

Master in “Calcolo scientifico”

II Livello

Università “La Sapienza”

ROMA

Anno Accademico: 2016-2017

SERIE TEMPORALI CLIMATICHE

Studente Dott.ssa Ig. Adelaide Nardolillo

Referente Prof. Camillo Cammarota

15/06/2018

Indice

1. Esempi di serie temporali	pag.2
2. Modello matematico delle serie in funzione del tempo	pag.6
2.1 Regressione	pag.6
2.1.a Caso polinomiale	pag.7
2.2.b Caso armonico	pag.9
2.2 Periodogramma	pag.9
2.3 Metodo SSA (Singular Spectrum Analysis)	pag.12
3. Analisi misure NASA	pag.15
3.1 La variabile Global Land-Ocean Temperature Index (GLOTI)	pag.17
3.1.a Metodo SSA	pag.18
3.2 La variabile Sud Hemisph. Station Temperature Index (SHSTI)	pag.21
3.2.a Metodo SSA	pag.22
3.3 La variabile North Hemisph. S. Temperature Index (NHSTI)	pag.25
3.3.a Metodo SSA	pag.26
3.4 Le serie Station Temperature Index (STI) in media annuale: STIAG (Annual mean Global), STISH (Sud Hemisphere) e STINH (North Hemisphere)	pag.28
3.5 SOI (Southern Oscillation Index)	pag.31
3.6 SLP (Sea Level Pression)	pag.35
4. Le serie temporali ENEA temperatura e pressione	pag.39
4.1 Apparatati di misura. Tipologie sensori	pag.39
4.2 Descrizione delle serie	pag.40
4.3 Metodo SSA	pag.42
5. Bibliografia	pag.54

1. Esempi di serie temporali

Il sistema climatico è caratterizzato da grandezze fisiche climatiche, che sono campionate come serie temporali con rappresentazione di variabili casuali. Esse sono essenzialmente la temperatura, la pressione e gli indici (es. SOI e STI) di deviazione rispetto al normale andamento stagionale. In particolare, gli indici sono delle stime delle anomalie rispetto alle temperature medie e pressioni, calcolate su un determinato periodo di riferimento. Tali stime sono effettuate sulla base degli articoli scientifici, che propongono dei metodi formulistici per individuarli. Le guide nazionali e la guida WMO¹ posso agevolare nella scelta del metodo più opportuno.

Le anomalie, di seguito, fanno riferimento ai dati combinati “land+surface air and sea surface water”, cioè ad una combinazione di misure di terra, atmosfera e superficie del mare. Esse sono delle medie mensili globali, emisfero nord e sud.

Il primo esempio² è:

Global Land-Ocean Temperature Index, in 0.01 degrees Celsius, dal 1880 al 2017 base period: 1951-1980 [2], indicata con GLOTI, fig. 1.1;

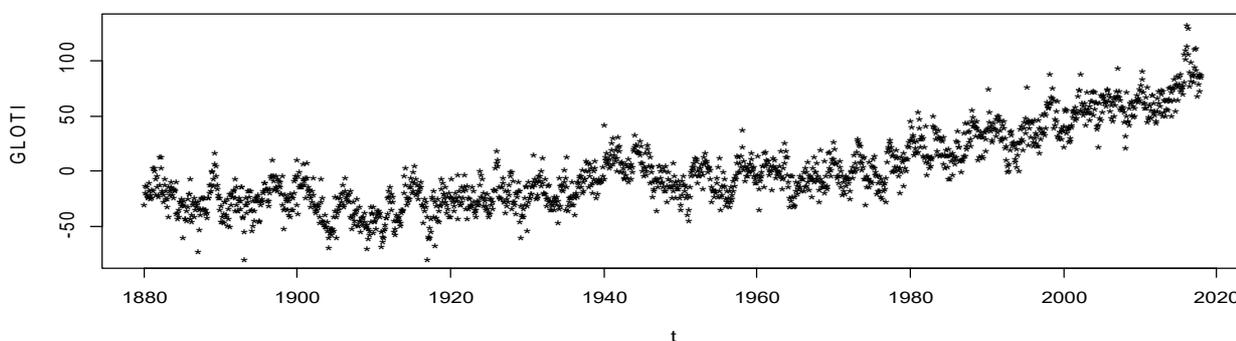


Fig. 1.1: GLOTI, misurato in °C in funzione del tempo in mesi.

Il secondo esempio è:

S. Hemisph. Station Temperature Index, in 0.01 degrees Celsius, dal 1880 al 2017 base period: 1951-1980³, indicata con SHSTI, fig. 1.2;

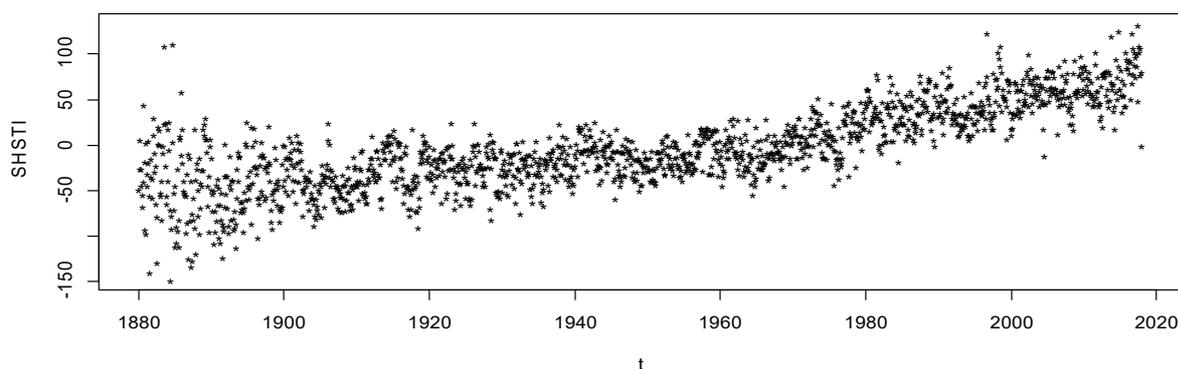


Fig. 1.2: SHSTI, misurato in °C in funzione del tempo.

1. Vedi n.7 bibliografia;

2. www.data.giss.nasa.gov; sources: GHCN-v3 1880-11/2017 (meteorological stations only), using elimination of outliers and homogeneity adjustment;

3. sources: GHCN-v3 1880-11/2017 + SST: ERSST v5 1880-11/2017, using elimination of outliers and homogeneity adjustment.

N. Hemisph. Station Temperature Index, in 0.01 degrees Celsius, dal 1880 al 2017 base period: 1951-1980 [3], indicata con NHSTI, fig. 1.3;

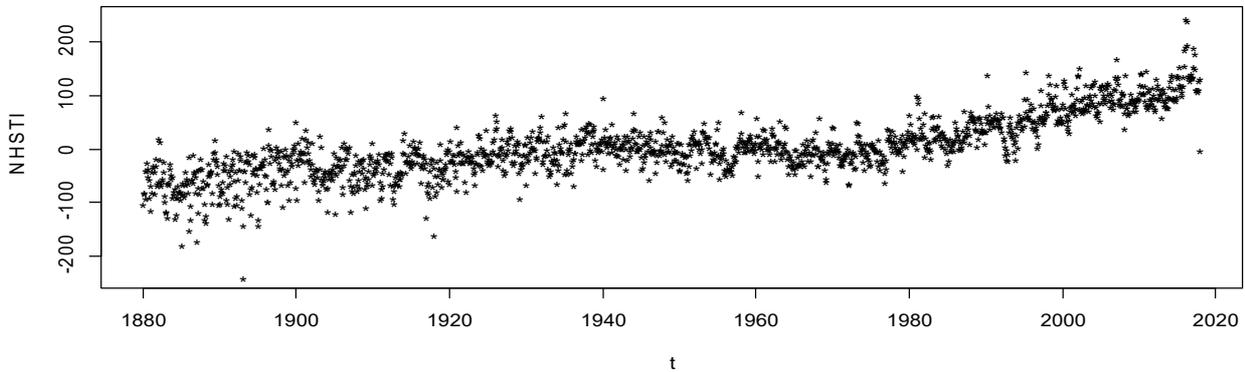


Fig. 1.3: NHSTI, misurato in °C in funzione di t.

La seconda sequenza di grandezze è relativa a medie annuali e, quindi, ha un'elaborazione maggiore rispetto alle prime tre. Esse fanno riferimento a precise stazioni metereologiche di riferimento:

Annual Mean Station Temperature Index, in 0.01 degrees Celsius, dal 1880 al 2017 selected zonal means [3], indicata con AMSTI, nelle grandezze di STIAG, indice di temperatura in media annuale globale, fig. 1.4, STIANH, media annuale emisfero boreale, fig. 1.5, STIASH, media annuale emisfero australe, fig.1.6, dati NASA.

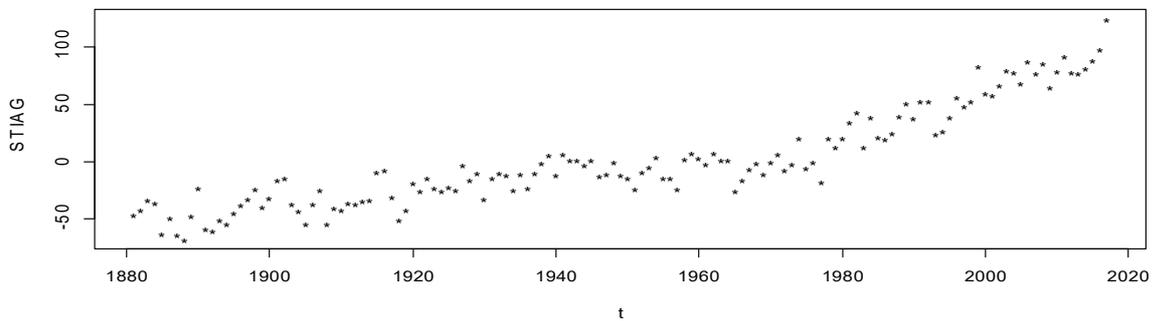


Fig. 1.4: STIAG, misurato in °C in funzione del tempo.

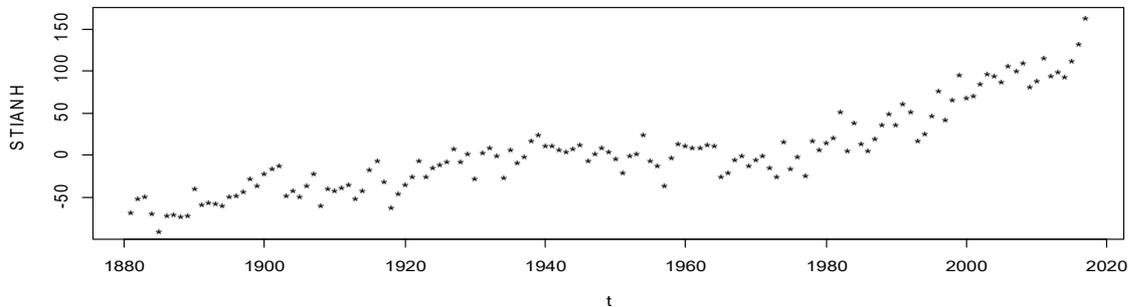


Fig. 1.5: STIANH, misurato in °C in funzione del tempo.

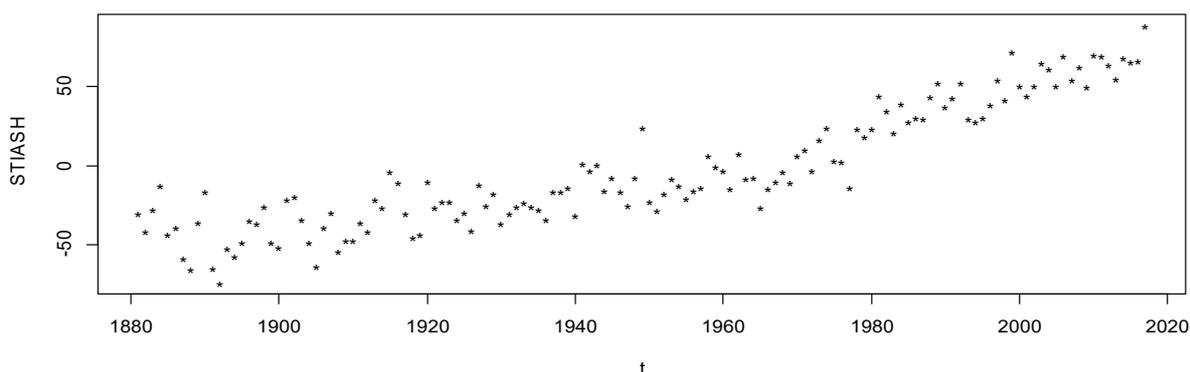


Fig. 1.6: STIASH, misurato in °C in funzione del tempo.

Le medie sono considerate mensili nei primi 3 grafici, mentre è preso in considerazione il valor medio annuale negli altri 3 grafici. Il set di dati nel primo caso è di 1656, nel secondo caso è di 137.

La temperatura e i suoi indici dipendono dallo spazio e dal tempo: $T(r,t)$. Dal punto di vista temporale, è interessante effettuare delle osservazioni sulle medie utilizzate e sul numero di dati a disposizione.

Il quinto esempio riguarda il SOI, Southern Oscillation Index, fig. 1.7, utilizzato nello studio di El Niño⁴, su 804⁵ dati:

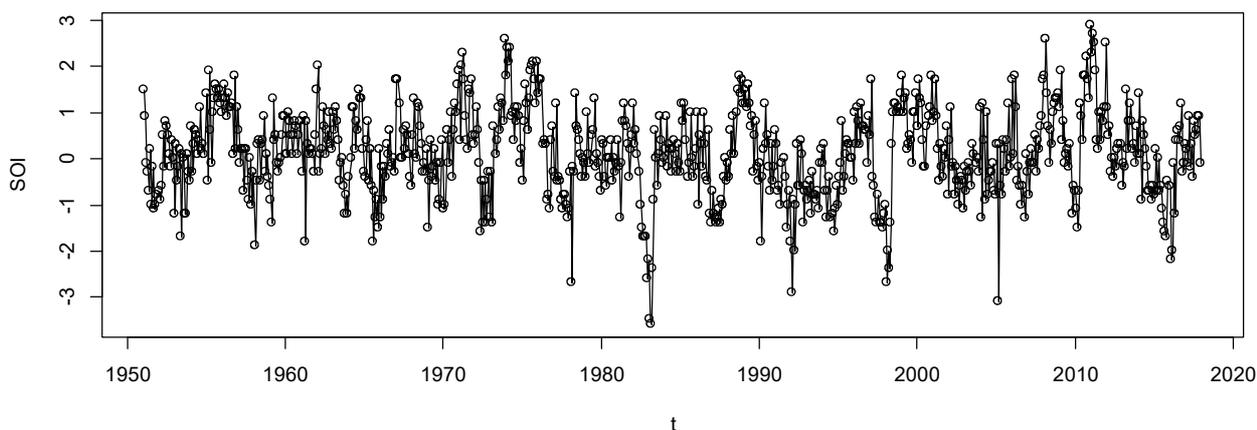


Fig. 1.7: SOI dal 1951 al 2017 in funzione del tempo in mesi.

Il SOI è la differenza tra l'indice di livello di pressione del mare (SLP) di Tahiti e lo SLP di Darwin, in Australia, calcolato nel seguente modo:

$$SOI = \frac{ST - SD}{MSD}$$

$$ST = \frac{SLP(\text{attuale Tahiti}) - SLP(\text{medio Tahiti})}{\sigma(\text{Tahiti})} \quad \text{con N1 numero di dati Tahiti;}$$

4. Vedi n.19 bibliografia.

5. Dati NASA, www.data.giss.nasa.gov;

$$SD = \frac{SLP(\text{attuale Darwin}) - SLP(\text{medio Darwin})}{\sigma(\text{Darwin})} \quad N2 \text{ numero di dati Darwin};$$

$$MSD = \sqrt{\frac{\sum_i (STI_i - SD_i)^2}{N}} ;$$

$$N = N1 + N2.$$

La fase negativa del SOI rappresenta un periodo sotto la norma a Tahiti e uno sopra a Darwin. Prolungati periodi di SOI negativo coincidono con anomalie di caldo e, quindi, episodi di El Niño. Prolungati periodi di SOI positivo coincidono con anomalie di freddo e episodi della Niña.

Il sesto esempio è lo SLP, indice di pressione, fig. 1.8, misurato in hPa:

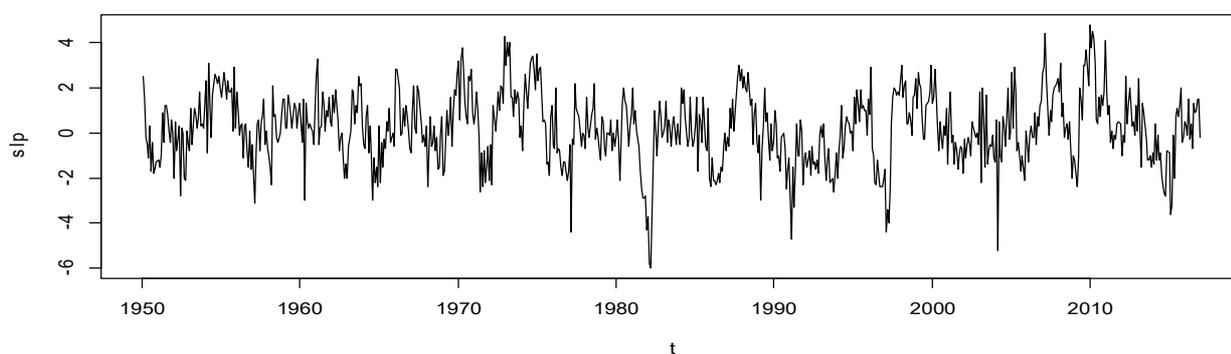


Fig. 1.8: SLP di Darwin dal 1951 al 2017 in funzione del tempo in mesi.

2. Modello matematico delle serie in funzione del tempo

La relazione tra due grandezze x e t può essere individuata per descrivere, dal punto di vista fisico-matematico, il sistema considerato⁶. Le descrizioni riguardano la poliedricità del fenomeno. Posso effettuare una descrizione formulistica, ad esempio, tramite la regressione⁷, oppure posso scomporre nelle componenti costituenti i dati sperimentali, quali il trend, la stagionalità e il rumore, che dovrebbe rappresentare l'errore commesso nell'esperimento. Il concetto di numero è alla base di ogni analisi. Tratto, in sintesi, la regressione polinomiale, armonica e il metodo SSA.

2.1 Regressione

La serie temporale $x(t)$, con $t=1, \dots, N$, può essere espressa (1):

$$x(t) = f(t) + w_t \quad (1).$$

L'espressione di w_t è il rumore:

$$w_t = x(t) - f(t).$$

La funzione $f(t)$ può essere descritta tramite (2):

$$f(t) = f_0(t) + f_1(t) + \dots + f_k(t) \quad (2)$$

ed, il modello utilizzato è detto lineare multivariato. Posso rappresentare la serie (3):

$$x_t = \beta_0 + \beta_1 x_{t,1} + \dots + \beta_k x_{t,k} + w_t \quad (3),$$

in forma matriciale: (4)

$$\begin{pmatrix} x_1 \\ \dots \\ x_N \end{pmatrix} = \begin{pmatrix} 1^0 & 1^1 & 1^2 & \dots & 1^k \\ 1^0 & 2^1 & 2^2 & \dots & 2^k \\ \dots & \dots & \dots & \dots & \dots \\ 1^0 & N^1 & N^2 & \dots & N^k \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_N \end{pmatrix}$$

$$x = X \beta + w \quad (5).$$

La soluzione è (6):

$$\hat{\beta} = (X^T X)^{-1} X^T x \quad (6).$$

6 Vedi n.12 bibliografia.

7 Vedi n.3 bibliografia.

In questo caso, la funzione $f(t)$ è un polinomio di grado, determinabile dal grafico dei dati sperimentali. La regressione, che ho utilizzato nell'elaborazione dei valori NASA è il polinomio di 3° grado. Suppongo che $f(t)=a+bt+ct^2+dt^3$. Pongo:

$$Q = \sum_{i=1}^N (x_i - (a + bt + ct^2 + dt^3))^2 \quad (7);$$

minimizzando Q (7) ricavo a, b, c e d .

Il metodo dei minimi quadrati⁸ è un'approssimazione dei coefficienti, quando non si conosce la distribuzione teorica di $x(t)$ e esso ha $x(t)$ con errori su t percentualmente molto minori degli errori su x . Nel caso in questione, la t è ipotizzata esatta. Quindi, sicuramente l'errore su x è maggiore di quello su t .

Il minimo è (8):

$$\frac{\partial Q}{\partial a} = \sum_{i=1}^N (x_i - a - bt - ct^2 - dt^3)(-2) = 0 \quad (8);$$

$$\frac{\partial Q}{\partial b} = \sum_{i=1}^N ((x_i - a - bt - ct^2 - dt^3)(2)(-t)) = 0;$$

$$\frac{\partial Q}{\partial c} = \sum_{i=1}^N ((x_i - a - bt - ct^2 - dt^3)(2)(-t^2)) = 0;$$

$$\frac{\partial Q}{\partial d} = \sum_{i=1}^N ((x_i - a - bt - ct^2 - dt^3)(2)(-t^3)) = 0;$$

quindi, (9):

$$\sum_{i=1}^N (x_i - a - bt - ct^2 - dt^3) = 0 \quad (9);$$

$$\sum_{i=1}^N (x_i t - at - bt^2 - ct^3 - dt^4) = 0;$$

$$\sum_{i=1}^N (x_i t^2 - at^2 - bt^3 - ct^4 - dt^5) = 0;$$

$$\sum_{i=1}^N (x_i t^3 - at^3 - bt^4 - ct^5 - dt^6) = 0;$$

$$\begin{pmatrix} \sum_{t=1}^N 1 & \sum_{t=1}^N t & \sum_{t=1}^N t^2 & \sum_{t=1}^N t^3 \\ \sum_{t=1}^N t & \sum_{t=1}^N t^2 & \sum_{t=1}^N t^3 & \sum_{t=1}^N t^4 \\ \sum_{t=1}^N t^2 & \sum_{t=1}^N t^3 & \sum_{t=1}^N t^4 & \sum_{t=1}^N t^5 \\ \sum_{t=1}^N t^3 & \sum_{t=1}^N t^4 & \sum_{t=1}^N t^5 & \sum_{t=1}^N t^6 \end{pmatrix} * \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = \begin{pmatrix} \sum_{t=1}^N x_t \\ \sum_{t=1}^N x_t t \\ \sum_{t=1}^N x_t t^2 \\ \sum_{t=1}^N x_t t^3 \end{pmatrix}$$

$$T * A = H \quad (11).$$

La soluzione è (12):

$$A = T^{-1} H \quad (12).$$

Nell'applicazione⁹ conviene calcolare le varie sommatorie e individuare i valori dei coefficienti.

La matrice T è simmetrica e il calcolo dell'inversa è semplificato:

$$D_{ij} = \frac{(-1)^{i+j} \det(T_{ij})}{\det(T)} ; T^{-1} = D^t \quad (13).$$

I minori T_{ij} sono uguali per gli elementi simmetrici, quindi, il calcolo si riduce a 10. La matrice trasposta D^t coincide con D. Il coefficiente calcolato è: (14)

$$\begin{aligned} a = & \frac{\sum x_t (\sum t^2 (\sum t^4 \sum t^6 - (\sum t^5)^2) - \sum t^3 (\sum t^3 \sum t^6 - \sum t^5 \sum t^4) + \sum t^4 (\sum t^3 \sum t^5 - (\sum t^4)^2))}{\det(T)} + \\ & + \frac{\sum x_t t (\sum t (\sum t^4 \sum t^6 - (\sum t^5)^2) - \sum t^2 (\sum t^3 \sum t^6 - \sum t^5 \sum t^4) + \sum t^3 (\sum t^3 \sum t^5 - (\sum t^4)^2))}{\det(T)} + \\ & + \frac{\sum x_t t^2 (\sum t (\sum t^3 \sum t^6 - \sum t^4 \sum t^5) - \sum t^2 (\sum t^2 \sum t^6 - (\sum t^4)^2) + \sum t^3 (\sum t^2 \sum t^5 - \sum t^3 \sum t^4))}{\det(T)} + \\ & + \frac{\sum x_t t^3 (\sum t (\sum t^3 \sum t^5 - (\sum t^4)^2) - \sum t^2 (\sum t^2 \sum t^5 - \sum t^4 \sum t^3) + \sum t^3 (\sum t^2 \sum t^4 - (\sum t^3)^2))}{\det(T)}. \end{aligned}$$

Il controllo della validità della regressione effettuata avviene tramite il t.test dei residui w_t della regressione, cioè pongo l'ipotesi che il valor medio dei residui uguale a zero sia accettabile, tramite l'utilizzo di una funzione di distribuzione t.student, essendo nel caso di varianza di essi non nota. La condizione verificata è quella base del metodo proposto.

9 Vedi n.11 bibliografia.

$$r = \frac{\sum_{t=1}^N (x_t - \bar{x})(t - \bar{t})}{\sqrt{\sum_{t=1}^N (x_t - \bar{x})^2 \sum_{t=1}^N (t - \bar{t})^2}} \quad (15).$$

Se $r = 0$, non c'è correlazione lineare; se $r = 1$ o -1 , allora c'è correlazione lineare e la funzione, che le descrive, è una retta con pendenza positiva o negativa a seconda del valore di r . Il coefficiente r circa 1 indica la presenza di una dipendenza lineare più altri contributi, in generale, si considera una funzione $f(t)$ polinomiale.

2.1.b Caso armonico

La funzione $g(t)$ (16) può essere di tipo trigonometrico, con frequenze f_i :

$$g(t) = \sum_{i=1}^m (a_1(f_i) \cos(2\pi t f_i) + a_2(f_i) \sin(2\pi t f_i)) \quad (16);$$

Una frequenza f (17) produce una $g(t)$:

$$\begin{aligned} g(t) &= A \cos(2\pi f t + \varphi) + c = a_1 \cos(2\pi f t) + a_2 \sin(2\pi f t) + a = \\ &= g_0(t) + g_1(t) + g_2(t) \quad (17). \end{aligned}$$

2.2 Periodogramma

La trasformata discreta di Fourier determina (18):

$$a_1(f_i) = \frac{2}{N} \sum_{t=1}^N x_t \cos(2\pi t f_i) \quad \text{e} \quad a_2(f_i) = \frac{2}{N} \sum_{t=1}^N x_t \sin(2\pi t f_i) \quad (18).$$

Definisco periodogramma¹¹ la quantità (19):

$$P_i = \frac{1}{N} \left(\sum_{t=1}^N x_t \cos(2\pi f_i t) \right)^2 + \frac{1}{N} \left(\sum_{t=1}^N x_t \sin(2\pi f_i t) \right)^2 = (|a_1(f_i)|^2 + |a_2(f_i)|^2) \frac{N}{4} \quad (19).$$

Esso rappresenta lo spettro o densità spettrale stimata della serie temporale x_t .

Le grandezze, che trattano della dipendenza tra i valori della serie nei tempi s e t , sono l'autocovarianza (20) e

l'autocorrelazione (21): $\gamma(s, t) = \text{cov}(x_s, x_t) = E[(x_s - \mu_s)(x_t - \mu_t)] \quad (20);$

che nel caso stazionario:

$$1. \quad \gamma(s, t) = E[(x_s - \mu)(x_t - \mu)] \quad ;$$

¹⁰ Vedi n.9 bibliografia.
¹¹ Vedi n.16 bibliografia.

2. dipende dalla differenza $|s-t|=h$ ed è indicata: $\gamma(s,t)=\gamma(h)$;

3. per $s=t$: $\gamma(s,t)=\text{var}(x_t)=\sigma^2$.

L'autocorrelazione è:

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}} \quad (21);$$

nel caso stazionario: $\rho(s,t) = \frac{\gamma(s,t)}{\sigma^2}$.

Il legame con la densità spettrale è la trasformata (22):

$$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i \omega h} f(\omega) d\omega \quad (22);$$

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h} .$$

La stima dell'autocovarianza (23) e autocorrelazione (24) è:

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}) \quad (23);$$

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} \quad (24).$$

Considero i coefficienti $a_1(f_i)$ e $a_2(f_i)$, li sommo e moltiplico per una costante (25):

$$(a_1(f_i) + i a_2(f_i)) \frac{\sqrt{N}}{2} = d(f_i) \quad (25);$$

dove con d intendo la trasformata di Fourier discreta (26), che è collegata alla stima dell'autocovarianza:

$$d(f_i) = \sum_{h=-(N-1)}^{N-1} \hat{\gamma}(h) e^{-2\pi i f_i h} \quad (26).$$

Il periodogramma (27) è legato all'autocovarianza:

$$P_i = |d(f_i)|^2 = \left| \sum_{h=-(N-1)}^{N-1} \hat{\gamma}(h) e^{-2\pi i f_i h} \right|^2 = \left| \sum_{h=-(N-1)}^{N-1} \hat{\gamma}(h) \cos(2\pi f_i h) \right|^2 + \left| \sum_{h=-(N-1)}^{N-1} \hat{\gamma}(h) \sin(2\pi f_i h) \right|^2 ; (27)$$

$$\hat{a}_1(f_i) = \left(\sum_{h=-(N-1)}^{N-1} \hat{y}(h) \cos(2\pi f_i t) \right) \frac{\sqrt{N}}{2} \quad (28);$$

$$\hat{a}_2(f_i) = \left(\sum_{h=-(N-1)}^{N-1} \hat{y}(h) \sin(2\pi f_i t) \right) \frac{\sqrt{N}}{2} \quad (29).$$

Nelle applicazioni, ho ricavato le frequenze f_i dal periodogramma della grandezza in funzione di esse, quindi, ho effettuato la regressione con il seno e coseno e ho trovato i coefficienti a_j . Il trend individuato segue l'andamento dei dati in modo migliore all'aumentare del numero delle frequenze. Tuttavia, il segnale è contenuto in una parte dei dati, mentre c'è anche la stagionalità e il rumore, contenuto in essi. È importante, quindi, considerare le frequenze opportune per non introdurre il rumore nel trend. La misura della predicibilità di una serie y_s da una serie x_t è effettuata tramite la covarianza (30) e la correlazione (31):

$$\gamma_{xy}(s, t) = \text{cov}(x_s, y_t) = E[(x_s - \mu_x)(y_t - \mu_y)] \quad (30);$$

$$\rho_{xy}(s, t) = \frac{\gamma_{xy}(s, t)}{\sqrt{\gamma_x(s, s) \gamma_y(t, t)}} \quad (31).$$

SSA è Singular Spectrum Analysis¹², esso è un metodo di regressione, che permette di analizzare le caratteristiche del sistema fisico studiato, nel caso in particolare la temperatura, la pressione e gli indici che scaturiscono.

La serie temporale¹³ $x(t)$ con $t=1, \dots, N$ e $x(t)$ reale è ricavata sperimentalmente o, in seguito, ad elaborazione dai dati sperimentali, considero il metodo SSA per studiarla.

1. Costruzione della matrice traiettoria X

Fisso una finestra temporale L e considero le righe della matrice costituite dai vettori $v_1=(x_1, \dots, x_L)$, il vettore $v_2=(x_2, \dots, x_{L+1})$ ecc. $v_{N-L+1}=(x_{N-L+1}, \dots, x_N)$. Definisco la matrice detta traiettoria X , con $M = N-L+1$: (32)

$$X = \frac{1}{\sqrt{M}} \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_{N-L+1} \end{pmatrix} = \frac{1}{\sqrt{M}} \begin{pmatrix} x_1, \dots, x_L \\ x_2, \dots, x_{L+1} \\ \dots \\ x_{N-L+1}, \dots, x_N \end{pmatrix}$$

Gli elementi della X sono legati dalla relazione: $X_{ij} = x_{i+j-1}$ con $1 \leq i \leq L$ e $1 \leq j \leq N-L+1$.

2. Matrice “lagged-covarianza” S

La matrice S è (33):

$$S = X^T X \quad (33);$$

Essa è simmetrica e è una rappresentazione matriciale della covarianza.

3. Diagonalizzazione di S

Essendo la matrice S reale e simmetrica posso scrivere (34): $S = E \Lambda E^T$ (34);

dove Λ è diagonale. Essa è detta decomposizione spettrale. Gli elementi diagonali di Λ sono gli autovalori di S , soluzioni dell'equazione $\det(S - \lambda I) = 0$. I λ_i sono L , quanto la dimensione della matrice quadrata S : $L \times L$ e quanto la finestra temporale scelta. La matrice E è costituita dagli autovettori e_i .

Gli autovalori di S sono il quadrato degli autovalori generalizzati di X , matrice traiettoria della serie temporale. Ciò si dimostra considerando: $X = D H E^T$ con H diagonale. S è data da (35):

$$S = (D H E^T)^T (D H E^T) = E H D^T D H E^T = E H^2 E^T: H^2 = \Lambda \text{ cvd (35).}$$

12 Vedi n.10 bibliografia.
13 Vedi n.17 bibliografia.

Considero $A=XE$ e nelle componenti (36):

$$a_i^k = \sum_{j=1}^L x_{i+j-1} e_j^k \quad (36).$$

Ricavo (37):

$$x_{i+j-1} = \sum_{k=1}^L a_i^k e_j^k \quad (37).$$

Individuo gli elementi costituenti la serie, potendo selezionare gli indici interessanti degli autovalori.

Esempio

Considero la serie temporale: $x(t) = (-1.6, 1.0, 0.1, 1.2, -1.4, 0.6, 0.2, 0.9, -0.3, 1.6)$

con $t=(20,40,60,80,100,120,140,160,180,200)$ indicati $t=1, \dots, 10$.

Calcolo X e S:

$L=3$;

$v1 = (x1, \dots, x3) = (-1.6, 1.0, 0.1)$

$v2 = (x2, \dots, x4) = (1.0, 0.1, 1.2)$

$v3 = (x3, \dots, x5) = (0.1, 1.2, -1.4)$

$v4 = (x4, \dots, x6) = (1.2, -1.4, 0.6)$

$v5 = (x5, \dots, x7) = (-1.4, 0.6, 0.2)$

$v6 = (x6, \dots, x8) = (0.6, 0.2, 0.9)$

$v7 = (x7, \dots, x9) = (0.2, 0.9, -0.3)$

$v8 = (x8, \dots, x10) = (0.9, -0.3, 1.6)$ $M=10-3+1=8$

$(3 \times 3) = (3 \times 8)(8 \times 3)$

$$X = \frac{1}{\sqrt{8}} \begin{pmatrix} v1 \\ v2 \\ v3 \\ v4 \\ v5 \\ v6 \\ v7 \\ v8 \end{pmatrix} = \begin{pmatrix} -1.6, 1.0, 0.1 \\ 1.0, 0.1, 1.2 \\ 0.1, 1.2, -1.4 \\ 1.2, -1.4, 0.6 \\ -1.4, 0.6, 0.2 \\ 0.6, 0.2, 0.9 \\ 0.2, 0.9, -0.3 \\ 0.9, -0.3, 1.6 \end{pmatrix} \frac{1}{\sqrt{8}}$$

$$S = X^T X = \frac{1}{8} \begin{pmatrix} -1.6 & 1.0 & 0.1 & 1.2 & -1.4 & 0.6 & 0.2 & 0.9 \\ 1.0 & 0.1 & 1.2 & -1.4 & 0.6 & 0.2 & 0.9 & -0.3 \\ 0.1 & 1.2 & -1.4 & 0.6 & 0.2 & 0.9 & -0.3 & 1.6 \end{pmatrix} \begin{pmatrix} -1.6 & 1.0 & 0.1 \\ 1.0 & 0.1 & 1.2 \\ 0.1 & 1.2 & -1.4 \\ 1.2 & -1.4 & 0.6 \\ -1.4 & 0.6 & 0.2 \\ 0.6 & 0.2 & 0.9 \\ 0.2 & 0.9 & -0.3 \\ 0.9 & -0.3 & 1.6 \end{pmatrix} = \begin{pmatrix} 1.0225 & -0.48375 & 0.4075 \\ -0.48375 & 0.71375 & -0.34375 \\ 0.4075 & -0.34375 & 0.90875 \end{pmatrix}$$

Determino gli autovalori e gli autovettori di S:

pag.14

$\det(S-\lambda I) = 0$; $\lambda_1 = 1.7234927$; $\lambda_2 = 0.5661709$; $\lambda_3 = 0.3553363$;

```
> a=eigen(S)
> a$val
[1] 1.7234927 0.5661709 0.3553363
```

```
> E=a$vec
> E
      [,1] [,2] [,3]
[1,] 0.6670381 -0.5263764 0.5272457
[2,] -0.5057914 0.1996649 0.8392312
[3,] 0.5470240 0.8264755 0.1330525
```

Verifico l'esattezza di quanto effettuato, moltiplicando la matrice E per Λ diagonale e la trasposta di E:

$E\Lambda E^T =$

$$\begin{pmatrix} 0.6670381 & -0.5263764 & 0.5272457 \\ -0.5057914 & 0.1996649 & 0.8392312 \\ 0.5470240 & 0.8264755 & 0.1330525 \end{pmatrix} * \begin{pmatrix} 1.723493 & 0.0000000 & 0.0000000 \\ 0.0000000 & 0.5661709 & 0.0000000 \\ 0.0000000 & 0.0000000 & 0.3553363 \end{pmatrix} * \begin{pmatrix} 0.6670381 & -0.5057914 & 0.5470240 \\ -0.5263764 & 0.1996649 & 0.8264755 \\ 0.5272457 & 0.8392312 & 0.1330525 \end{pmatrix} =$$

$$= \begin{pmatrix} 1.02250 & -0.48375 & 0.40750 \\ -0.48375 & 0.71375 & -0.34375 \\ 0.40750 & -0.34375 & 0.90875 \end{pmatrix} \text{ cvd.}$$

Calcolo i coefficienti a_i^k :

$A = X * E =$

$$\begin{pmatrix} -1.51835 & 1.1245147 & 0.0089433 \\ 1.2728878 & 0.4853607 & 0.7708318 \\ -1.3060795 & -0.9701055 & 0.8735285 \\ 1.8367681 & -0.4152972 & -0.4623973 \\ -1.1279234 & 1.022021 & -0.2079948 \\ 0.7913862 & 0.4679351 & 0.6039409 \\ -0.4859118 & -0.1735195 & 0.8208415 \\ 1.6273101 & 0.7887226 & 0.4356358 \end{pmatrix} * \frac{1}{\sqrt{8}}$$

Ricostruisco la matrice traiettoria della serie temporale e, quindi, gli elementi della serie di partenza:

$X_r = A * E^T =$

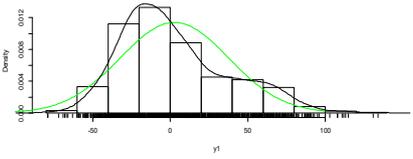
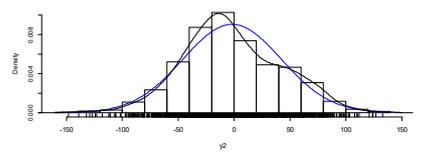
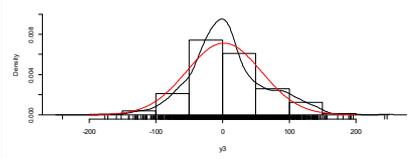
$$\begin{pmatrix} -1.5999999 & 1.0000000 & 0.0999999 \\ 1.0000000 & 0.0999999 & 1.2 \\ 0.1 & 1.2000001 & -1.4000001 \\ 1.1999999 & -1.4000001 & 0.6000001 \\ -1.3999999 & 0.6000000 & 0.1999999 \\ 0.6000000 & 0.1999999 & 0.9000000 \\ 0.2000000 & 0.9000000 & -0.3000001 \\ 0.9000000 & -0.3000001 & 1.6 \end{pmatrix} * \frac{1}{\sqrt{8}}$$

La serie viene ricostruita fedelmente, e, quindi, questo metodo si dimostra molto preciso nell'utilizzazione e, soprattutto, nella scomposizione delle parti di una serie temporale.

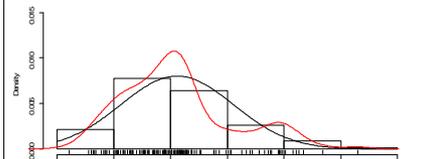
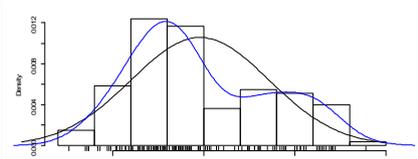
Le serie temporali sono analizzate da un punto di vista statistico¹⁴, tab. 3.1, 3.2 e 3.3, quindi, utilizzo gli strumenti matematici caratteristici in esse, quali intervallo di confidenza e funzione di distribuzione. Essendo il numero di dati piccolo rispetto all'intervallo temporale, occorre considerare come intervallo di confidenza, indicato nella guida del WMO¹⁵ (World Meteorologic Organization), che è (38):

$$x_m \pm t \frac{\sigma}{\sqrt{N}} \quad (38);$$

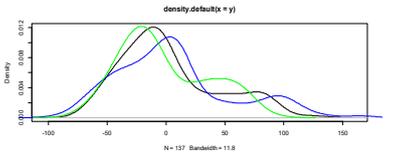
dove la t rappresenta il valore della t.student.

N=1656 GLOTI=(2.72± 0.01)°C	SHSTI=(-1.71 ± 0.10)° C	NHSTI=(6.84± 0.06)°C
 <p>Fig. 3.1: Densità di probabilità della variabile GLOTI, rapportata alla curva normale in verde. Sono indicati i punti sperimentali sull'asse delle ascisse.</p>	 <p>Fig. 3.2: Densità di probabilità della variabile SHSTI, rapportata alla curva normale in blu. Sono indicati i punti sperimentali sull'asse delle ascisse</p>	 <p>Fig. 3.3: Densità di probabilità della variabile NHSTI, rapportata alla curva normale in rosso. Sono indicati i punti sperimentali sull'asse delle ascisse</p>
<pre>> shapiro.test(GLOTI)</pre> <p><i>Shapiro-Wilk normality test</i></p> <p>data: GLOTI</p> <p>W = 0.95748, p-value < 2.2e-16</p> <p>La variabile non è gaussiana.</p>	<pre>> shapiro.test(SHSTI)</pre> <p><i>Shapiro-Wilk normality test</i></p> <p>data: SHSTI</p> <p>W = 0.99289, p-value = 3.657e-07</p> <p>La variabile non è gaussiana.</p>	<pre>> shapiro.test(NHSTI)</pre> <p><i>Shapiro-Wilk normality test</i></p> <p>data: NHSTI</p> <p>W = 0.9794, p-value = 1.094e-14</p> <p>La variabile non è gaussiana.</p>

Tab. 3.1: Variabili GLOTI, SHSTI e NHSTI sono analizzate statisticamente tramite valor medio, errore, gaussianità e funzione di distribuzione.

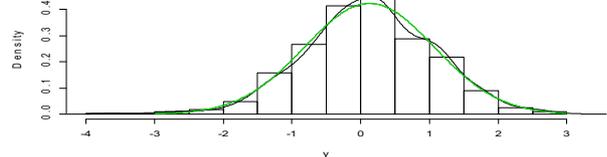
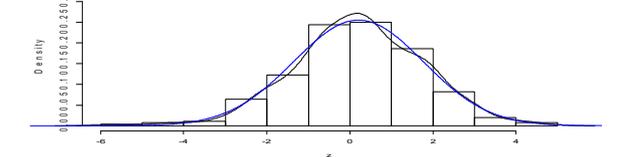
N=137 STIAG = (2.12±0.14)°C	STIANH = (5.82±0.01)°C	STIASH = (-2.03±0.16)°C
 <p>Fig. 3.4: Densità di probabilità della variabile STIAG in verde, rapportata alla curva normale in nero. Sono indicati i punti sperimentale sull'asse delle ascisse</p>	 <p>Fig. 3.5: Densità di probabilità della variabile STIANH in rosso, rapportata alla curva normale in nero. Sono indicati i punti sperimentale sull'asse delle ascisse</p>	 <p>Fig. 3.6: Densità di probabilità della variabile STIASH in blu, rapportata alla curva normale in nero. Sono indicati i punti sperimentale sull'asse delle ascisse</p>
<pre>> shapiro.test(STIAG)</pre> <p><i>Shapiro-Wilk normality test</i></p> <p>data: STIAG</p> <p>W = 0.93561, p-value = 6.237e-06</p>	<pre>> shapiro.test(STIANH)</pre> <p><i>Shapiro-Wilk normality test</i></p> <p>data: STIANH</p> <p>W = 0.94513, p-value = 3.064e-05</p>	<pre>> shapiro.test(STIASH)</pre> <p><i>Shapiro-Wilk normality test</i></p> <p>data: STIASH</p> <p>W = 0.94751, p-value = 4.655e-05</p>

14. Vedi n.2 bibliografia.
15. Vedi n.18 bibliografia.

<p>Le variabili non sono gaussiane, come evidente nel test d'ipotesi.</p>	 <p>Fig. 3.7: Rappresento le funzioni di distribuzione delle 3 variabili ed ottengo in blu density (z=STIASH), in verde la density (x=STIANH), in nero la STIAG, che sembra la somma delle altre due.</p>	<p>Il grafico riporta le distribuzioni tra loro ed evidenzia un andamento simile che è dovuto alle variazioni del flusso di calore, di conseguenza la non gaussianità delle variabili è spiegabile dal punto di vista sperimentale.</p>
---	---	--

Tab. 3.2: Le variabili STIAG, STIANH e STIASH sono studiate statisticamente, e il rapporto tra loro.

Le variabili in media annuale hanno delle distribuzioni più vicine all'andamento gaussiano, come evidenzia il test di normalità, in quanto, sono medie di dati mensili e, quindi, più prossime all'astrazione fisica del significato della grandezza. Le variazioni sperimentali delle grandezze sono evidenti in funzioni di distribuzione non normali, ma di strane forme.

<p>N=804 SOI=(0.13±0.01) hPa</p>	<p>SLP=(0.20±0.02) hPa</p>
 <p>Fig. 3.8: Densità di probabilità della variabile SOI, rapportata alla curva normale in verde.</p>	 <p>Fig. 3.9: Densità di probabilità della variabile SLP, rapportata alla curva normale in blu.</p>
<pre>> shapiro.test(SOI) Shapiro-Wilk normality test data: SOI W = 0.99347, p-value = 0.001419 La variabile SOI non è gaussiana.</pre>	<pre>> shapiro.test(SLP) Shapiro-Wilk normality test data: SLP W = 0.99408, p-value = 0.003014 La variabile SLP non è gaussiana.</pre>

Tab. 3.3: Le variabili SOI e SLP sono studiate e evidenziano una non gaussianità, ma con un p poco minore di 0.05.

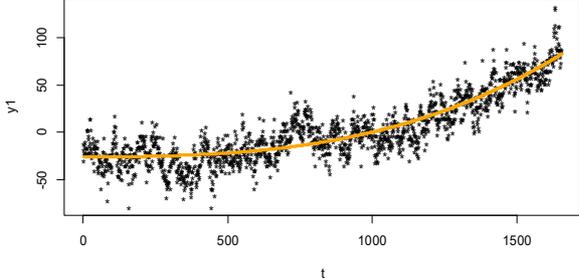
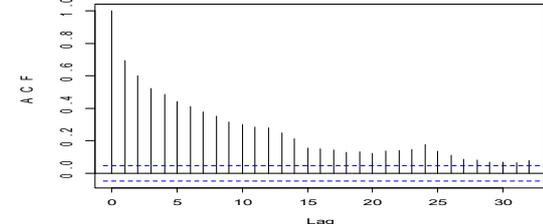
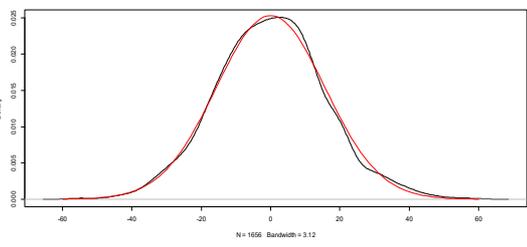
Gli indici di pressione sono ancora più vicini all'andamento¹⁶ auspicato, in quanto la pressione atmosferica risente meno delle variazioni sperimentali, ma è più costante nel tempo.

La correlazione tra gli indici di temperatura mensili:

```
> cor(y1,y2) [1] 0.8418985
> cor(y1,y3) [1] 0.9152682
> cor(y2,y3) [1] 0.7215055
```

mostra che c'è una dipendenza lineare forte tra le serie. Nelle trattazioni delle variabili considero il tempo in mesi.

Considero la regressione polinomiale di 3° grado, che dovrebbe descrivere l'andamento della variabile nel tempo tab.3.1.a. Innanzitutto, osservo che la superficie della Terra¹⁷ è di 510 milioni di km². L'emisfero nord ha 100 milioni di km² di continenti e 154 milioni di km² di mari e oceani; nell'emisfero sud, la terra ferma è di 48 milioni di km², mentre l'acqua è di circa 200 milioni di km². L'aria ricopre l'intero pianeta. Dell'energia solare in arrivo sulla Terra circa 340 W/m², il calore irradiato dal suolo e dalle acque è di 238 W/m². Le caratteristiche delle variabili sono legate proprio a questi elementi e ai flussi di calore in essi, prodotti dall'effetto di correnti marine e oceaniche, venti, onde, maree, movimento del Sole, della Luna, dei pianeti ecc. e loro variazioni. La GLOTI (y₁) comprende le caratteristiche di temperatura complessive di tutta la Terra ed ha, quindi, un comportamento legato alla "somma" dei comportamenti delle due variabili dell'indice di temperatura Nord e Sud.

<p>Regressione polinomiale di 3° grado</p>	<p>L'acf e il t.test evidenziano le caratteristiche della regressione effettuata.</p>
 <p>Fig. 3.1.1: Rappresentazione dei dati e della regressione polinomiale in arancio.</p>	<pre>> t.test(r1) One Sample t-test data: r1 t = 0.0072757, df = 1655, p-value = 0.9942 alternative hypothesis: true mean is not equal to 0 95 percent confidence interval: -0.7565521 0.7621857 sample estimates: mean of x 0.002816833</pre> <p>Il t.test evidenzia che il valor medio dei residui è nullo.</p>
<pre>> l=lm(y1~t+t^2+t^3) > l lm(formula = y1 ~ t + t^2 + t^3) Coefficients: (Intercept) t t^2 t^3 -2.582e+01 3.911e-04 4.933e-06 2.085e-08</pre>	 <p>Fig. 3.1.2: L'acf dei residui.</p>
<pre>> shapiro.test(r1) Shapiro-Wilk normality test data: r1 W = 0.99811, p-value = 0.054</pre> <p>I residui della regressione polinomiale sono gaussiani.</p>	 <p>Fig. 3.1.3: Densità di probabilità di residui della regressione polinomiale, rapportata alla normale in rosso.</p>

Tab. 3.1.a: La regressione polinomiale della serie y1.

Il valor medio dei residui è nullo. L'acf dei residui presenta delle oscillazioni studiabili con il metodo SSA, proposto nel par. 2.3. La figura 3.1.4 evidenzia lo spettro dei residui e le caratteristiche di esso. Sono presenti due picchi essenziali, uno intorno allo zero e un'altro a circa 0.08 (1/12, frequenza annuale). Le frequenze rilevanti si raggruppano in questo intervallo, infatti le periodicità sono maggiori di 1 anno. Il resto è considerabile come rumore¹⁸. Distinguere il rumore dal segnale è importante, in quanto le informazioni sono contenute nel segnale e il rumore potrebbe far individuare elementi sbagliati. Ogni variabile ha uno spettro diverso, perché ha caratteristiche armoniche diverse. Le periodicità sono contenute in esso.

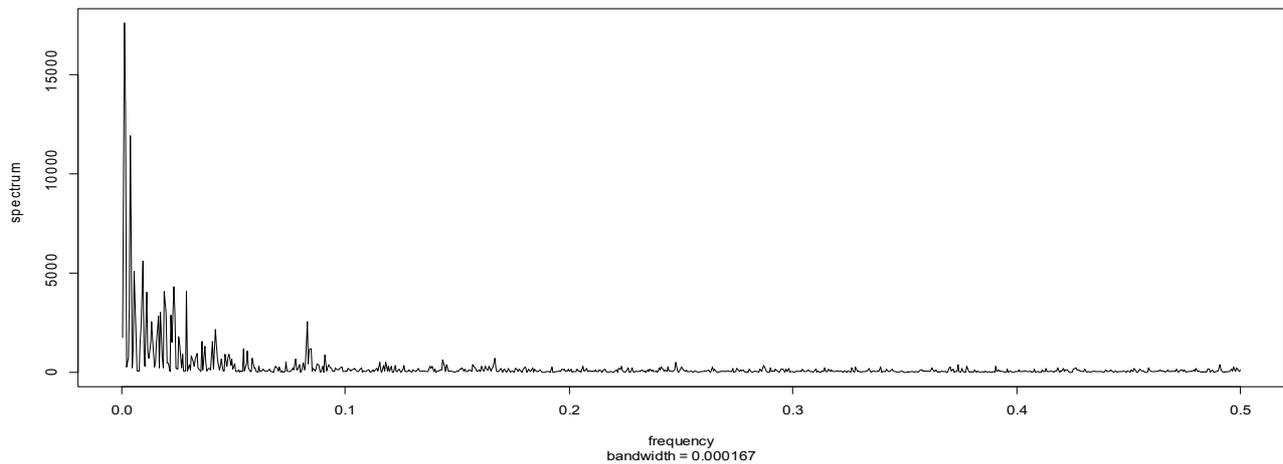


Fig. 3.1.4: Lo spettro dei residui della regressione polinomiale.

3.1.a Metodo SSA

```
> s1 = ssa(y1);
```

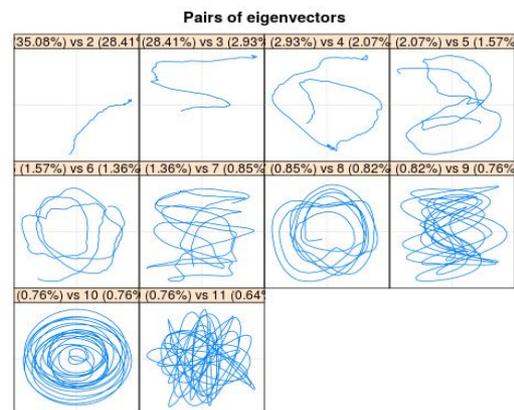
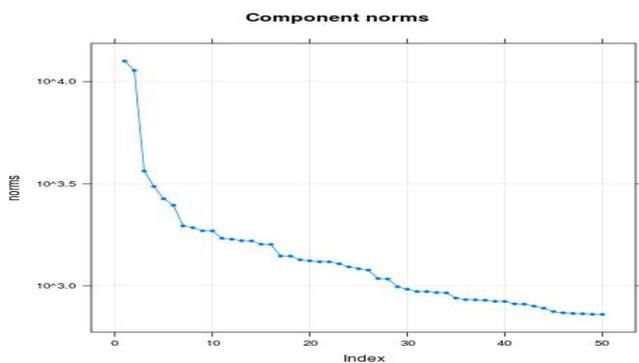


Fig. 3.1.5: Gli autovalori e le coppie di autovettori della serie y1.

```
> r1 = reconstruct(s1, groups=list(trend=c(1:6), sea1=c(7:8), sea2=c(9:10), rumore=c(11:50))
```

```
> trend=r1$trend > sea1=r1$sea1 > sea2=r1$sea2 > rumore=r1$rumore
```

```
> h=trend+sea1+sea2+rumore
```

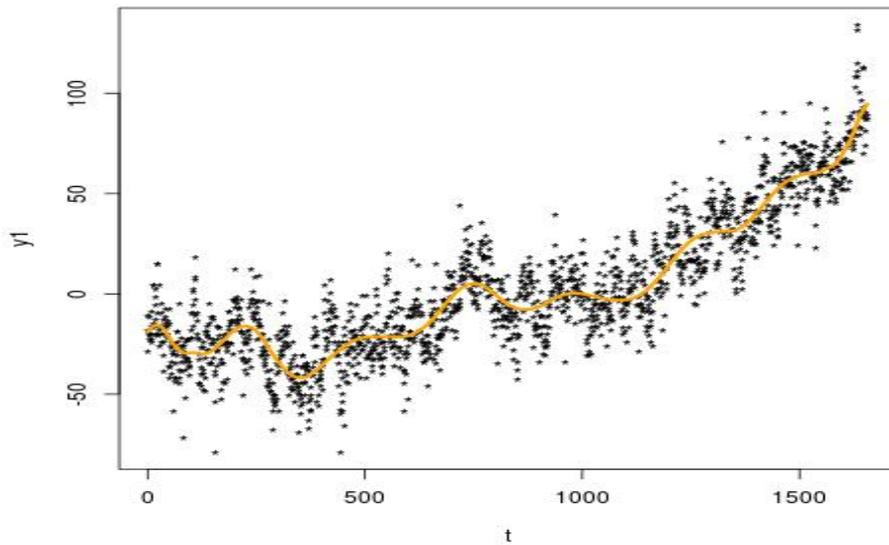


Fig. 3.1.6: Il trend segue l'andamento dei dati sperimentali, tramite 6 oscillazioni.

Rappresento il rumore e le stagionalità, fig. 3.1.7:

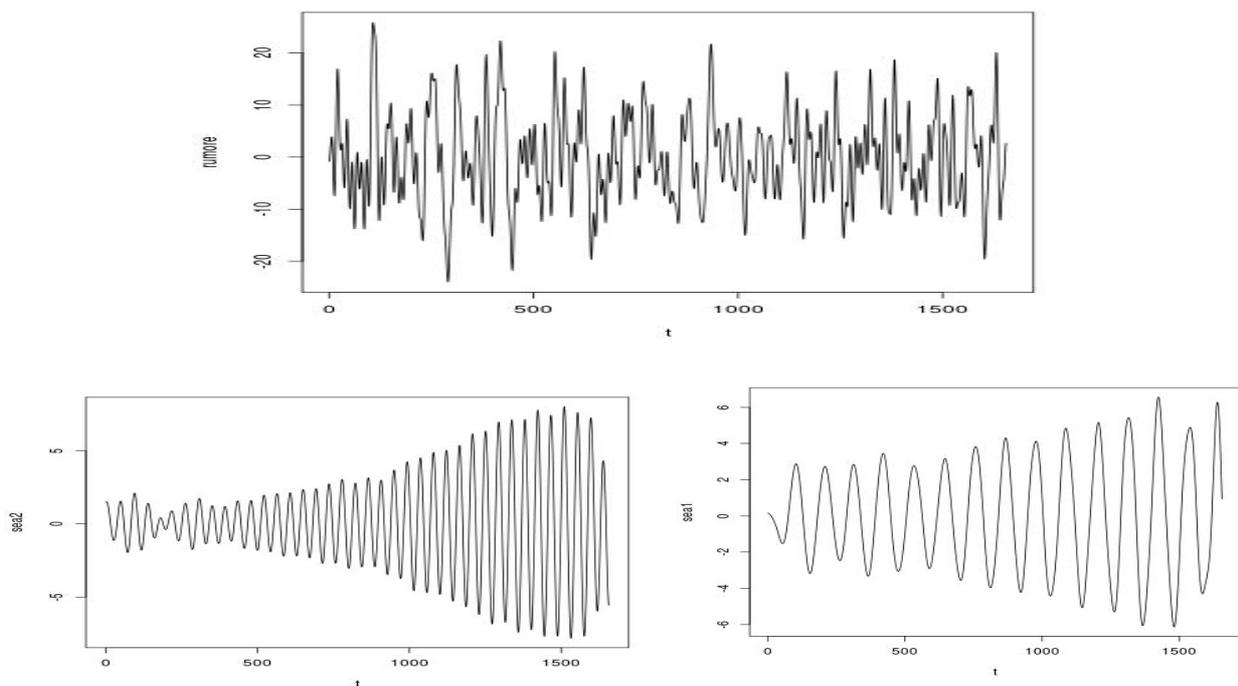


Fig. 3.1.7: I grafici in ordine orario rappresentano il rumore e le componenti della stagionalità.

Il trend produce una regressione, che segue le varie oscillazioni dei valori sperimentali. Le componenti della periodicità manifestano caratteristiche di vario genere: diurna, notturna, invernale, primaverile, estiva, autunnale, annuale, dei venti, delle correnti d'acqua salata e dolce ecc.. Il rumore potrebbe essere analizzato nelle componenti, ma è legato alle variazioni di calore, quindi lo definirei un rumore essenzialmente termico.

19 Vedi n.1 bibliografia.

Analizzo i residui, fig. 3.1.8:

```
> h=trend+sea1+sea2+rumore
```

```
> res1=y1-h
```

```
> t.test(res1) One Sample t-test
```

```
data: res1
```

```
t = 0.060215, df = 1655, p-value = 0.952
```

```
alternative hypothesis: true mean is not equal to 0
```

```
95 percent confidence interval:
```

```
-0.4484841 0.4768935
```

```
sample estimates:
```

```
mean of x
```

```
0.01420467
```

```
> shapiro.test(res1) Shapiro-Wilk normality test
```

```
data: res1
```

```
W = 0.99608, p-value = 0.0002861
```

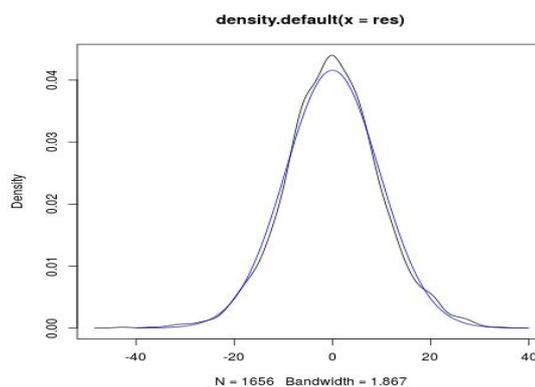
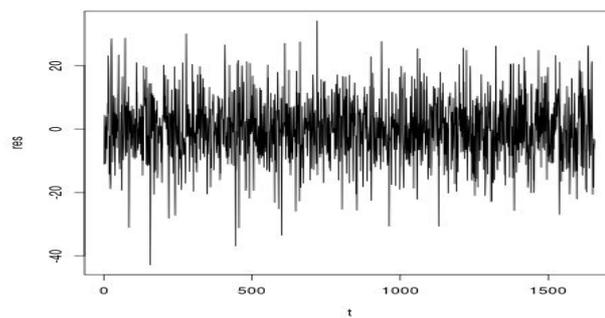


Fig. 3.1.8: I residui e la funzione di distribuzione confrontata con la curva normale.

I residui hanno valor medio zero e non sono gaussiani. L'acf è buono, fig.3.1.9. La regressione è buona.

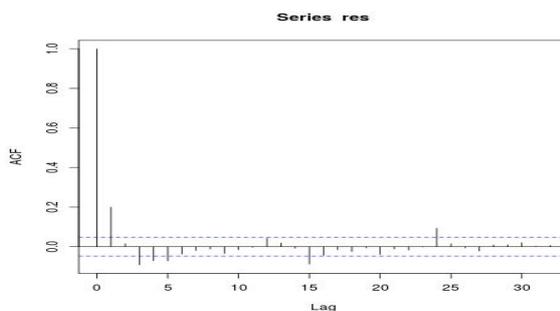
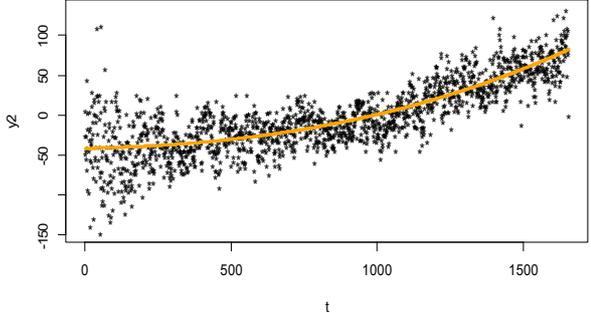
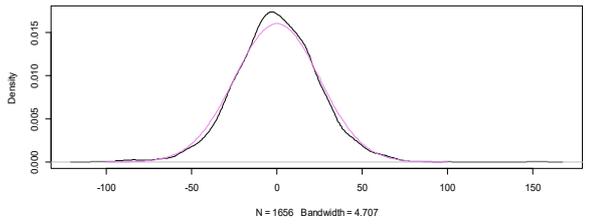
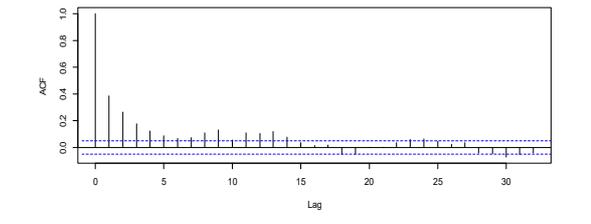


Fig. 3.1.9: L'acf dei residui, che resta abbastanza nei limiti dell'intervallo blu di valori.

3.2. La variabile Sud Hemisph. Station Temperature Index (SHSTI)

La SHSTI (y_2) è studiata nella tab. 3.2.a, essa evidenzia un andamento parabolico, ma per maggiore precisione è utilizzata la regressione di 3° grado.

<p>Regressione polinomiale di 3° grado</p> <pre>> l2=lm(y2~t+t^2+t^3)</pre> <pre>> l2 lm(formula = y2 ~ t + t^2 + t^3)</pre> <p>Coefficients:</p> <table border="1"> <thead> <tr> <th>(Intercept)</th> <th>t</th> <th>t²</th> <th>t³</th> </tr> </thead> <tbody> <tr> <td>-4.175e+01</td> <td>6.811e-03</td> <td>2.713e-05</td> <td>8.420e-09</td> </tr> </tbody> </table>	(Intercept)	t	t ²	t ³	-4.175e+01	6.811e-03	2.713e-05	8.420e-09	<p>Considero la regressione polinomiale perché l'andamento è essenzialmente parabolico.</p>  <p>Fig. 3.2.1: Regressione polinomiale di 3° grado sui dati sperimentali.</p>
(Intercept)	t	t ²	t ³						
-4.175e+01	6.811e-03	2.713e-05	8.420e-09						
<pre>> t.test(r2) One Sample t-test</pre> <p>data: r2</p> <p>$t = 0.008866$, $df = 1655$, $p\text{-value} = 0.992$</p> <p>alternative hypothesis: true mean is not equal to 0</p> <p>95 percent confidence interval: -1.194406 1.205253</p> <p>sample estimates: mean of x 0.005423509</p> <p>Il valor medio dei residui è zero.</p>	 <p>Fig. 3.2.2: La funzione di distribuzione non è gaussiana, infatti rapportata alla curva normale in violetto osservo poca coincidenza.</p>								
<pre>> shapiro.test(r2)</pre> <p>Shapiro-Wilk normality test</p> <p>data: r2</p> <p>$W = 0.98348$, $p\text{-value} = 7.173e-13$</p> <p>I residui non sono gaussiani.</p>	 <p>Fig. 3.2.3: L'acf dei residui è abbastanza buono.</p>								

Tab. 3.2.a: Regressione parabolica della variabile y_2 .

Lo spettro di y_2 e dei residui sono simili, solo che i picchi sono più evidenti nello spettro di questi. Analizzo lo spettro dei residui della regressione polinomiale. Esso ha evidenti frequenze su tutto l'intervallo (0,0.5), che sono dette frequenze di Rayleigh e sono date da $f_i=i*1/\Delta t$ con $1/\Delta t=0.001$, che è la risoluzione delle frequenze, $i=1...M$. Nell'emisfero Sud ci sono delle periodicità, legate alle correnti d'acqua marina, maree e ai venti, che non risentono soltanto dell'andamento annuale o di più anni, quindi, sono presenti frequenze anche oltre 0.1, cioè con periodi più piccoli di 12 mesi. Tuttavia, i picchi di esse non sono particolarmente elevati, fig. 3.2.4.

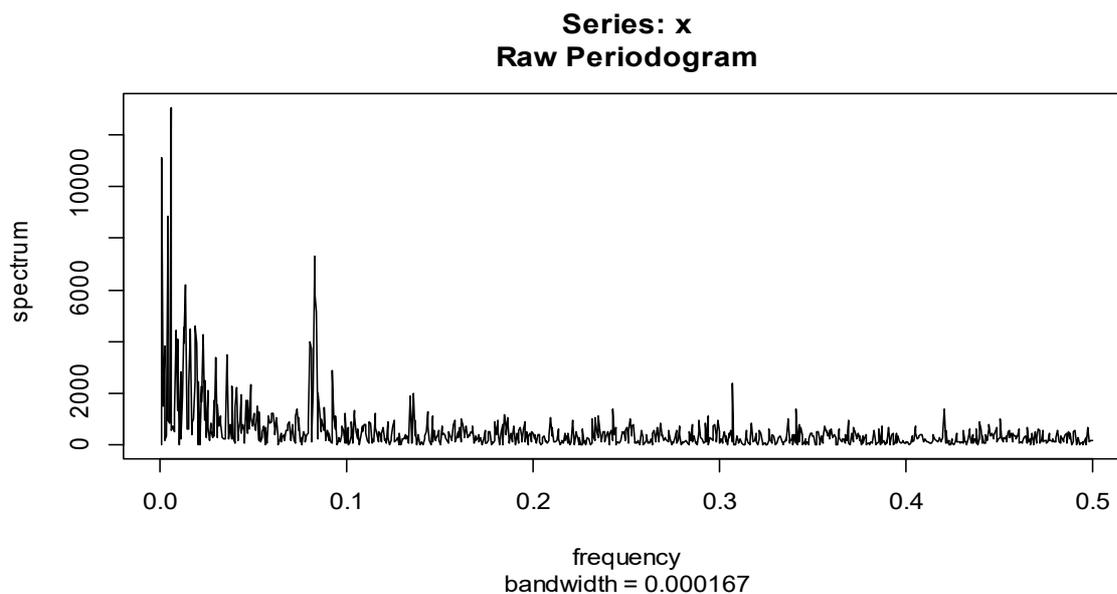


Fig. 3.2.4: Lo spettro dei residui della regressione polinomiale di 3° grado. Esso è definibile spettro continuo.

Considero le frequenze che hanno picco al di sopra di 5000; esse sono limitate all'intervallo (0,0.1). Lo spettro dei residui individua le seguenti:

$$f_1=0.001; f_2=0.004; f_3=0.006; f_4=0.014; f_5=0.083.$$

Le frequenze sono concentrate intorno ad uno o più picchi nell'intervallo (0,0.02) e intorno a 0.08. Il picco più alto $f_5=0.0829$, che è la frequenza annuale. I periodi corrispondenti sono:

$$\tau_1=83,3 \text{ anni}; \tau_2=20,83 \text{ anni}; \tau_3=13,9 \text{ anni}; \tau_4=6 \text{ anni}.$$

Le periodicità legate all'emisfero Sud²⁰ sono molto più variabili, tali da coprire tutto lo spettro, in quanto la grande componente di acqua produce variazioni molto maggiori. Il segnale risulta più rumoroso.

3.2.a Metodo SSA

>s2=ssa(y2)

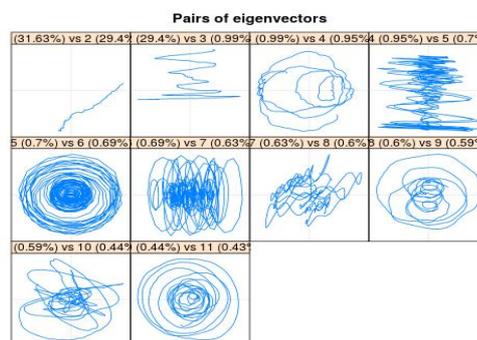
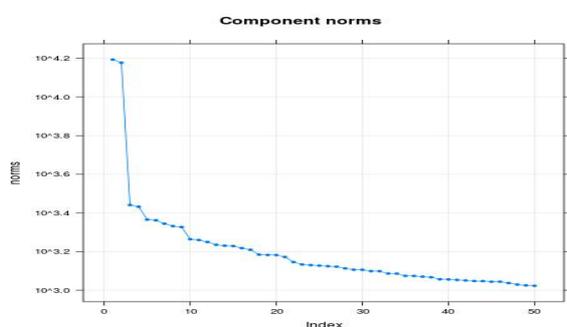


Fig. 3.2.5: Gli autovalori e le coppie di autovettori della serie y_2 .

```
>r2=reconstruct(s2,groups=list(trend1=c(1:4), trend2=c(7), sea1=c(5:6),sea2=c(8:9), rumore=c(10:50)))
```

Il trend1 e i dati sperimentali sono, fig.3.2.6:

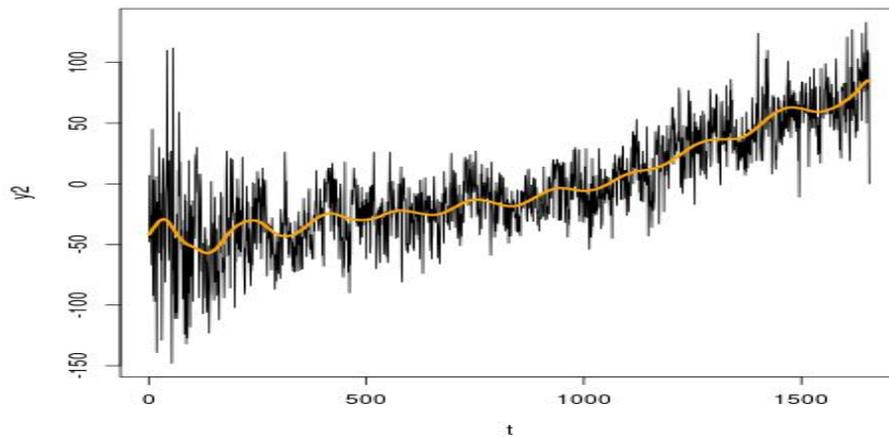


Fig. 3.2.6: Il trend individua le oscillazioni dei dati sperimentali, ma è costruito in modo da tralasciare il rumore notevolmente presente in queste misure.

La prima componente del trend presenta delle oscillazioni, circa decennali, ed è crescente. La seconda componente, fig. 3.2.4, del trend contiene una parte molto oscillante, quasi a sembrare del rumore, e una parte oscillante con periodicità di circa 50 anni. La stagionalità ha andamento a spirale, fig. 3.2.7.

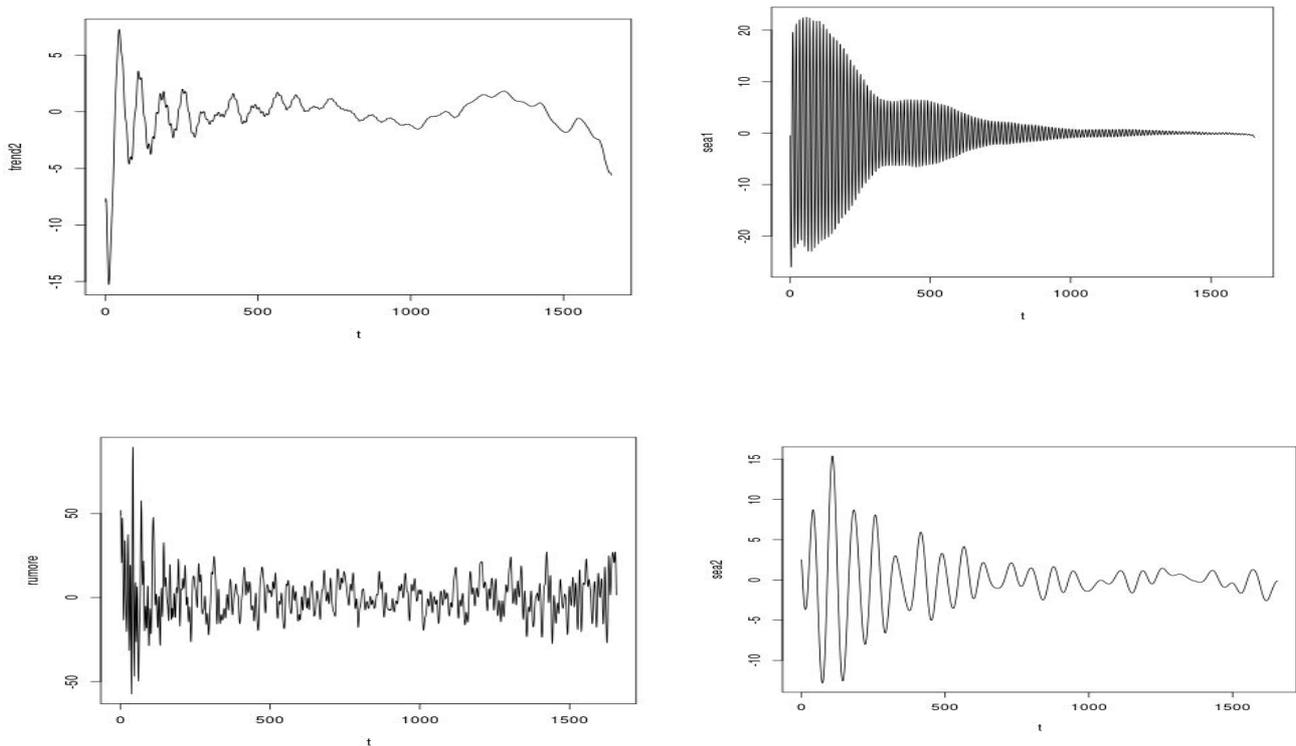


Fig. 3.2.7: I quattro grafici rappresentano: la seconda componente del trend, le due componenti di stagionalità e il rumore.

Gli autovalori, fig. 3.2.5, non mostrano la natura effettiva della componente, occorre analizzare il significato grafico²¹. Infatti, questa variabile ha elementi di rumore nei primi autovalori dello spettro. L'emisfero Sud presenta delle variazioni di calore²² molto maggiori per le caratteristiche dell'acqua marina, ma soprattutto per la presenza di un gran numero di correnti con periodicità proprie, variazioni di percorso e di temperatura. Inoltre le maree, legate alla Luna, hanno un'influenza rilevante.

Analizzo i residui, fig. 3.2.8:

```
> res2=y2-(trend+trend2+sea1+sea2+rumore)
> t.test(res2) One Sample t-test
data: res2
t = -0.11012, df = 1655, p-value = 0.9123
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-0.9374980 0.8378281
```

sample estimates:

mean of x

-0.04983494

```
> shapiro.test(res2) Shapiro-Wilk normality test
```

data: res2

W = 0.98196, p-value = 1.4e-13

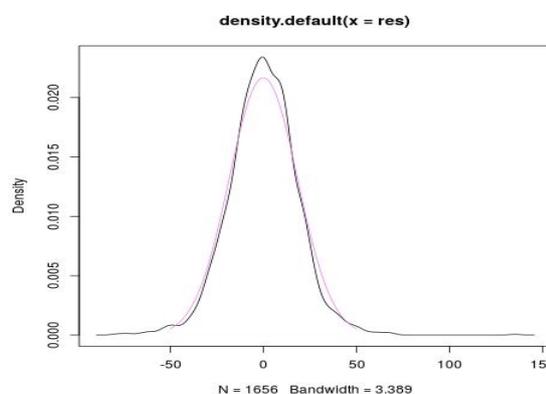
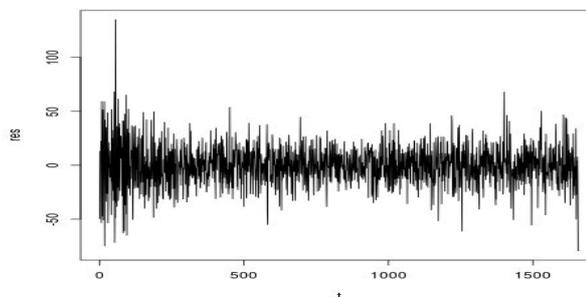


Fig. 3.2.8: I residui e la funzione di distribuzione confrontata con la gaussiana in violetto. Noto che la normale ha un picco meno alto della funzione di distribuzione dei residui.

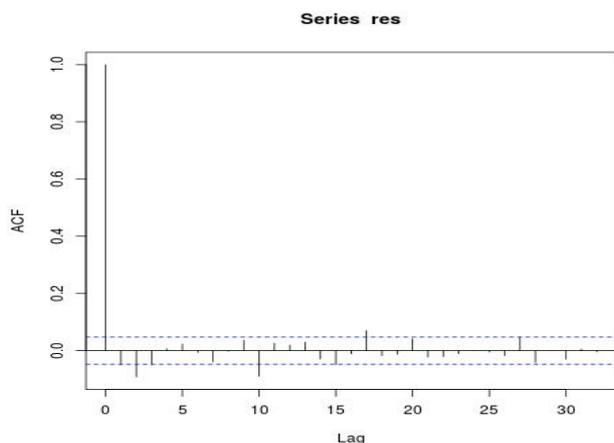
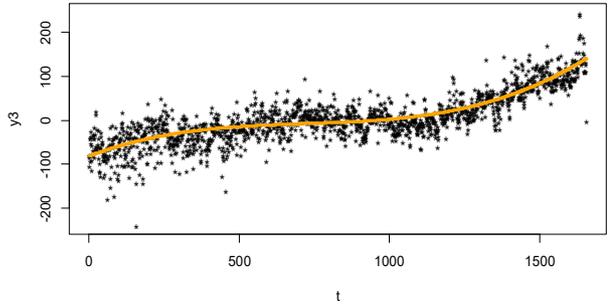
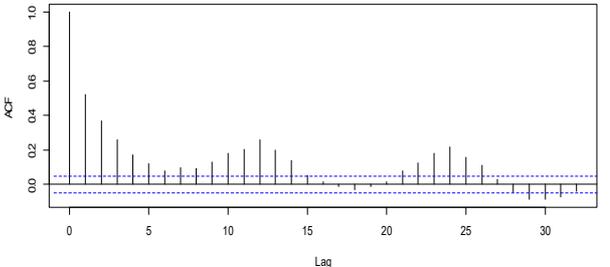
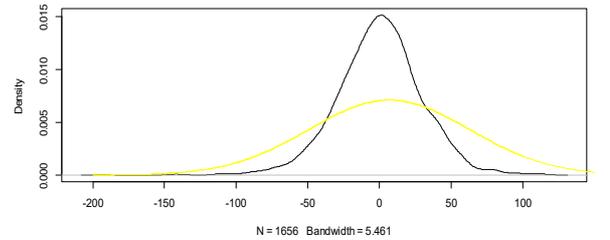


Fig. 3.2.9: L'acf dei residui di y_2 evidenzia un buon andamento nei limiti, in qualche punto c'è dell'oscillazione analizzabile.

Il valor medio dei residui è zero e non sono gaussiani, fig. 3.2.8. L'acf è buono, fig. 3.2.9. La regressione è buona.

21 Vedi n.15 bibliografia.
22 Vedi n.8 bibliografia.

La regressione della variabile NHSTI (y_3), tab. 3.3.a, è caratterizzata da una “zona di equilibrio” tra il contributo della terraferma e del mare²³.

<p>Regressione polinomiale di 3° grado</p> <pre>> l=lm(y3~t+t^2+t^3)</pre> <p>> l=lm(formula = y3 ~ t + t^2 + t^3)</p> <p>Coefficients:</p> <p>(Intercept) t t^2 t^3</p> <p>-8.181e+01 2.621e-01 -3.317e-04 1.539e-07</p>	<p>Considero la regressione polinomiale cubica perché l'andamento presenta un incurvamento centrale.</p>  <p>Fig. 3.3.1: Rappresento i dati e la regressione polinomiale in arancio.</p>
<p>Il valor medio dei residui è nullo:</p> <pre>> t.test(r3) One Sample t-test</pre> <p>data: r3</p> <p>t = 0.053388, df = 1655, p-value = 0.9574</p> <p>alternative hypothesis: true mean is not equal to 0</p> <p>95 percent confidence interval: -1.407991 1.486786</p> <p>sample estimates: mean of x 0.03939725</p>	 <p>Fig. 3.3.2: L'acf dei residui presenta delle oscillazioni.</p>
<pre>> shapiro.test(r3)</pre> <p>Shapiro-Wilk normality test</p> <p>data: r3</p> <p>W = 0.98371, p-value = 9.18e-13</p> <p>I residui $r_3=y_3-y$ non sono gaussiani, ma hanno valor medio nullo.</p>	 <p>Fig. 3.3.3: Rapporto la funzione di distribuzione dei residui con una normale in giallo.</p>

Tab. 3.3.a: Regressione polinomiale della serie y_3 .

Analizzo lo spettro dei residui, fig. 3.3.4:

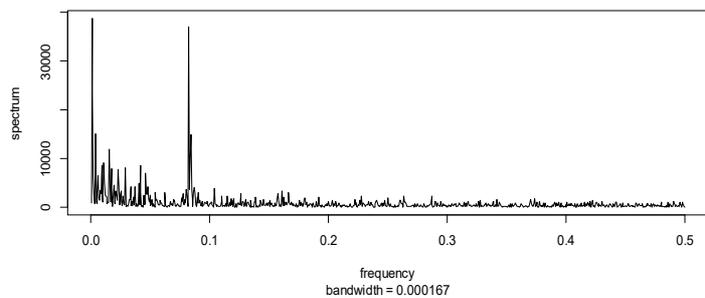


Fig. 3.3.4: Spettro dei residui della regressione polinomiale di 3° grado.

Nello spettro dei residui, fig. 3.3.4, ci sono 2 frequenze fondamentali di altezza quasi uguale, che sono circa 0.001 e 0.08, frequenza annuale. Al di fuori di questo intervallo c'è essenzialmente rumore. Le frequenze al di sopra di 5000 di altezza sono: $f_1=0.001$; $f_2=0.004$; $f_3=0.006$; $f_4=0.009$; $f_5=0.011$; $f_6=0.016$; $f_7=0.017$; $f_8=0.023$; $f_9=0.029$; $f_{10}=0.042$; $f_{11}=0.046$; $f_{12}=0.083$; $f_{13}=0.085$. I periodi corrispondenti sono: $\tau_1=83,3$ anni; $\tau_2= 20,83$ anni; $\tau_3=13,9$ anni; $\tau_4=9,26$ anni; $\tau_5= 7,58$ anni; $\tau_6=5,21$ anni; $\tau_7=4,9$ anni; $\tau_8=3,62$ anni; $\tau_9=2,87$ anni; $\tau_{10}=1,98$ anni; $\tau_{11}=1,8$ anni; $\tau_{12}= 1$ anni; $\tau_{13}=0,98$ anni. L'emisfero Nord è costituito in maggior parte di terraferma, e le periodicità dominanti sono essenzialmente maggiori di 1 anno. Esse potrebbero essere legate al moto di rivoluzione dei pianeti del sistema solare, quali Urano (84 anni), Saturno (29 anni), Giove (12 anni), Marte (687 giorni) e Venere (224 giorni). Il segnale²⁴ è meno rumoroso e le frequenze nello spettro sono ben definite, le variazioni sono minori dell'emisfero Sud.

3.3.a Metodo SSA

```
>s3=ssa(y3);
```

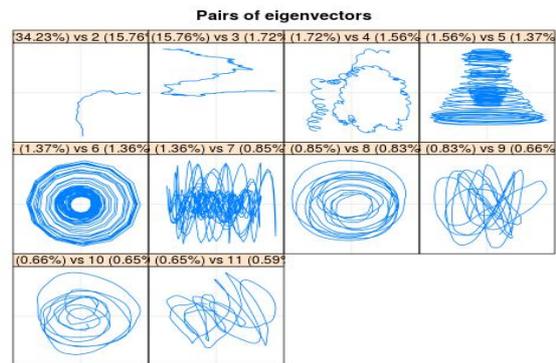
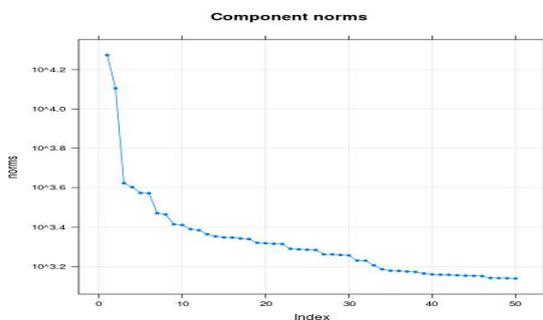


Fig. 3.3.5: Gli autovalori di y_3 sono distinti nella parte che rappresenta il rumore e la parte alta che individua il trend e la stagionalità. Gli autovettori sono accoppiati e rappresentano gli andamenti di base.

```
>r3=reconstruct(s3,groups=list(trend=c(1:4),sea1=c(5:6),sea2=c(7:8),rumore=c(9:50)));
```

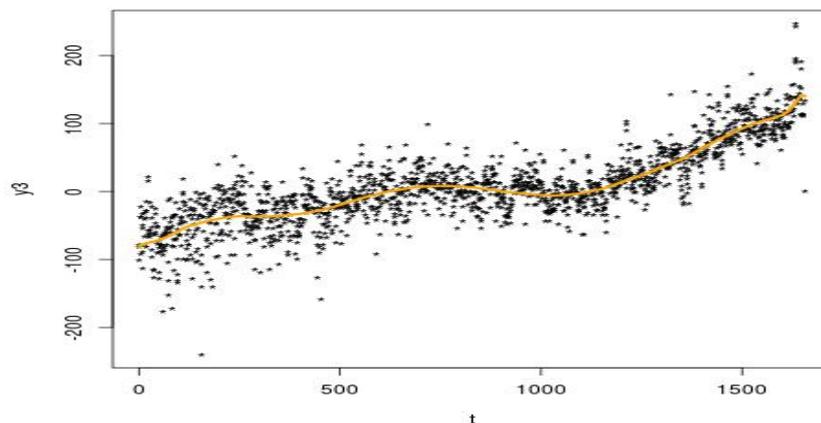


Fig. 3.3.6: Il trend suddivide i dati in tre periodi di circa 50 anni.

Il trend1 è crescente e manifesta 3 oscillazioni. Il sea1 è spirale, il sea2 è oscillazione modulata, fig. 3.3.7. La presenza di acqua, circa confrontabile con la quantità di terraferma, determina delle variazioni nella variabile y_3 . Tuttavia, la regolarità è dominante.

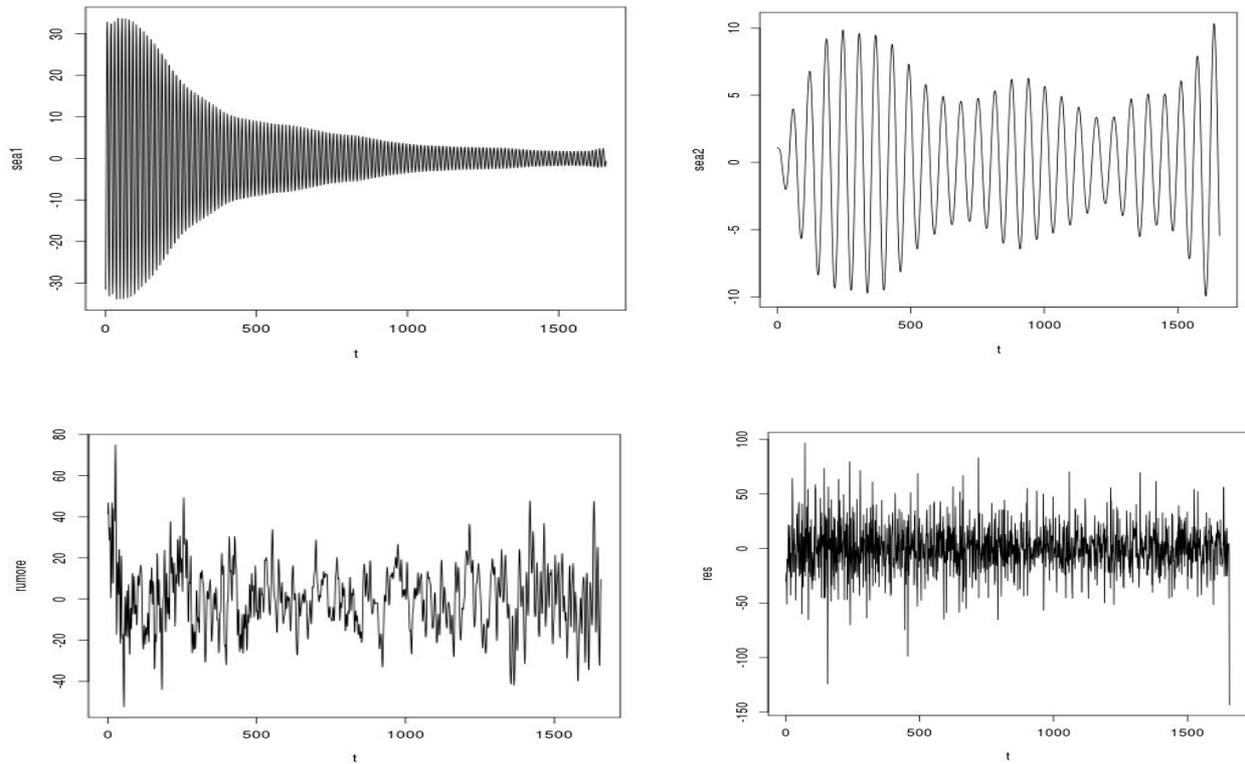


Fig. 3.3.7: Le due componenti della stagionalità hanno oscillazioni diverse. Il rumore è quasi oscillante. I residui possono presentare delle componenti studiabili.

Analizzo i residui, fig. 3.3.7 e fig. 3.3.8:

```
> res3=y3-h
> t.test(res3) One Sample t-test
data: res3
t = -0.30007, df = 1655, p-value = 0.7642
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-1.1689986 0.8587698
sample estimates:
mean of x
-0.1551144
> shapiro.test(res3) Shapiro-Wilk normality test
data: res3
W = 0.97244, p-value < 2.2e-16
```

Il valor medio dei residui è zero. L'acf è buono.

La regressione è buona.

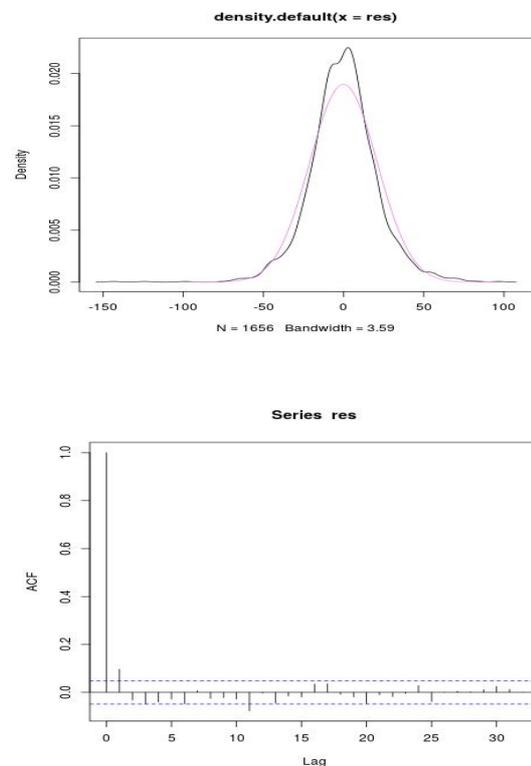
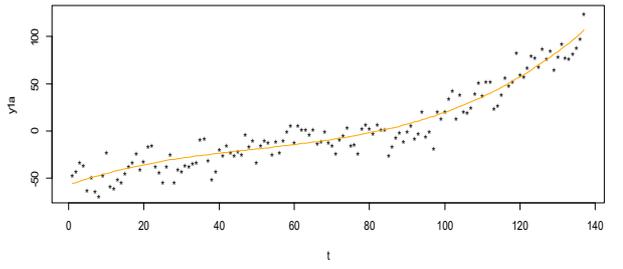
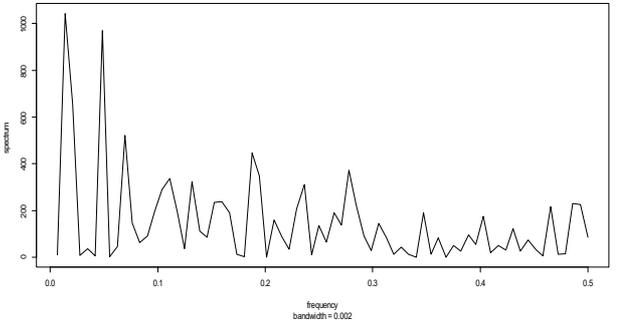
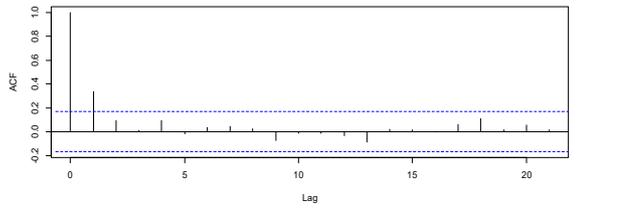


Fig. 3.3.8: La funzione di distribuzione dei residui ha e picchi, la normale violetto ha un'ampiezza

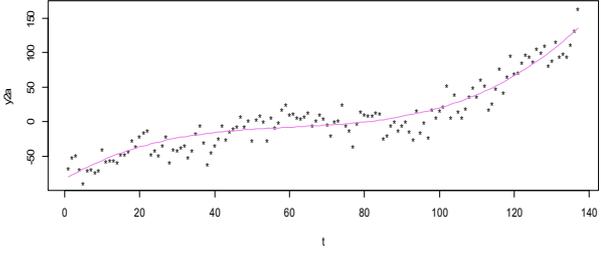
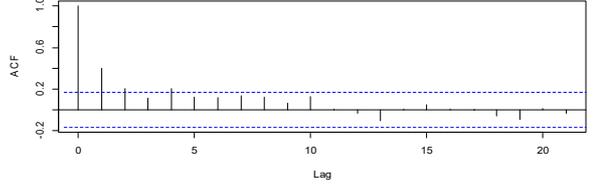
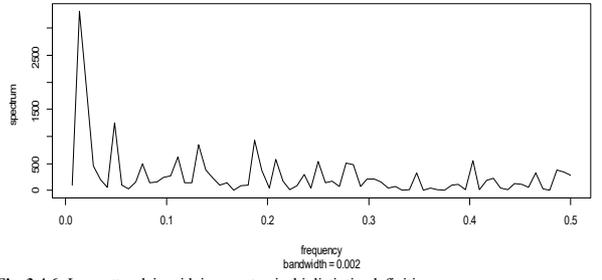
maggiore, ma altezza minore. L'acf dei residui è migliore dei precedenti.

3.4 Le serie Station Temperature Index (STI) in media annuale: STIAG (Annual mean Global), STISH (Sud Hemisphere) e STINH (North Hemisphere)

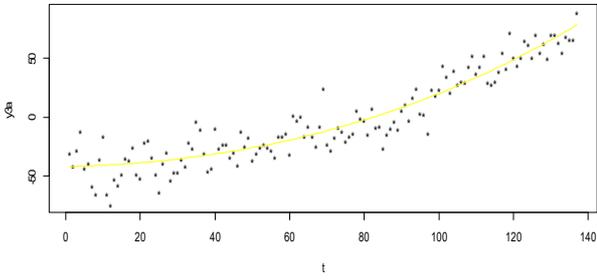
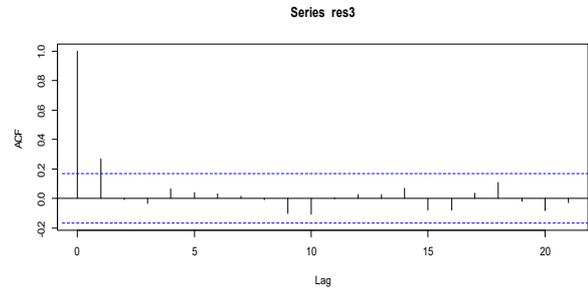
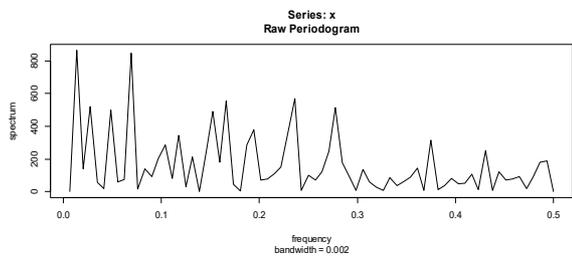
Le serie temporali STI sono mediate su un arco temporale di un anno, ottengo le serie STIAG (y), STISH (x) e STINH (z), che sono analizzate nella regressione polinomiale, nelle tab. 3.4.a, 3.4.b e 3.4.c. L'andamento è più pulito e riproduce il comportamento dell'indice, analizzato nel par. 3.3. La regressione polinomiale produce un acf buono e residui gaussiani. Le variabili sono più vicine all'astrazione fisica delle suddette. Il tempo è espresso in anni.

<p>La serie climatica y è polinomiale.</p> <pre>> p1=lm(y~t+t^2+t^3) > p1=lm(formula = y ~ t + t^2 + t^3) Coefficients: (Intercept) t t^2 t^3 -5.724e+01 1.414e+00 -1.956e-02 1.311e-04</pre>	<p>N=137</p>  <p>Fig. 3.4.1: I valori sperimentali e la regressione polinomiale in arancio.</p>
<pre>> res=y-p1 > t.test(res) One Sample t-test data: res t = 0.0039673, df = 136, p-value = 0.9968 alternative hypothesis: true mean is not equal to 0 95 percent confidence interval: -2.086793 2.095183 sample estimates: mean of x 0.004 > shapiro.test(resy) Shapiro-Wilk normality test data: resy W = 0.99093, p-value = 0.5211</pre> <p>I residui sono gaussiani.</p> <p>La regressione polinomiale di 3° grado è buona.</p>	<p>Le frequenze dello spettro dei residui oltre 400 di intensità sono: f1=0.015; f2=0.05; f3=0.07; f4=0.19.</p> <p>Le periodicità sono di: 5, 14, 20 e 67 anni.</p>  <p>Fig. 3.4.2: Lo spettro dei residui nell'intervallo (0,0.5).</p>  <p>Fig. 3.4.3: L'acf dei residui è ottimo.</p>

Tab. 3.4.a: Regressione polinomiale di y.

<p>La serie z è polinomiale. Effettuo la regressione.</p>	<p>N=137</p>
<pre>> p2=lm(z~t+t^2+t^3) > p2=lm(formula = z ~ t + t^2 + t^3) Coefficients: (Intercept) t t^2 t^3 -8.374e+01 3.224e+00 -4.899e-02 2.713e-04</pre>	 <p>Fig. 3.4.4: I dati sperimentali e la regressione polinomiale in violetto.</p>
<pre>> res=z-p2 > t.test(res) One Sample t-test data: res t = 0.0092132, df = 136, p-value = 0.9927 alternative hypothesis: true mean is not equal to 0 95 percent confidence interval: -2.829382 2.855869 sample estimates: mean of x 0.01324342</pre>	 <p>Fig. 3.4.5: L'acf dei residui è buono.</p> <p>La regressione polinomiale è buona.</p>
<p>I residui sono gaussiani:</p> <pre>> shapiro.test(res) Shapiro-Wilk normality test data: res W = 0.98584, p-value = 0.1706</pre>	<p>Le due frequenze più alte sono: f1=0.014; f2=0.049.</p>  <p>Fig. 3.4.6: Lo spettro dei residui presenta picchi distinti e definiti.</p> <p>Le periodicità sono di: 20 e 71 anni.</p>

Tab. 3.4.b: Regressione polinomiale di z.

<p>La serie climatica x è polinomiale. Effettuo la regressione.</p>	<p>N=137</p>
<pre>> p3=lm(x~t+t^2+t^3) > p3=lm(formula = x ~ t + t^2 + t^3) Coefficients: (Intercept) t t^2 t^3 -4.212e+01 8.802e-02 4.177e-03 1.164e-05</pre>	 <p>Fig. 3.4.7: I valori sperimentali e la regressione polinomiale in giallo.</p>
<pre>> res3=x-p3 > t.test(res3) One Sample t-test data: res3 t = 0.0051338, df = 136, p-value = 0.9959 alternative hypothesis: true mean is not equal to 0 95 percent confidence interval: -2.161702 2.172955 sample estimates: mean of x 0.00562644</pre>	 <p>Fig. 3.4.8: L'acf dei residui è ottimo. La regressione polinomiale è buona.</p>
<pre>> shapiro.test(res3) Shapiro-Wilk normality test data: res3 W = 0.99247, p-value = 0.6822</pre> <p>I res3 sono gaussiani.</p>	 <p>Fig. 3.4.9: Lo spettro dei residui evidenzia 2 picchi di pari altezza. Nella fig.3.4.9 trovo 2 frequenze fondamentali: f1=0.014; f2=0.069. Le periodicità sono di: 71 e 14 anni.</p>

Tab. 3.4.c: Regressione polinomiale di x.

Le variabili che descrivono El Nino²⁵ sono essenzialmente rappresentanti l'oscillazione del clima, quindi, la regressione più opportuna è armonica, il numero degli elementi della serie temporale è N=804. Le variazioni della pressione atmosferica sono dovute a perturbazioni meteorologiche al livello del mare. La circolazione della troposfera è costituita da tre celle di Hadley, di Ferrel e polare, che determinano il clima globale. Dallo spettro del SOI, fig. 3.5.1, ricavo le frequenze di intensità maggiore o circa uguale a 5:

$$f_1=0.003; f_2=0.007; f_3=0.017; f_4=0.023; f_5=0.028; f_6=0.033; f_7=0.038; f_8=0.049; f_9=0.054; f_{10}=0.112.$$

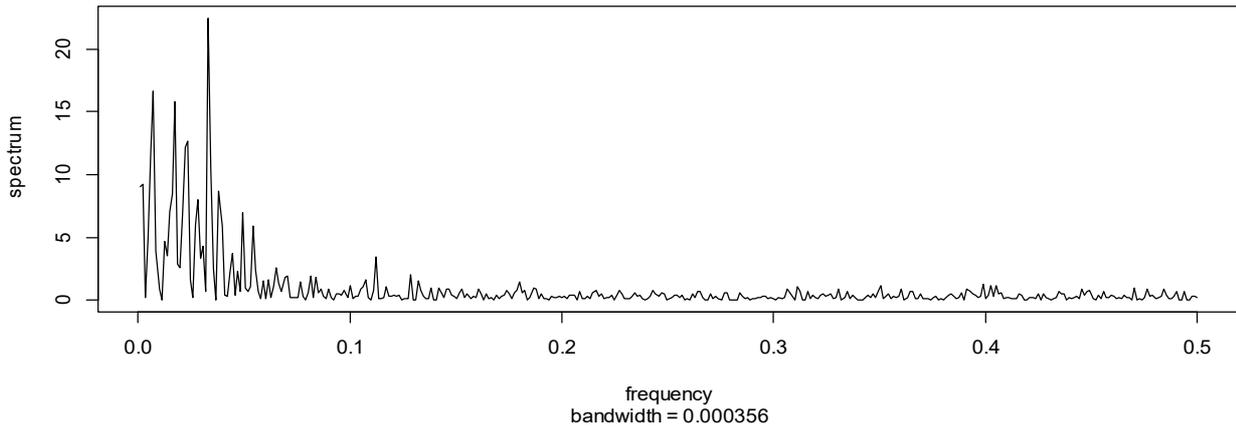


Fig. 3.5.1: Lo spettro del SOI.

I periodi corrispondenti sono:

$$\tau_1=27,8 \text{ anni}; \tau_2= 11,9 \text{ anni}; \tau_3 =4,9 \text{ anni}; \tau_4= 3,62 \text{ anni}; \tau_5=2,98 \text{ anni}; \tau_6=2,53 \text{ anni}; \tau_7=2,2 \text{ anni}; \tau_8=1,7 \text{ anni}; \tau_9= 1,54 \text{ anni}; \tau_{10}= 9 \text{ mesi}.$$

La regressione armonica con 10 frequenze:

$$> l=lm(soi~z1+w1+z2+w2+z3+w3+z4+w4+z5+w5+z6+w6+z7+w7+z8+w8+z9+w9+z10+w10)$$

$$> l \text{ lm(formula = soi ~ z1 + w1 + z2 + w2 + z3 + w3 + z4 + w4 + z5 + w5 + z6 + w6 + z7 + w7 + z8 + w8 + z9 + w9 + z10 + w10)}$$

Coefficients:

(Intercept)	z1	w1	z2	w2	z3	w3	z4	w4	z5	w5	z6
0.133589	0.108635	-0.028964	0.261646	0.027849	0.255948	-0.061417	-0.309748	-0.059198			
	w6	z7	w7	z8	w8	z9	w9	z10	w10		
	-0.032792	0.036806	0.162841	-0.006136	0.147796	0.137007	-0.054871	0.128878	-0.031868		

>summary(l):

Residuals: Min 1Q Median 3Q Max

Coefficients: Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.133589 0.029921 4.465 9.20e-06 ***

z1 0.108635 0.042300 2.568 0.010406 * w1 -0.028964 0.041564 -0.697 0.486103 z2 0.261646 0.041352 6.327 4.19e-10 ***

w2 0.027849 0.041961 0.664 0.507092 z3 0.255948 0.041482 6.170 1.09e-09 *** w3 -0.061417 0.041663 -1.474 0.140850

z4 -0.309748 0.041541 -7.456 2.36e-13 *** w4 -0.059198 0.041554 -1.425 0.154671 z5 -0.054283 0.041506 -1.308 0.191316

w5 0.074107 0.041544 1.784 0.074837 . z6 0.270251 0.041494 6.513 1.32e-10 *** w6 -0.032792 0.041548 -0.789 0.430200

z7 0.036806 0.041492 0.887 0.375323 w7 0.162841 0.041556 3.919 9.68e-05 *** z8 -0.006136 0.041573 -0.148 0.882700

w8 0.147796 0.041476 3.563 0.000388 *** z9 0.137007 0.041555 3.297 0.001021 ** w9 -0.054871 0.041480 -1.323 0.186274

z10 0.128878 0.041505 3.105 0.001971 ** w10 -0.031868 0.041523 -0.767 0.443029

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8322 on 783 degrees of freedom

Multiple R-squared: 0.242, Adjusted R-squared: 0.2227

F-statistic: 12.5 on 20 and 783 DF, p-value: < 2.2e-16

> trend=l\$fitted;

Nel grafico, fig. 3.5.2, sono rappresentati i dati con la regressione armonica, detta trend:

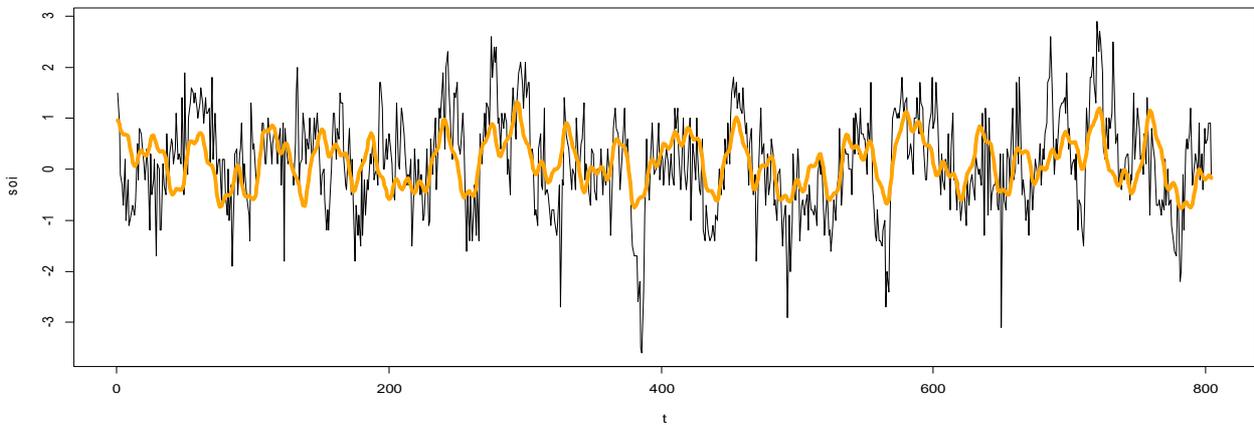


Fig. 3.5.2: I valori sperimentali e il trend effettuato.

> res=l\$residuals

> acf(res)

> t.test(res) One Sample t-test

data: res

t = 2.0047e-15, df = 803, p-value = 1

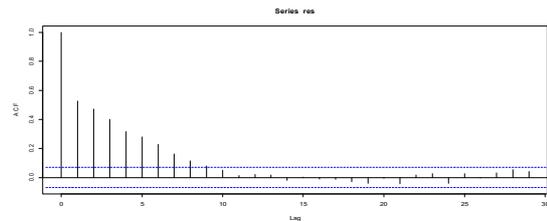


Fig. 3.5.3: L'acf dei residui presenta della regressione nella parte iniziale e un buon andamento nella seconda parte.

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

-0.05689162 0.05689162

sample estimates: mean of x 5.810142e-17

Il valor medio dei residui è zero e di ordine di grandezza molto piccolo, circa 10^{-17} .

> shapiro.test(res) Shapiro-Wilk normality test

data: res

W = 0.99626, p-value = 0.05302

I residui sono gaussiani, fig. 3.5.4.

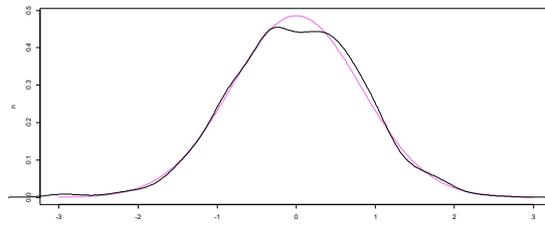


Fig. 3.5.4: La funzione di distribuzione dei residui è abbastanza aderente alla curva normale in violetto, ma non è piccata opportunamente; tuttavia l'andamento è gaussiano.

La regressione armonica conduce ad un buon trend, rapportabile a quello ottenuto con il metodo SSA.

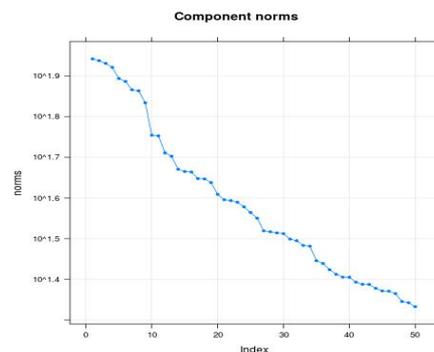
Applico il metodo, con finestra temporale pari a 50 al SOI, fig. 3.5.5:

>s=ssa(soi,l=50)

>r=reconstruct(s,groups=list(trend=c(1:13),sea=c(14:18),rumore=c(19:50)))

>trend=r\$trend; >sea=r\$sea; rumore=r\$rumore;

>h=trend+sea+rumore;



Rappresento il trend con i dati sperimentali, fig. 3.5.6:

Fig. 3.5.5: Gli autovalori del SOI con finestra temporale pari a 50.

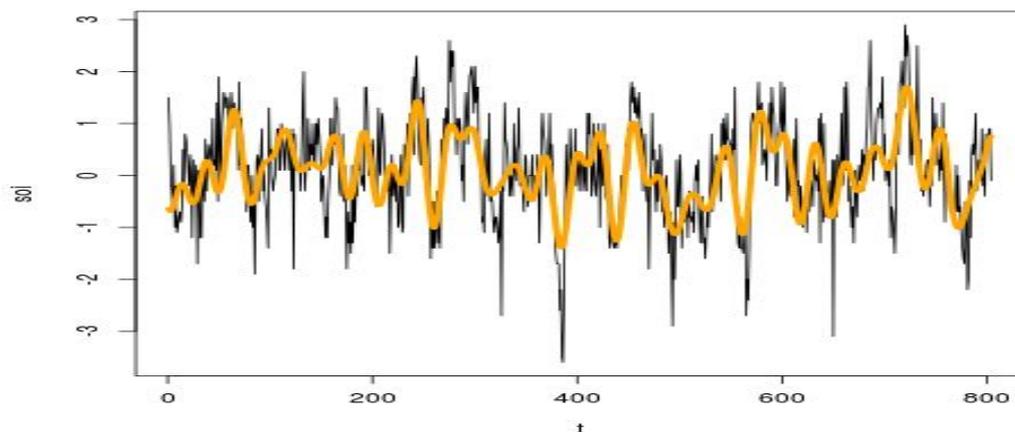


Fig. 3.5.6: I valori sperimentali con il trend individuato.

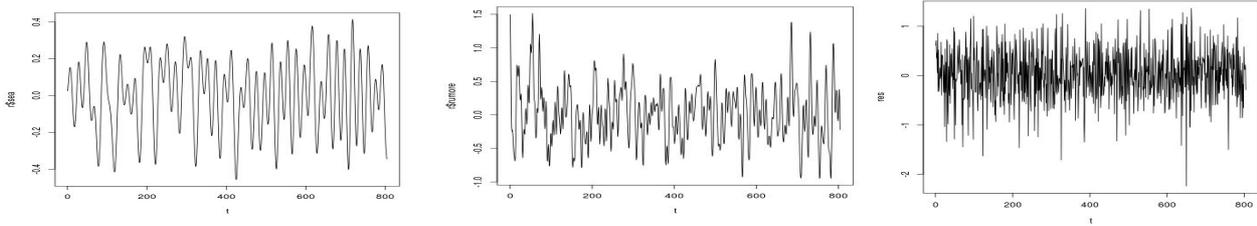


Fig. 3.5.7: La stagionalità è la composizione dei più contributi, il rumore e i residui.

```

> res=soi-h
> t.test(res)  One Sample t-test
data:  res
t = 1.6018, df = 803, p-value = 0.1096
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
-0.006626864  0.065419437
sample estimates:
mean of x  0.02939629
> shapiro.test(res)  Shapiro-Wilk normality test
data:  res
W = 0.99407, p-value = 0.002977
    
```

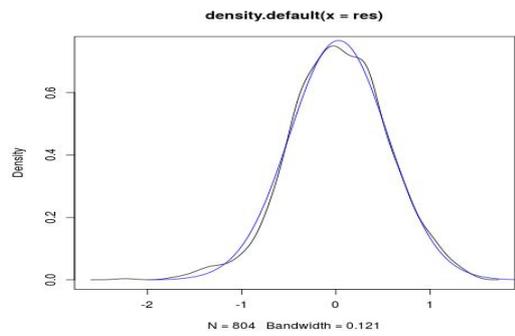


Fig. 3.5.8: La funzione di distribuzione dei residui e la curva normale in blu.

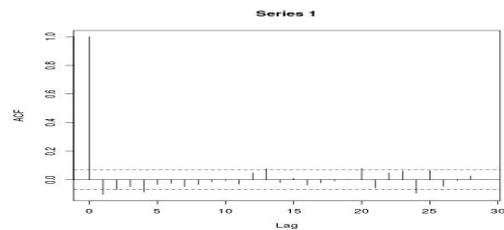


Fig. 3.5.9: L'acf dei residui è buono.

Il valor medio è zero, però, con un p circa 0.1, di poco maggiore di 0.05. I residui non sono gaussiani, fig.3.5.8. La regressione è buona. Il metodo SSA ha un acf dei residui, fig. 3.5.9, migliore della regressione armonica, fig. 3.5.2, e individua la forma del trend in modo efficace, fig. 3.5.6. La regressione armonica ha un migliore t.test dei residui e, anche, la gaussianità di essi, ma ha un trend peggiore a livello di andamento, è migliore in senso “statistico”, perché mantiene in modo ottimo la condizione di residui della regressione pari a zero. Tuttavia, per individuare l’andamento effettivo dei dati è opportuno utilizzare il metodo SSA.

Lo spettro dello SLP è, fig. 3.6.1:

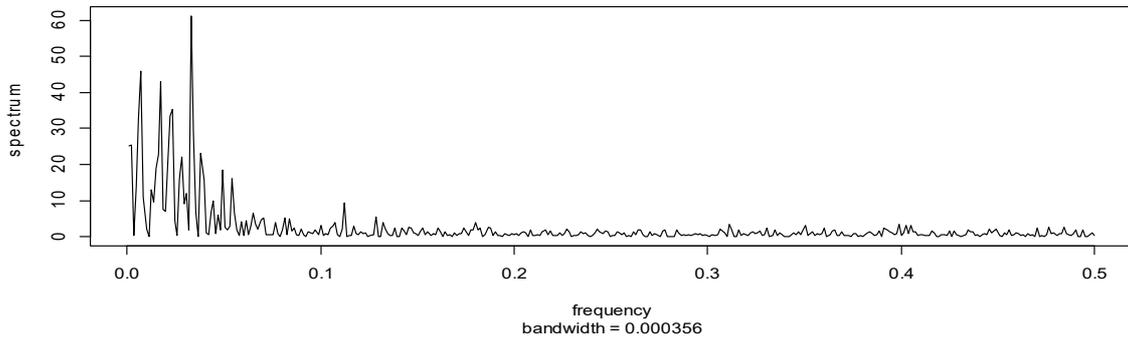


Fig. 3.6.1: Lo spettro dello SLP, le frequenze sono nell'intervallo (0,0.2).

Individuo le frequenze con intensità maggiore di 5:

$f_1=0.002; f_2=0.007; f_3=0.013; f_4=0.017; f_5=0.023; f_6=0.028; f_7=0.030; f_8=0.033; f_9=0.038; f_{10}=0.045; f_{11}=0.049; f_{12}=0.054; f_{13}=0.113.$

I periodi corrispondenti sono:

$\tau_1=41,7 \text{ anni}; \tau_2=11,9 \text{ anni}; \tau_3=6,4 \text{ anni}; \tau_4= 4,9 \text{ anni}; \tau_5=3,6 \text{ anni}; \tau_6=2,97 \text{ anni}; \tau_7=2,8 \text{ anni}; \tau_8=2,5 \text{ anni}; \tau_9=2,2 \text{ anni}; \tau_{10}= 1,85 \text{ anni}; \tau_{11}= 1,7 \text{ anni}; \tau_{12}= 1,54 \text{ anni}; \tau_{13}=8,8 \text{ mesi}.$

Il periodo di 11,9 anni potrebbe essere il tempo della rivoluzione di Giove, il periodo di 4,9 anni potrebbe essere il tempo di rotazione tra il piano lunare e quello dell'eclittica nella posizione ortogonale, ecc. La regressione della linea dei nodi è il fenomeno determinante nello SLP, in quanto è la frequenza f_5 di intensità maggiore rispetto a tutte le altre. Effettuo la regressione armonica, fig. 3.6.2:

$$l = \text{lm}(\text{formula} = \text{slp} \sim z1 + w1 + z2 + w2 + z3 + w3 + z4 + w4 + z5 + w5 + z6 + w6 + z7 + w7 + z8 + w8 + z9 + w9 + z10 + w10 + z11 + w11 + z12 + w12 + z13 + w13)$$

Coefficients:

(Intercept)	z1	w1	z2	w2	z3	w3	z4	w4	
0.16508	-0.36162	0.22268	0.44898	0.05340	-0.06938	-0.33114	0.41442	-0.07722	
	z5	w5	z6	w6	z7	w7	z8	w8	z9
-0.51888	-0.08795	-0.10343	0.13845	0.02656	0.08587	0.45774	-0.06049	0.06951	
	w9	z10	w10	z11	w11	z12	w12	z13	w13
0.26122	-0.10777	-0.12356	-0.01065	0.25576	0.22091	-0.08216	-0.07967	0.07048	

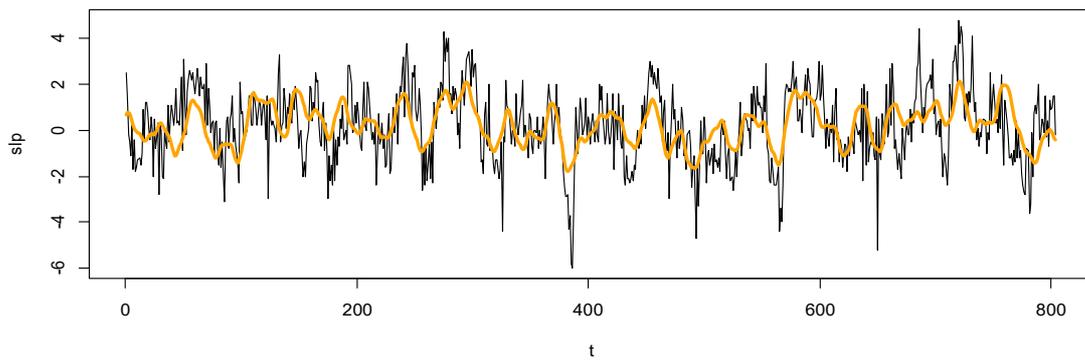


Fig. 3.6.2: I valori sperimentali e il trend.

> t.test(res) One Sample t-test

data: res

$t = 1.5492e-15$, $df = 803$, $p\text{-value} = 1$

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval: -0.09089405 0.09089405

sample estimates: mean of x 7.173568e-17

> shapiro.test(res) Shapiro-Wilk normality test

data: res

$W = 0.99713$, $p\text{-value} = 0.1666$

Il valor medio dei residui è zero.

I residui sono gaussiani, fig.3.6.3.

L'acf è buono, fig.3.6.3. La regressione è buona.

Confronto la densità di probabilita' del SOI, in nero, con quella dello SLP, in blu, fig. 3.6.4:

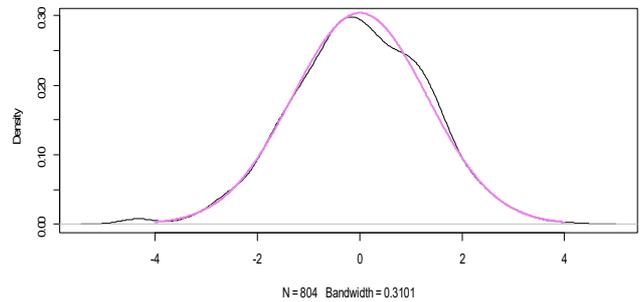
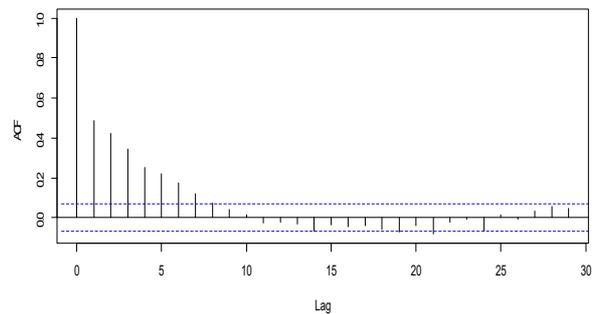


Fig. 3.6.3: L'acf dei residui e la funzione di distribuzione dei residui che si adatta alla curva normale in violetto.

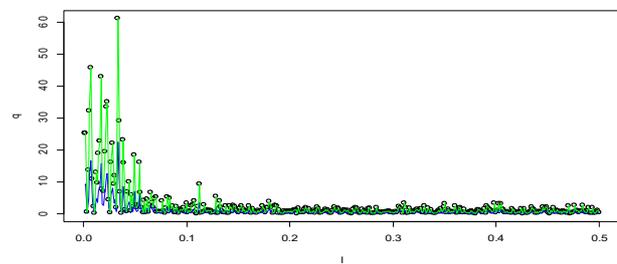
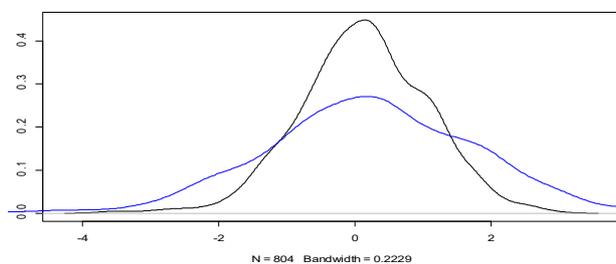


Fig. 3.6.4: Le funzioni di distribuzione e lo spettro del SOI e dello SLP.

Esse sembrano proporzionali e, così, lo spettro di un fattore circa 3.

```
>s=ssa(slp,L=50)
```

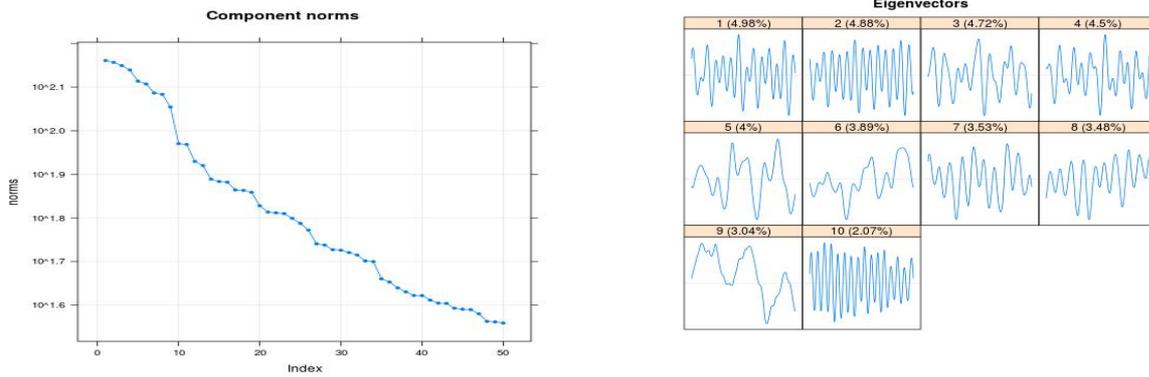


Fig. 3.6.5: Gli autovalori dello SLP e gli autovettori. Gli andamenti di questi sono oscillanti in modo vario.

```
>r=reconstruct(s,groups=list(trend=c(1:13), seasonality=c(14:18),rumore=c(19:50)))
```

```
>trend=r$trend; sea=r$seasonality; rumore=r$rumore;
```

```
>k=trend+sea+rumore
```

Rappresento i dati e il trend, fig. 3.6.6:

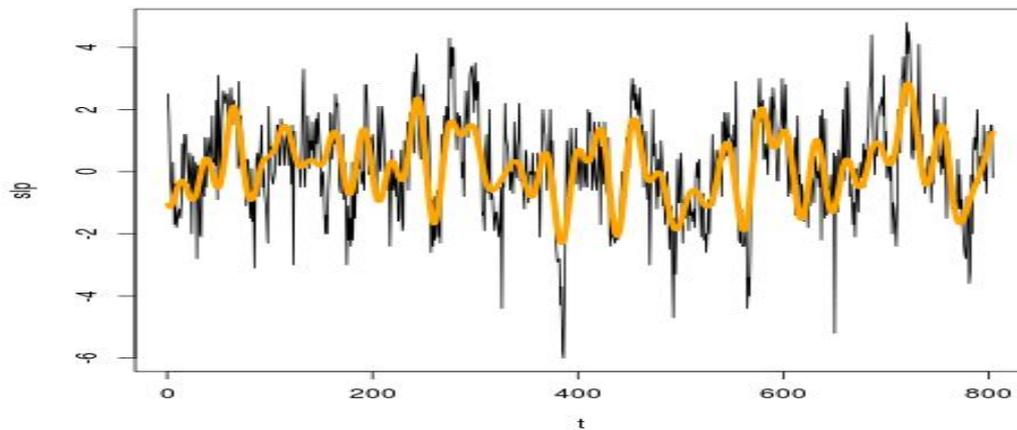


Fig. 3.6.6: I valori sperimentali e il trend dello SLP. Gli andamenti sono abbastanza coincidenti.

Rappresento le periodicità, come “somma di contributi”, ed il rumore, fig. 3.6.7:

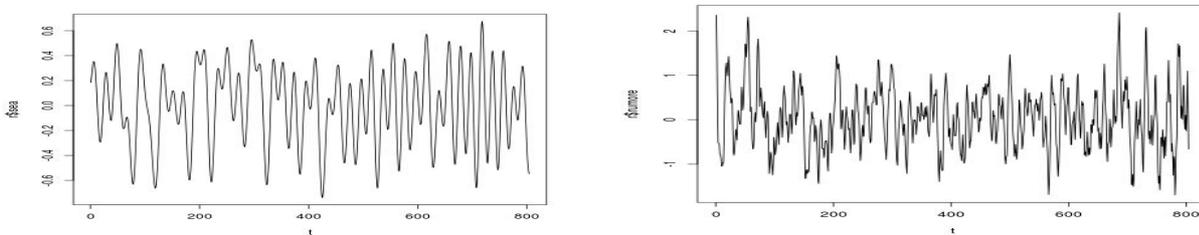


Fig. 3.6.7: La stagionalità e il rumore dello SLP.

Rappresento k e i dati sperimentali, la serie ricostruita è abbastanza fedele a quella iniziale, come evidente in fig. 3.6.7:

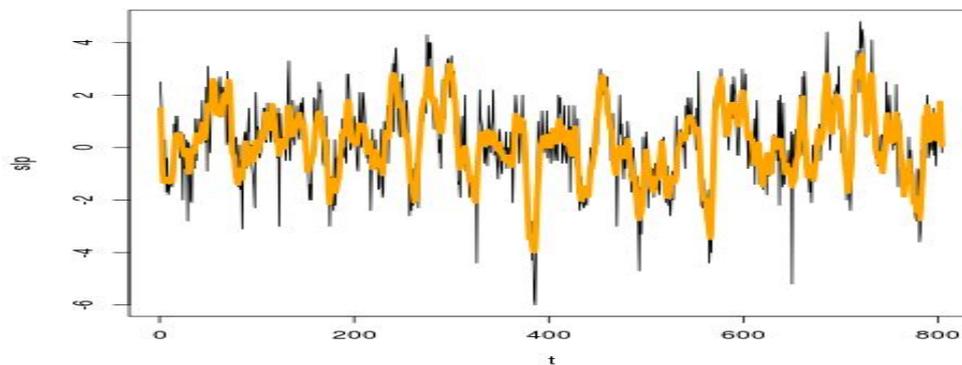


Fig. 3.6.7: I valori sperimentali e la serie ricostruita tramite il metodo SSA.

Analizzo i residui:

```
> res=slp-h
```

```
> t.test(res) One Sample t-test
```

```
data: res
```

```
t = 1.6002, df = 803, p-value = 0.1099
```

```
alternative hypothesis: true mean is not equal to 0
```

```
95 percent confidence interval: -0.0109436 0.1075004
```

```
sample estimates: mean of x 0.04827839
```

```
> shapiro.test(res) Shapiro-Wilk normality test
```

```
data: res W = 0.99516, p-value = 0.01214
```

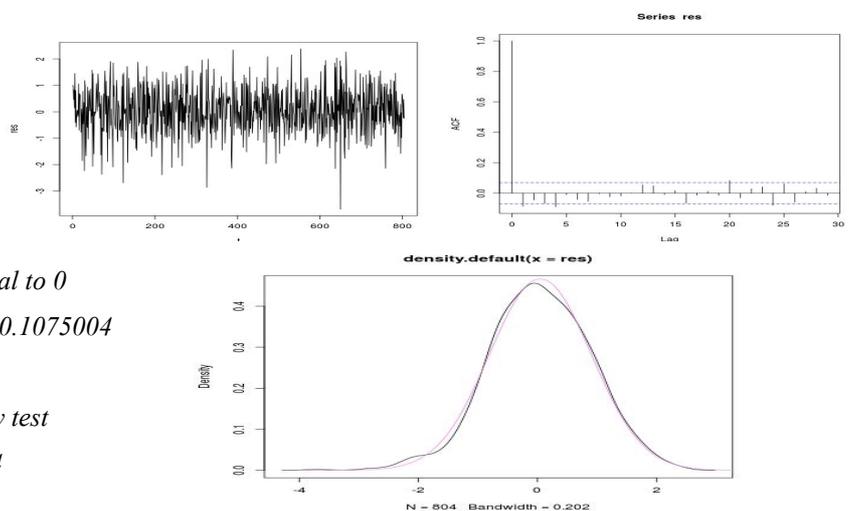


Fig. 3.6.8: I residui nella prima immagine, l'acf dei residui nella seconda è ottimo, la funzione di distribuzione non aderisce alla normale.

Il valor medio dei residui è zero. I residui non sono gaussiani, fig. 3.6.8. La regressione è ottima. Il I metodo è più efficace a livello di valor medio dei residui, mentre il II metodo è più efficace a livello di acf. I residui, nel primo caso, sono gaussiani, nel secondo non gaussiani. Il trend, ottenuto con il metodo SSA, è migliore rispetto a quello dell'analisi armonica. Il metodo SSA risulta ottimo per effettuare trend e individuare stagionalità per grandezze fisico-climatiche semplici o indici. Di conseguenza, nel capitolo 4, applicherò solo quest'ultimo metodo.

I dati sperimentali sono forniti dall'ENEA Frascati (RM) gruppo di ricerca Artale, Marullo e misurati dalla stazione di Osservazioni Climatiche ENEA di Lampedusa.

4.1 Apparat di misura: tipologie sensori

L'apparato sperimentale consiste di una boa oceanica, fig.4.1.1, (35.49° N, 12.47° E), a circa 3,3 miglia a SO di Lampedusa, che è utilizzata per studiare interazioni aria-mare e verificare le osservazioni satellitari.

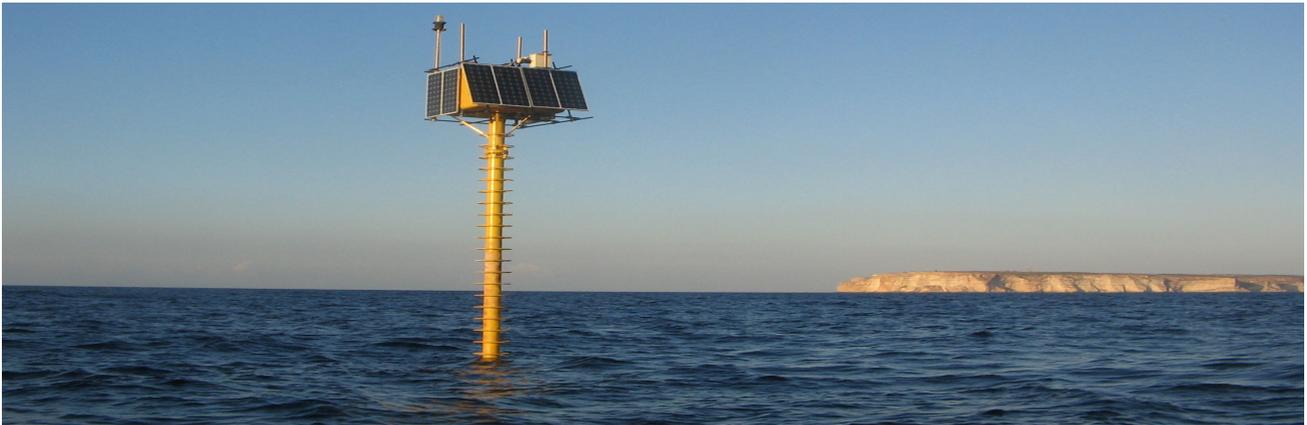


Fig. 4.1.1: Boa oceanica di tipo "meda elastica". Sullo sfondo si vede l'isola di Lampedusa.

Essa²⁶ è costituita di una torretta, di un tubo superiore emerso, un tubo porta galleggiante (in acqua), di 3 galleggianti, di 2 tubi collegati di 273x12.7 e 355x12.7, cavi metallici che la collegano al basamento in cemento, posto sul fondale marino. Sulla torretta ci sono: 2 radiometri e 1 livella elettronica, collegati ad un sistema di acquisizione. Il sistema di approvvigionamento energetico è con pannelli fotovoltaici e di accumulo di energia. Successivamente, è stata aggiunta una stazione meteorologica, spettrometri e dei sensori per misurare per 40 m, di cui 7 m emersi e il resto in mare. Il fondale in questo punto è di 74 m ed è un punto d'incontro tra masse d'aria+acqua europee e africane. Ci sono i sensori barometrici e le sonde multiparametriche²⁷, fig.4.2. Il primo è tipo trasduttore capacitivo al silicio, monta più di un trasduttore di pressione per consentire una certa ridondanza e tramite l'interfaccia, oltre a richiedere i dati si possono programmare ad esempio: le unità in cui convertire i dati (Caratteristiche: Campo di misura: 500, 1100 hPa; Uscita Elettrica: seriale RS232/TTL o RS485/422, velocità da 300 a 9600 baud ; Ripetibilità: ± 0.03 hPa; Risoluzione: 0.01 hPa; Non linearità: ± 0.05 hPa; Intercambiabilità: entro la precisione complessiva; Intervallo: temperatura $-40 \div 60$ °C, umidità 10,100 %; Precisione Complessiva: ± 0.3 hPa; Caratteristiche: alimentazione: 10, 35 Vcc; Elettriche: (derivata direttamente dall'unità di acquisizione); Consumo: standby: 0.1 mA, massimo: 25 mA; Massima distanza da un acq.: RS232: 15 m, RS485/RS422: 1 Km; Costanza nel tempo: ± 0.1 hPa per anno; Intervallo ritaratura: 2 anni; Protezione: zener veloce; Disponibilità ricambi: 1 mese; Accessibilità: completa a tutte le parti componenti; Dimensioni: 145x1138x65 mm; Peso: 0,95 Kg). La seconda può contenere i seguenti sensori: pH, conducibilità elettrica, potenziale redox, ossigeno disciolto, temperatura e livello idrometrico. Essa può essere immersa fino ad una profondità di 30 metri e collegata all'unità in superficie tramite un cavo autoportante di alimentazione e trasmissione dei dati (Caratteristiche di temperatura: Tecnica di misura: termoresistenza Pt1000 a norme DIN; Campo di misura: $-5 \div 60$ °C ; Precisione: $\pm 0,2$ °C; Risoluzione: 0,1 °C).

26 Vedi n.20 bibliografia.

27 www.ispra.it.

I valori sperimentali sono presi minuto per minuto dal 7/4/2017 al 9/11/2017. Le grandezze fisiche²⁸ considerate sono T_1 e P_1 , misurate per $h = 1\text{m}$ e T_2 e P_2 , per $h = 2\text{m}$. Le serie sono non gaussiane e hanno intervallo di confidenza e medie:

$$T_1 = (16.5 \pm 0.7) ^\circ\text{C}$$

$$P_1 = (1.2 \pm 0.3) \text{ decibar}$$

$$T_2 = (16.4 \pm 0.6) ^\circ\text{C}$$

$$P_2 = (2.0 \pm 0.3) \text{ decibar}$$

Gli andamenti di queste grandezze sono, fig. 4.2.1:

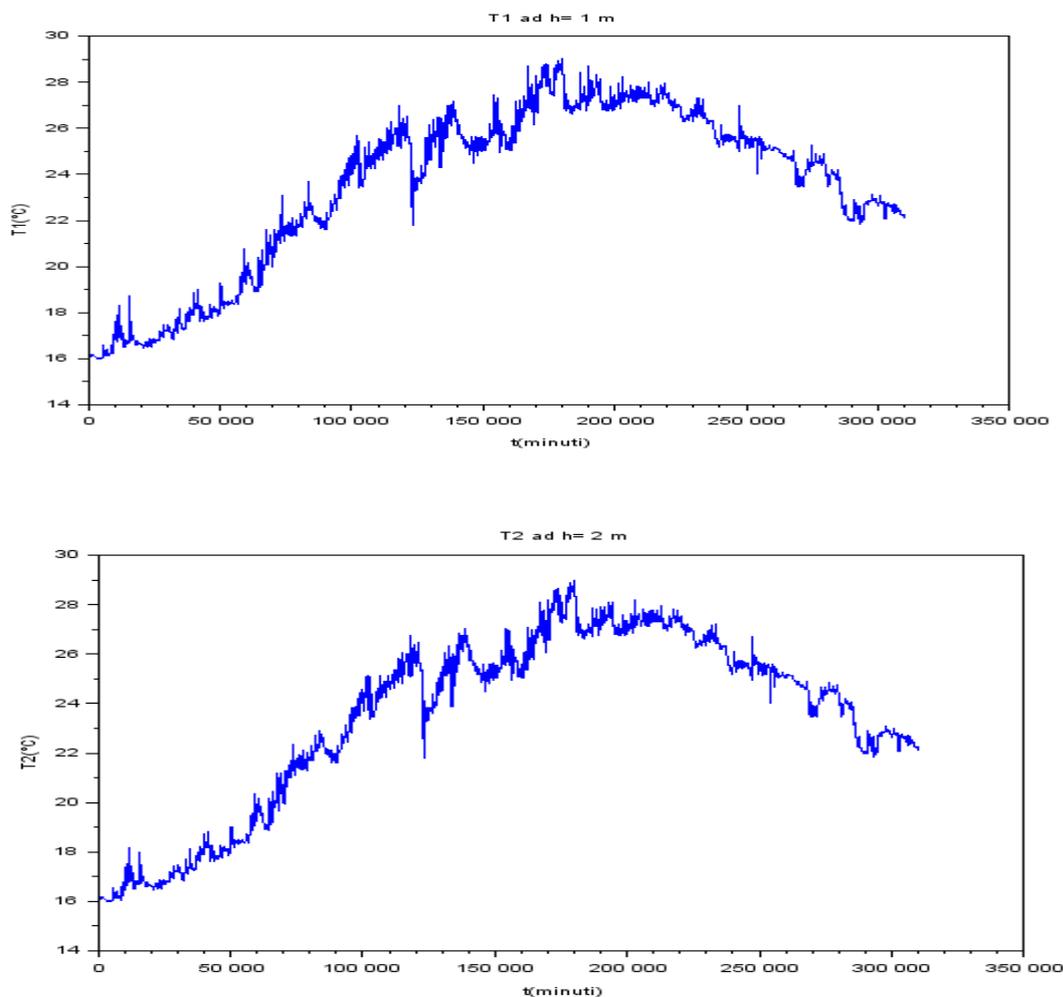


Fig. 4.2.1: Il primo grafico rappresenta la T_1 in funzione del tempo in minuti, sono circa 350000 valori. Il secondo grafico rappresenta la T_2 di poco diversa dalla T_1 .

Le temperature, a diverse altezze sotto la superficie del mare, sono circa uguali, anche se all'aumentare della profondità nell'idrosfera la temperatura dovrebbe diminuire. La differenza di 1 m produce variazioni, essenzialmente sulle cifre decimali, poco apprezzabili graficamente. Nell'elaborazione considero, quindi, la differenza tra esse, per indagare sulle diversità e sulle cause.

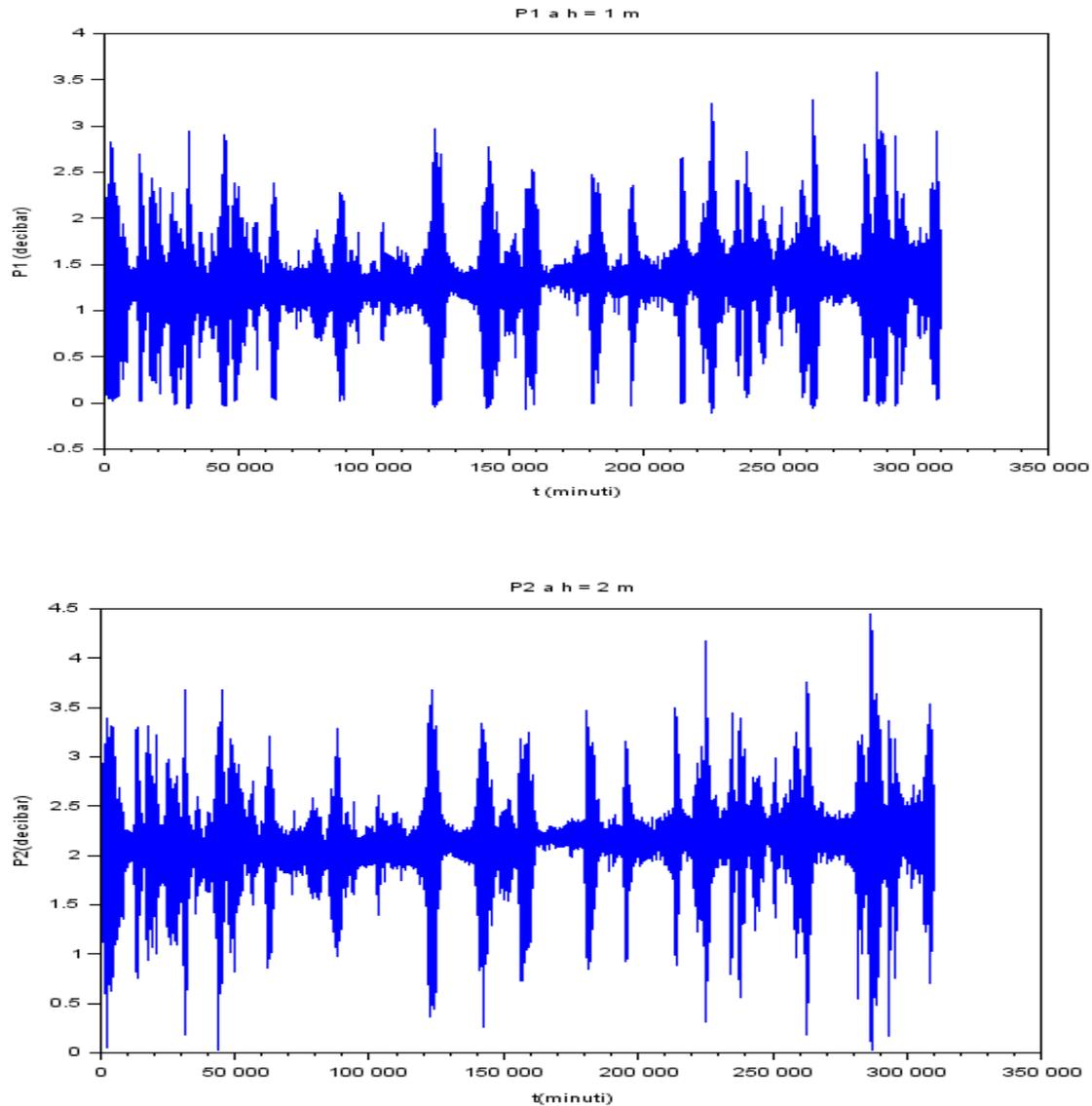


Fig. 4.2.2: Il primo grafico rappresenta la P_1 in decibar in funzione del tempo in minuti. Il secondo grafico è la P_2 .

Le pressioni, fig. 4.2.2, a differenti altezze sono visibilmente diverse, in quanto oscillano intorno a valori medi diversi, quali 1 decibar e 2 decibar. Tale unità di misura è stata considerata in quanto essa corrisponde alla profondità; ad 1 m la pressione è circa 1 decibar, a 2 m è circa 2 decibar. L'oscillazione è, invece, simile. Gli intervalli, in cui i valori non sono stati presi sperimentalmente, sono stati ricostruiti, effettuando la media dei valori precedenti. Successivamente, essendo N circa 300000 eccessivo, ho mediato i valori su un'arco temporale orario, ottenendo circa 5000 ore. Le medie riproducono fedelmente gli andamenti delle variabili generali. L'analisi statistica è effettuata, tramite il metodo SSA sulle medie. Il trend e la stagionalità individuano le caratteristiche climatiche della zona di Frascati, nella terraferma, quindi, la componente legata all'alternarsi delle stagioni è dominante. Infatti, la forma ad arco della temperatura rappresenta proprio l'andamento annuale di essa. La regione di brusco calo sulla curva è prodotta da un periodo di perturbazioni meteorologiche tra 2000 e 3000 ore. La pressione segue, anch'essa, un arco di ampiezza maggiore con notevole oscillazione. In realtà, manifesta un andamento periodico annuale ad arco. I risultati sono in accordo con i risultati Ispra. Il tempo, di seguito, è espresso in ore.

Rappresento la serie T_1 con la regressione SSA:

```
> s1=ssa(T1);
```

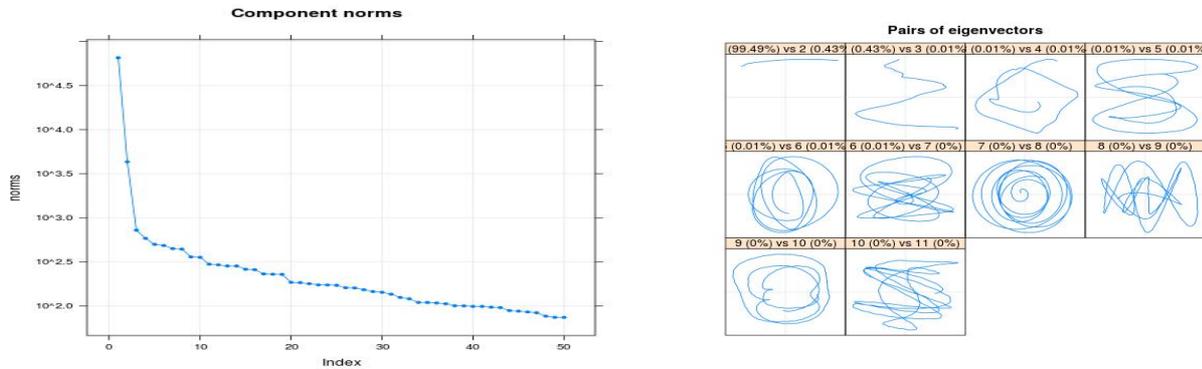


Fig. 4.3.1: Autovalori e autovettori di T_1 .

```
> r1=reconstruct(s1,groups=list(trend=c(1:4),sea1=c(5,6),sea2=c(7,8),sea3=c(9,10),rumore=c(11:50)));
```

```
> trend1=r1$trend; sea1=r1$sea1; sea2=r1$sea2; sea3=r1$sea3; rumore=r1$rumore;
```

Il trend e i dati sperimentali sono, fig.4.3.2:

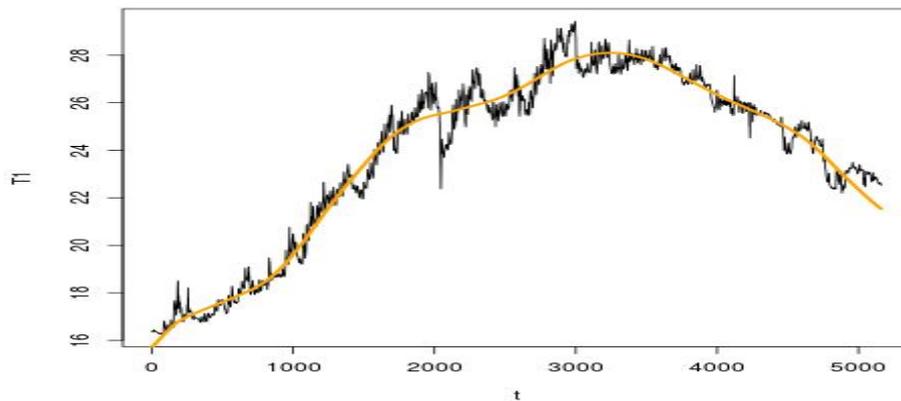
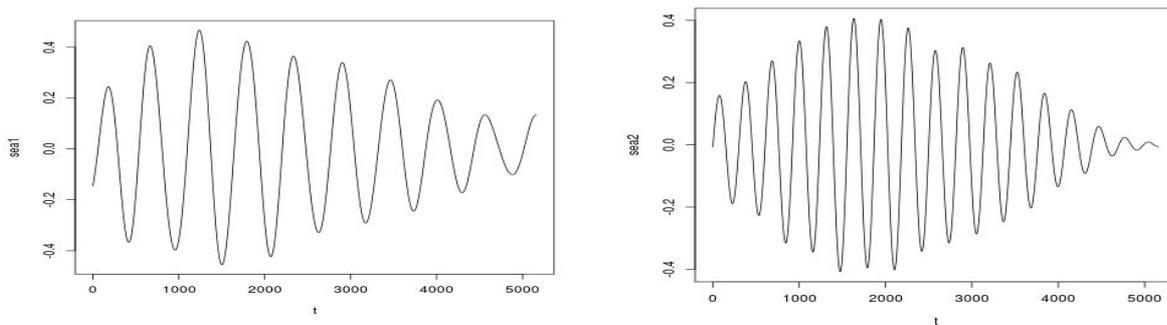


Fig. 4.3.2: I valori sperimentali e il trend.

Le componenti della stagionalità sono, fig. 4.3.3:



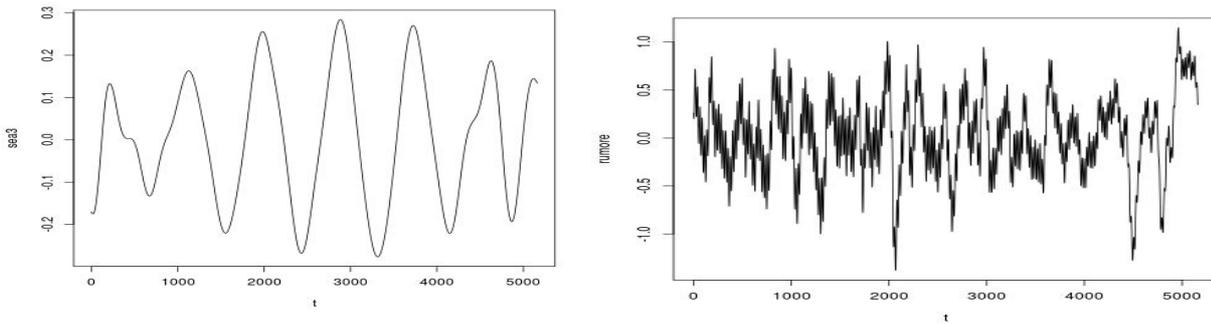


Fig. 4.3.3: Le tre componenti della periodicità e il rumore.

Il valor medio dei residui è zero, essi non sono gaussiani, fig. 4.3.4.

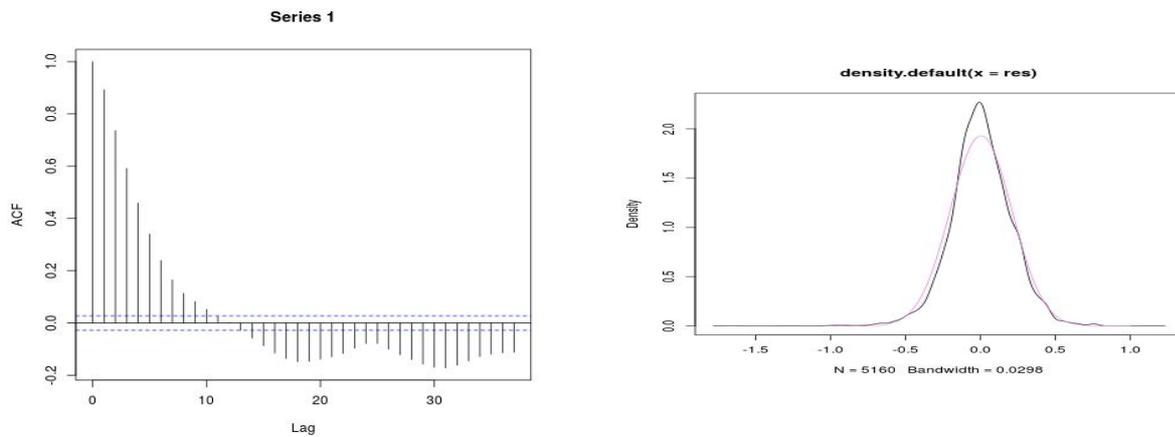


Fig. 4.3.4: L'acf dei residui presenta dell'oscillazione. La funzione di distribuzione dei residui paragonata alla curva normale in violetto.

Lo spettro della serie T_1 evidenzia le frequenze: $f_1=0.000095$, $f_2=0.0006$, $f_3=0.0017$, $f_4=0.003$ e $f_5=0.0417$.

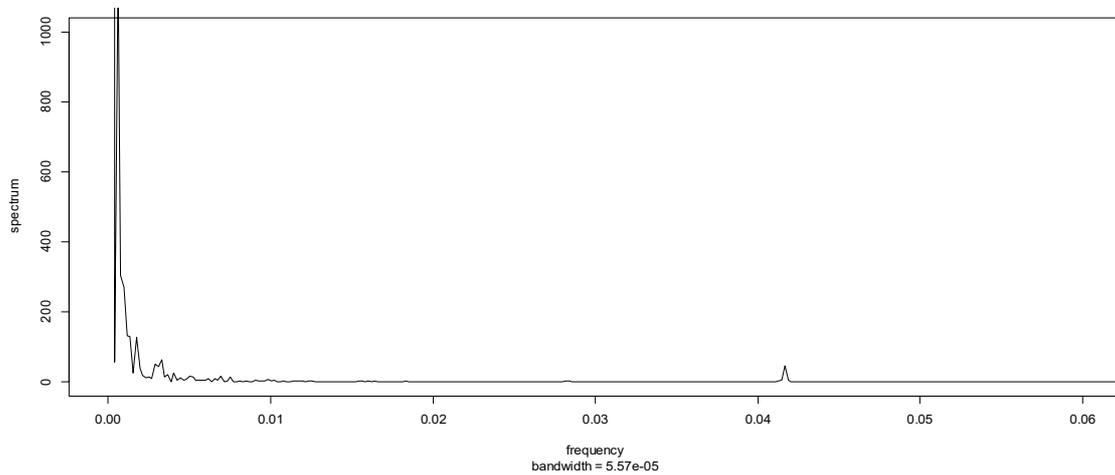


Fig. 4.3.5: Lo spettro di T_1 con picchi intorno allo zero e a 0.04.

Ricavo i periodi:

$\tau_1=1,2$ anni, circa 1 anno (periodicità annuale); $\tau_2=69,4$ giorni; $\tau_3=24$ giorni (circa mese sidereo, 27 giorni); $\tau_4=14$ giorni (distanza temporale tra novilunio e plenilunio); $\tau_5=24$ ore (periodicità giornaliera).

> l lm(formula = T1 ~ z1 + w1)

Coefficients:

(Intercept)	z1	w1
18.259	-3.226	8.754

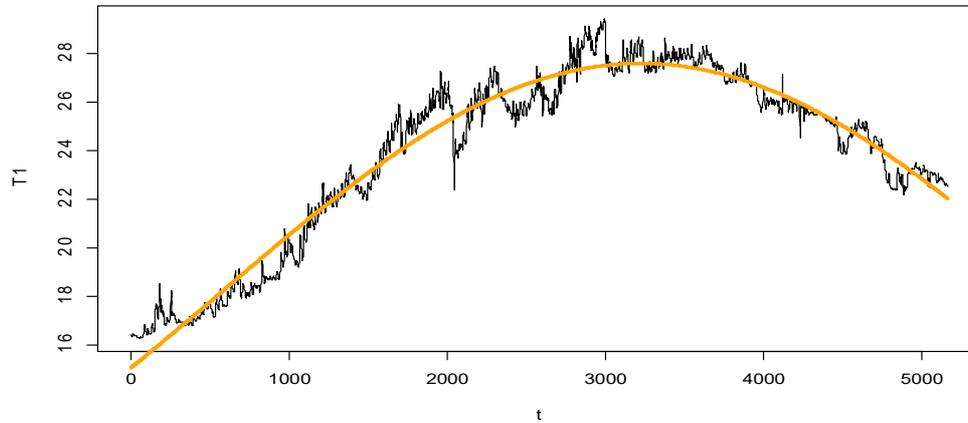


Fig. 4.3.6: La regressione armonica con la frequenza più elevata sui valori sperimentali.

La frequenza f_1 è la frequenza oraria annuale ($1/(365*24)$), che fa comprendere l'andamento generale.

> l lm(formula = T1 ~ z1 + w1 + z2 + w2)

Coefficients:

(Intercept)	z1	w1	z2	w2
18.1059	-3.2099	8.9638	0.5252	0.1167

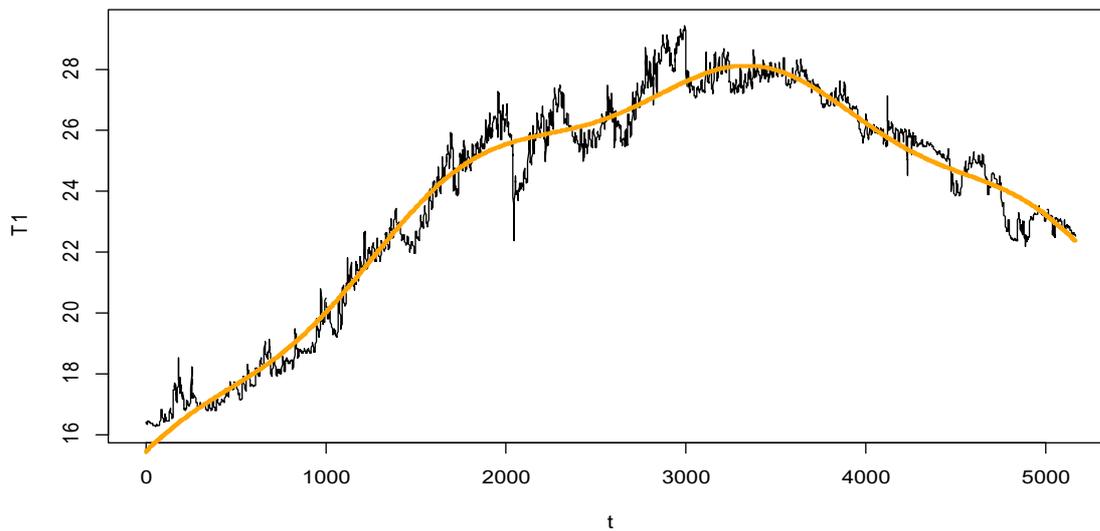


Fig. 4.3.7: La regressione armonica con f_1 e f_2 ha andamento molto simile al trend del metodo SSA.

La frequenza f_2 è $1/(69*24)$, circa 0.0006, che è la periodicità di 69 giorni, circa 2 mesi, che caratterizza l'andamento stagionale.

> l = lm(formula = T1 ~ z1 + w1 + z2 + w2 + z3 + w3 + z4 + w4 + z5 + w5)

Coefficients:

(Intercept)	z1	w1	z2	w2	z3
18.14494	-3.22234	8.91587	0.53559	0.12998	0.24811
	w3	z4	w4	z5	w5
0.01388	0.10513	-0.19234	0.03668	-0.17475	

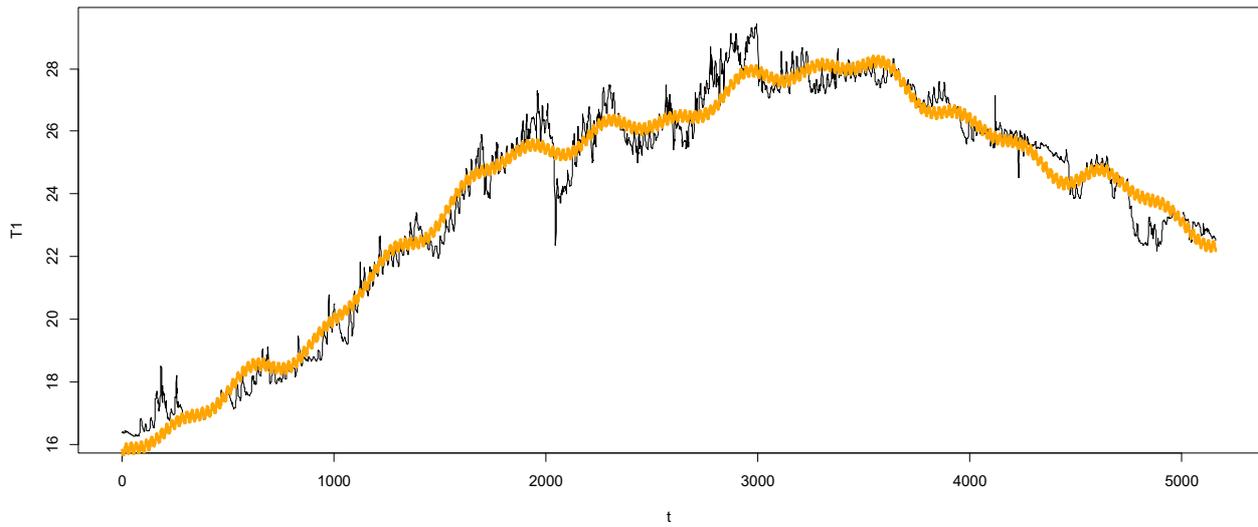


Fig. 4.3.8: La regressione armonica con 5 frequenze.

La regressione, fig. 4.3.8, effettuata rappresenta la periodicità del mese sidereo, del plenilunio e delle maree diurne.

> t.test(res) One Sample t-test

data: res

t = 3.9341e-15, df = 5159, p-value = 1

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval: -0.01509567 0.01509567

sample estimates: mean of x 3.029352e-17

Il trend, fig. 4.3.2, descrive l'andamento annuale e l'alternarsi delle stagioni. Tuttavia, la temperatura è influenzata dal moto di rivoluzione della Terra, dal moto di rotazione della Terra e i moti della Luna, in particolare, il plenilunio. Gli effetti della Luna sul mare sono noti tramite le maree diurne con periodicità di 24 ore. E' interessante notare l'importanza del plenilunio, che produce variazioni nel flusso di calore e, quindi, della temperatura del Mar Mediterraneo.

```
>s2=ssa(P1);
```

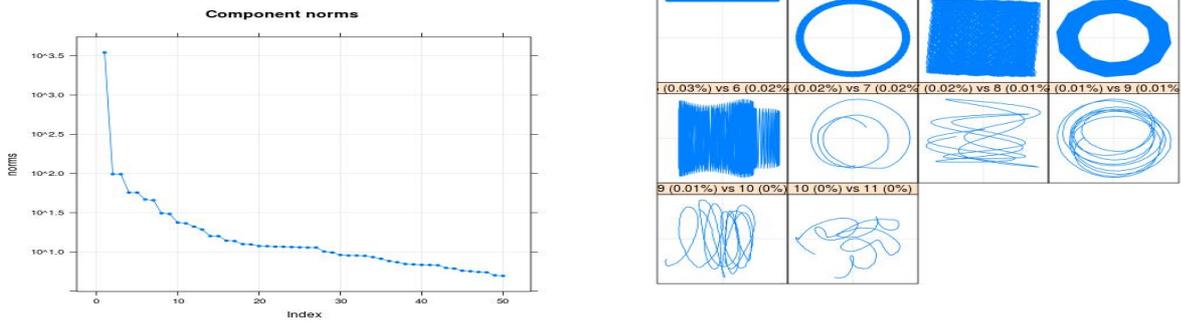


Fig. 4.3.9: Gli autovalori e autovettori di P_1 .

```
>r2=reconstruct(s2,groups=list(trend=c(1,10,11,12,13),sea1=c(2:3),sea2=c(4:5),sea3=c(6:7),sea4=c(8:9),rumore=c(14:50)));
```

```
>trend2=r2$trend; sea1=r2$sea1; sea2=r2$sea2; sea3=r2$sea3; sea4=r2$sea4; rumore2=r2$rumore;
```

Rappresento il trend e i dati sperimentali, fig. 4.3.10:

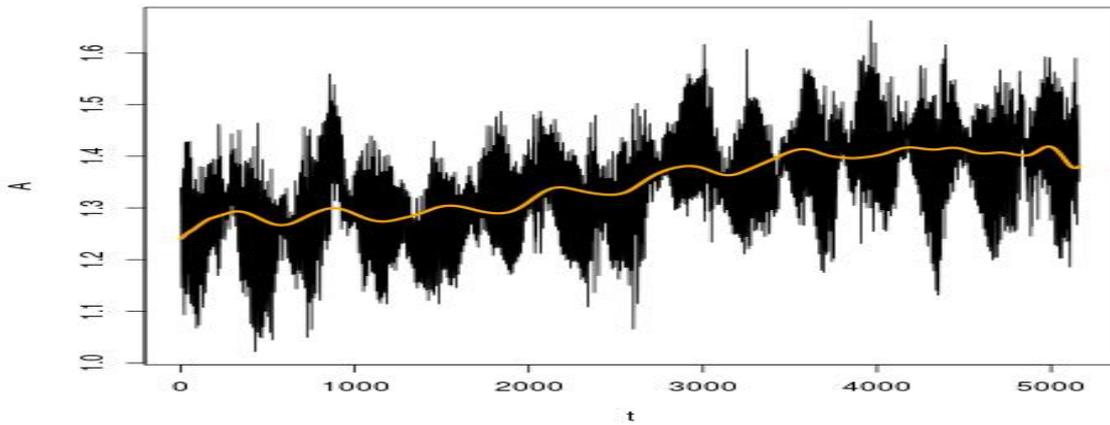


Fig. 4.3.10: I valori sperimentali e il trend in arancio.

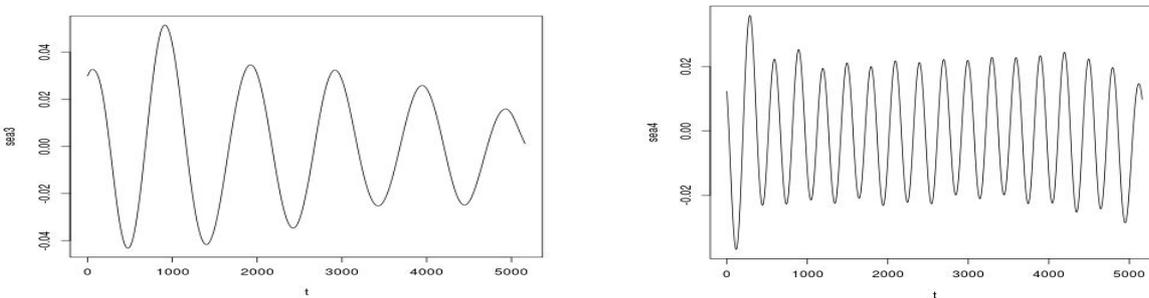


Fig. 4.3.11: La terza e la quarta componente della stagionalità, la prima e la seconda stagionalità hanno frequenze particolari.

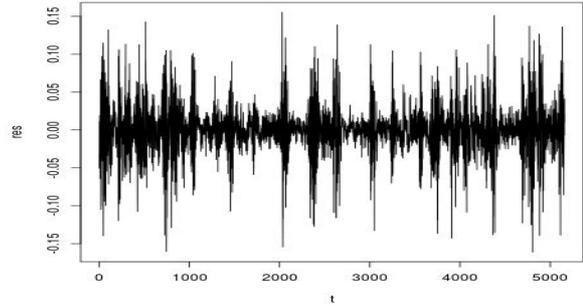
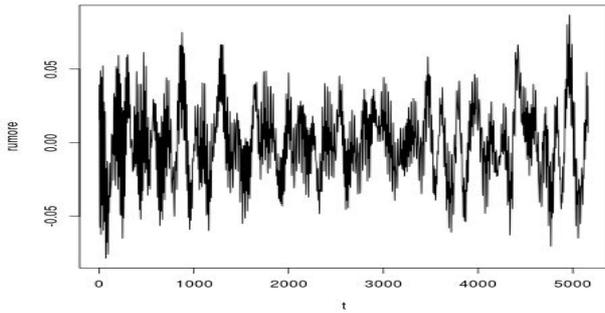


Fig. 4.3.12: Il rumore e i residui della regressione molto prossimi allo zero.

I residui della regressione sono, fig. 4.3.12:

```
> res = P1 - (trend + sea1 + sea2 + sea3 + sea4 + rumore)
```

```
> t.test(res) One Sample t-test
```

data: res

$t = -0.020677$, $df = 5159$, $p\text{-value} = 0.9835$

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval: -0.0008267018 0.0008094451

sample estimates: mean of x -8.628349e-06

Il valor medio dei residui è zero. L'acf è buono.

La regressione è buona, fig. 4.3.13.

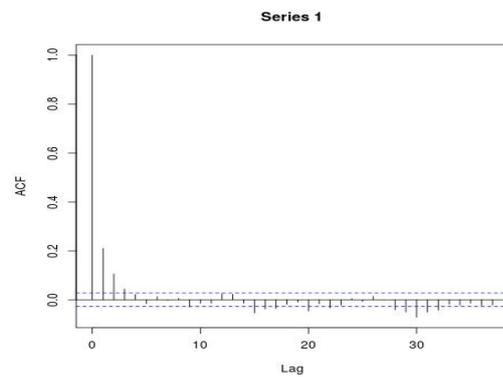
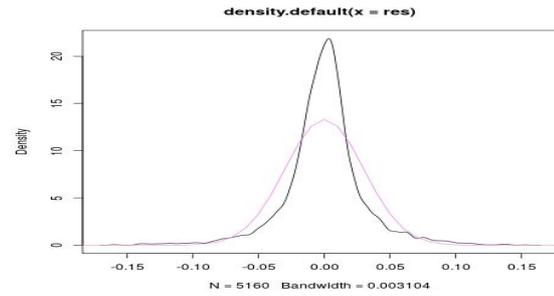


Fig. 4.3.13: La funzione di distribuzione evidenzia dei residui non gaussiani. L'acf ha qualche oscillazione, ma è buono.

```
>s3=ssa(T2);
```

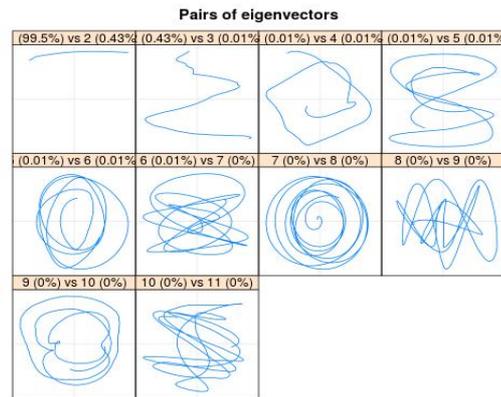
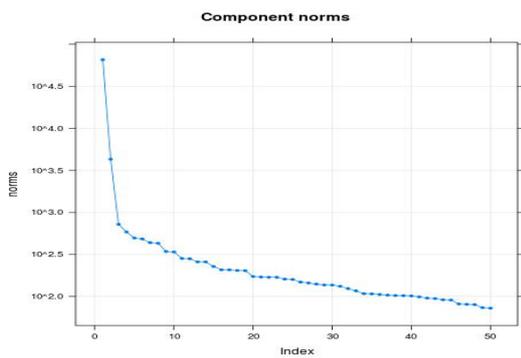


Fig. 4.3.14: Gli autovalori e gli autovettori di T_2 .

```
>r3=reconstruct(s3,groups=list(trend=c(1:4),sea1=c(5:6),sea2=c(7:8),sea3=c(9:10),sea4=c(11:12),sea5=c(13:14),
rumore=c(15:50)));
```

```
>trend3=r3$trend; sea1=r3$sea1;sea2=r3$sea2; sea3=r3$sea3, sea4=r3$sea4; sea5=r3$sea5; rumore=r3$rumore;
```

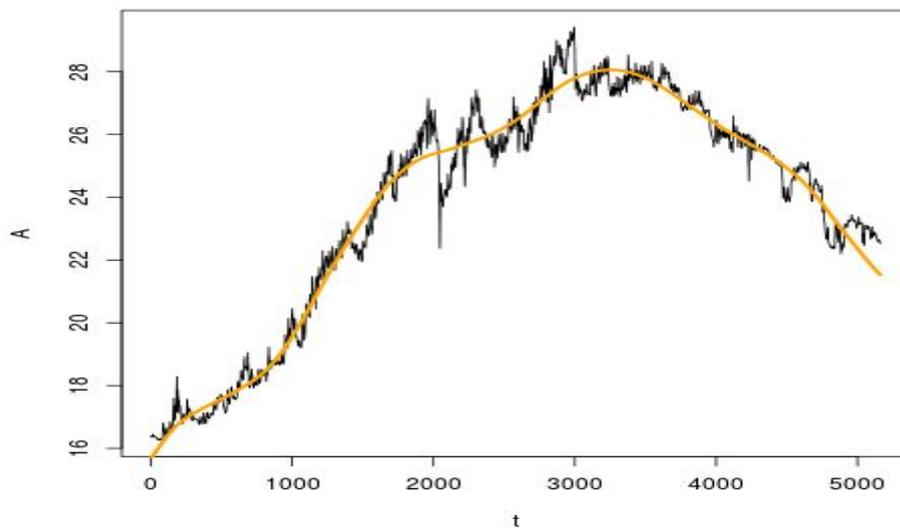


Fig. 4.3.15: I valori sperimentali e il trend.

Rappresento il trend, fig. 4.3.15, in primis, per evidenziare il buon accordo con i dati. Le componenti del sea e il rumore evidenziano le stagionalità e il rumore, dovuto alla misurazione in sé e alle variazioni del flusso di calore.

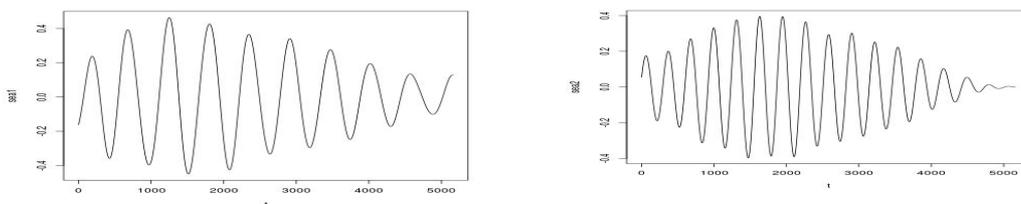


Fig. 4.3.16: Le due componenti della stagionalità visibilmente modulate.

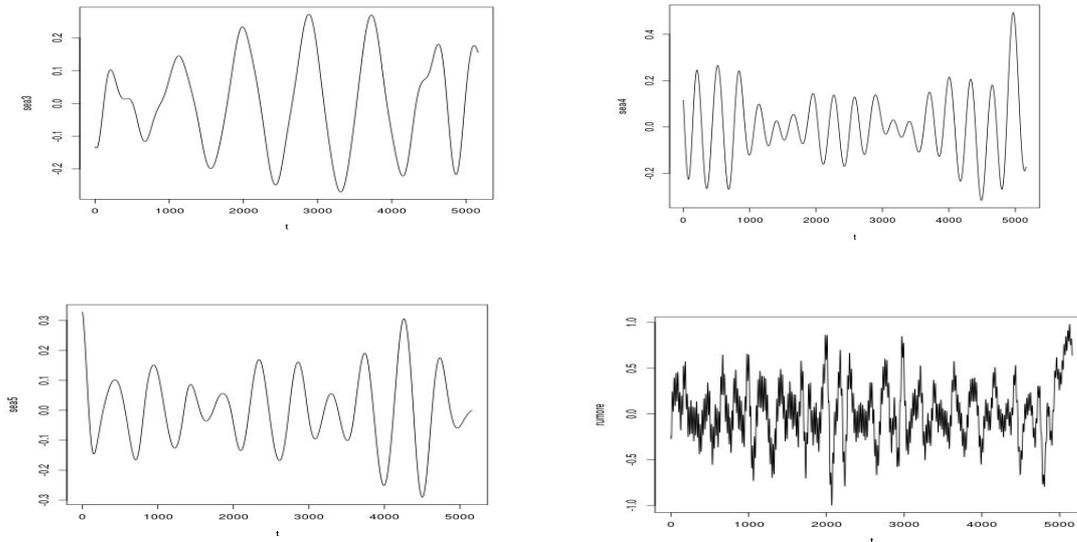


Fig. 4.3.17: La terza, quarta, la quinta componente della stagionalità e il rumore.

`> t.test(res)`

One Sample t-test

data: res

t = 1.6358, df = 5159, p-value = 0.1019

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval: -0.0008912855 0.0098733984

sample estimates: mean of x 0.004491056

La regressione è buona.

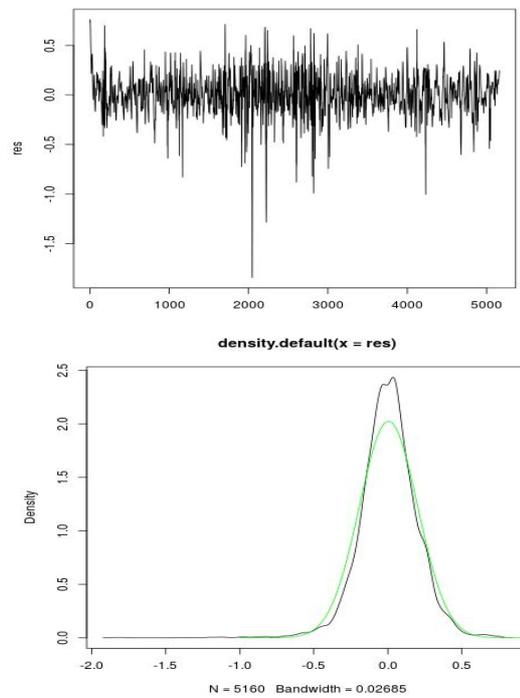


Fig. 4.3.18: I residui della regressione e la funzione di distribuzione dei residui confrontata con la gaussiana.

Il valor medio di T_1 è maggiore di quello di T_2 , perché all'aumentare della profondità diminuisce la temperatura, determinata dal calore proveniente dal Sole. Ciò è valido, in generale, a meno di correnti calde, nella zona poco al di sotto del livello superficiale. Per individuare le differenze sottraggo: $T_1 - T_2$ (A). Ottengo il seguente grafico:

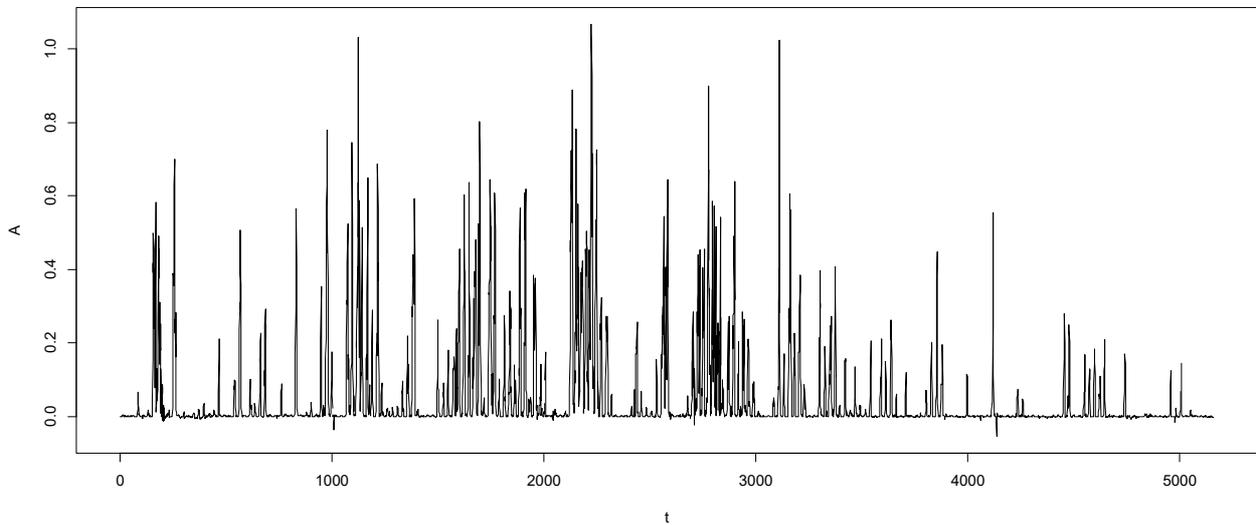


Fig. 4.3.19: $T_1 - T_2$ in funzione di t , noto che è maggiore di zero, tranne in qualche raro punto.

`> t.test(A) One Sample t-test`

`data: A`

`t = 28.851, df = 5159, p-value < 2.2e-16`

`alternative hypothesis: true mean is not equal to 0`

`95 percent confidence interval: 0.04507680 0.05164929`

`sample estimates: mean of x 0.04836305`

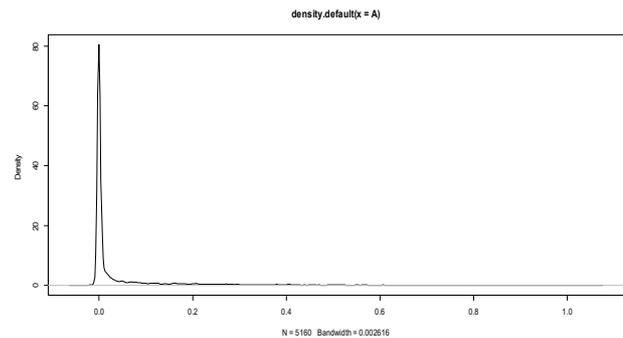


Fig. 4.3.20: La funzione di distribuzione della differenza tra temperature.

Il valor medio non è zero, le due temperature sono distinte. La correlazione di A è 1, la covarianza è 0.01449937. La funzione di distribuzione è molto piccata, sembrando una t .student. L'acf presenta dell'oscillazione, fig. 4.3.21, ciò potrebbe significare, che ci sono delle periodicità caratteristiche.

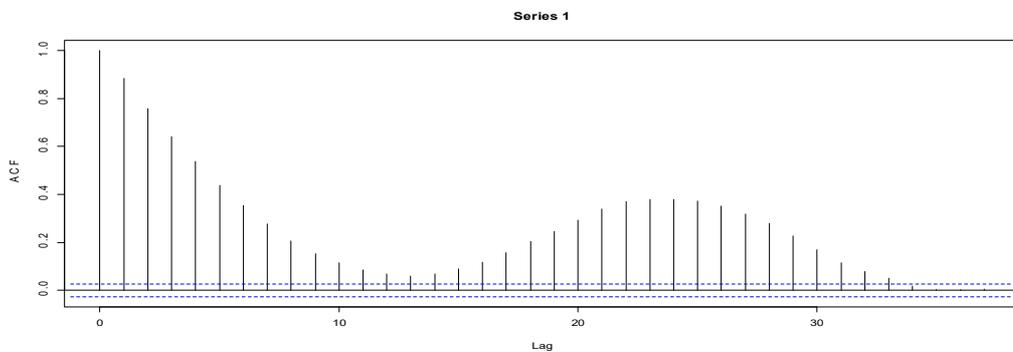


Fig. 4.3.21: L'acf della differenza tra temperature.

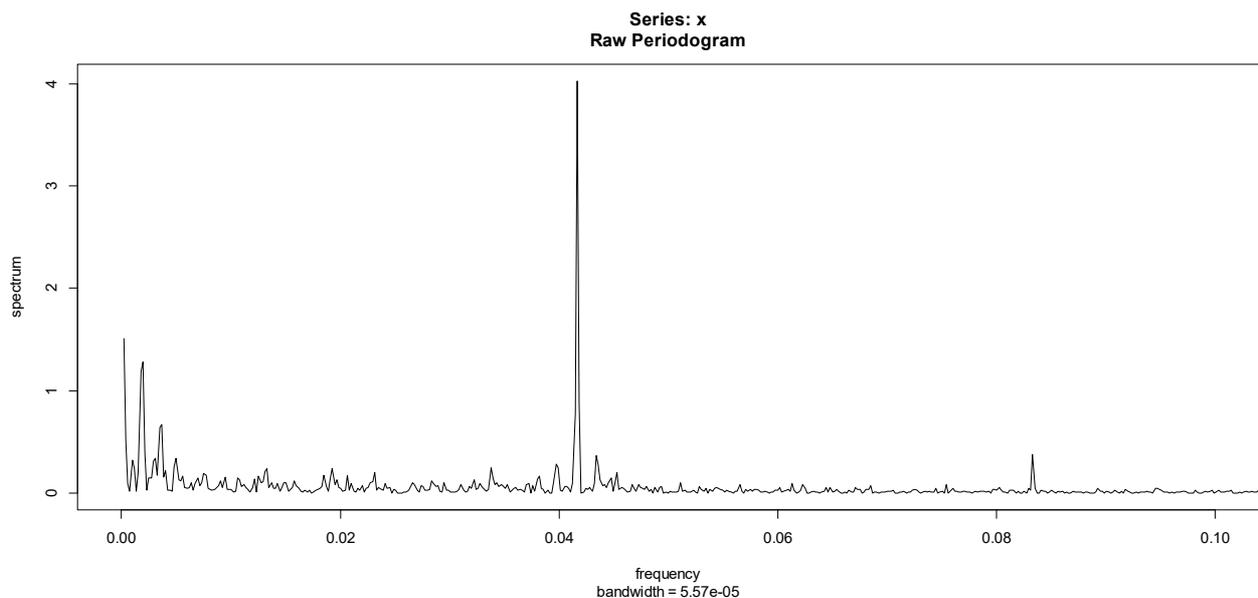


Fig. 4.3.22: Lo spettro di T1-T2.

Le frequenze sono:

$$f1=0.0018; f2=0.0034; f3=0.0416; f4=0.0833.$$

Il periodo è:

$$\tau_1=23,1 \text{ giorni}; \tau_2=12,3 \text{ giorni}; \tau_3=24 \text{ ore}; \tau_4=12 \text{ ore}.$$

I 24 giorni sono circa un mese sidereo lunare. Le 24 ore sono legate alla marea diurna, le 12 ore alla marea semidiurna. Di conseguenza, le variazioni tra le due temperature sono legate alle periodicità del movimento Terra-Sole e Terra-Luna. La varianza della T₁ è maggiore della varianza della T₂:

$$> \text{var}(T1)=13.13525;$$

$$> \text{var}(T2)=13.06996;$$

La varianza dei residui è:

$$> \text{var}(res1)=0.3740484;$$

$$> \text{var}(res2)=0.3440525$$

In quanto, le variazioni della temperatura al diminuire della profondità aumentano.

`>s=ssa(P2)`

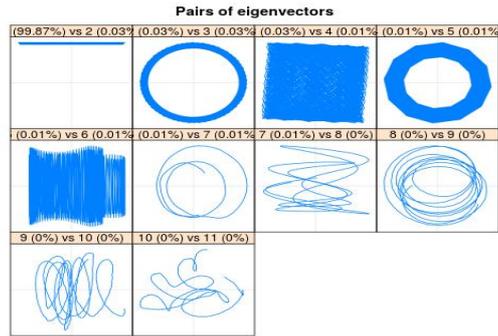
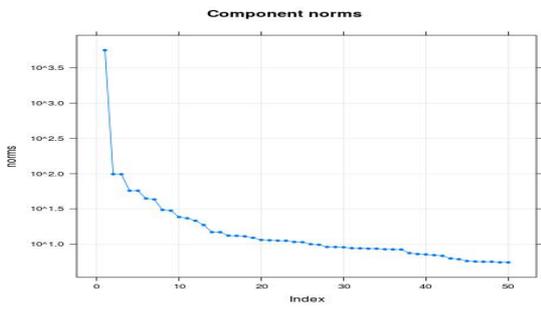


Fig. 4.3.23 : Gli autovalori e gli autovettori di P_2 .

`>r=reconstruct(s,groups=list(trend=c(1,10,11,12,13),sea1=c(2:3),sea2=c(4:5),sea3=c(6:7),sea4=c(8:9),
rumore=c(14:50)))`

Il trend sui dati è, fig. 4.3.24:

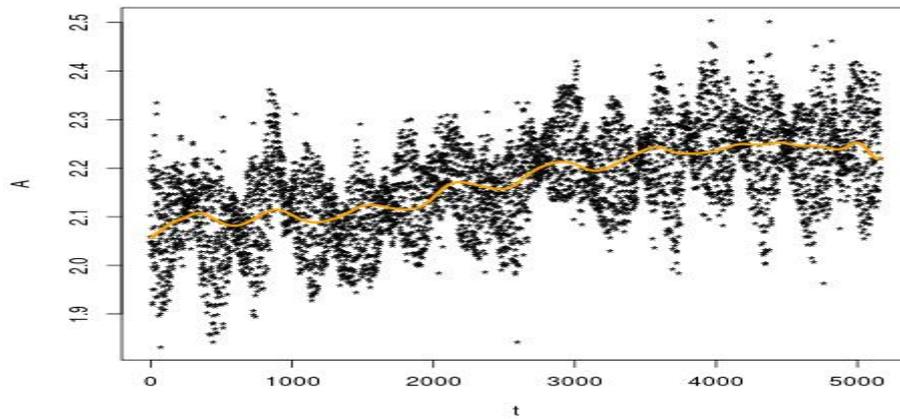


Fig. 4.3.24: I valori sperimentali e il trend con intersezione dell'asse y in circa 2.

Rappresento i dati sperimentali con trend+sea3+sea4, fig.4.3.25.

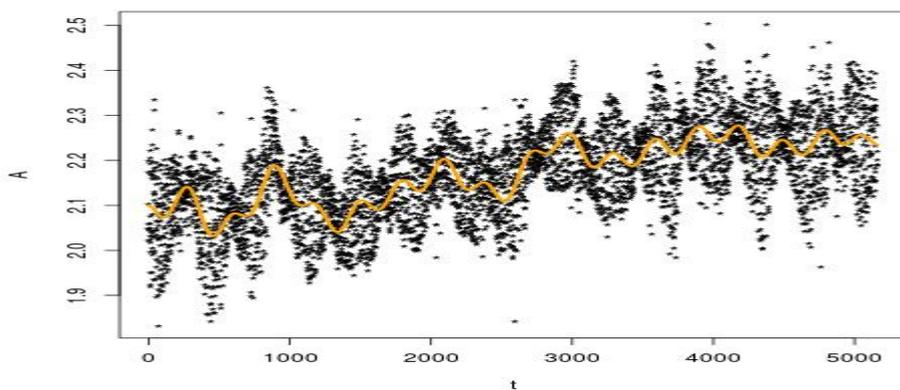


Fig. 4.3.25: I valori sperimentali e il trend che ha oscillazioni corrispondenti ad ogni picco della curva.

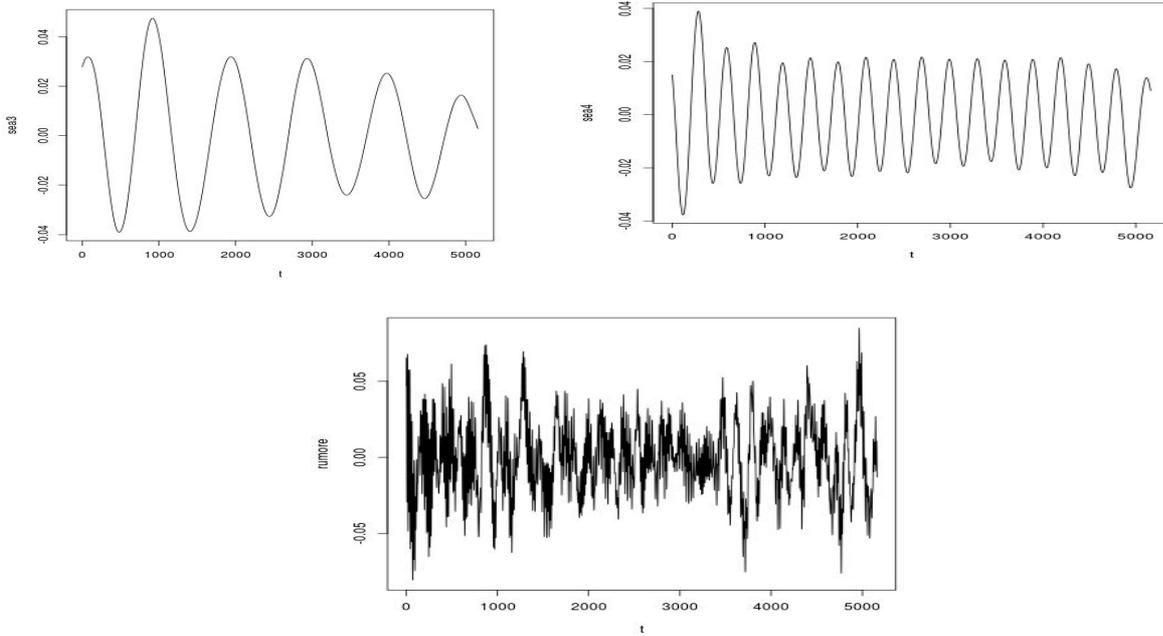


Fig. 4.3.26: La terza, la quarta componente della stagionalità e il rumore.

I residui sono, fig.4.3.27:

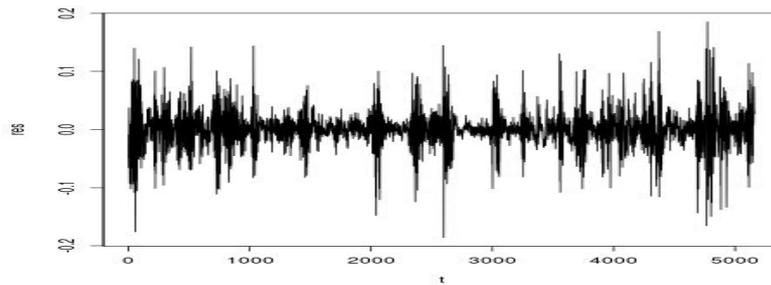


Fig. 4.3.27: I residui della regressione sono di circa 10^{-5} .

```
> t.test(res)
One Sample t-test
data: res
t = -0.15128, df = 5159, p-value = 0.8798
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval: -0.0008305771 0.0007115720
sample estimates: mean of x -5.950255e-05
```

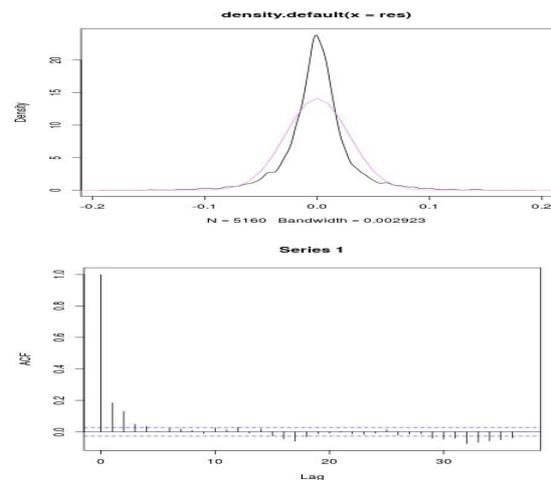


Fig. 4.3.28: La funzione di distribuzione dei residui rapportata alla curva normale. L'acf dei residui è buona.

L'acf dei residui, fig. 4.3.28, è buono. Il valor medio dei residui è zero, con un p circa 0.9. I residui non sono gaussiani. La regressione è buona.

1. **Frequency analysis, modulation and noise**, Stanford Goldman, McGraw-Hill Book company, inc. 1948 New York;
2. **Using R for Introductory Statistics**, John Verzani, Chapman & Hall/Crc, 2000;
3. **Probabilità, variabili aleatorie e processi stocastici**, A. Papoulis, Celum stellatum Boringhieri, 1973, Torino;
4. **Elements of large-sample theory**, E.L. Lehmann, Springer, 1999;
5. **Improvements to NOAA's historical Merged Land Ocean Surface Temperature Analysis (1880,2006)**, T.M. Smith, Reynold, (2007);
6. **Un sistema nazionale unico per i dati climatologici**, ARPA, Desiato, 2007, SCIA, su www.arpa.it;
7. **Guide to climatological practices**, WMO, 2011;
8. **Encyclopedia of global environmental change: The Earth system: physical and chemical dimensions of global environmental change**; Micheal Mac Cracken, Jhon S. Perry, WILEY, CANADA,2002;
9. **Methods of correlation and regression. Linear and curvilinear**, Madercai Ezekiel, Karl A. Fox, New York, John Wiley&Sons, Inc.1959;
10. **Singular spectrum analysis. A new tool in time series analysis**, James B. Elsner and Anastasias A. Tsonis, Plenum Press, New York, 1996;
11. **Algebra lineare**, Serge Lang, Bollati Boringhieri, Torino, 1970;
12. **Introduzione all'elaborazione dei dati sperimentali**, C.Cametti, A. Di Biasio, CISU(RM),1994;
13. **La scienza, La Terra**, vol.3, La biblioteca di Repubblica, Grandi opere di UTET cultura, Istituto Geografico De Agostini, Novara, 2005;
14. **Random signals and noise**, Wilbur B. Davenport and William L. Root, Mac Graw Hill Book Company, Inc., NY, 1958;
15. **Una rivoluzione nella teoria dei numeri**, Gauss, Geni della Matematica, RBA Italia, 2017;
16. **Time series analysis and its applications with R examples**, Third edition, R. H. Shumway, D.S. Stoffer, Springer text in Statistics, New York, 2010;
17. **Basic Singular Spectrum Analysis and forecasting with R**, Nina Golyandina, Anton Korobeynikov, Department of Statistical Modelling, Faculty of Mathematics and Mechanics, St. Petersburg State University, Universitetskyypr28, St.Petersburg 198504, Russia;
18. **Guide to Meteorological Instruments and Methods of Observation**, 2014 edition Updated in 2017, WMO-No. 8;

19. **El Nino and Southern Oscillation Multiscale and global and regional impacts**, H.Diaz and V.Markgraf
Cambrige University Press, 2000 Usa;

20. **La meda elastica di Lampedusa: caratteristiche tecniche**, V. ARTALE e coll., RT/2016/13/ENEA.

Ringrazio la NASA per la disponibilità dei dati sperimentali utilizzati e l'ENEA, in particolare la stazione di Lampedusa e i prof. Artale e Marullo per il lavoro di misurazione.

Ringrazio tutti coloro che mi hanno aiutata nelle difficoltà, in particolare, prof. Cammarota, che mi ha seguita con pazienza.

Ringrazio il Dipartimento di Matematica di La Sapienza.

Adelaide