

## 4. I dati sperimentali

### 4.1 Introduzione

La serie esaminata, (Ceschia, 1998), è relativa alle temperature registrate alle ore 9 presso l'Osservatorio Malignani a Udine-Castello. Le temperature, espresse in gradi Celsius, sono state rilevate con la precisione strumentale di un decimo di grado. La serie è stata sottoposta ad accurati test di consistenza ed a verifiche di omogeneità di misurazione nel corso degli anni.

La serie di osservazioni meteorologiche, iniziata a Udine-Castello ad opera di A. Malignani nel 1894, continua ormai da oltre cent'anni ad opera di tre generazioni della stessa famiglia.

La stazione è situata sul lato Sud-Ovest del colle del Castello di Udine, alla quota di 136 m sul livello del mare. I termometri a mercurio erano difesi “*contro l'irradiazione solare e notturna da una torre e una tettoia*” (Malignani, 1939). Ora la vecchia stazione storica è stata trasformata in una stazione automatica dotata dei più moderni sensori ed è gestita in collaborazione con il Dipartimento di Fisica dell'Università di Udine.

## 4.1 Introduzione

Nonostante la cura con cui i membri della Famiglia Malignani hanno seguito nel corso degli anni la raccolta dei dati dell'Osservatorio, la serie presenta alcuni buchi prodotti per lo più da cause di forza maggiore. La maggior parte dei dati mancanti, infatti, corrisponde al periodo che va dall'ottobre del 1917 all'inizio del 1919, ossia tra la disfatta di Caporetto e la fine della prima guerra mondiale.

Nel presente lavoro ci si è limitati ad utilizzare il più lungo spezzone completo della serie che va dal 1° febbraio 1920 al 31 dicembre 1996. Le principali caratteristiche della serie sono riassunte nella Tabella 4.1; la serie è rappresentata nella Figura 4.1.

<b>Ora di rilevazione</b>	<b>9:00</b>
<b>Periodo di campionamento</b>	<b>24 h</b>
<b>Quantità dei dati</b>	<b>28094</b>
<b>Minimo</b>	<b>-13.5</b>
<b>Massimo</b>	<b>32.0</b>
<b>Media</b>	<b>12.35</b>
<b>Deviazione standard</b>	<b>7.94</b>

TABELLA 4.1. Principali caratteristiche della serie studiata.

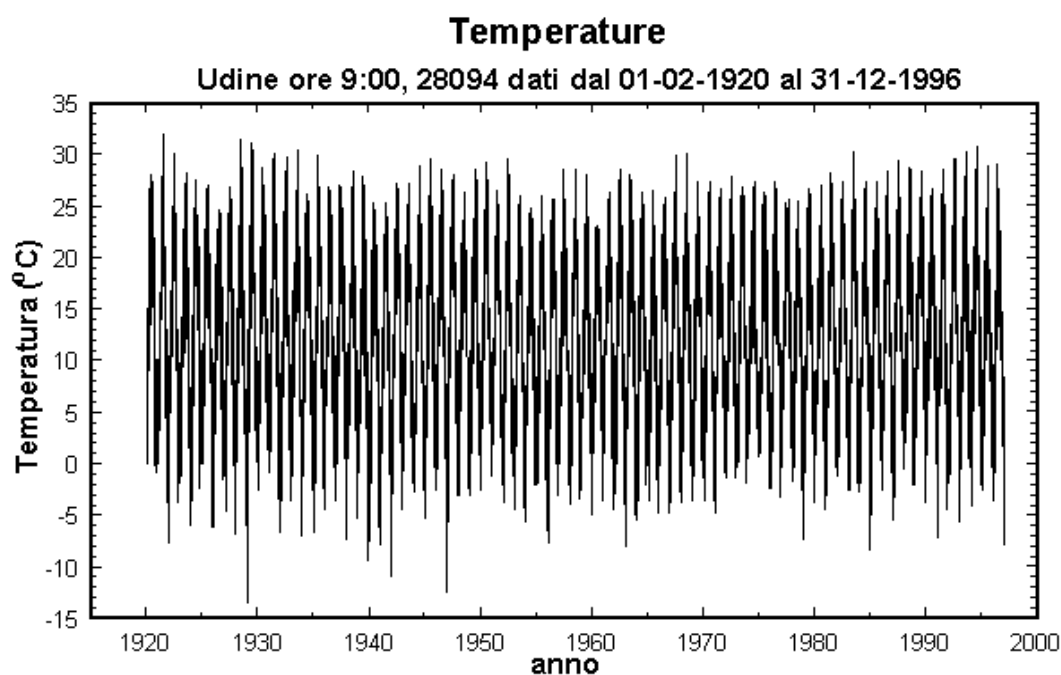


FIGURA 4.1 La serie di temperature di Udine-Castello.

#### 4. I dati sperimentali

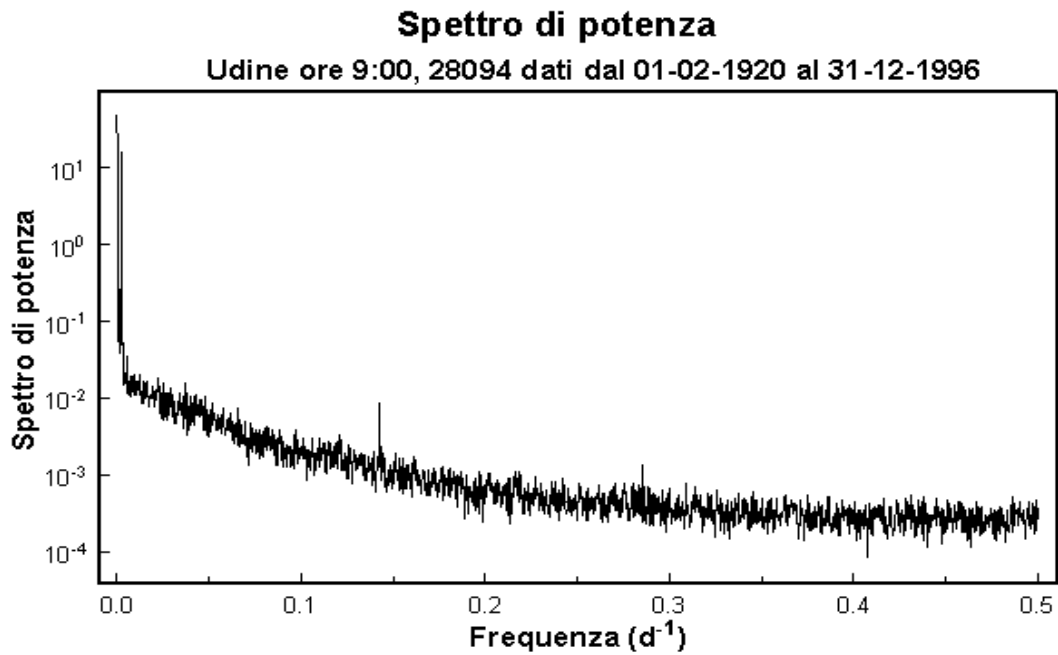


FIGURA 4.2 Spettro di potenza della serie di temperature.

Nella Figura 4.2 è rappresentato lo spettro di potenza della serie in esame. Nell'ingrandimento della zona a bassa frequenza di Figura 4.3 si vede come, a parte il valore massimo relativo alla componente a frequenza nulla (costante), sia presente il picco relativo all'andamento annuale.

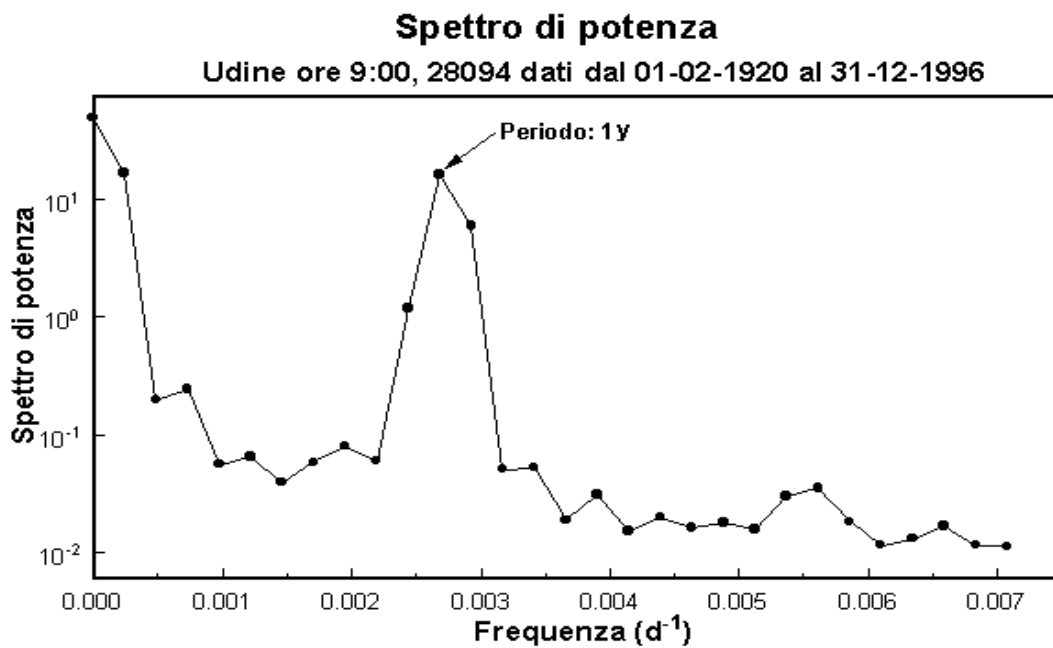


FIGURA 4.3 Ingrandimento della zona a bassa frequenza dello spettro di Figura 4.2.

## 4.1 Introduzione

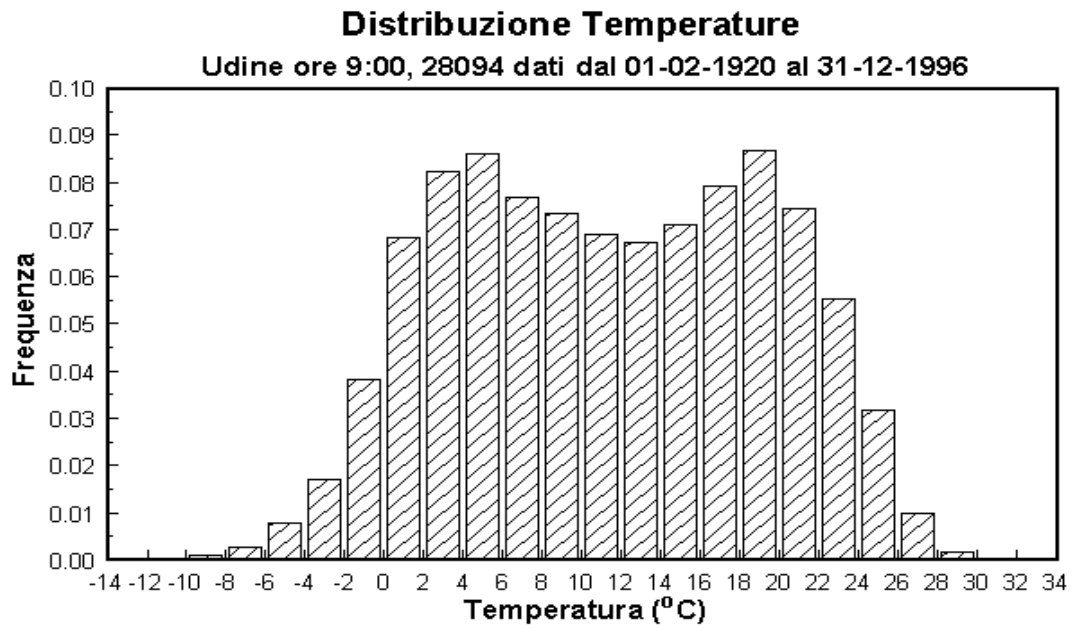


FIGURA 4.4 Distribuzione della serie di temperature.

Nelle Figure 4.4 e 4.5 sono rappresentate rispettivamente la distribuzione delle temperature e l'autocorrelazione della serie.

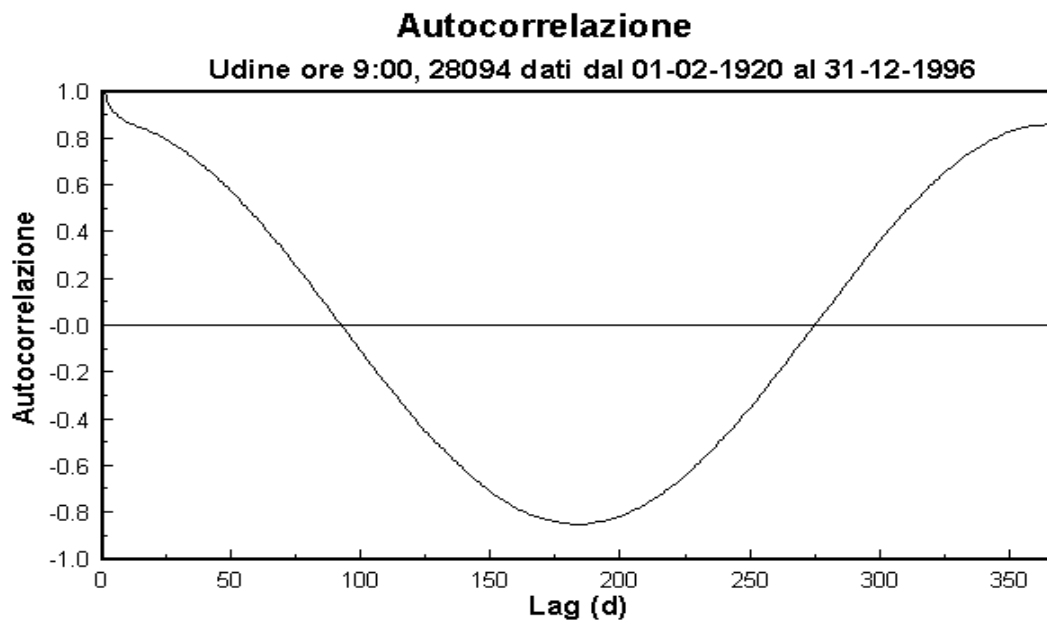


FIGURA 4.5 Autocorrelazione della serie di temperature.

## 4.2 Denoising: il metodo delle wavelet

Una serie temporale può possedere caratteristiche caotiche le quali possono essere nascoste dalla presenza di rumore. Se l'intensità del rumore è sufficientemente alta, il metodo di Takens non consente la ricostruzione corretta dell'attrattore. In generale il rumore presente nelle serie sperimentali possiede una dimensione molto maggiore di quella del sistema dinamico soggiacente e pertanto può mascherarne le caratteristiche (Cuomo et al., 1993). Diversi sono i tentativi di sviluppare metodi pratici per eliminare il rumore presente in una serie (Sauer, 1992; Grassberger et al., 1993; Kostelich and Schreiber, 1993). Alcuni di questi metodi (Donoho, 1993; Saito, 1994) si basano sulla recente tecnica di analisi e sintesi dei dati nota come *Wavelet Transform* (Daubechies, 1988; Mallat, 1989; Kaiser, 1994) che è stata adottata anche in questo lavoro.

### 4.2.1 Continuous Wavelet Transform

La *continuous wavelet transform* di un segnale  $f(t)$  è definita dalla relazione:

$$\tilde{f}(a,b) = \int_{-\infty}^{+\infty} \Psi_{a,b}^*(t) f(t) dt \quad (4.1)$$

ove

$$\Psi_{a,b}(t) \equiv |a|^{-p} \Psi\left(\frac{t-b}{a}\right) \quad (4.2)$$

è una funzione a valori complessi e media nulla appartenente a  $L^2(\mathbf{R})$ ; il valore di  $p$  è irrilevante e ci si atterrà alla convenzione prevalente ponendo  $p=1/2$  (Kaiser, 1994).

La scelta di  $\Psi$  è largamente arbitraria e per il presente lavoro è stata considerata la *Mexican hat function* (Figura 4.6) che è l'opposto della derivata seconda della distribuzione normale a media nulla e varianza unitaria:

$$\Psi(t) = \frac{(1-t^2)e^{-t^2/2}}{\sqrt{2\pi}}. \quad (4.3)$$

## 4.2 Denoising: il metodo delle wavelet

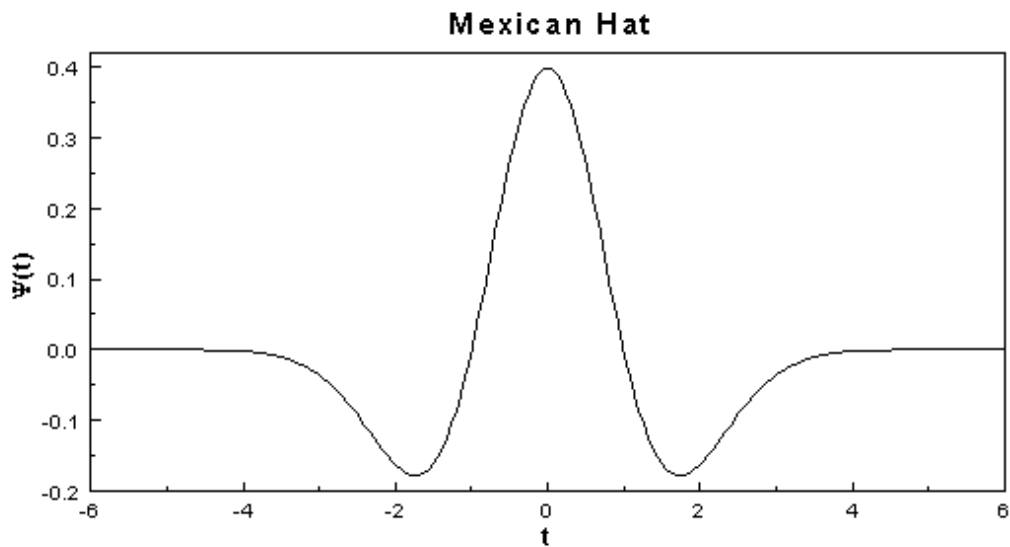


FIGURA 4 6 Diagramma della funzione di equazione 4.3.

Si dimostra (Kaiser, 1994) che, in tal caso,  $\tilde{f}(a,b)$  rappresenta il dettaglio di secondo ordine di  $f$  alla scala  $a$ , nel senso che  $\tilde{f}(a,b)$  è la derivata seconda di una “media mobile” di  $f$  calcolata mediante una gaussiana traslata (parametro  $b$ ) e dilatata (parametro  $a$ ).

In altre parole, un dato coefficiente della trasformata, ossia un dato elemento  $\tilde{f}(a_0,b_0)$  di  $\tilde{f}(a,b)$ , sarà in modulo tanto maggiore quanto più il segnale risulti variabile in un intorno di  $b_0$ , con un “periodo” intorno a  $4a_0$ .

L’antitrasformata della (4.1) è data dalla:

$$f_1(t) = \int_0^{+\infty} a^{2p-3} da \int_{-\infty}^{+\infty} \Psi_{a,b}(t) \tilde{f}(a,b) db, \quad (4.4)$$

in effetti la  $f_1(t)$  differisce dal segnale originale  $f$  nel senso che il procedimento di analisi e sintesi sottrae ad ogni valore di  $f_1(t)$  la media locale di  $f$  calcolata intorno a  $t$ .

Il ruolo giocato dal parametro  $a$  è dunque quello di analizzare il segnale a diverse scale della variabile temporale (a piccoli valori di  $a$  corrispondono scale più fini); nell’ipotesi quindi che il segnale sia affetto da rumore ad alta frequenza, risulta evidente come, sostituendo al limite inferiore di integrazione della variabile  $a$  nella

#### 4. I dati sperimentali

(4.4) una soglia  $a_0 > 1$ , sia possibile ottenere un segnale depurato dal rumore (*denoising*).

#### 4.2.2. Descrizione dell'algoritmo

L'analisi dei dati è stata implementata mediante integrazione numerica della (4.1) con passo di integrazione unitario, avendo scelto l'unità di tempo pari alla distanza (costante) tra due dati successivi della serie temporale:

$$\tilde{f}(a,b) = \frac{1}{\sqrt{a}} \sum_i \Psi\left(\frac{t_i - b}{a}\right) f(t_i). \quad (4.5)$$

Per ovviare agli effetti di bordo causati dalla lunghezza finita del segnale si sono imposte condizioni al contorno cicliche: tale scelta, seppur arbitraria, si è rivelata alquanto efficace. Il parametro  $a$  varia da 1 ad un valore  $a_{\max}$  determinato dall'utente,  $b$  varia da 1 a  $b_{\max}$  pari al numero di dati della serie. Nella Figura (4.7) è rappresentata una serie di valori della variabile  $x$  del sistema di Lorenz con il 50% di rumore bianco additivo (curva tratteggiata) e la stessa serie priva di rumore (curva

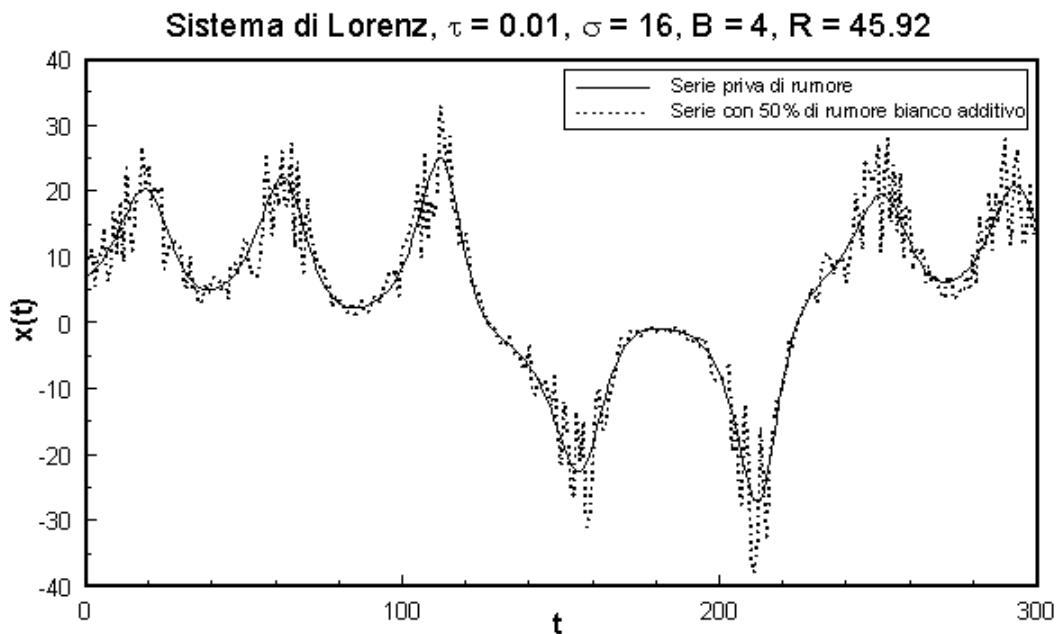


FIGURA 4.7 Variabile  $x$  del sistema di Lorenz con 50% di rumore bianco additivo (curva tratteggiata) e la stessa serie priva di rumore (curva continua).

## 4.2 Denoising: il metodo delle wavelet

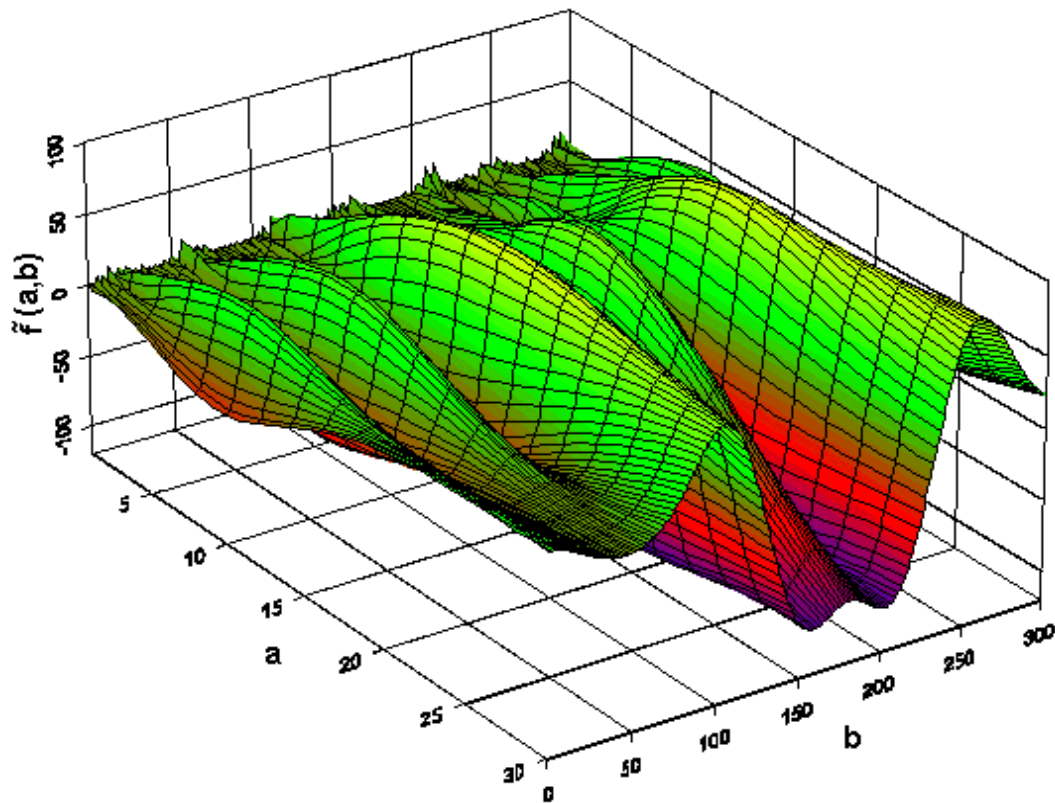


FIGURA 4.8 Superficie di equazione  $z = \tilde{f}(a,b)$  (4.5) relativa alla serie rumorosa di Figura (4.7).

continua). La serie di Lorenz rumorosa è stata usata per valutare l'efficacia del metodo. Nella figura (4.8) è rappresentata la superficie di equazione  $z = \tilde{f}(a,b)$  relativa alla serie rumorosa di Figura (4.7).

La sintesi (ricostruzione) è stata ottenuta mediante integrazione numerica della (4.4):

$$f_1(t) = \sum_a \sqrt{a} \sum_b \Psi\left(\frac{t_i - b}{a}\right) \tilde{f}(a,b) \quad (4.6)$$

$b$  varia da 1 a  $b_{\max}$  come per l'analisi, mentre gli estremi di integrazione per  $a$  sono forniti in input. Ciò consente, come opzione, di effettuare il denoising del segnale.

Il vettore  $f_1(t)$ , risultato della sintesi, viene fornito come input ad un secondo modulo il quale aggiunge ad ogni elemento di  $f_1(t)$  la media locale calcolata attorno a  $t$ ; questo procedimento viene ripetuto diverse volte per diverse larghezze ( $w$ ) della finestra sulla quale si calcola la media locale. Si ottengono così diversi vettori per



#### 4. I dati sperimentali

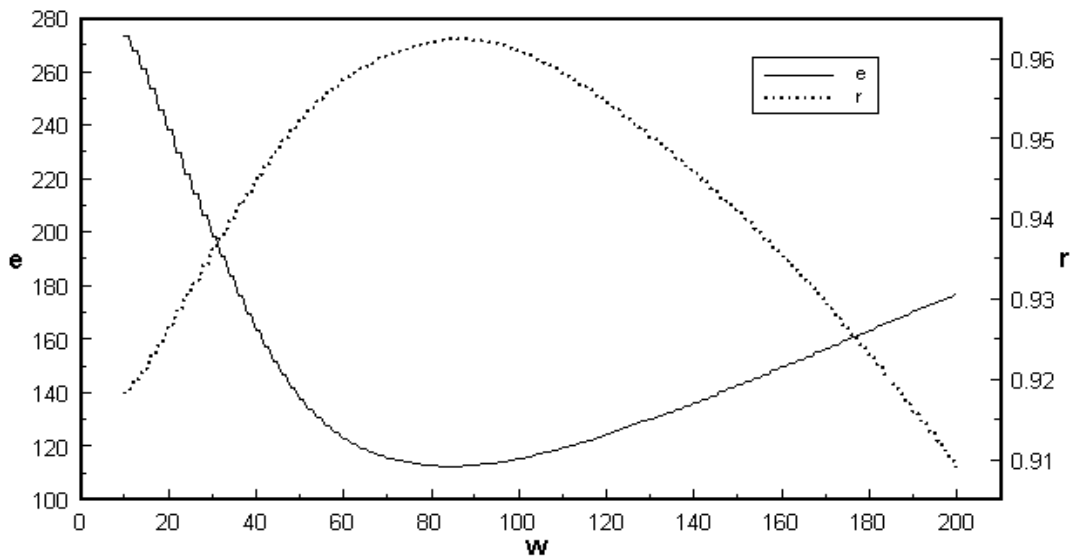


FIGURA 4.9 Errore quadratico medio ( $e$ ) e coefficiente di correlazione ( $r$ ) tra la serie rumorosa originale e quella ricostruita  $f_1(t)$  in funzione della larghezza della finestra ( $w$ ) sulla quale si calcola la media mobile aggiunta a ciascun punto di  $f_1(t)$ .

ciascuno dei quali vengono calcolati il coefficiente di correlazione ( $r$ ) e l'errore quadratico medio ( $e$ ) rispetto al segnale originale.

Nella figura 4.9 sono rappresentati i valori di  $r$  ed  $e$  in funzione della larghezza  $w$  della finestra; è così possibile operare una scelta di  $w$  tale da rendere minimo l'errore quadratico medio o da rendere massimo il coefficiente di correlazione (non sempre tali condizioni si ottengono in corrispondenza ad un unico valore di  $w$ ). In corrispondenza a tale scelta si ottiene infine il segnale ricostruito sommando ad ogni elemento di  $f_1(t)$  la media locale calcolata attorno a  $t$ .

Il procedimento descritto è stato applicato alla serie rumorosa di Figura 4.7: nell'analisi il parametro  $a$  assumeva valori compresi tra  $a_{\min} = 1$  e  $a_{\max} = 30$ ; il *denoising* è stato ottenuto annullando nella ricostruzione i coefficienti  $\tilde{f}(a,b)$  con  $a = 1$  e  $a = 2$ . Come si vede nella Figura 4.9, il valore ottimale della larghezza della finestra sulla quale calcolare la media locale è risultato pari a 84.

Nella Figura 4.10 vengono confrontate la serie *denoised* con quella priva di rumore. La validità del procedimento è confermata dal fatto che l'applicazione del metodo dei falsi vicini alla serie *denoised* consente di determinare la corretta

## 4.2 Denoising: il metodo delle wavelet

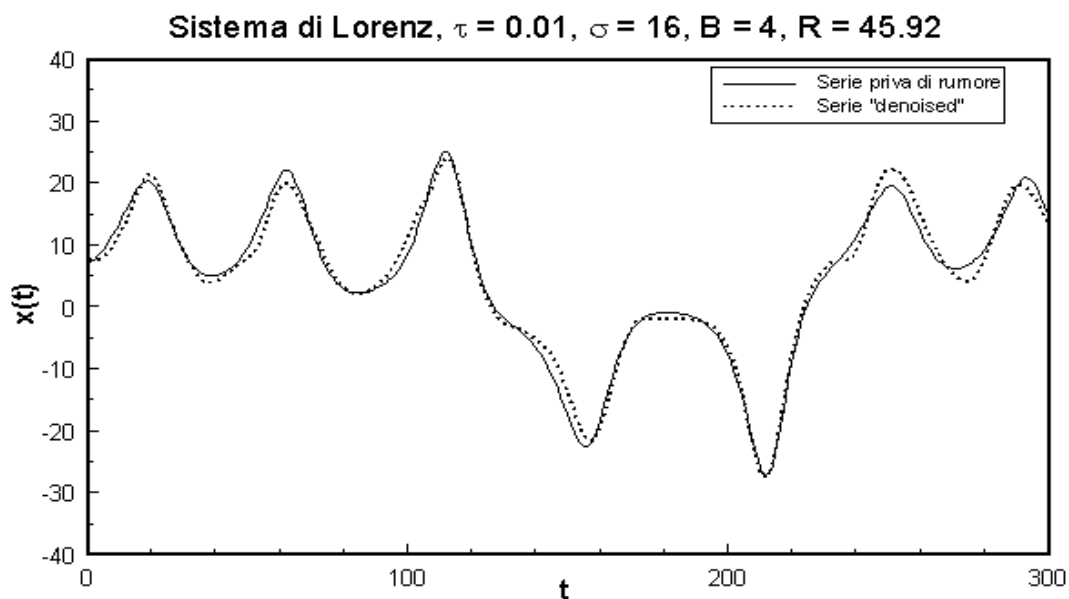


FIGURA 4.10 Confronto tra la serie *denoised* e quella priva di rumore di Figura 4.7.

dimensione di embedding come si vede nella Figura 4.11, anche se il rumore residuo fa sì che il numero di falsi vicini corrispondente a  $m = 3$  sia leggermente maggiore di zero anziché nullo.

Il metodo descritto è stato applicato alla serie delle temperature adottando l'approccio più conservativo: il *denoising* è stato ottenuto annullando nella ricostruzione i coefficienti  $\tilde{f}(a,b)$  con  $a = 1$ , ossia solo quelli alla frequenza più

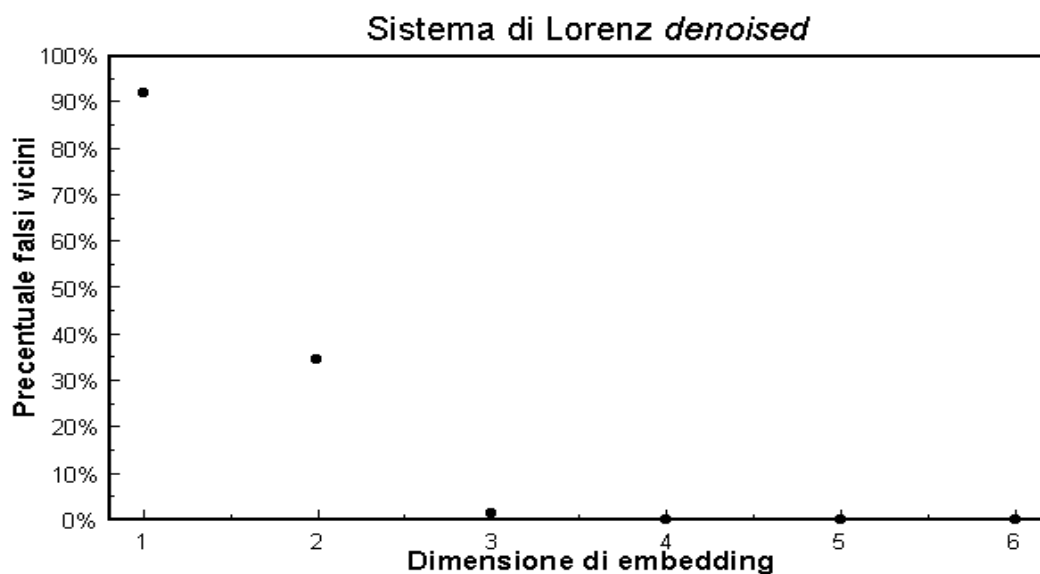


FIGURA 4.11 Falsi vicini per la serie *denoised* di Figura 4.7.

#### 4. I dati sperimentali

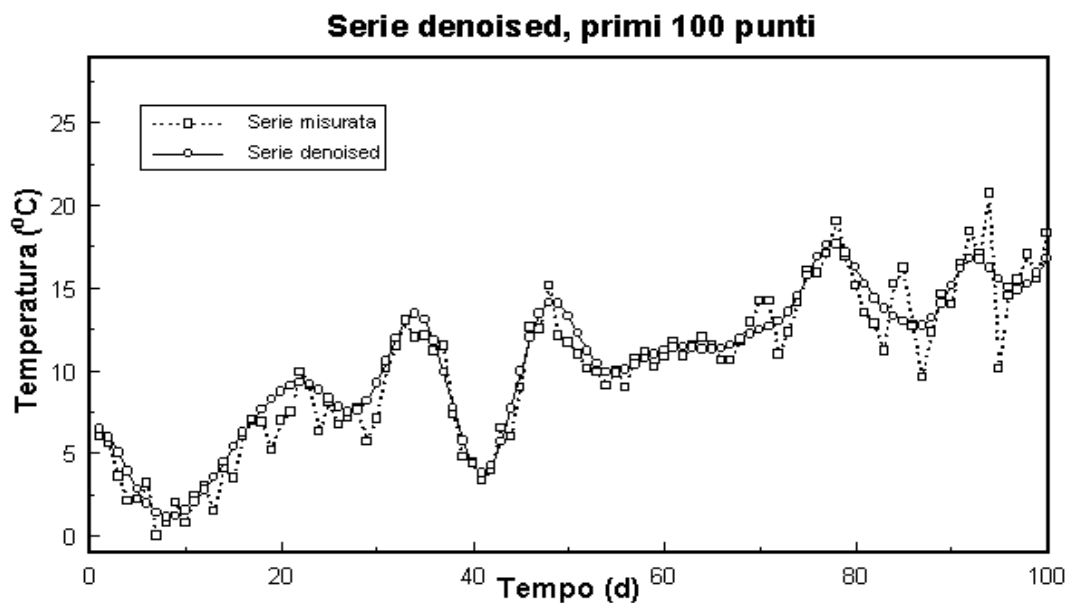


FIGURA 4.12 Confronto tra la serie di temperature misurata e quella *denoised*.

alta.

Nella Figura 4.12 sono rappresentati i primi 100 punti della serie *denoised* e quelli della serie misurata.

Per verificare che la componente sottratta col procedimento di *denoising* è

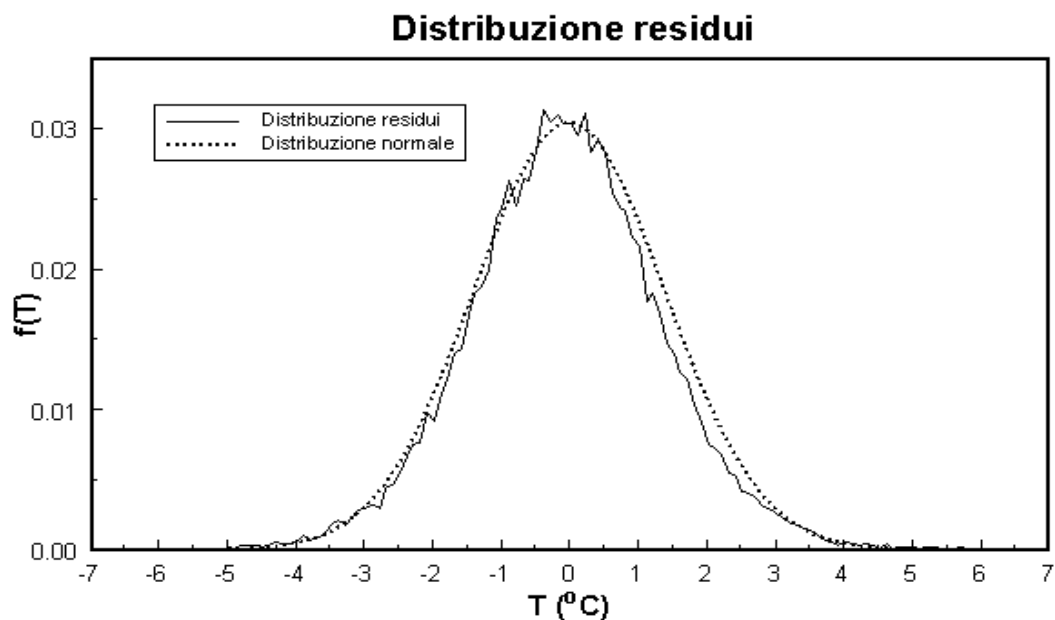


FIGURA 4.13 Confronto tra la distribuzione dei residui e quella normale.

## 4.2 Denoising: il metodo delle wavelet

effettivamente un rumore bianco, la distribuzione dei residui (differenze tra la serie misurata e quella *denoised*) è stata confrontata (Figura 4.13) con la distribuzione normale avente la stessa media ( $\mu = 0.00$ ) e varianza ( $\sigma^2 = 1.95$ ) dei residui. Come si vede dalla figura, il buon accordo tra la distribuzione dei residui e quella normale conferma la natura casuale della componente sottratta.