Analisi spettrale delle serie temporali

Traduzione in italiano, redatta ed adattata da Federico Lombardo, dal testo in lingua greca tratto da:

Koutsoyiannis, D., *Lecture notes on Stochastic Methods in Water Resources*, Edition 3, 100 pages, National Technical University of Athens, Athens, 2007.

1. La trasformata di Fourier

L'identità di Eulero: $e^{i\theta} = \cos\theta + i\ sen\theta$ (i unità immaginaria)

La trasformata di Fourier: $F(\xi) = \int_{-\infty}^{\infty} f(x)\ e^{-i\ 2\ \pi\ \xi\ x} dx$

La trasformata inversa di Fourier: $f(x) = \int_{-\infty}^{\infty} F(\xi)\ e^{i\ 2\ \pi\ \xi\ x} d\xi$

La variabile x è solitamente il tempo [s] mentre ξ rappresenta la frequenza [Hz]. Se la funzione $f(x)$ è reale (codominio reale: $i = 0$) ed è pari ($f(x) = f(-x)$), allora esiste una forma semplificata per la trasformata $F(\xi)$:

$$F(\xi) = \int_{-\infty}^{\infty} f(x)\ \cos(2\ \pi\ \xi\ x)\ dx = 2\int_{0}^{\infty} f(x)\ \cos(2\ \pi\ \xi\ x)\ dx$$

Allo stesso modo, la trasformazione inversa è semplificata come segue:

$$f(x) = \int_{-\infty}^{\infty} F(\xi)\ \cos(2\ \pi\ \xi\ x)\ d\xi = 2\int_{0}^{\infty} F(\xi)\ \cos(2\ \pi\ \xi\ x)\ d\xi$$

Nota: in letteratura sono spesso utilizzate anche le seguenti formule per definire la trasformata e l'antitrasformata di Fourier, scritte in termini della frequenza angolare $\omega = 2\ \pi\ \xi$ [rad/s]:

$$F(\omega) = \int_{-\infty}^{\infty} f(x)\ e^{-i\ \omega\ x} dx\ , \quad f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\ e^{i\ \omega\ x} d\omega$$

2. La trasformata finita di Fourier

Se la funzione $f(x)$ è periodica di periodo 1, allora la trasformata di Fourier è nulla per qualsiasi valore non intero di ξ. Questo dà luogo alla definizione di un caso particolare della trasformata di Fourier tale per cui la variabile x varia in un intervallo ripetuto periodicamente (come ad es.: *[–1/2, 1/2]*), mentre la variabile ξ assume i valori interi k. Quindi la funzione $f(x)$ è definita come la trasformazione inversa della trasformata finita di Fourier F_k (invece di $F(\xi)$) ed data da una sommatoria, invece che da un integrale:

$$F_k = \int_{-1/2}^{1/2} f(x)\ e^{-i\ 2\ \pi\ k\ x} dx\ , \quad f(x) = \sum_{k=-\infty}^{\infty} F_k\ e^{i\ 2\ \pi\ k\ x}$$

Se la funzione *f(x)* è reale (codominio reale: *i = 0*) ed è pari (*f (x) = f (–x)*), allora sarà egualmente reale e pari anche la F_k e potrà essere espressa nella seguente forma semplificata:

$$F_k = \int_{-1/2}^{1/2} f(x) \cos(2 \pi k x) \, dx = 2 \int_{0}^{1/2} f(x) \cos(2 \pi k x) \, dx$$

Allo stesso modo, la trasformazione inversa è semplificata come segue:

$$f(x) = \sum_{k=-\infty}^{\infty} F_k \cos(2 \pi k x) = F_0 + 2 \sum_{k=1}^{\infty} F_k \cos(2 \pi k x)$$

3. Varianti della trasformata finita di Fourier

A seconda dell'intervallo di definizione della funzione *f (x)*, possono scaturire diverse varietà di trasformate finite di Fourier. Vengono riportati due esempi che saranno utili in seguito:

a) Trasformata finita del coseno definita nell'intervallo *[0, 1/2]*.

   Supponiamo che la funzione reale *f (x)* sia definita nell'intervallo *[0, 1/2]*. Definiamo la trasformata finita del coseno, per ogni intero *k* (una funzione pari di *k*), come segue:

$$F_k^c = \int_{0}^{1/2} f(x) \cos(2 \pi k x) \, dx$$

   Se si suppone di estendere l'intervallo di definizione della *f(x)* anche a valori negativi, per simmetria, si ricava semplicemente che $F_k^c = (1/2)F_k$, dove F_k è la trasformata finita di Fourier nell'intervallo *[–1/2, 1/2]*; da quanto precede risulta anche che la trasformata inversa è pari a *2 f(x)*, ovvero:

$$f_k^c(x) = 2 \ f(x) = 2 \sum_{k=-\infty}^{\infty} F_k^c \cos(2 \pi k x) = 2 \ F_0 + 4 \sum_{k=1}^{\infty} F_k^c \cos(2 \pi k x)$$

b) Trasformata finita di Fourier e la sua inversa definite nell'intervallo *[0, 1]*:

$$F_k^I = \int_{0}^{1} f(x) \, e^{-i \, 2 \, \pi \, k \, x} dx \ , \quad f(x) = \sum_{k=-\infty}^{\infty} F_k^I \, e^{i \, 2 \, \pi \, k \, x}$$

4. Trasformata discreta di Fourier

Supponiamo di avere a disposizione *n* osservazioni (dati) acquisite in un periodo complessivo *t* ad intervalli regolari $\Delta t = t \,/\, n$. La frequenza di campionamento (sampling frequency) è pari a $\xi_s = 1 \,/\, \Delta t = n \,/\, t$, mentre la cosiddetta risoluzione di frequenza (frequency resolution) è pari a

$\Delta\xi = 1 / t = \xi_s / n$. Indichiamo gli n dati per mezzo della successione x_m, dove $m = 0, ..., n - 1$ (e non $1, ..., n$). Si osservi che la variabile m / n è definita nell'intervallo $[0, 1)$ (più precisamente, in $[0, 1 - 1 / n]$), così per analogia con la trasformata finita di Fourier F^I, dopo la sostituzione, la discretizzazione e la trasformazione della somma integrale ($x \to m/n$, $f(x) \to x_m$, $dx \to 1/n$), si può definire la trasformata discreta di Fourier (Discrete Fourier Transform - DFT):

$$u_k = \frac{1}{n} \sum_{m=0}^{n-1} x_m \, e^{-i\ 2\ \pi\ k\ m/n}, \quad k = 0, \ ..., \ n-1$$

La trasformata inversa è pari a:

$$x_m = \sum_{k=0}^{n-1} u_k \, e^{i\ 2\ \pi\ k\ m/n}, \quad m = 0, \ ..., \ n-1$$

Commento: la variabile $m \, \Delta t$ rappresenta il tempo, mentre la variabile $k \, \Delta\xi = (k / n) \, \xi_s := \xi_k$ rappresenta la frequenza, così la DFT può essere indicata invece che con u_k, con $u_{k/n}$ o u_{ξ_k}. In via semplificativa e senza perdere di generalità, si può porre $\Delta t = 1 = \xi_s$, quindi il tempo e la frequenza diventano rispettivamente pari a m e k / n.

Nota: spesso in letteratura si trovano delle relazioni differenti, quali:

$$u_k = \sum_{m=0}^{n-1} x_m \, e^{-i\ 2\ \pi\ k\ m/n}, \quad x_m = \frac{1}{n} \sum_{k=0}^{n-1} u_k \, e^{i\ 2\ \pi\ k\ m/n}$$

5. **Commenti sulla trasformata discreta di Fourier**

a) Tipicamente i dati x_m ($m = 0, ..., n - 1$) sono dei numeri reali. Tuttavia le trasformazioni u_k ($k = 0, ..., n - 1$) sono dei numeri complessi, ovvero:

$$u_k = v_k + i \, w_k$$

dove v_k e w_k sono numeri reali. La trasformazione inversa applicata a dei numeri complessi u_k porta ad x_m termini reali.

b) Qualora i dati x_m siano dei numeri reali, valgono le seguenti relazioni:

$$k = 0: \qquad v_0 = \bar{x} \ (\text{media degli } x_m), \ w_0 = 0$$

$$1 \le k \le n-1: \qquad v_k = v_{n-k}, \ w_k = -w_{n-k}$$

vale a dire che la parte reale è simmetrica rispetto al punto $n / 2$, mentre la parte immaginaria è antisimmetrica. Quindi per qualsiasi n, $w_{n/2} = 0$.

c) Le relazioni di simmetria suggeriscono che gli m numeri reali x_m risultino, a seguito della trasformazione, in m numeri reali indipendenti v_k e w_k.

d) Le stesse relazioni suggeriscono che, se si conoscono le u_k per $k \leq n / 2$, risultano quindi automaticamente determinate le restanti. In altri termini, le frequenze $k / n \leq 0,5$ determinano interamente la trasformazione, mentre le altre frequenze non aggiungono alcuna informazione. La frequenza limite di *0,5* è conosciuta come la frequenza di Nyquist.

e) Se $r_k = |u_k| = \left(v_k^2 + w_k^2\right)^{1/2}$ è una misura del numero complesso u_k, vale quanto segue:

$$\frac{1}{n} \sum_{m=0}^{n-1} x_m^2 = \sum_{k=0}^{n-1} r_k^2, \qquad s^2 = \frac{1}{n} \sum_{m=0}^{n-1} \left(x_m - \bar{x}\right)^2 = \sum_{k=1}^{n-1} r_k^2 = \frac{1}{n} \sum_{1 \leq k \leq n/2} p_k^2 - \frac{p_{n/2}^2}{2n}$$

dove $p_k^2 := 2\ n\ r_k^2$. Il termine $\dfrac{p_{n/2}^2}{2n}$ esiste solo se n è un numero pari.

f) La seconda di queste equazioni ci consente di analizzare la dispersione del campione s^2 nelle singole componenti r_k^2 che corrispondono alle frequenze k / n da $\Delta\xi = 1 / n$ fino a $\xi_N = 0,5$ (la frequenza *0* ($k = 0$) corrisponde alla media e non presenta dispersione). Una superiorità significativa di una r_k^2 sulle altre rivela un comportamento periodico del processo con frequenza k / n (periodo n / k).

g) L'andamento di $p_k^2 = 2\ n\ r_k^2$ al variare di k (per $1 \leq k \leq n / 2$) o, più comunemente, della frequenza k / n (per $1 / n \leq k / n \leq 0,5$), è chiamato periodogramma.

h) Il calcolo della DFT può essere effettuato utilizzando direttamente la formula di definizione ma il processo risulta lento. Nel caso in cui il numero di dati n sia una potenza di *2* (ad esempio *64*, *128*, *256*, *512*), esiste un diffuso algoritmo di calcolo veloce (Fast Fourier Transform - FFT), che fornisce oggi la possibilità di operare senza restrizioni sul numero di dati.

6. Spettro di potenza

Per un processo stocastico X_t a tempo discreto $t = 0, 1, \ldots$, sia detta funzione di autocovarianza $\gamma_m := Cov\ [X_t, X_{t+m}]$, $m = 0, \pm 1, \ldots$; la funzione di autocovarianza è la trasformata inversa finita di Fourier di una funzione chiamata spettro di potenza $s(\xi)$ con ξ definita nell'intervallo $[0, 1/2]$. Poiché γ_m è una funzione pari, si ha:

$$s(\xi) = 2 \sum_{m=-\infty}^{\infty} \gamma_m \cos(2\ \pi\ m\ \xi) = 2\ \gamma_0 + 4 \sum_{m=1}^{\infty} \gamma_m \cos(2\ \pi\ m\ \xi)$$

La relazione inversa è $\gamma_m = \int_0^{1/2} s(\xi) \cos(2\ \pi\ m\ \xi)\ d\xi$

Da ciò segue che l'area sottesa dallo spettro di potenza è pari alla varianza γ_0 del processo. In alternativa, lo spettro di potenza può essere definito in base alla funzione di autocorrelazione presentando così l'area pari a *1*.

Se viene usata la funzione di autocovarianza stimata da una serie storica x_t ($t = 0, \ldots, n - 1$), in accordo con lo stimatore standard si ha:

$$g_m = \frac{1}{n} \sum_{m=0}^{n-m} \left(x_t - \bar{x}\right) \left(x_{t+m} - \bar{x}\right)$$

Si dimostra che lo spettro di potenza $s(\xi)$ coincide con il periodogramma per ogni frequenza discreta $\xi = k / n$, per gli interi positivi $k \leq n / 2$ (non applicabile a $\xi = 0 = k$).

Per i processi stocastici reali risulta $s(\xi) \geq 0$ per ogni ξ. Qualunque deviazione computazionale da questa regola deve essere attribuita ad errori di calcolo oppure all'incoerenza della successione γ_m (matrici di covarianza definite negative).

7. Calcolo dello spettro di potenza secondo la sua definizione

Si suppongano noti n valori della funzione di autocovarianza γ_m per $m = 0, \ldots, n - 1$. Teoricamente, l'applicazione dell'equazione che definisce lo spettro di potenza può essere effettuata per ogni valore della frequenza ξ; basandosi sulla definizione dell'equazione si ottiene (assumendo $\gamma_m = 0$ per $m \geq n$):

$$s(\xi) = 2\ \gamma_0 + 4 \sum_{m=1}^{n-1} \gamma_m \cos(2\ \pi\ m\ \xi)$$

Tuttavia, i risultati non sono sempre coerenti per ogni ξ, quindi si effettua il calcolo usando le frequenze discrete. Si distinguono i seguenti casi:

1) Calcolo di $s(\xi)$ nei punti $\xi_k = k / n$ (proprio come nel periodogramma per un numero uguale di n valori):
   a) per $k = 0, \ldots, n / 2$ con n pari;
   b) per $k = 0, \ldots, (n - 1) / 2$ con n dispari.
2) Calcolo di $s(\xi)$ in punti con un passo più fitto che in 1): $\xi_k = k / (2n - 2)$ per $k = 0, \ldots, n - 1$
3) Calcolo di $s(\xi)$ in punti con un altro passo: $\xi_k = k / (2n - 1)$ per $k = 0, \ldots, n - 1$

Nei casi 2) e 1a) in cui l'ultima frequenza (cioè per $k = n - 1$ e $k = n / 2$ rispettivamente) è pari a *0,5*, è preferibile calcolare lo spettro di potenza mediante un'equazione leggermente differente, esplicitando cioè l'ultimo termine:

$$s(\xi) = 2\ \gamma_0 + 4 \sum_{m=1}^{n-2} \gamma_m \cos(2\ \pi\ m\ \xi) + 2\ \gamma_{n-1} \cos\left[2\ \pi\ (n-1)\ \xi\right]$$

L'inversione viene effettuata convertendo la corrispondente trasformata finita di Fourier in una somma, ovvero la seguente relazione, dove *s(0,5)* esiste nei casi 2) e 1a):

$$\gamma_m = \xi_1 \left[\frac{s(0) + (-1)^m \ s(0,5)}{2} + \sum_{0<\xi_k<0,5} s(\xi_k) \ \cos(2 \ \pi \ m \ \xi_k) \right]$$

8. Calcolo dello spettro di potenza tramite DFT o FFT

Si supponga ancora che la funzione autocovarianza γ_m sia nota per n valori di $m = 0, \ldots, n-1$. Se $n-1$ è una potenza di 2 (anziché n, come nell'analisi di dati standard), lo spettro di potenza può essere stimato (riducendo drasticamente il tempo di calcolo) utilizzando i punti della FFT $\xi_k = k / (2n-2)$ per $k = 0, \ldots, n-1$ (come nel precedente caso 2)).

Considerando $\gamma_m = 0$ per $m \geq n$, si definisce la successione δ_m, simmetrica rispetto al punto $n-1$, per $m = 0, \ldots, 2n-3$, come segue:

$$\delta_m = \begin{cases} 4 \ (n-1) \ \gamma_m & \text{se} \quad m \leq n-1 \\ \delta_{2(n-1)-m} & \text{altrimenti} \end{cases}$$

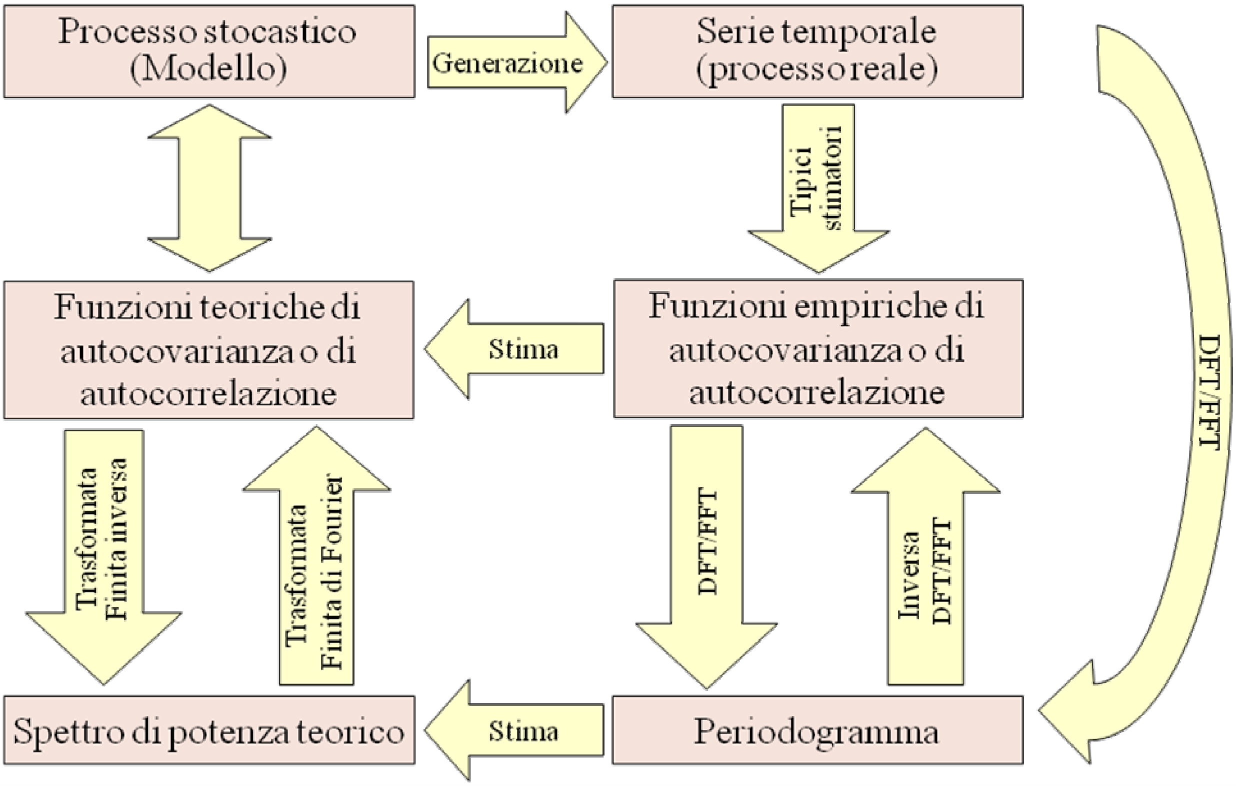
I valori di u_k della DFT e della FFT nei punti $k = 0, \ldots, n-1$ sono i valori richiesti di $s(\xi_k)$ per frequenze comprese tra 0 e $0,5$.

La successione inversa di u_k ottenuta mediante le inverse della DFT o FFT fornisce i valori di δ_m che determinano l'autocovarianza:

$$\gamma_m = \frac{\delta_m}{[4 \ (n-1)]}$$

per $m = 0, \ldots, n-1$.

9. Relazioni processo / serie storica, autocovarianza e spettro di potenza / periodogramma

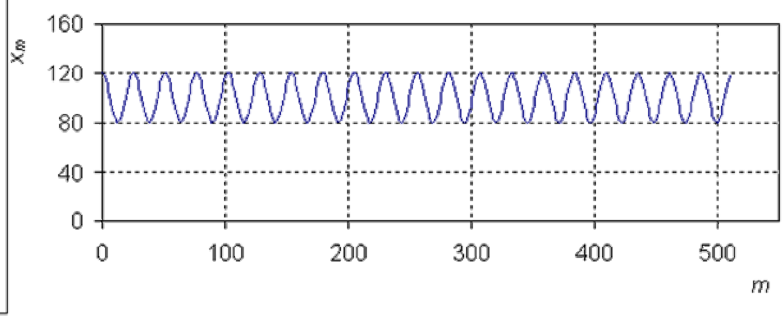


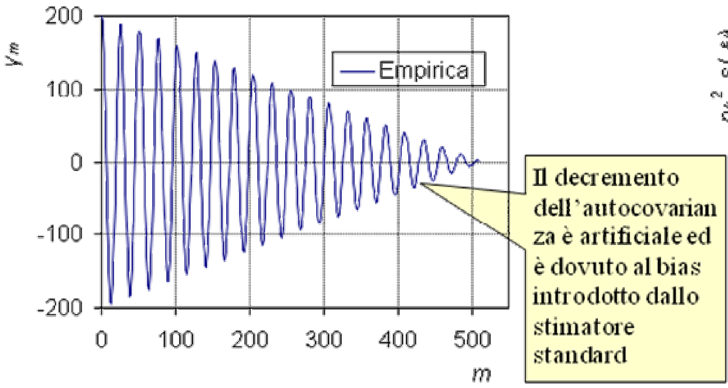
10. Esempi di analisi spettrale

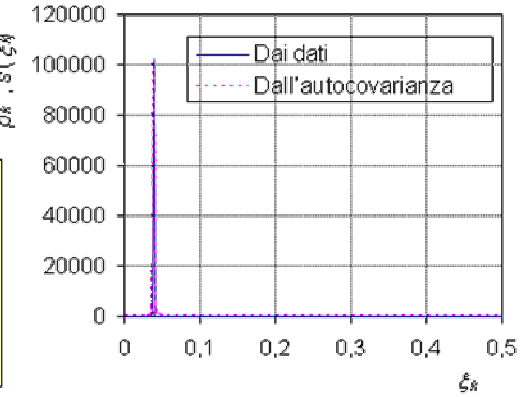
a) *Segnale periodico di un moto armonico*

Parametri del modello	
Livello medio del segnale	100
Ampiezza del segnale	20
Frequenza	0,039
Numerosità campione	512

Nota: il modello stocastico è deterministico quindi non è definita la funzione di autocovarianza teorica

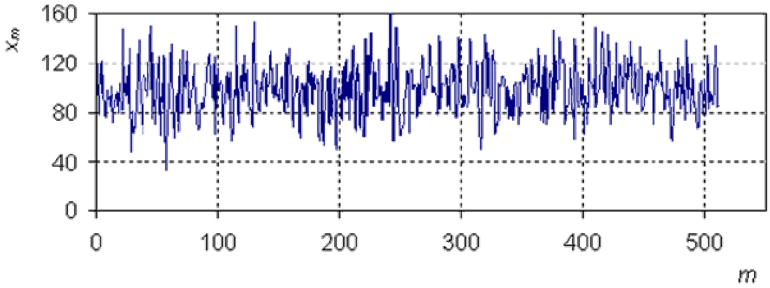


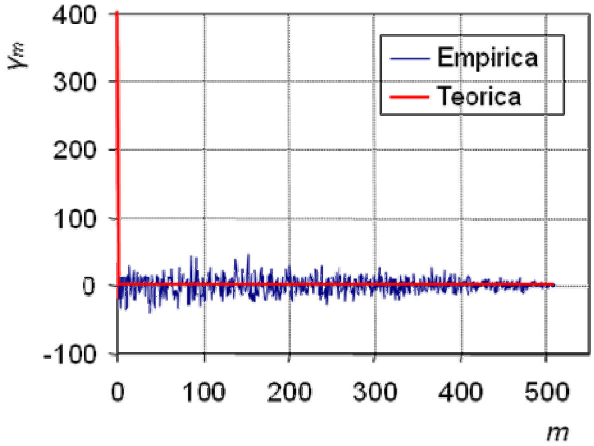


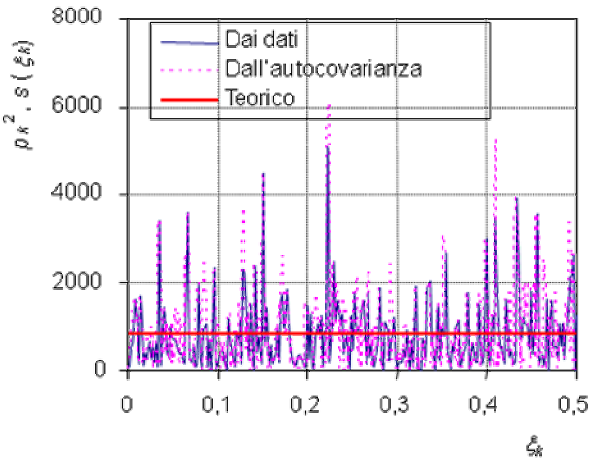


b) *Rumore bianco*

Parametri del modello	
Media teorica	100
Deviazione standard teorica	20
Numerosità campione	512

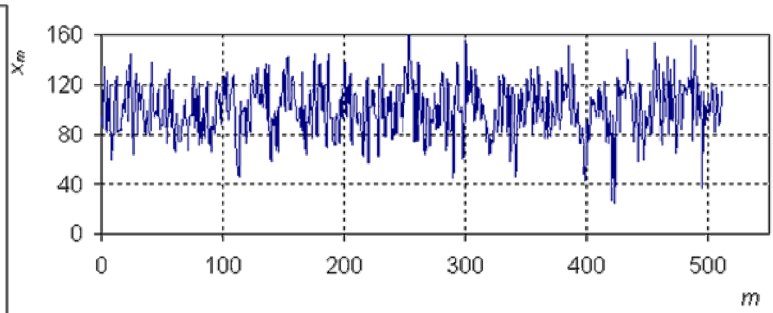


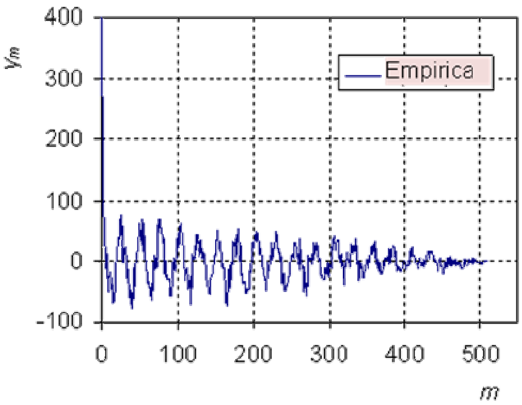


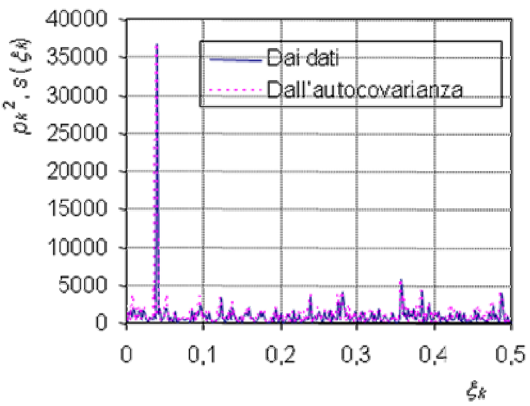


c) <u>*Segnale periodico con rumore bianco*</u>

Parametri del modello	
Livello medio del segnale	100
Ampiezza del segnale	10
Frequenza	0,039
Dev. standard del rumore	20
Numerosità campione	512
Nota: il modello non è stazionario e quindi la funzione di autocovarianza teorica è funzione del tempo	

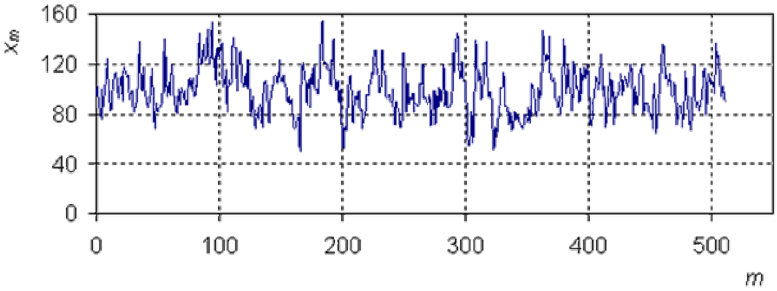


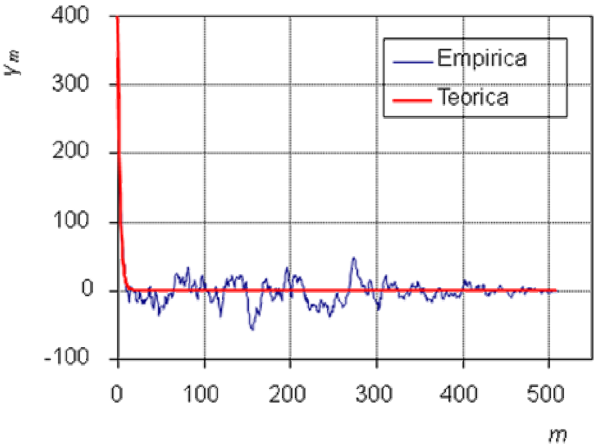


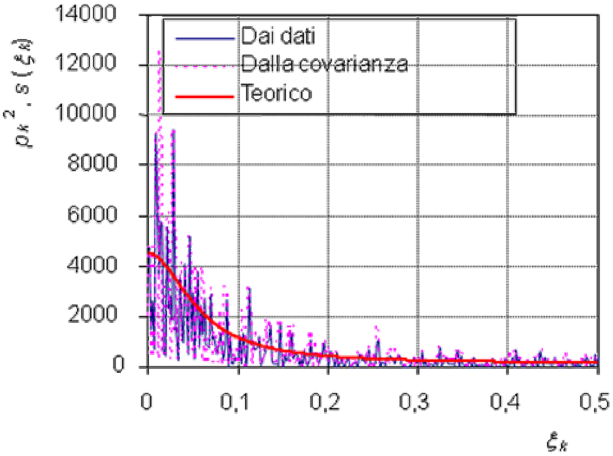


d) <u>*Modello di Markov (persistenza a breve termine)*</u>

Parametri del modello	
Media teorica	100
Deviazione standard teorica	20
Autocorrelazione teorica lag 1	0,7
Numerosità campione	512







e) Modello FGN (persistenza a lungo termine)

Parametri del modello	
Media teorica	100
Deviazione standard teorica	20
Coefficiente di Hurst teorico	0,7
Numerosità campione	512