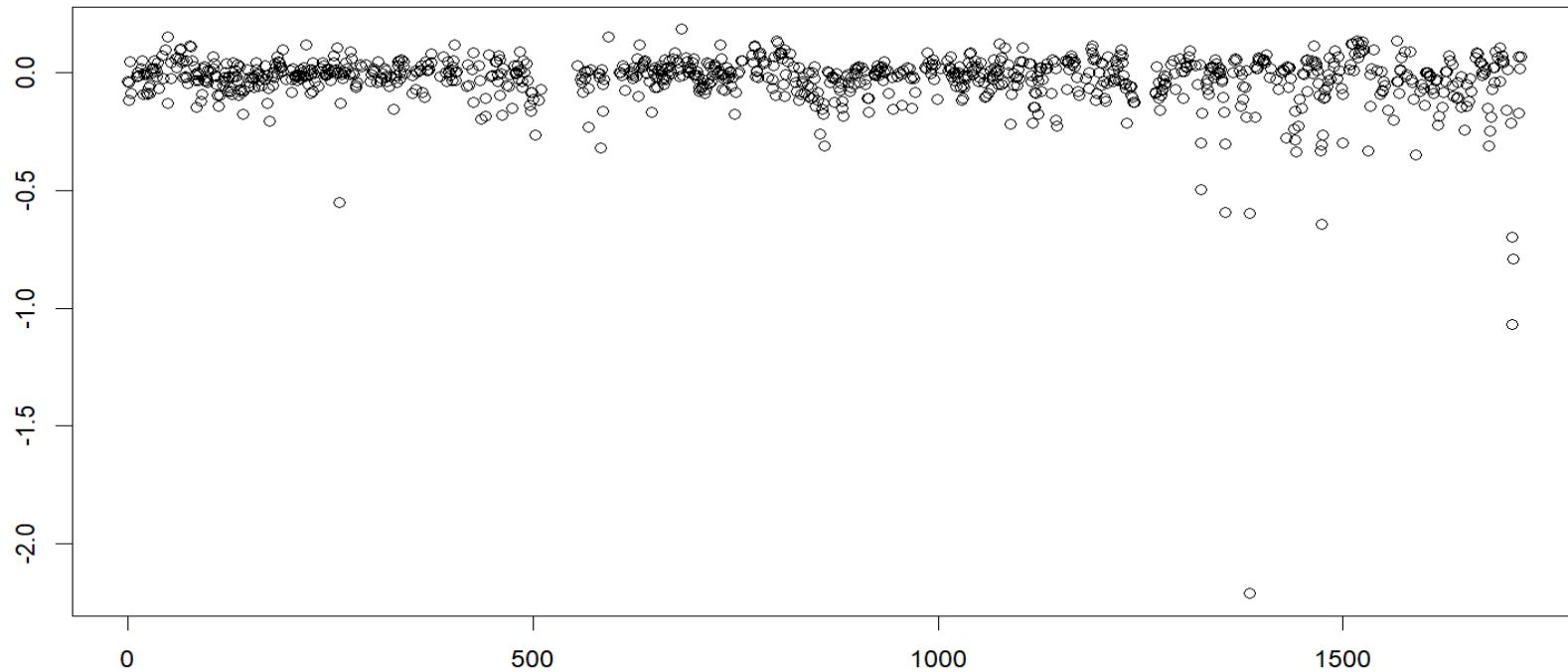
**Spline:** funzione polinomiale a tratti

Data una serie temporale  $\{X_t\}_{t \in T}$  su  $[0, N]$ , il dominio è diviso in  $k$  intervalli  $[t_i, t_{i+1})$  con  $0 = t_0 < t_1 < t_2 < \dots < t_k = N$ . In ciascun intervallo si adatta un polinomio  $P_i$  di grado massimo  $n+1$ .

Per garantire la regolarità, imponiamo la continuità fino alla derivata  $n$ -esima nei nodi  $t_i$ . La funzione risultante è una spline in  $C^n$ , continua e connessa.

# Modellizzazione con spline di $x_o$

$x_O$  senza spline df=20



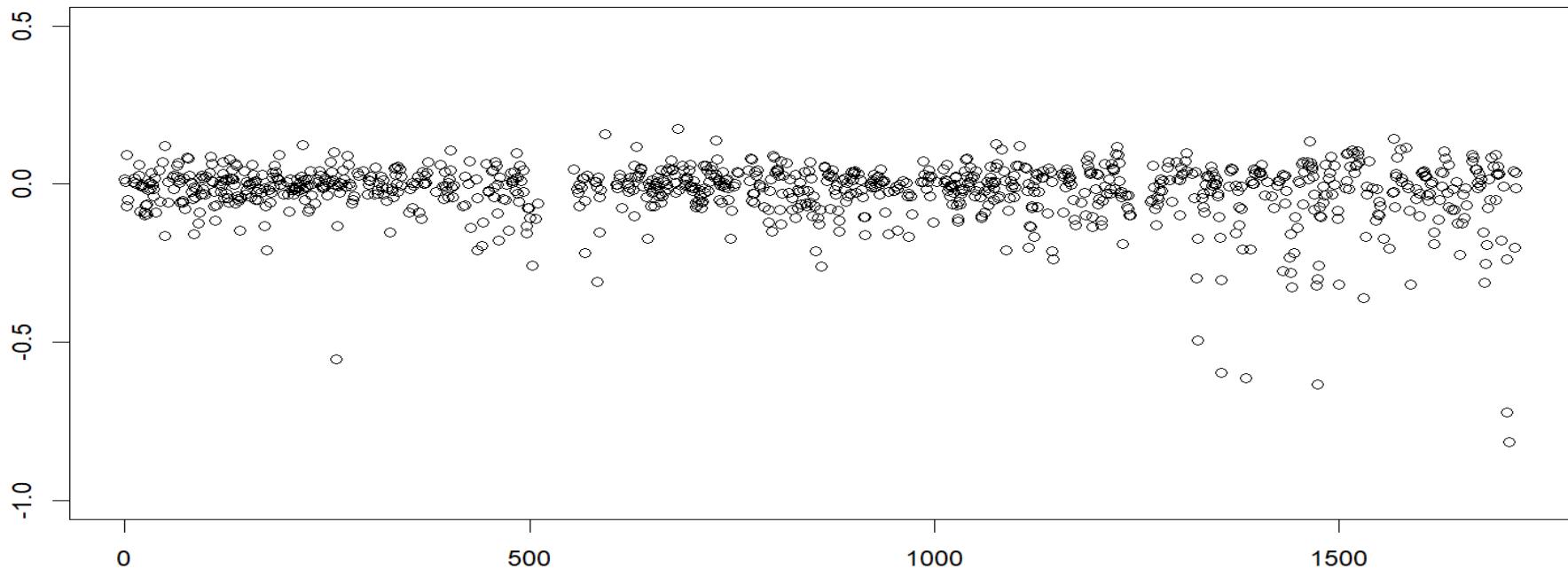
# Modellizzazione con LOESS

**LOcally Estimated Scatterplot Smoothing (LOESS):** regressione robusta locale

1. *Definizione della finestra locale:* per ogni punto  $t_i$ , si considera una finestra di centro  $t_i$  e ampiezza una frazione  $\alpha$  dei dati
2. *Assegnazione dei pesi:* i punti nella finestra ricevono pesi in base alla loro distanza da  $t_i$ , maggiore è la distanza, minore è il peso
3. *Regressione locale:* Weighted Least Square con i pesi identificati
4. Calcolo di  $t_{LOESS,i}$ : sostituzione di  $t_i$  con il corrispondente punto nella retta WLS

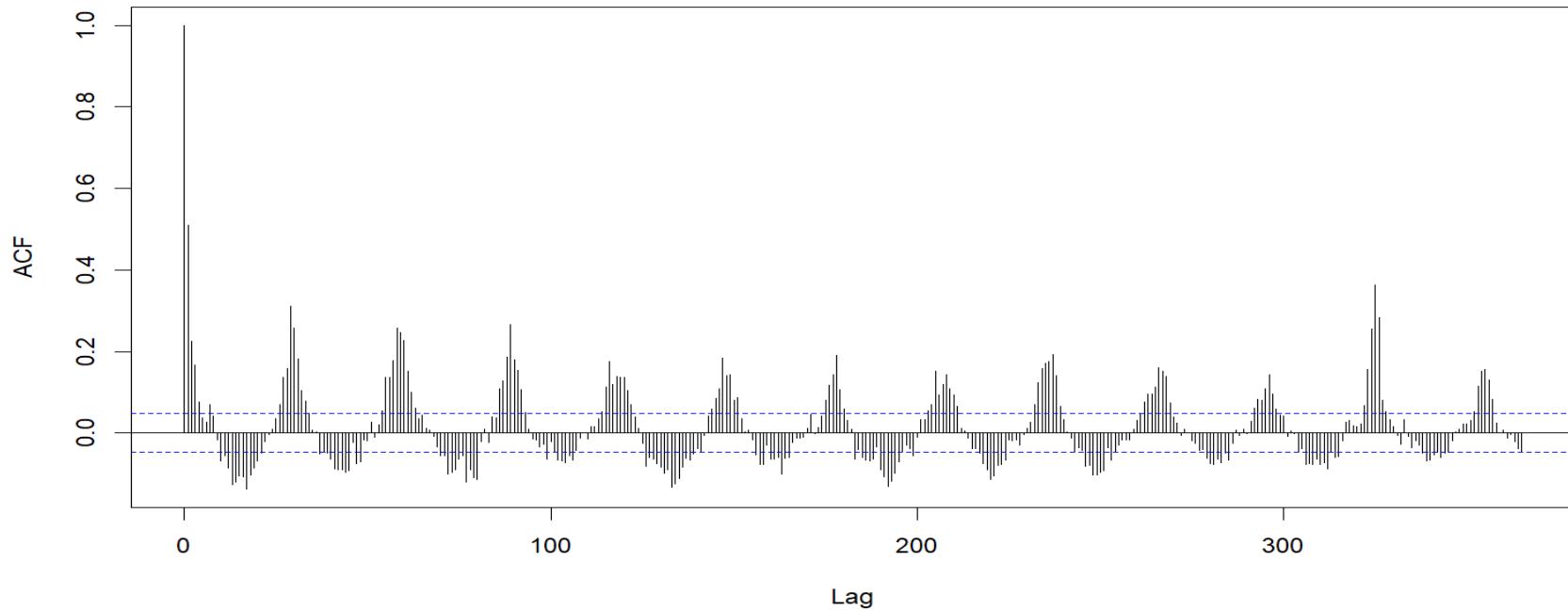
# Modellizzazione di $x_o$ con LOESS

$x_O$  de-LOESS



# Analisi dei residui

ACF x\_O de-LOESS span=10%



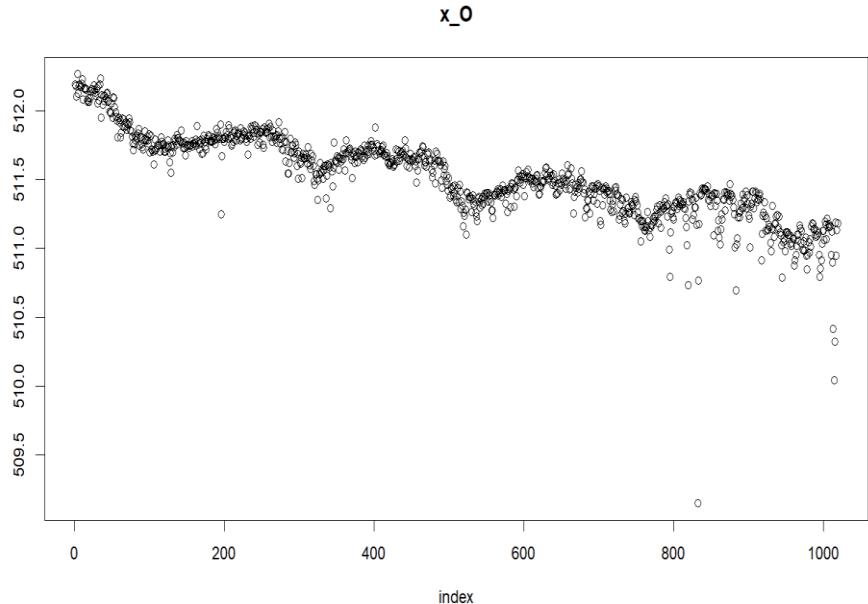
# Analisi in frequenza

- Analisi temporale

- Ricerca di un trend pluriennale
  - lineare
  - completo di stagionalità
- Analisi dei residui

- Analisi in frequenza

- Ricerca di frequenze significative
- Simulazioni per robustezza livello di significatività



# Periodogramma

## Periodogramma classico:

Data una serie storica  $\{X_t\}_{t \in T}$ , definiamo il periodogramma classico attraverso la trasformata di Fourier discreta, definita per frequenze  $\omega$ :

$$P_X(\omega) = \frac{1}{N} \left| \sum_{t=1}^N X_t e^{-i\omega t} \right|^2$$

Per serie non equispaziate, la funzione che descrive il campionamento temporale non è più una successione di  $\delta$  di Dirac

$$\mathcal{F}(\delta \times X) = \mathcal{F}(\delta)\mathcal{F}(X)$$

## Periodogramma di Lomb-Scargle

Ridefiniamo il periodogramma seguendo il ragionamento di Scargle.

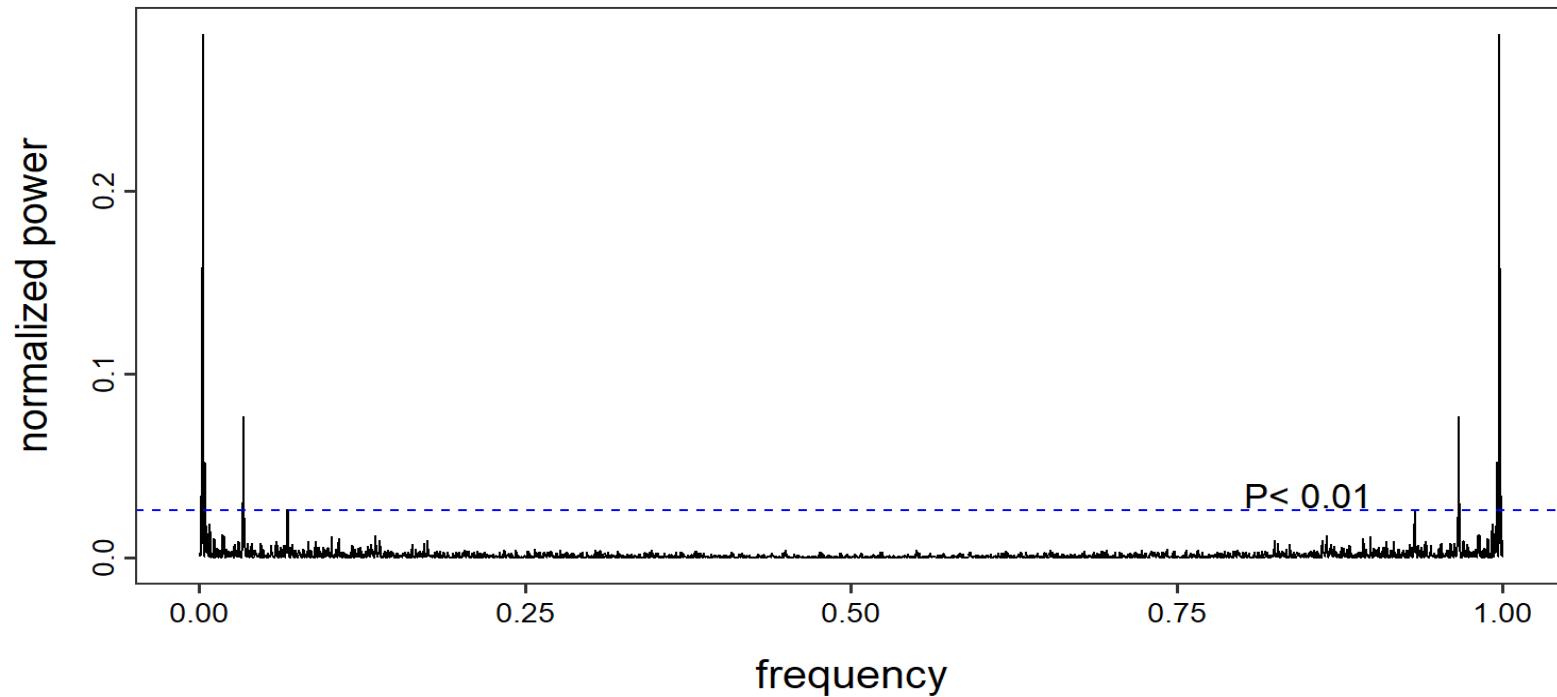
$$P_X(\omega) = \frac{1}{N} \left| \sum_{t=1}^N X_t e^{-i\omega t} \right|^2 = \frac{1}{N} \left[ \left( \sum_t X_t \cos(2\pi\omega t) \right)^2 + \left( \sum_t X_t \sin(2\pi\omega t) \right)^2 \right]$$

$$P_X(\omega) = \frac{A^2}{2} \left( \sum_t X_t \cos(2\pi\omega(t - \tau)) \right)^2 + \frac{B^2}{2} \left( \sum_t X_t \sin(2\pi\omega(t - \tau)) \right)^2$$

dove  $A$  e  $B$  sono costanti e  $\tau$  è una funzione che dipende solo dalla frequenza.

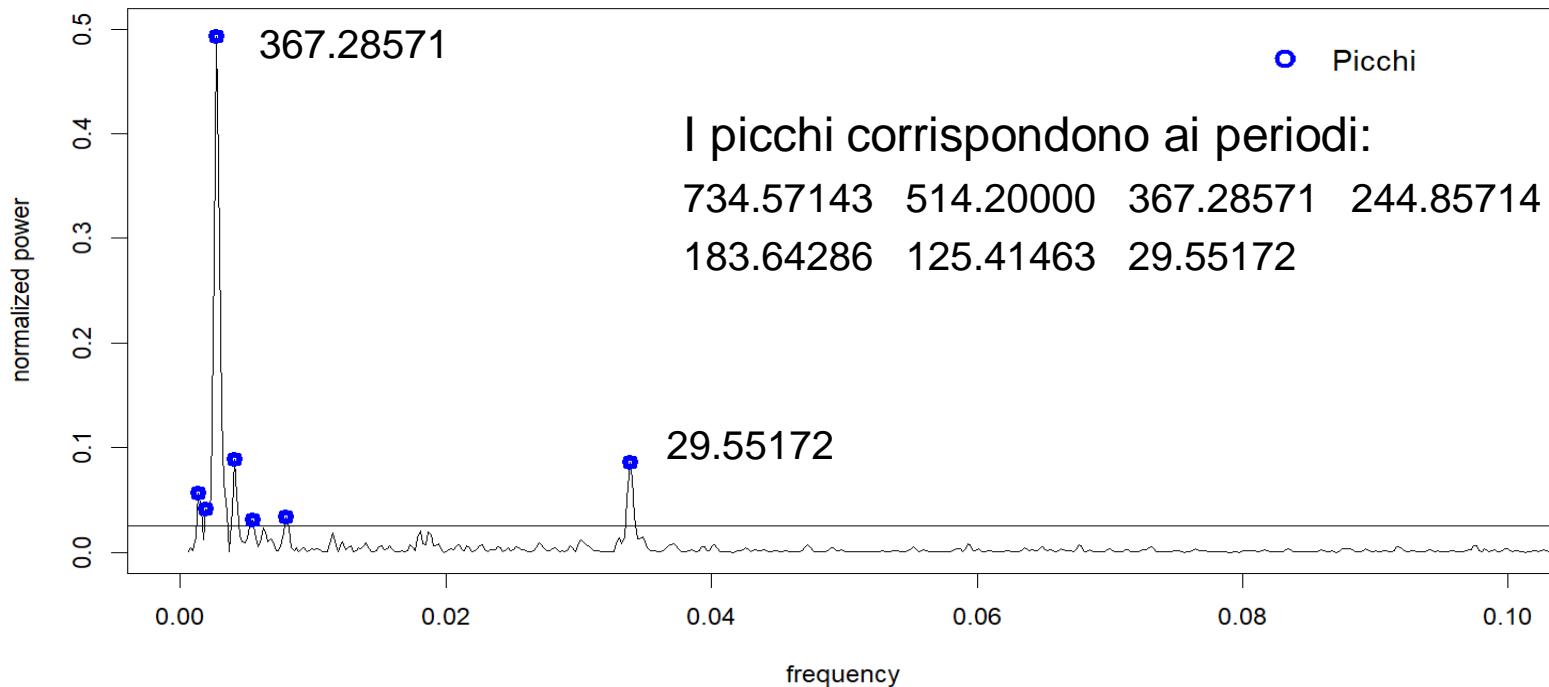
# Periodogramma di Lomb-Scargle di $x_o$

Lomb-Scargle Periodogram



# Periodogramma di Lomb-Scargle di $x_o$

LSP  $x_O$  rolling mean p=3



## Modellizzazione con signale periodico

Individuate due frequenze significative andiamo a approssimare i dati attraverso una regressione non-lineare ai minimi quadrati della funzione:

$$A_1 \cdot \sin(2 \cdot \pi \cdot freq_1 \cdot time + \phi_1) + A_2 \cdot \sin(2 \cdot \pi \cdot freq_2 \cdot time + \phi_2)$$

con parametri iniziali:

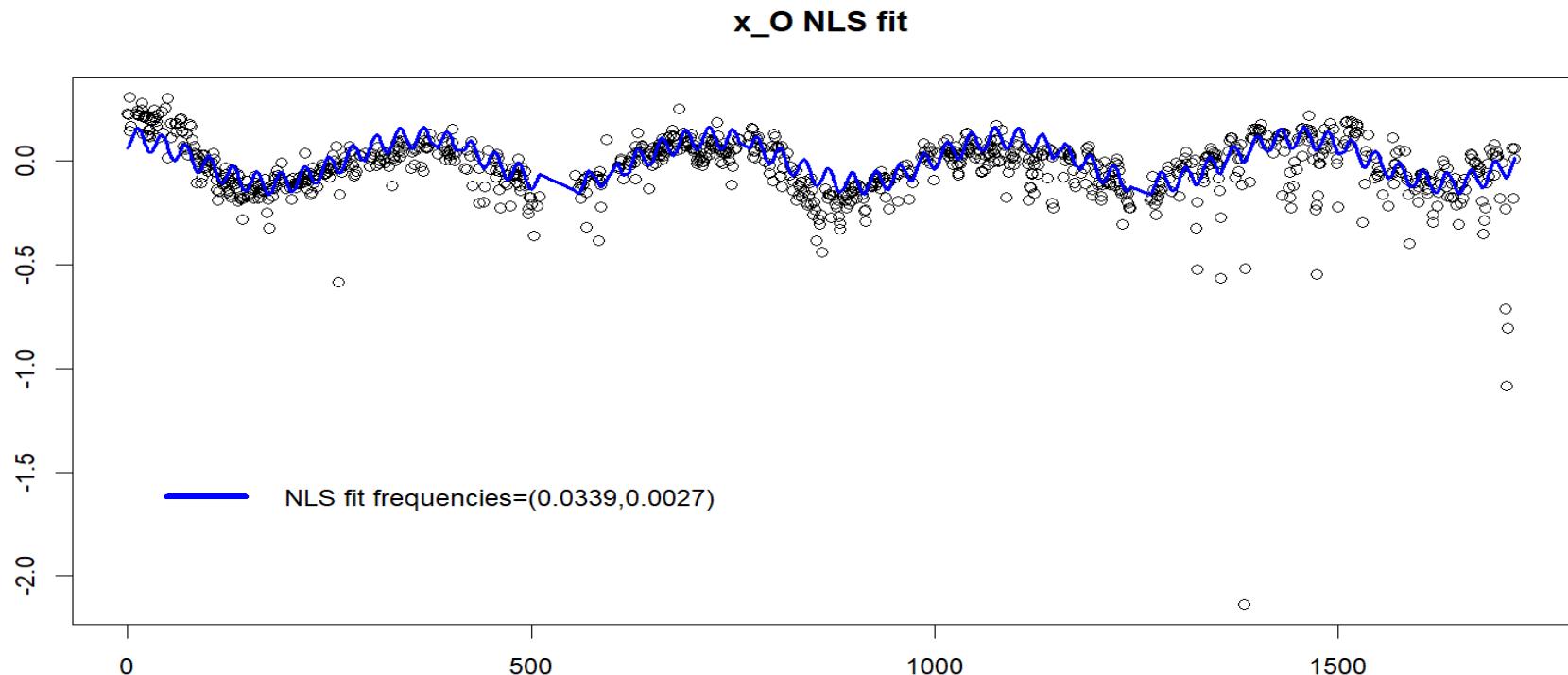
Parameter	Value
$A_1$	0.25
$freq_1$	0.0339
$\phi_1$	0
$A_2$	0.25
$freq_2$	0.0027
$\phi_2$	0

I parametri stimati sono:

Parameters:

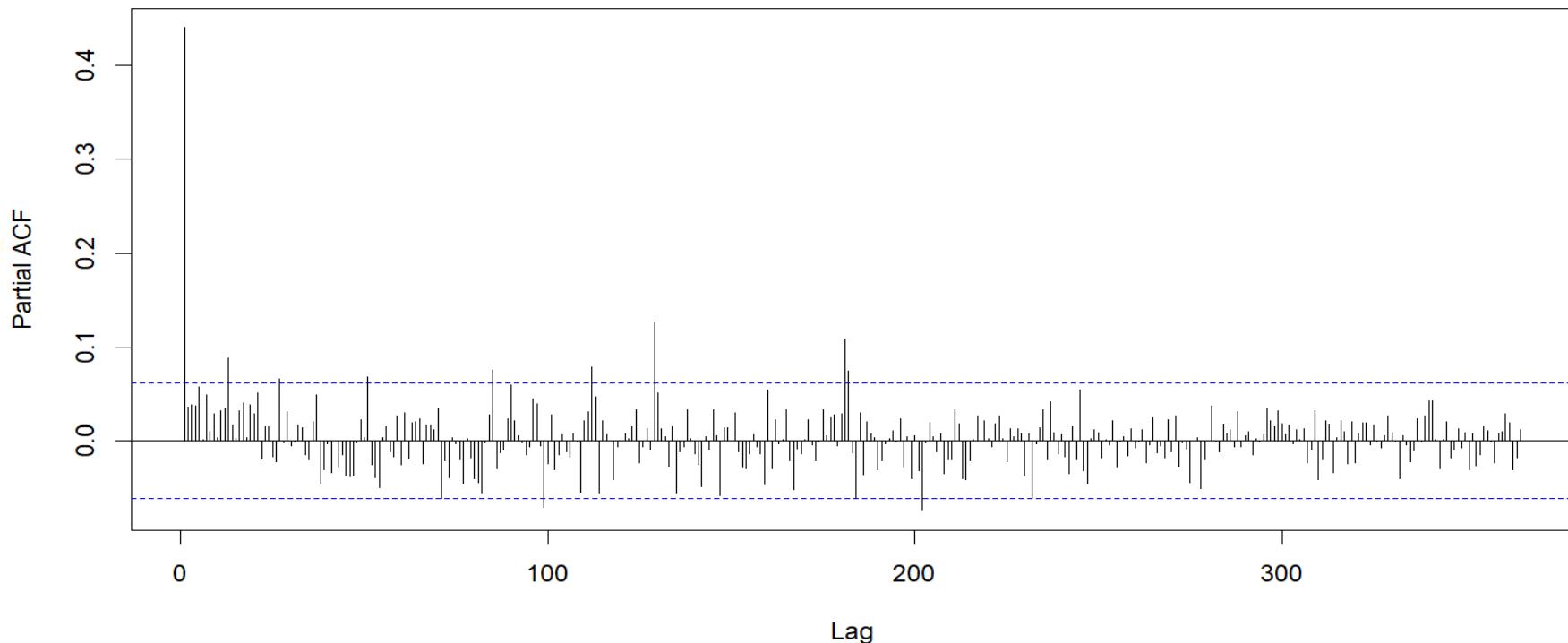
	Estimate	Std. Error	t value	Pr(> t )	
A1	5.288e-02	5.433e-03	9.734	< 2e-16	***
freq1	3.390e-02	3.159e-05	1073.179	< 2e-16	***
phi1	-1.172e+00	1.947e-01	-6.017	2.48e-09	***
A2	-1.104e-01	5.396e-03	-20.458	< 2e-16	***
freq2	2.743e-03	1.545e-05	177.541	< 2e-16	***
phi2	-1.459e+00	9.367e-02	-15.577	< 2e-16	***

# Modellizzazione di $x_o$ con segnale periodico



# Analisi dei residui

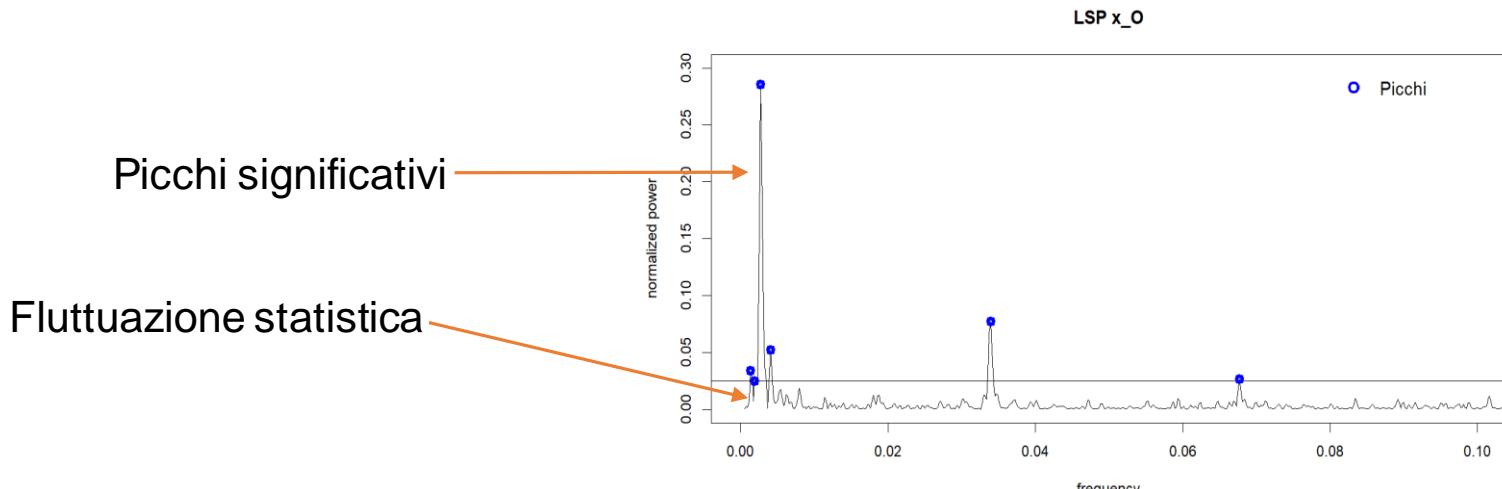
PACF x\_O de-NLS



# Analisi del livello di significatività

Il periodogramma Lomb-Scargle è stato stimato usando la funzione `lsp` di R  
(<https://cran.r-project.org/web/packages/lomb/index.html>)

Scelto un livello di significatività, `lsp` stima il livello di potenza corrispondente



Il metodo è dimostrato per serie storiche con rumore gaussiano  
(Baluev, MNRAS 385, 2008)

# Simulazioni

Simulazioni con:

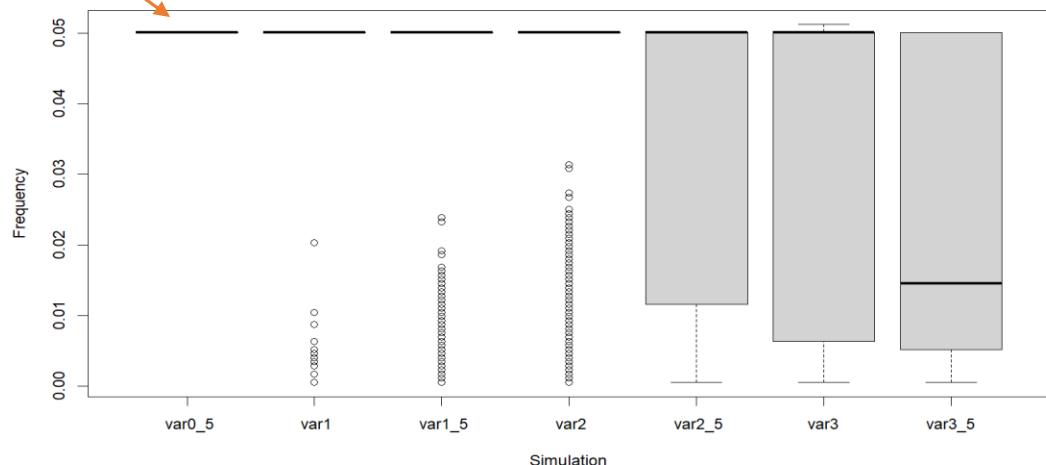
- segnale periodico, 1 frequenza nota, equispaziato con:
  - rumore bianco gaussiano
  - rumore AR(1)
  - rumore AR(2)e diversi SNR
- come sopra, ma con il campionamento temporale di ITVA01
- segnale periodico, 2 frequenze note, equispaziato con:
  - rumore bianco gaussiano

# Simulazioni

Sinusoide con una frequenza, *campionamento come ITVA01*  
Rumore AR(1)

amplitude = 1	$\sigma^2 = 0.5$	$\sigma^2 = 1$	$\sigma^2 = 1.5$	$\sigma^2 = 2$	$\sigma^2 = 2.5$	$\sigma^2 = 3$	$\sigma^2 = 3.5$
SNR	0.8485	0.6	0.4898	0.4242	0.3794	0.3464	0.3207
True frequency correctly detected	100%	100%	100%	99.87%	99.29%	97.65%	94.82%

Boxplot of frequencies with highest power, amplitude signal=1



GRAZIE PER  
L'ATTENZIONE!