



**Università di Pisa**

Corso di Laurea in  
Artificial Intelligence and Data Engineering

# **Analisi delle variazioni del contenuto di calore oceanico**

**Candidato**

Franco Terranova

ANNO ACCADEMICO 2021/2022

# INDICE

INTRODUZIONE.....	1
1.DATASET.....	1
1.1.CONTENUTO DEL DATASET.....	1
1.2.ESPLORAZIONE DEL DATASET.....	2
2.DECOMPOSIZIONE.....	3
2.1.ANALISI DEI RESIDUI.....	4
2.2.DECOMPOSIZIONE CON STAGIONALITA' NON STAZIONARIA.....	5
3.1.PREVISIONE CON HOLT-WINTERS.....	5
3.1.1.CONFRONTO CON PARAMETRI SCELTI MANUALMENTE.....	6
3.2.METODI REGRESSIVI PER SERIE STORICHE.....	7
3.2.1.YULE-WALKER.....	7
3.2.2.METODO DEI MINIMI QUADRATI.....	8
CONCLUSIONI.....	9

# INTRODUZIONE

A causa della sua elevata energia termica, l'oceano ha assorbito circa il 90% del riscaldamento globale. L'aumento del calore interno all'oceano è uno dei fattori maggiormente responsabili dell'innalzamento del livello globale del mare. Altri effetti catastrofici da esso generati sono lo sbiancamento dei coralli, lo scioglimento delle principali calotte glaciali terrestri e i cambiamenti nella salute e nella biochimica degli oceani. Il cambiamento climatico sta creando un nuovo fronte politico internazionale: quello delle piccole isole, particolarmente minacciate dall'innalzamento del livello del mare. Si ipotizza che l'analisi sia stata commissionata dall'ONU per la predizione delle variazioni del contenuto di calore oceanico con lo scopo di far fronte al problema dell'innalzamento del livello dei mari.

## 1. DATASET

L'analisi è stata svolta sul dataset della NASA relativo ai cambiamenti del contenuto di calore dell'oceano, scaricabile in formato JSON al seguente link: [https://climate.nasa.gov/system/internal\\_resources/details/original/2363\\_ECCO\\_V4r4\\_OHC\\_ZJ\\_YYYY-MM\\_anom.json](https://climate.nasa.gov/system/internal_resources/details/original/2363_ECCO_V4r4_OHC_ZJ_YYYY-MM_anom.json).

### 1.1. CONTENUTO DEL DATASET

Il dataset contiene informazioni sulle variazioni mensili del contenuto di calore oceanico per l'intera colonna d'acqua (dalla sommità al fondo dell'oceano), integrando osservazioni da satelliti e vari dispositivi di misurazione oceanica, inclusi strumenti di conducibilità-temperatura-profondità (CTD), galleggianti di profilazione Argo, batitermografi di consumo (XBT), array di ormeggio strumentati e profiler collegati al ghiaccio (ITP).

La densità areale del contenuto di calore oceanico tra due livelli di profondità è definita utilizzando un integrale definito:

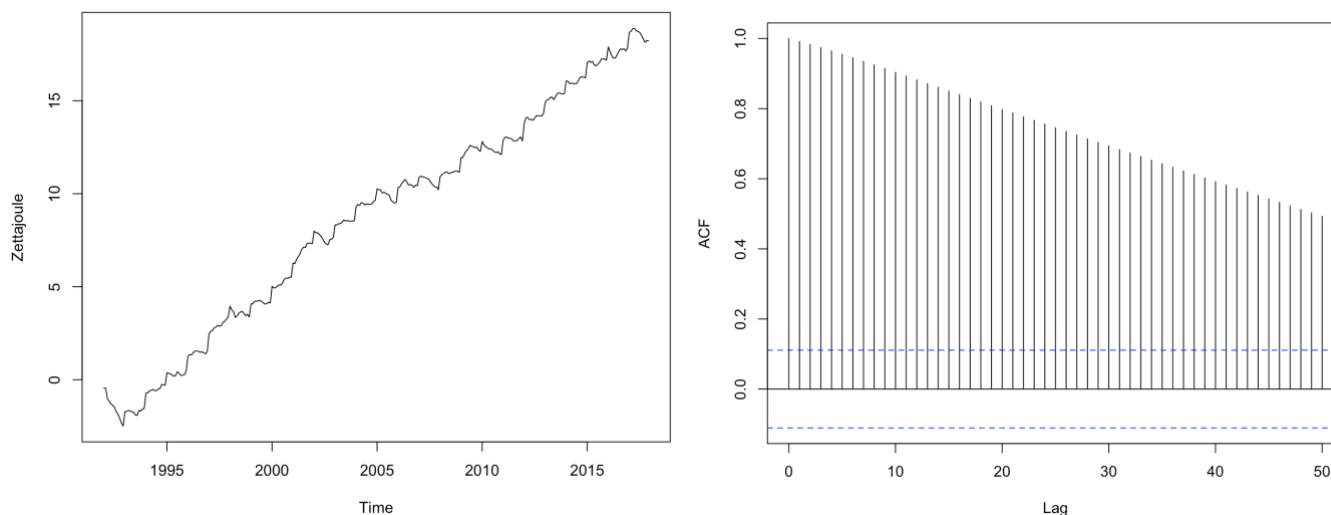
$$H = \rho c_p \int_{h_2}^{h_1} T(z) dz$$

dove  $\rho$  è la densità dell'acqua del mare,  $c_p$  è il calore specifico dell'acqua marina,  $h_2$  è la profondità inferiore,  $h_1$  è la profondità superiore e  $T(z)$  è il profilo di temperatura. Integrando questa densità sulla regione tridimensionale dell'oceano si ottiene il contenuto di calore totale.

La tabella analizzata contiene i dati dal gennaio 1992 al dicembre 2017 e le misurazioni presenti nel dataset sono espresse in zettajoule =  $10^{21}$  joule (un sestilione di joule).

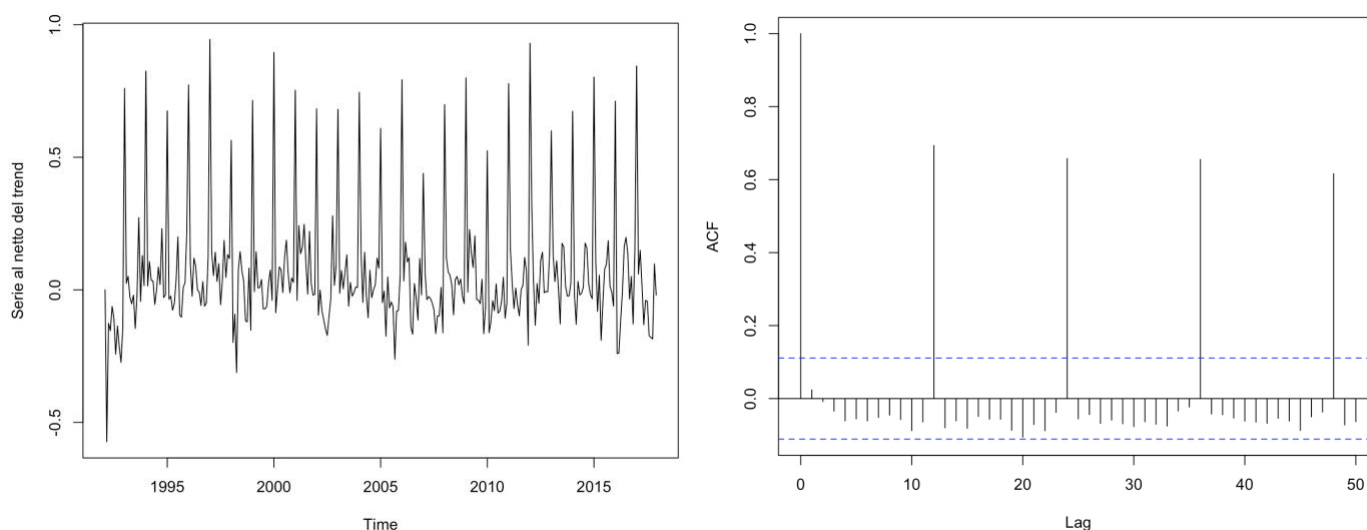
## 1.2. ESPLORAZIONE DEL DATASET

Effettuando una prima esplorazione dei dati di tipo grafico, notiamo la presenza di una possibile lieve stagionalità annuale e di un primo lieve trend discendente seguito da un forte trend ascendente. Notiamo un calo delle variazioni di calore in corrispondenza dell'inverno 1992, che può essere dovuto al calo delle temperature durante quell'inverno o all'iniziale scarsa accuratezza nella misurazione dei valori, visto che si trattava proprio del primo anno di misurazione.



La presenza di un trend ascendente dominante è confermata anche dalla funzione di autocorrelazione, che tuttavia non sembra evidenziare la presenza di stagionalità.

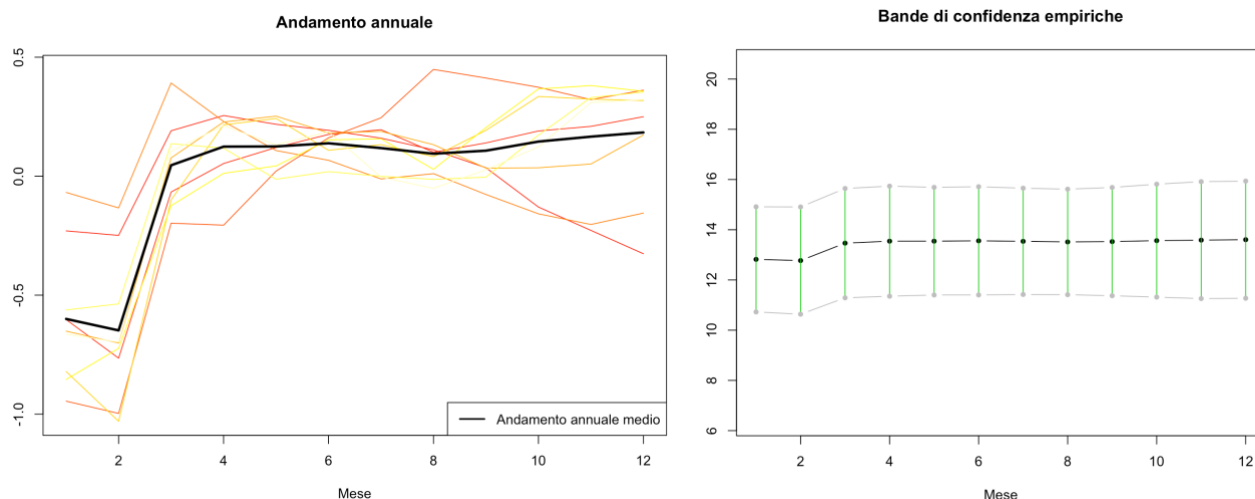
Osservando la serie al netto del trend scopriamo che quest'ultimo aveva nascosto la stagionalità, che risulta essere annuale proprio come era presumibile aspettarsi, viste le variazioni di temperatura nelle varie stagioni dell'anno.



Un'altra conferma può essere ottenuta confrontando gli andamenti in anni diversi.

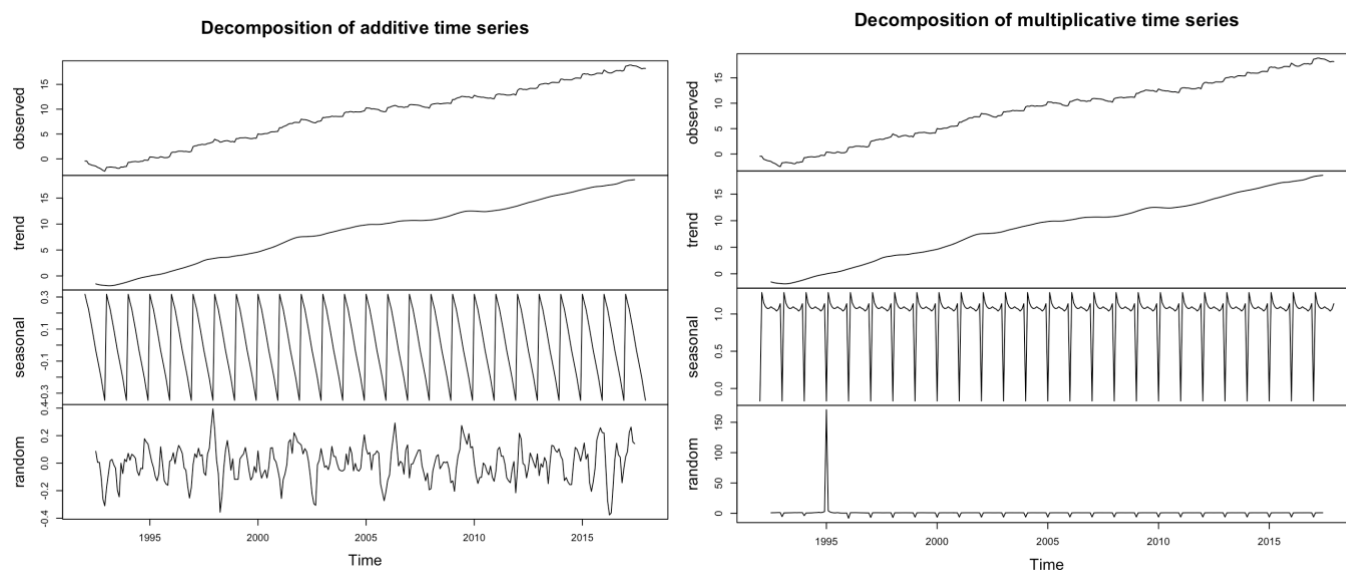
Nel confronto dei grafici dei diversi anni, dopo aver ricentrato i periodi, si evince un aumento delle variazioni tipicamente in corrispondenza del mese di marzo.

Nel grafico a destra troviamo invece le bande di confidenza empiriche a  $\pm$  una deviazione standard empirica.



## 2. DECOMPOSIZIONE

Confermata la presenza di stagionalità, iniziamo a decomporre la serie esaminando la decomposizione additiva e la decomposizione moltiplicativa, per capire quale riesce a catturare meglio il comportamento della serie.



La componente di rumore della decomposizione additiva sembra poco strutturata.

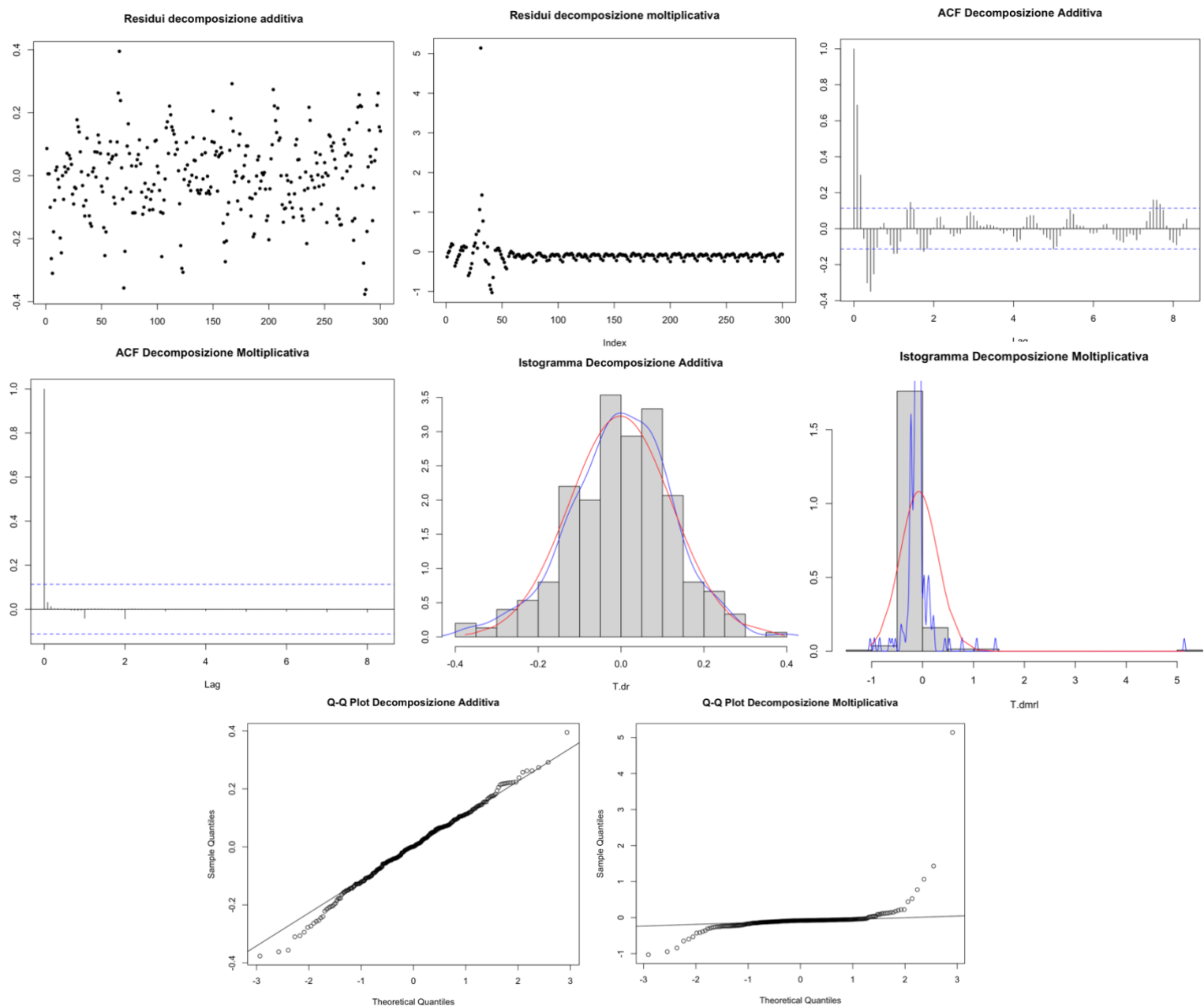
Si riscontra la presenza di una maggiore struttura, invece, nella componente di rumore della decomposizione moltiplicativa.

Inoltre, in corrispondenza del 1995, quest'ultima presenta un elevato picco.

È evidente, inoltre, come la componente di stagionalità estratta, risulta essere in entrambi i casi di lieve entità.

## 2.1. ANALISI DEI RESIDUI

Esploriamo i residui della decomposizione additiva e moltiplicativa della serie.



Varianza non spiegata

p-value del test di  
Shapiro-Wilk

Variabilità della funzione  
di autocorrelazione

Decomposizione additiva	0.0004	0.2194	0.2877
Decomposizione moltiplicativa	0.1539	$< 2.2e^{-16}$	0.2009

L'unico punto a favore della decomposizione moltiplicativa è dato dalla funzione di autocorrelazione, dove i valori risultano essere meno pronunciati e praticamente quasi tutti all'interno delle bande di confidenza, considerazione confermata anche dalla variabilità della funzione di autocorrelazione.

In ogni caso, l'analisi di residui suggerisce che la decomposizione additiva risulta essere molto più appropriata, vista la minore struttura presente nei residui additivi.

## 2.2. DECOMPOSIZIONE CON STAGIONALITA' NON STAZIONARIA

Si è tentato di effettuare una decomposizione con stagionalità non uniforme utilizzando valori di seasonal window dispari via via crescenti. Dall'analisi dei residui è emerso il miglioramento dei risultati all'aumentare della dimensione della finestra, ma quest'ultimi sono ben tuttavia peggiori rispetto a quelli ottenuti con la decomposizione con stagionalità uniforme. Per questo motivo si è deciso di continuare l'analisi prendendo in considerazione la decomposizione con stagionalità stazionaria.

### 3.1. PREVISIONE CON HOLT-WINTERS

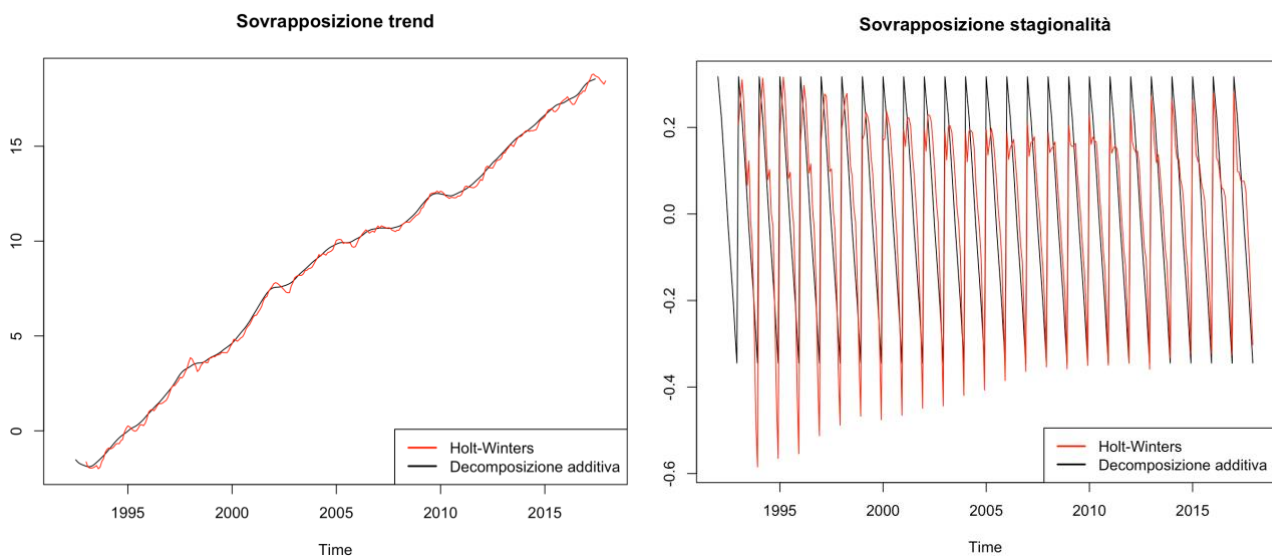
Applichiamo il metodo additivo di Holt-Winters con trend e stagionalità.

I parametri calcolati automaticamente dal metodo di Holt-Winters, minimizzando la varianza dei residui, forniscono un valore di  $\beta$  molto conservativo ( $\beta = 0.043676$ ), un valore di  $\alpha$  che predilige l'influenza del passato recente ( $\alpha = 0.87869$ ) e un valore di  $\gamma$  che tiene conto esclusivamente del passato recente ( $\gamma = 1$ ).

I valori forniti dal software indicano quindi un'analisi molto conservativa per quanto riguarda la direzione, mentre indicano un'analisi molto fedele al passato recente per quanto riguarda gli altri due valori.

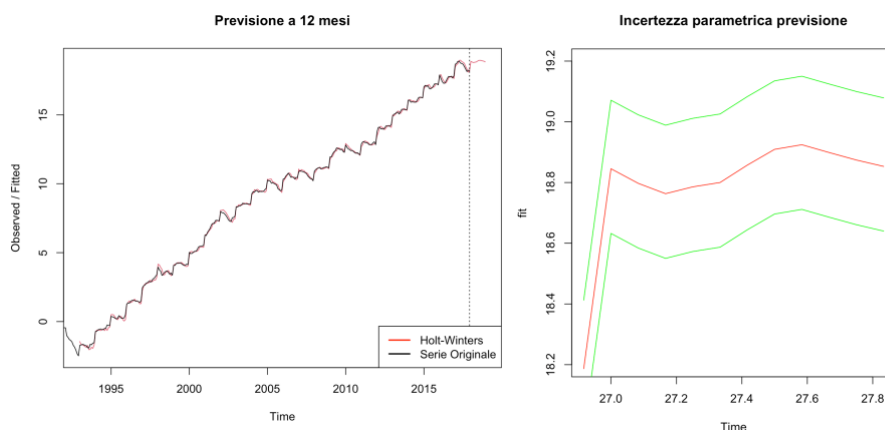
Per determinare le condizioni iniziali si è fatto uso di una regressione lineare che ha suggerito i valori  $l.start = -1.06$  e  $b.start = 0.04$ .

Sovrapponendo i risultati del modello additivo a quelli forniti dalla decomposizione additiva notiamo una forte somiglianza tra il grafico delle intercette e quello del trend, mentre nel grafico della stagionalità vediamo ampiezze leggermente diverse rispetto a quelle fornite dalla decomposizione.



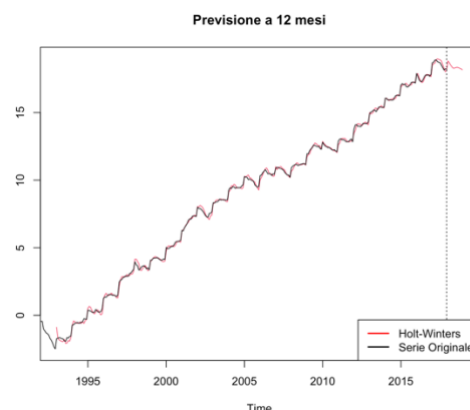
La previsione a 12 mesi effettuata con Holt-Winters suggerisce un aumento iniziale delle variazioni di calore con una successiva stabilizzazione, andamento molto simile a quello riscontrato negli anni precedenti.

In seguito ai risultati dell'analisi dei residui del modello additivo (non riportata nella relazione per semplicità), possiamo permetterci di non rigettare l'ipotesi di gaussianità, potendo quindi rappresentare l'incertezza della previsione per via parametrica.

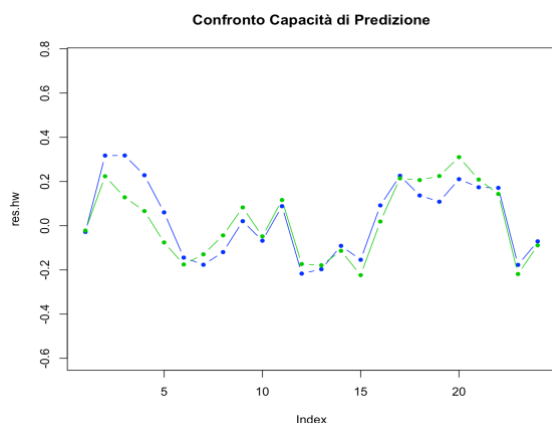


### 3.1.1. CONFRONTO CON PARAMETRI SCELTI MANUALMENTE

Si è tentato inoltre di esplorare una zona di parametri nell'intorno di quelli forniti dal software, preferendo tra questi, valori di  $\alpha$  e  $\gamma$  più conservativi rispetto ai precedenti ( $\alpha = 0.8$  e  $\gamma = 0.8$ ), a fronte di un valore di  $\beta$  meno conservativo rispetto al precedente ( $\beta = 0.2$ ). Le predizioni di questo modello risultano essere più ottimistiche rispetto a quelle del modello precedente, evidenziando un aumento iniziale simile ma con una successiva discesa seguita da una stabilizzazione.



Effettuando un confronto tra il modello con parametri scelti manualmente e il modello con parametri forniti dal software, dal confronto tra i residui (non riportato nella relazione per semplicità) è emerso come il modello con parametri forniti dal software risulta essere di gran lunga migliore rispetto all'altro, presentando tra le tante cose un valore di p-value del test di Shapiro-Wilk molto più alto (0.43 rispetto a 0.001). Il confronto tra le capacità di previsione mediante autovalidazione suggerisce invece risultati pressochè simili per i due modelli, leggermente migliori per il modello con parametri manuali.



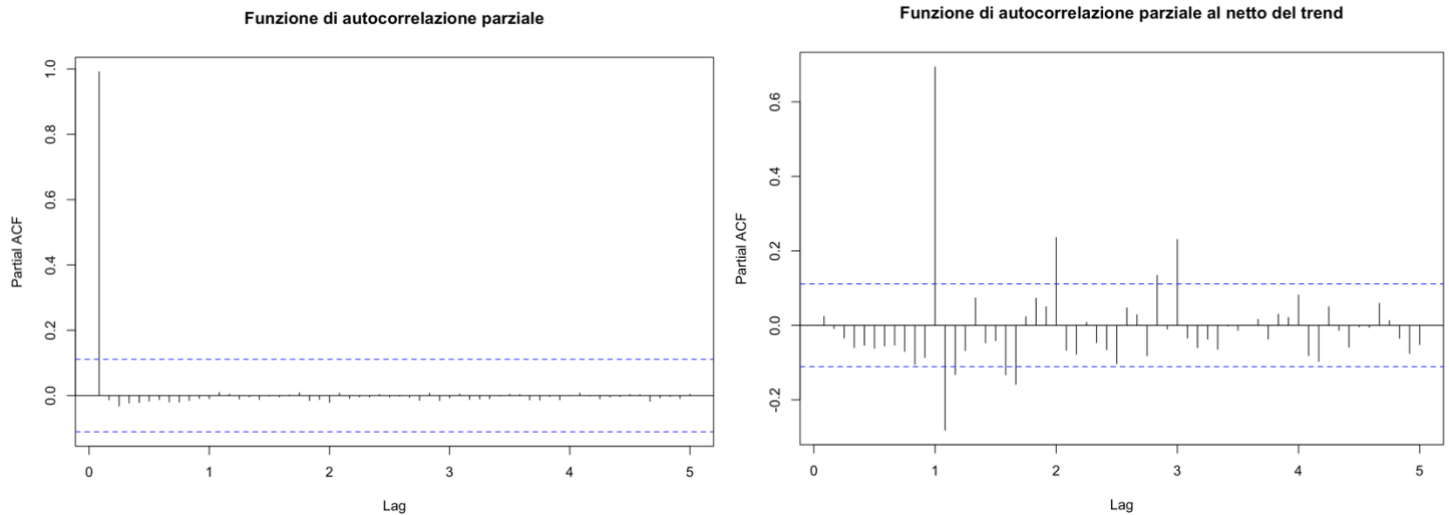
Errore Quadratico Medio	
Holt-Winters - Parametri forniti dal software	0.169
Holt Winters - Parametri manuali	0.162

Si è deciso comunque di preferire il modello con i parametri scelti dal software, vista la netta preferenza nel confronto tra i residui.



## 3.2. METODI REGRESSIVI PER SERIE STORICHE

La funzione di autocorrelazione parziale mostra una significativa dipendenza soltanto con il valore immediatamente precedente, mentre, sottraendo il trend, riscontriamo una dipendenza non nulla con i tre periodi precedenti e il valore immediatamente prima del periodo antecedente.



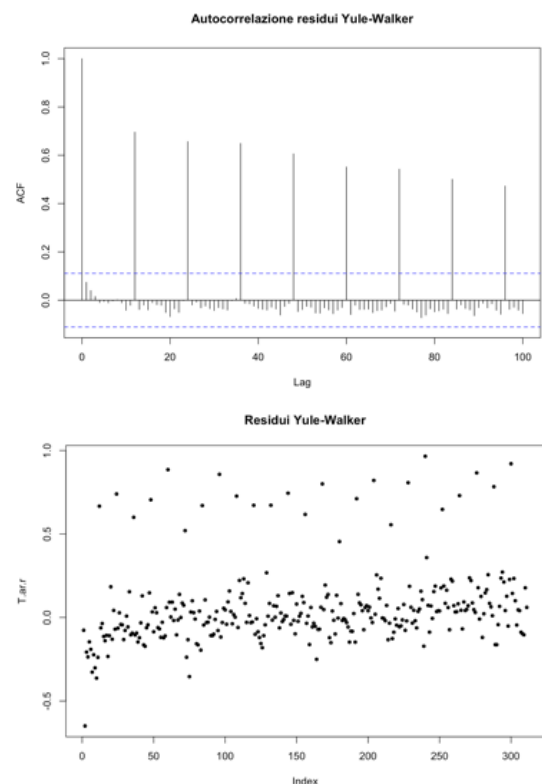
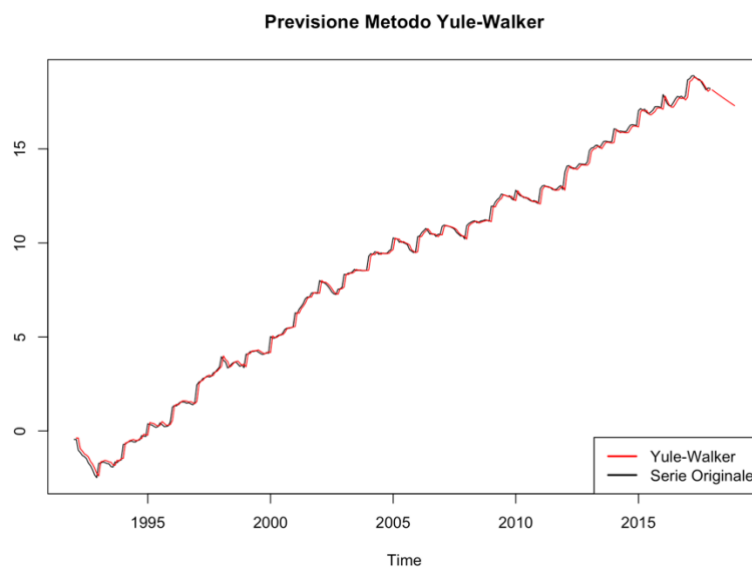
### 3.2.1. YULE-WALKER

Il modello di autoregressione di Yule-Walker decide di selezionare soltanto il lag precedente.

La previsione effettuata da questo modello risulta essere del tutto inappropriata, visto che una riduzione così netta delle variazioni, come quella prevista, risulta essere decisamente non probabile.

Questa affermazione viene confermata dai residui, che presentano una marcata struttura temporale.

Proprio per questo motivo si è deciso di non prendere in considerazione il modello di Yule-Walker.



## 3.2.2. METODO DEI MINIMI QUADRATI

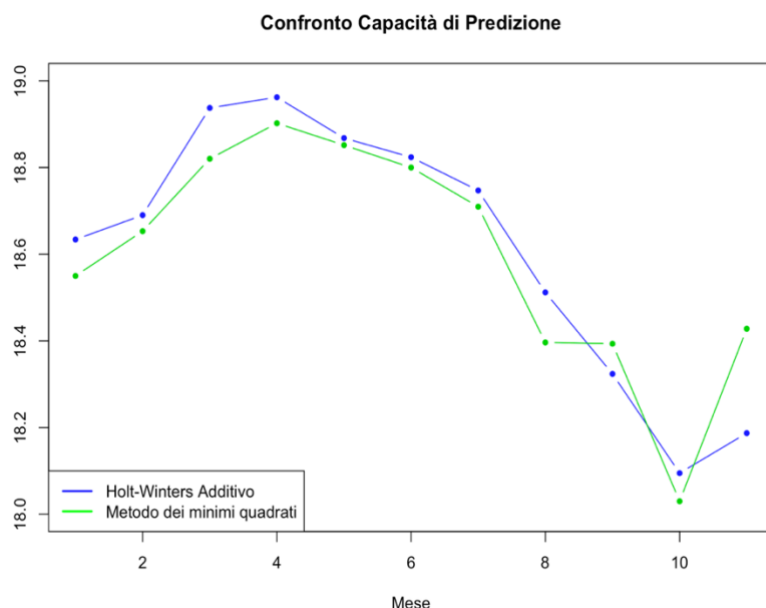
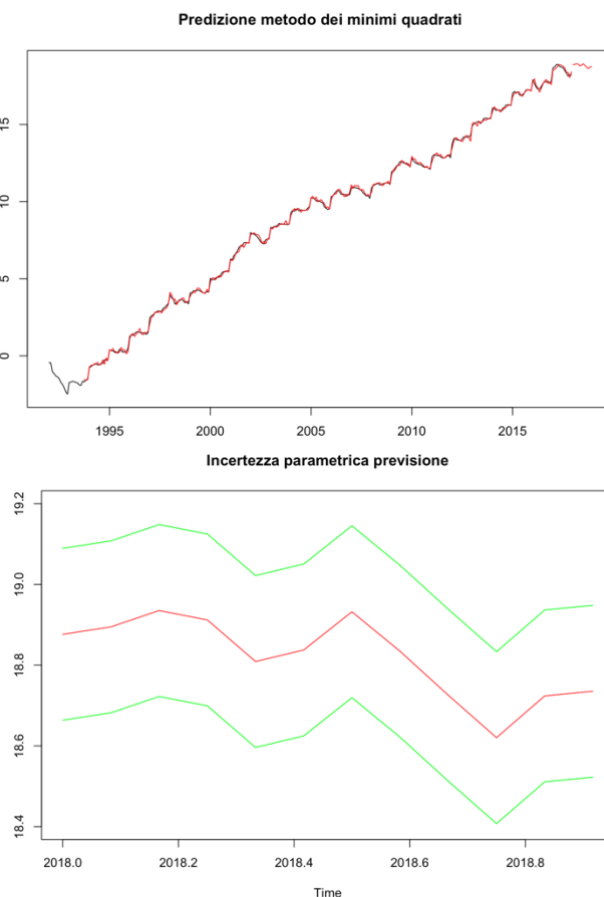
Il metodo dei minimi quadrati seleziona molti più lag, evidenziando 21 mesi come ritardo utile per l'autoregressione. Il risultato fornito dalla predizione risulta essere molto più probabile rispetto a quello fornito dal metodo di Yule-Walker, prevedendo un aumento immediato delle variazioni di calore all'inizio dell'anno, che tende poi a stabilizzarsi nei mesi restanti. L'andamento previsto risulta quindi essere molto simile a quello osservato negli anni precedenti.

I residui del modello (non riportati per semplicità) non presentano una struttura temporale e tra le tante cose hanno un p-value abbastanza elevato ( $p\text{-value} = 0.1961$ ), per questo motivo per la stima delle incertezze possiamo utilizzare un metodo parametrico.

Notiamo che la previsione rimane all'interno dei margini di previsione parametrici.

Possiamo notare, inoltre, come la previsione non discosta troppo da quella fornita da Holt-Winters.

Confrontando il metodo dei minimi quadrati con Holt-Winters (modello additivo con parametri forniti dal software), emerge una lieve preferenza per il metodo di Holt-Winters sia dal confronto tra i residui (non riportato nella relazione) che dal confronto tra le capacità di previsione tramite autovalidazione.



	Errore Quadratico Medio
Holt-Winters additivo	0.16
Metodo dei minimi quadrati	0.18

## CONCLUSIONI

In seguito alle analisi svolte, il metodo di Holt-Winters risulta essere quello più appropriato per la previsione. La previsione del metodo risulta essere molto simile all'andamento rilevato nei periodi precedenti, prevedendo un aumento della variabilità nei primi mesi dell'anno e una stabilizzazione nei mesi successivi.

Questa previsione risulta essere molto plausibile vista l'accelerazione del cambiamento climatico e del riscaldamento oceanico.

La serie studiata contiene dati fino all'inverno 2017 e la previsione effettuata riguarda l'anno 2018.

In seguito a successive ricerche, è emerso come per gli oceani il 2018 sia stato l'anno più caldo mai registrato, raggiungendo un accumulo record di calore.

La previsione effettuata risulta quindi essere coerente con quello accaduto nell'anno 2018.