

ANALISI DELLE EMISSIONI DI CO₂

G12 Memphis

Matteo Ballabio 1058828

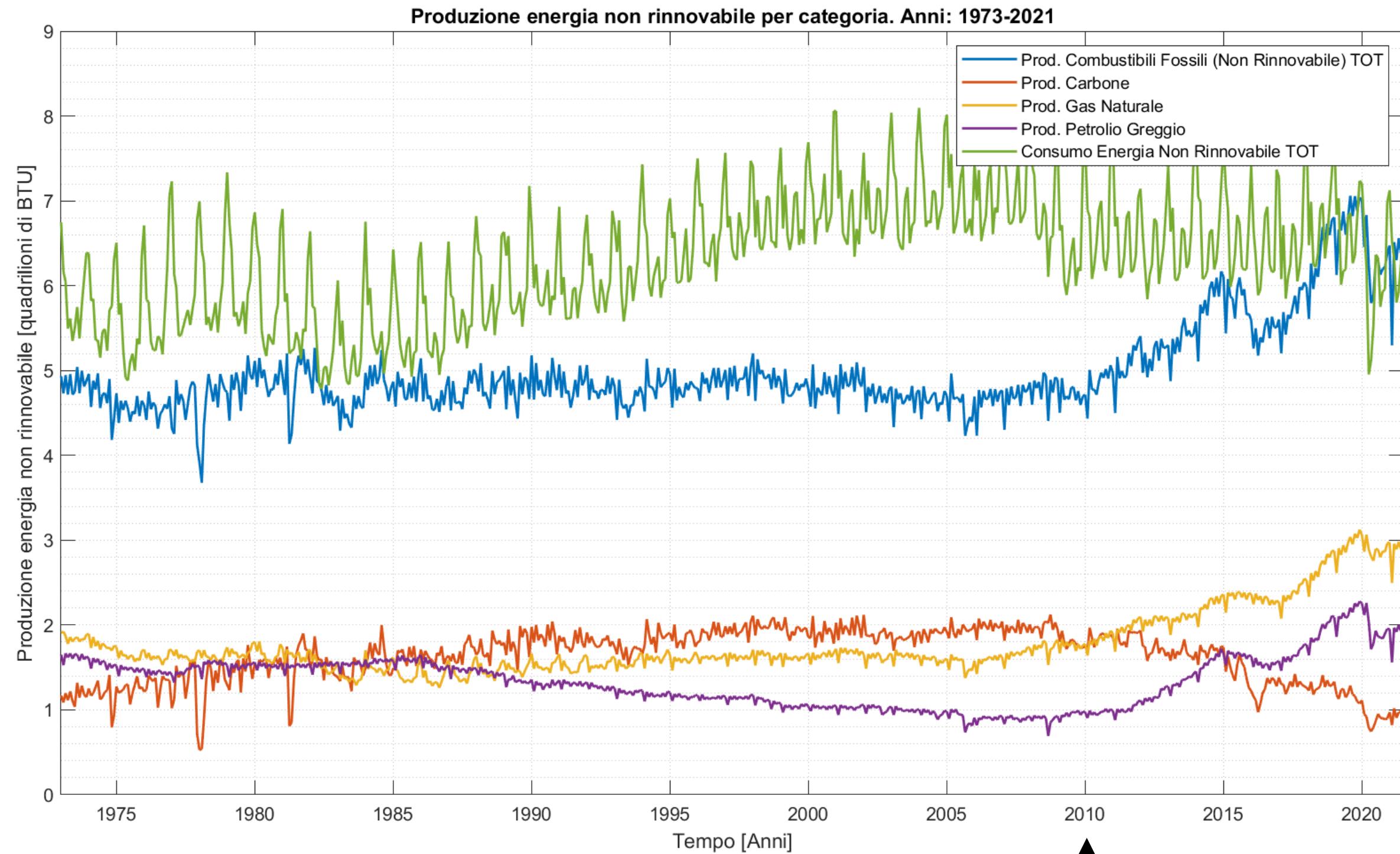
Matteo De Stefani 1080379

Lisa Galizzi 1059947

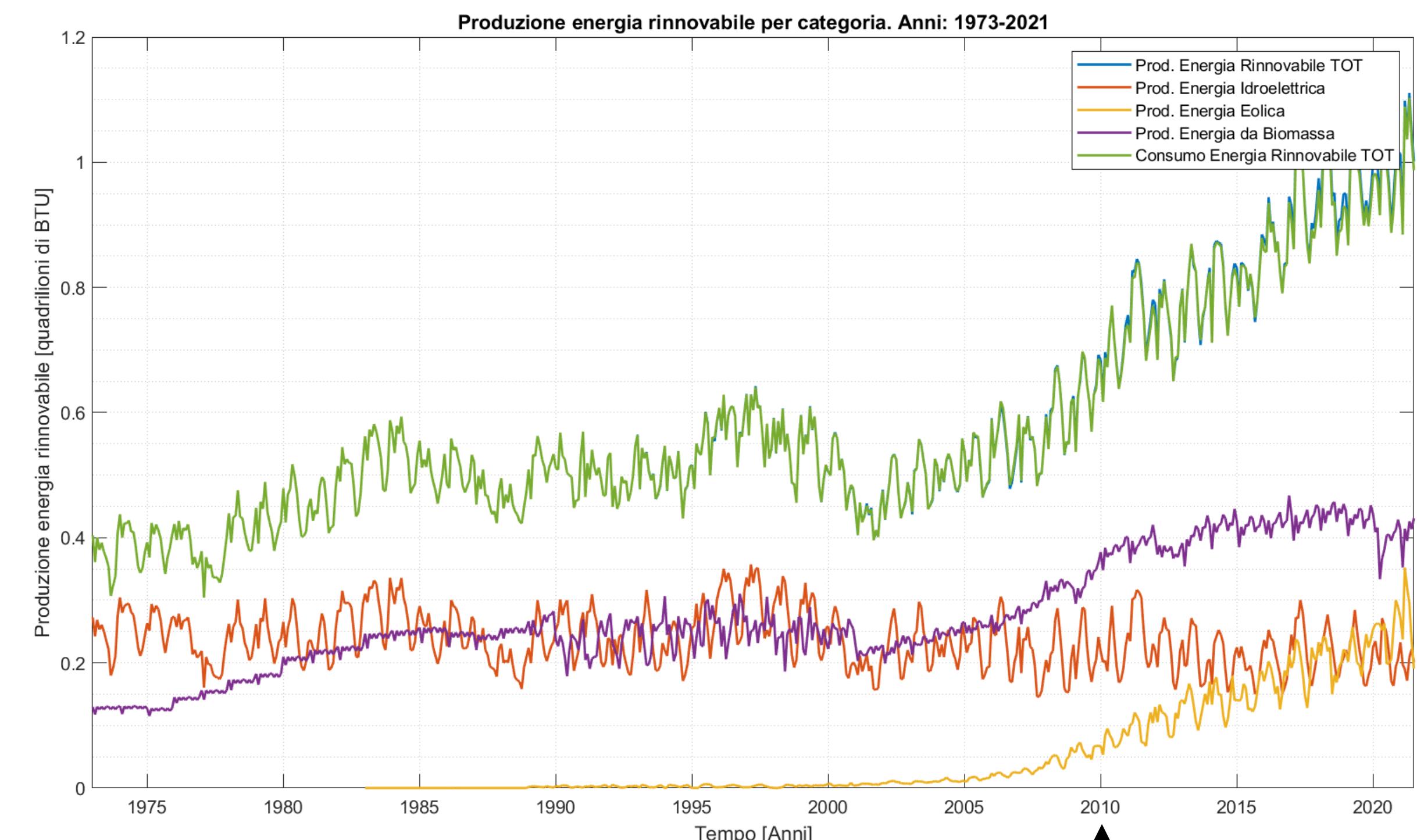
Vanessa Zani 1057577

ANALISI ESPLORATIVA

Energia non rinnovabile

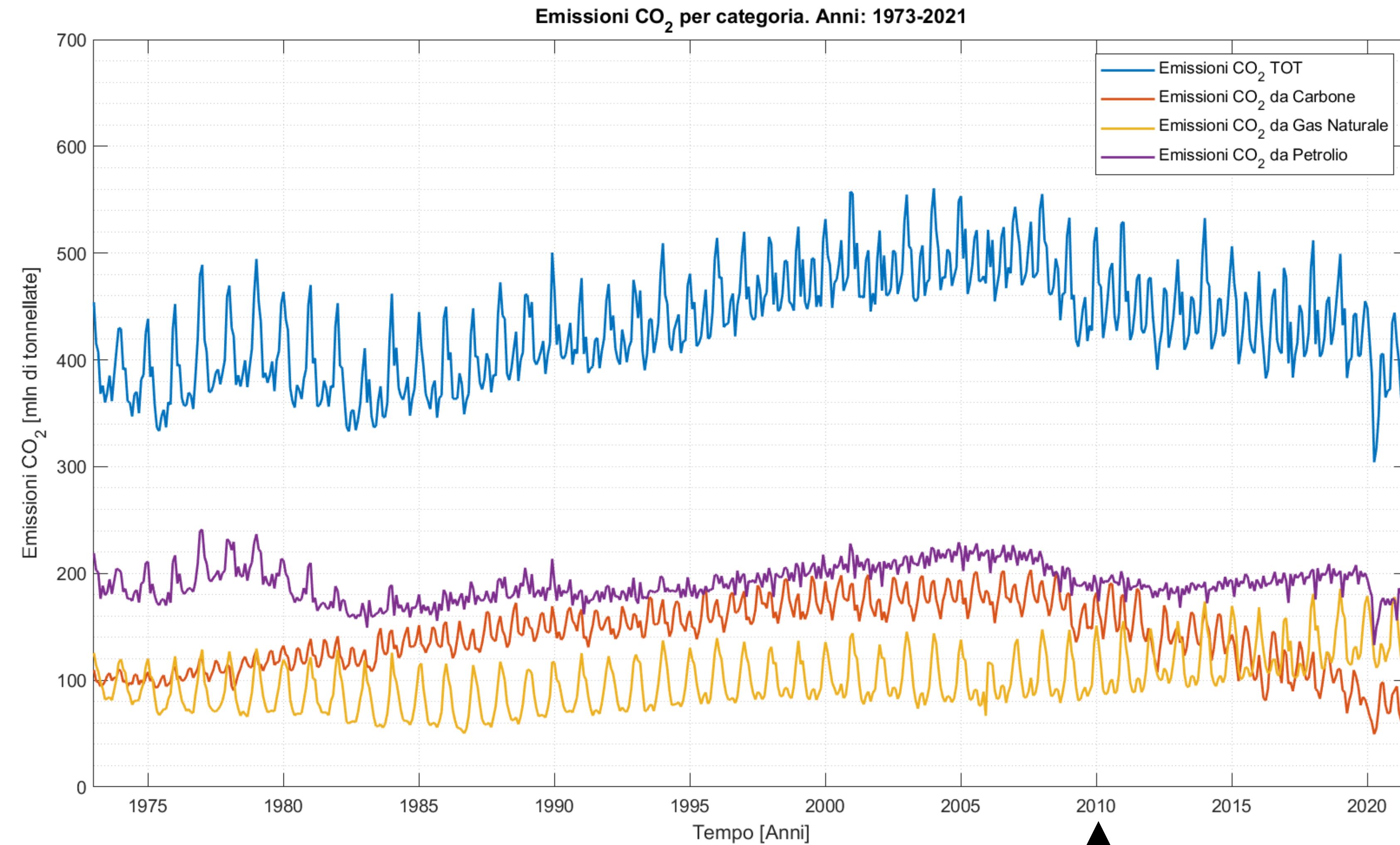


Energia rinnovabile



ANALISI ESPLORATIVA

Emissioni di CO₂



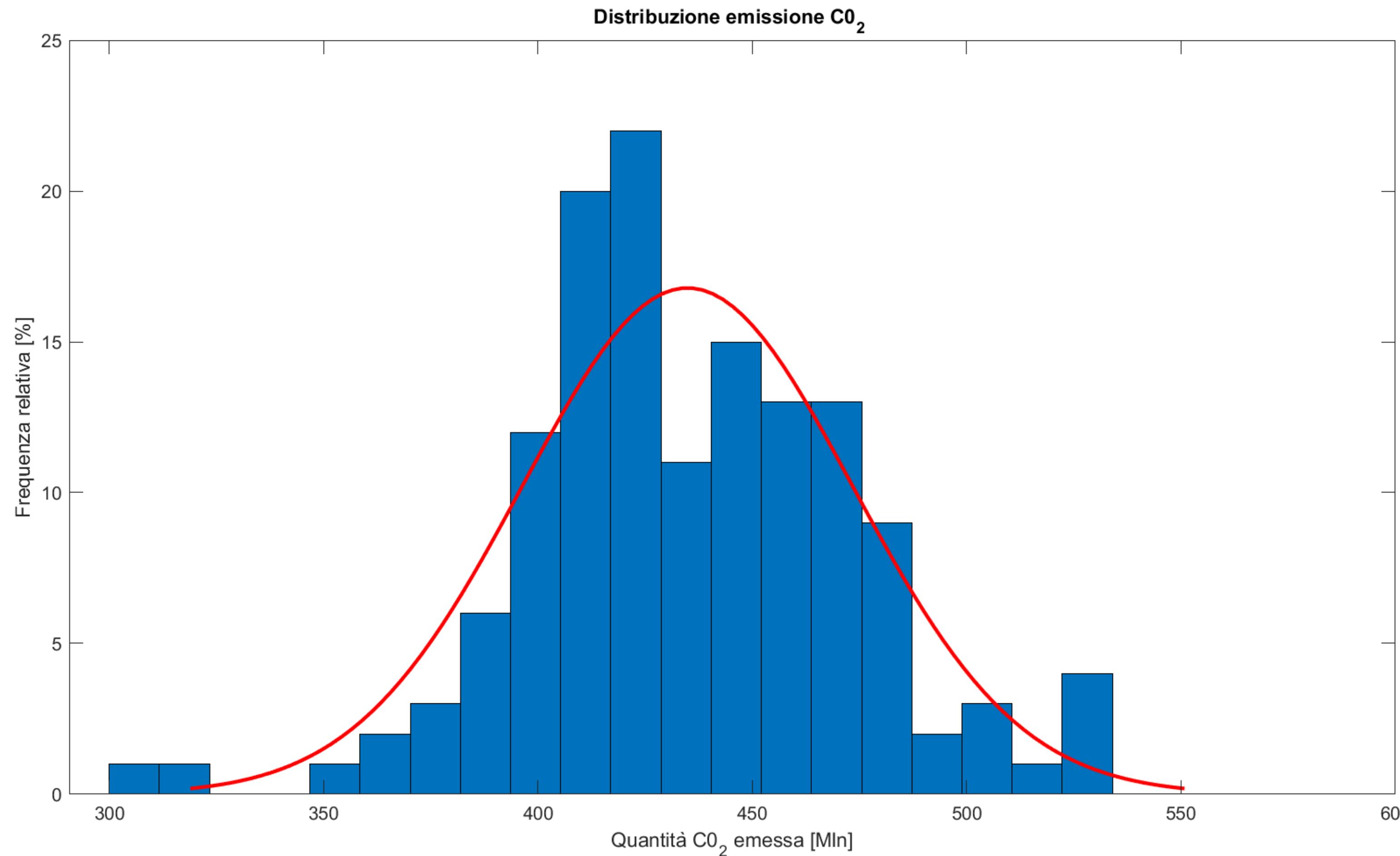
- 1. Stima delle emissioni di CO₂ basata su produzioni e consumi di fonti energetiche negli USA**
- 2. Previsione delle emissioni di CO₂ negli USA**
- 3. Relazione tra emissioni di CO₂ e HDD negli USA**
- 4. Stima delle anomalie delle temperature globali**

Dataset mensile
Gennaio 2010 - Luglio 2021

Dataset annuale
1949 - 2021

ANALISI ESPLORATIVA

Emissioni di CO₂ analisi della distribuzione



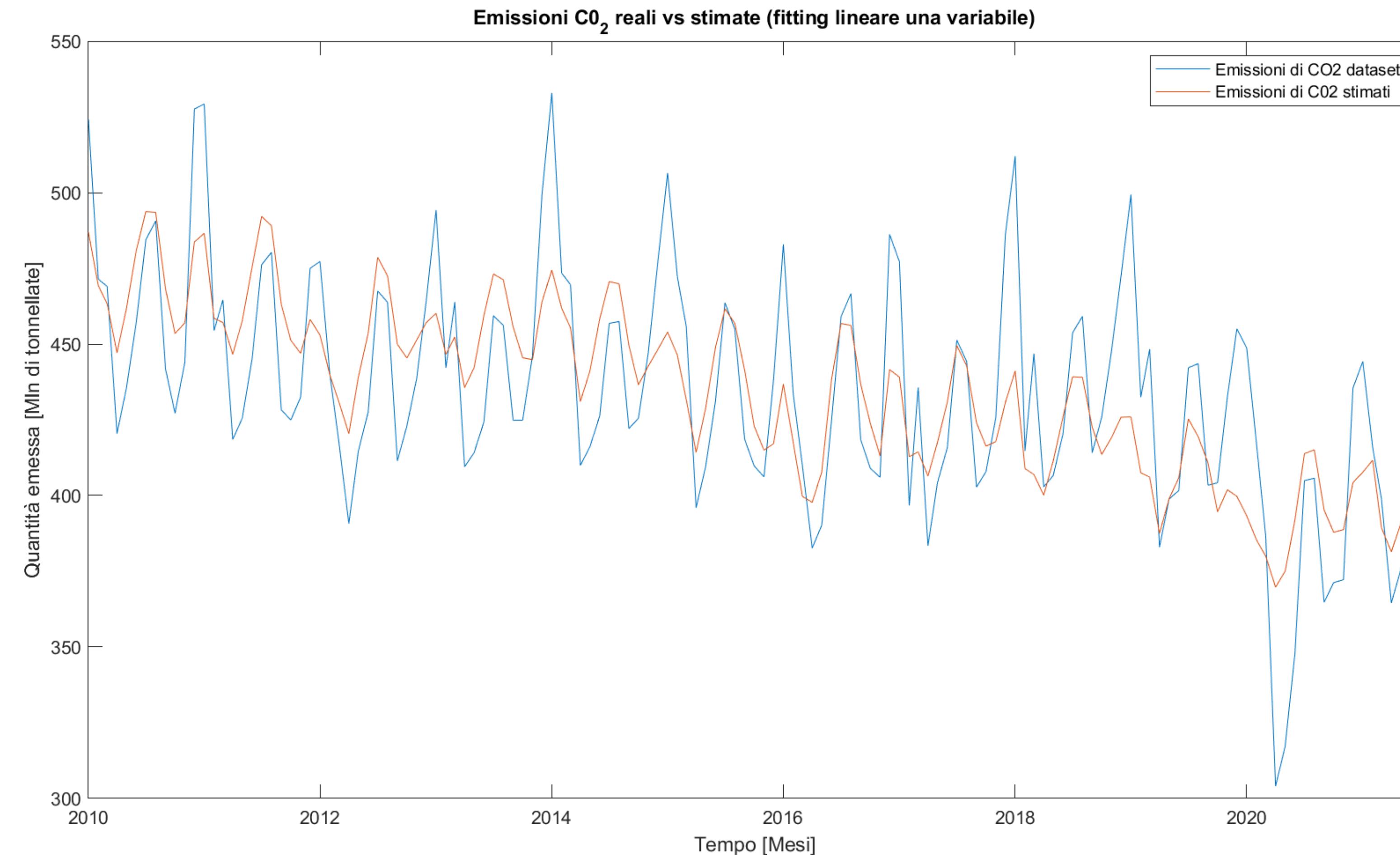
- L'ipotesi di normalità dei residui è confermata dal test di Lilliefors al 5% e da quello Bera-Jarque con livello di significatività dell'1% e del 5%



**Stima delle emissioni di CO₂ basata su produzioni
e consumi di fonti energetiche negli USA**

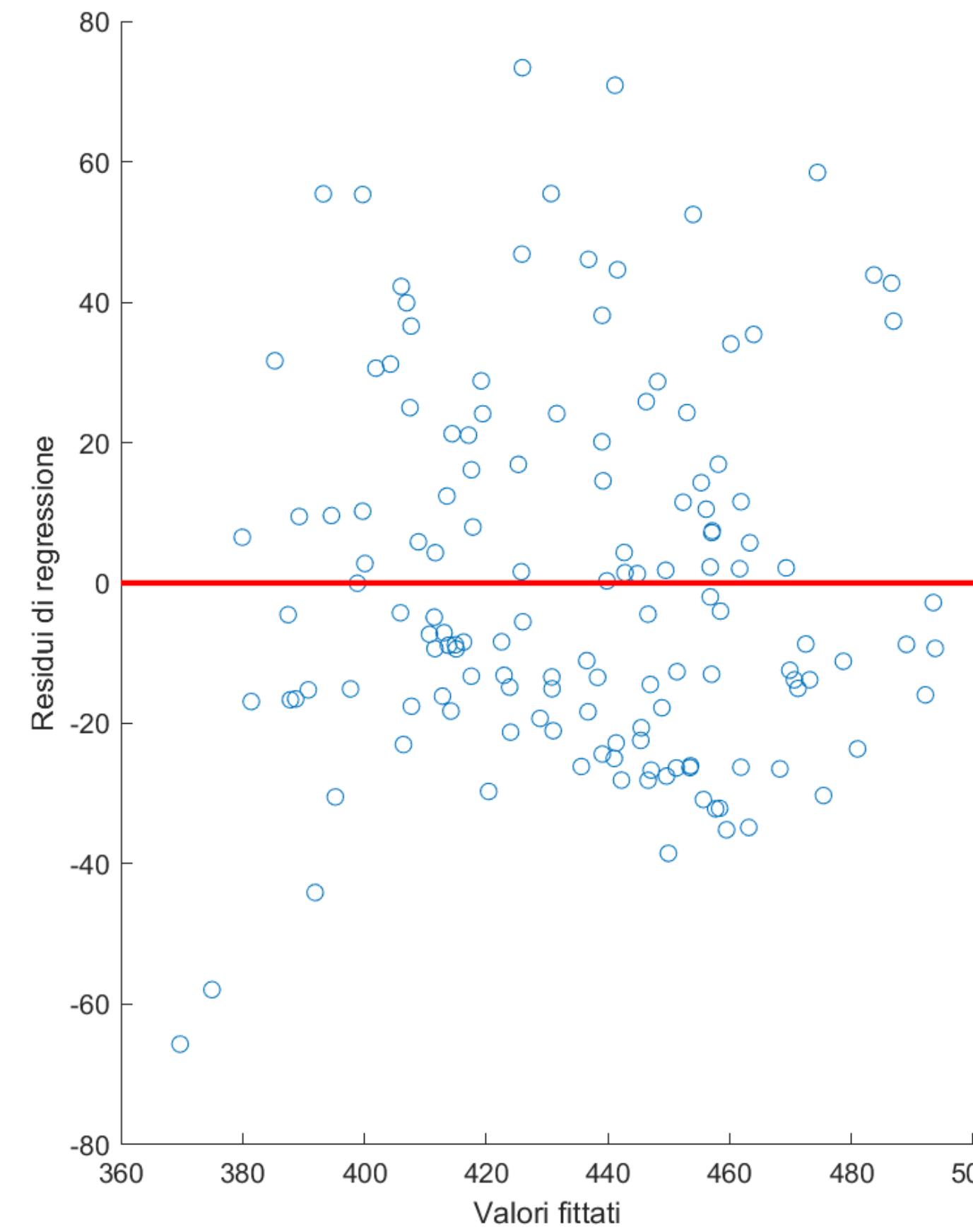
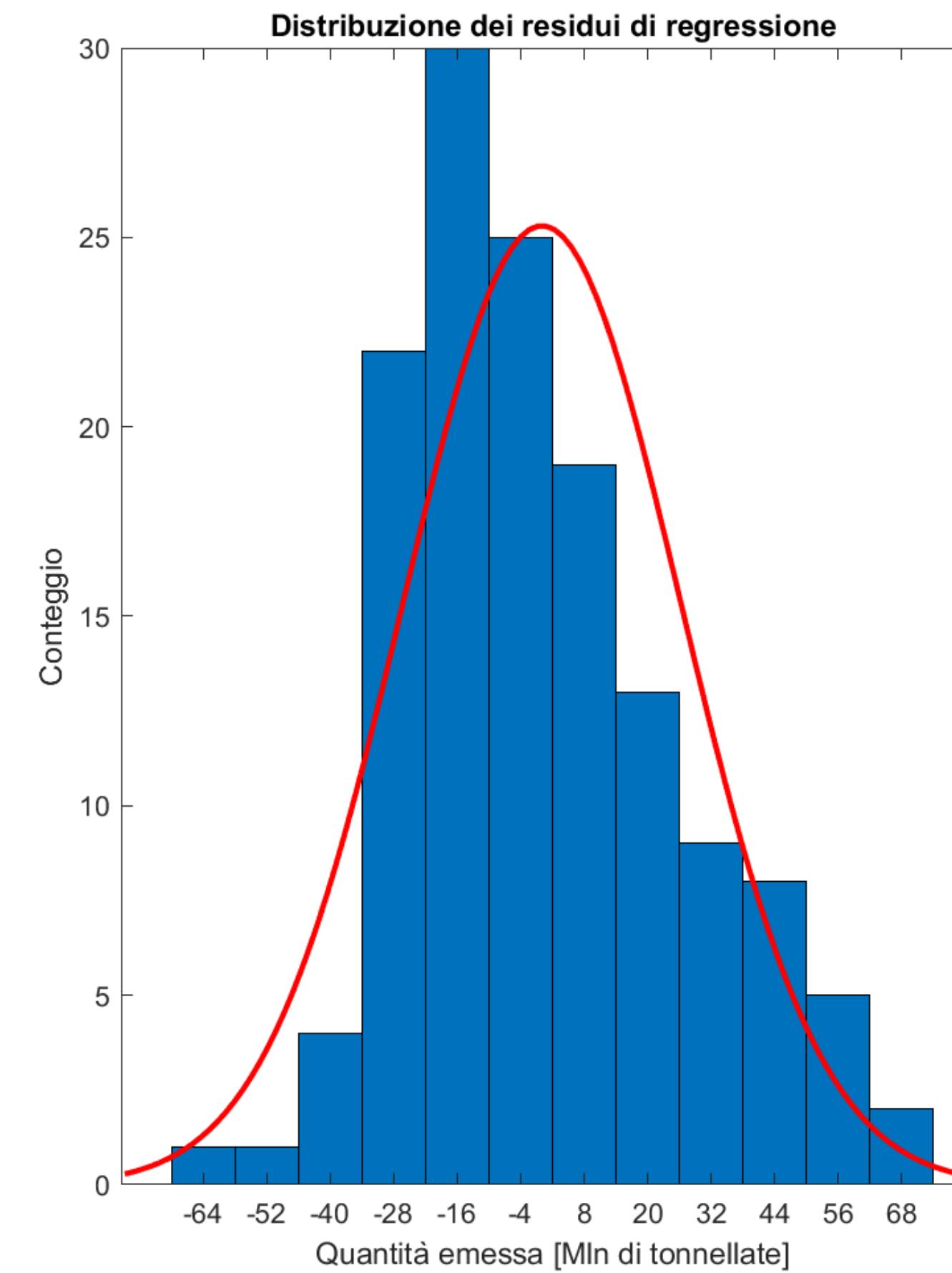
1.1 Scelta del regressore basata sulle varie correlazioni —> La produzione di carbone ha indice di correlazione di 0.73

1.2 Regressione semplice fra emissioni di CO₂ e produzione carbone totale



$R^2 \rightarrow 0.537$

1.3 Analisi dei residui

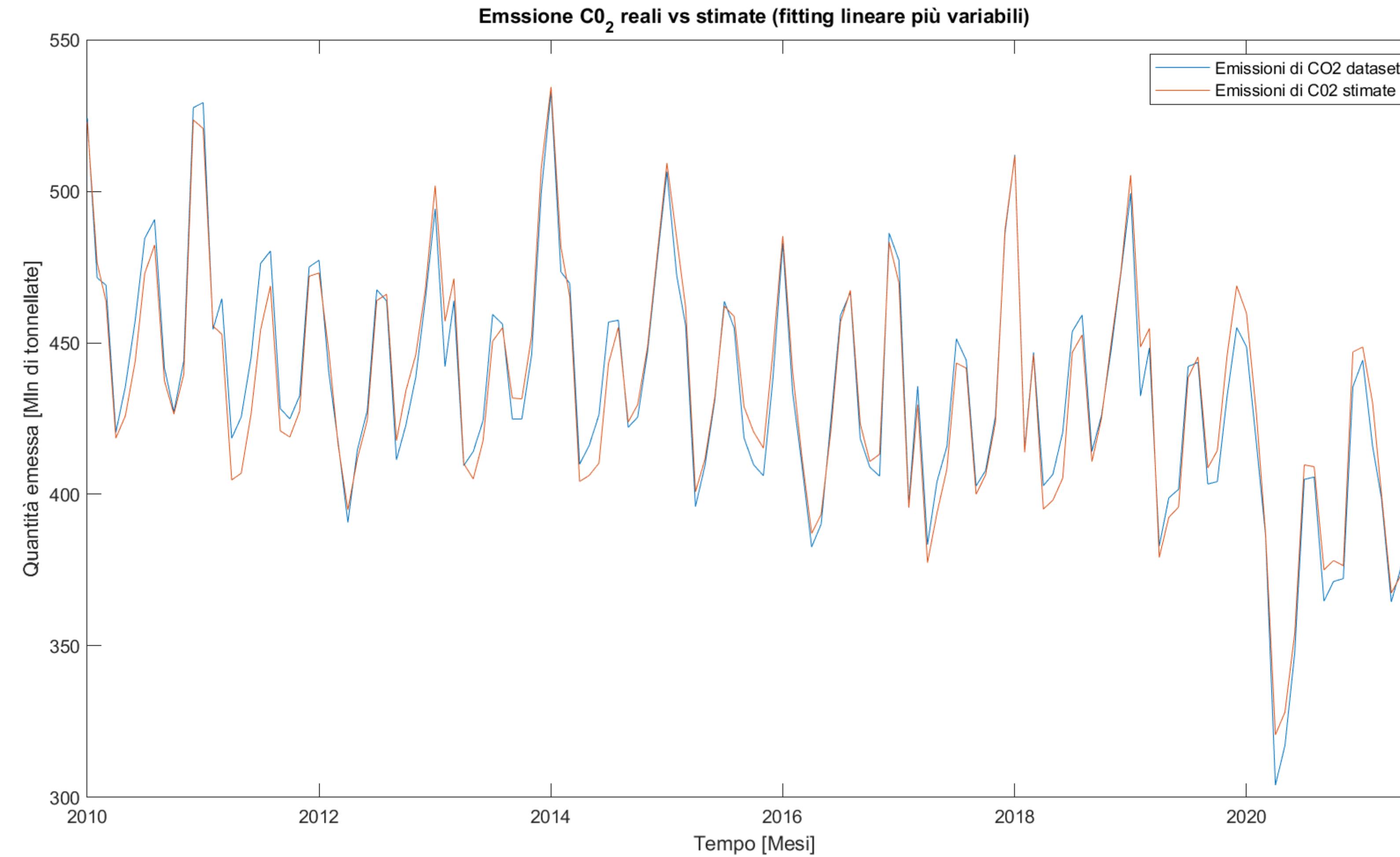


SK = 0.5687
K = 3.0127

- L'ipotesi di normalità dei residui è confermata dal test di Lilliefors al 5% e da quello Bera-Jarque con livello di significatività dell'1% e del 5%

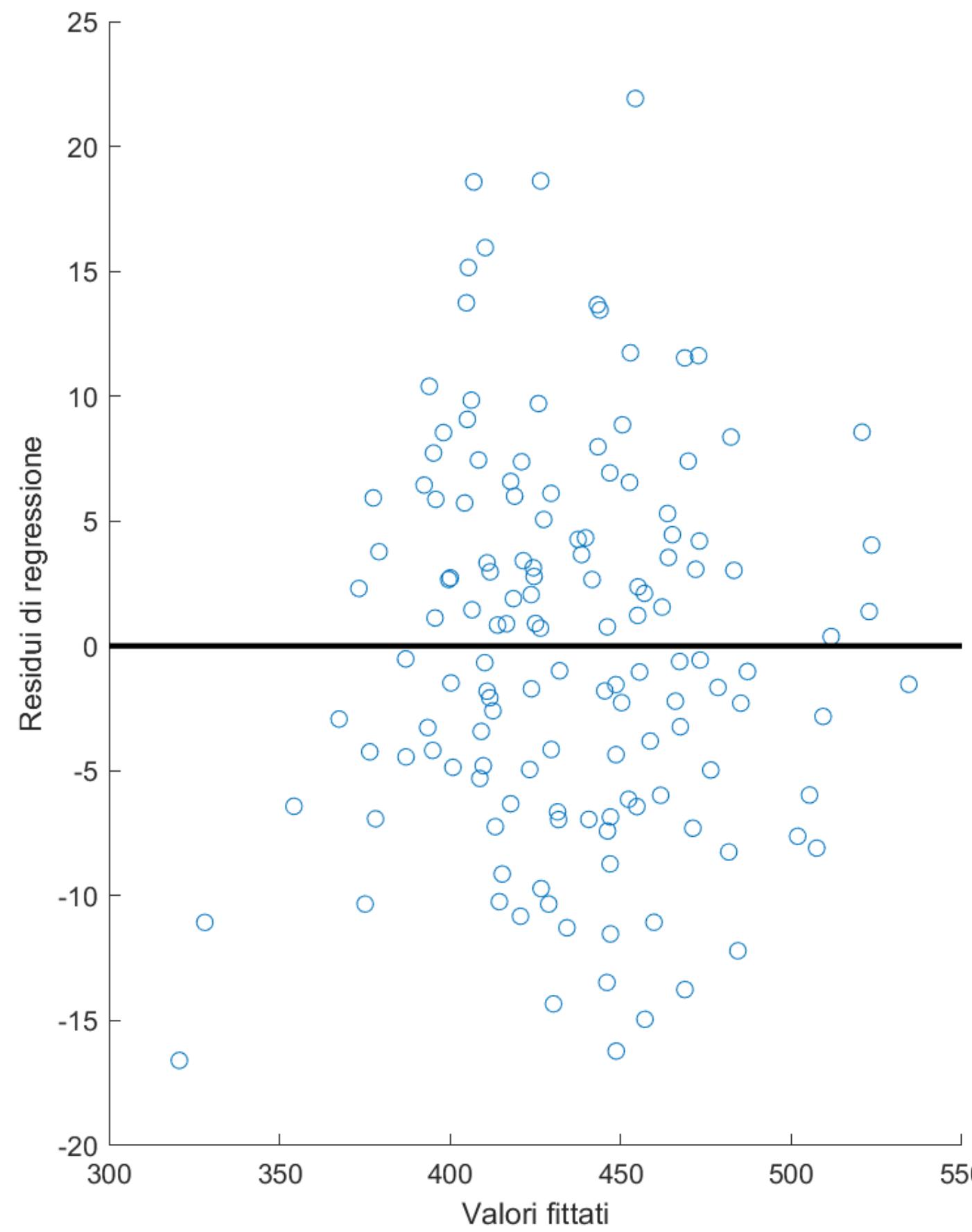
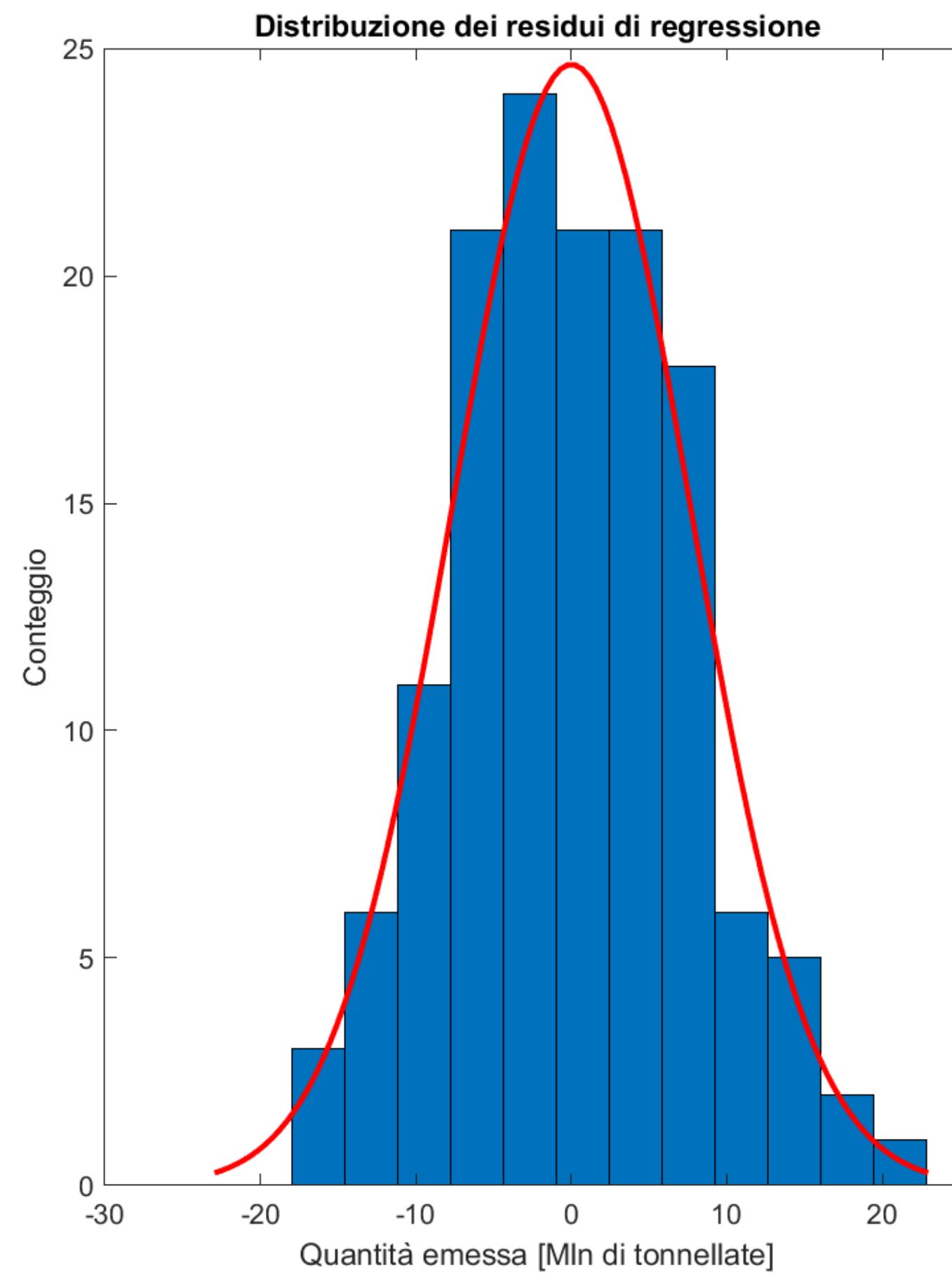
1.4 Regressione multipla con predittori aggregati:

- Produzione di energia rinnovabile totale
- Consumo energia rinnovabile e non rinnovabile totale



$R^2 \rightarrow 0.961$

1.5 Analisi dei residui

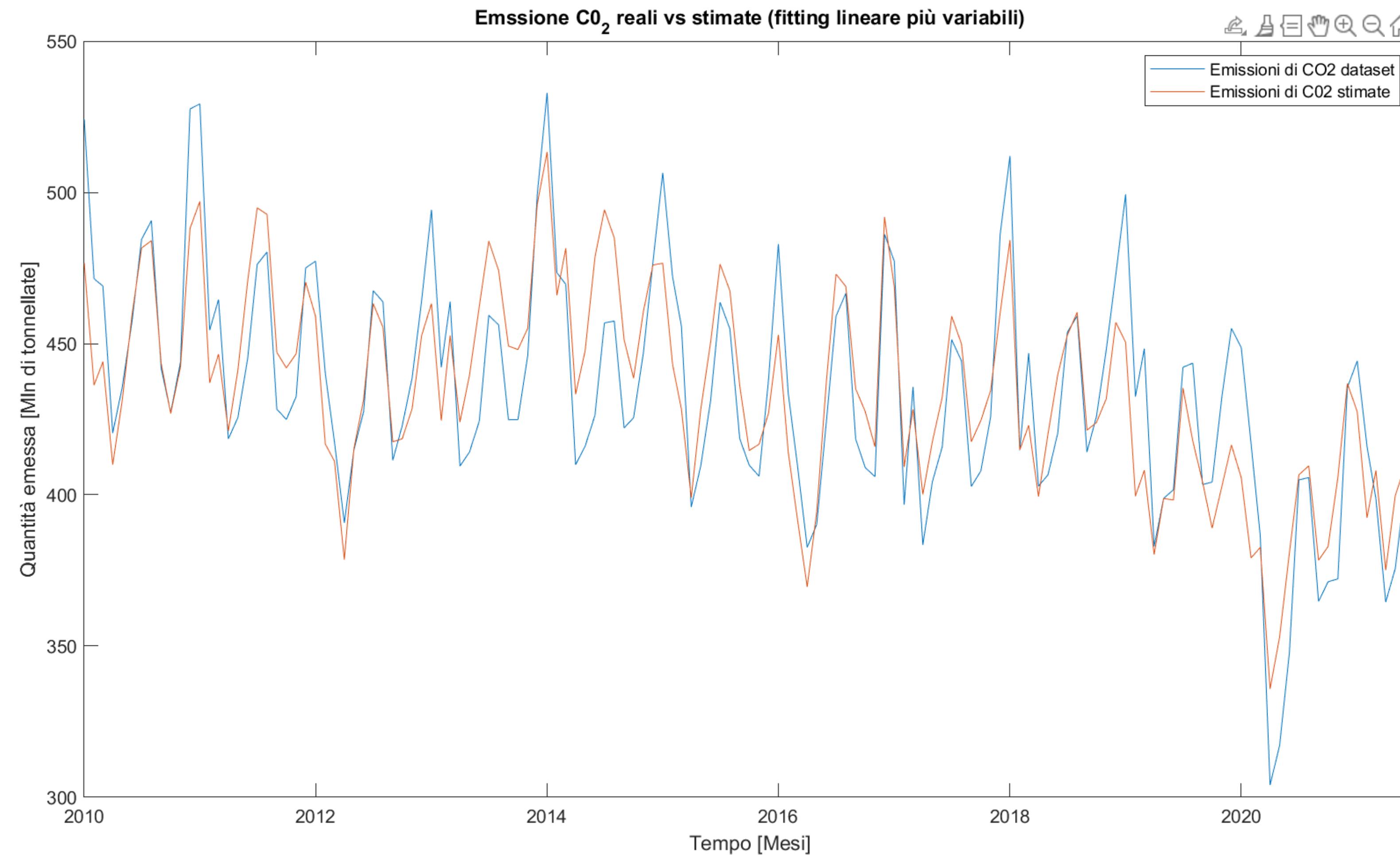


SK = 0.2170
K = 2.8723

- L'ipotesi di normalità dei residui è confermata dal test di Lilliefors al 5% e da quello Bera-Jarque con livello di significatività dell'1% e del 5%

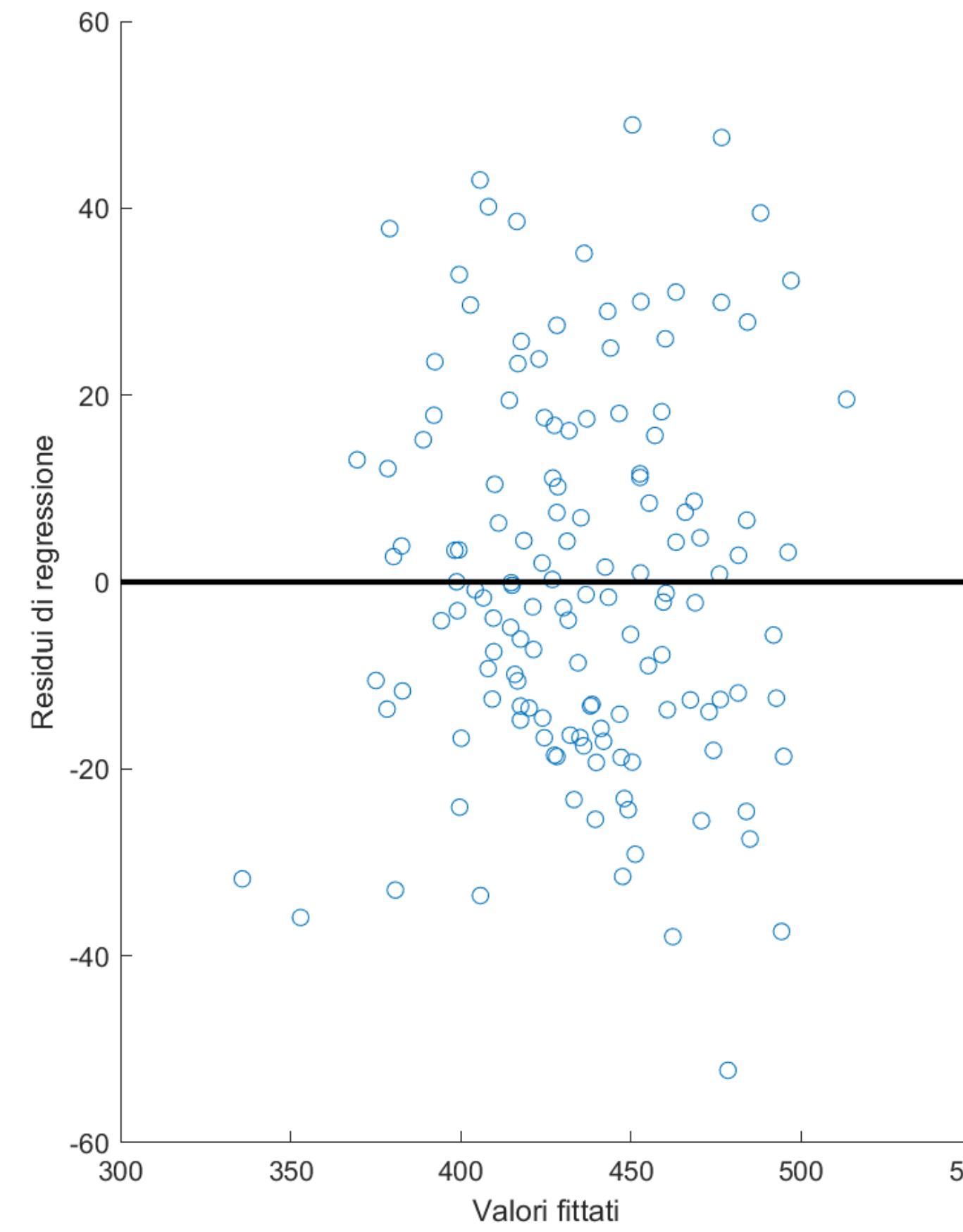
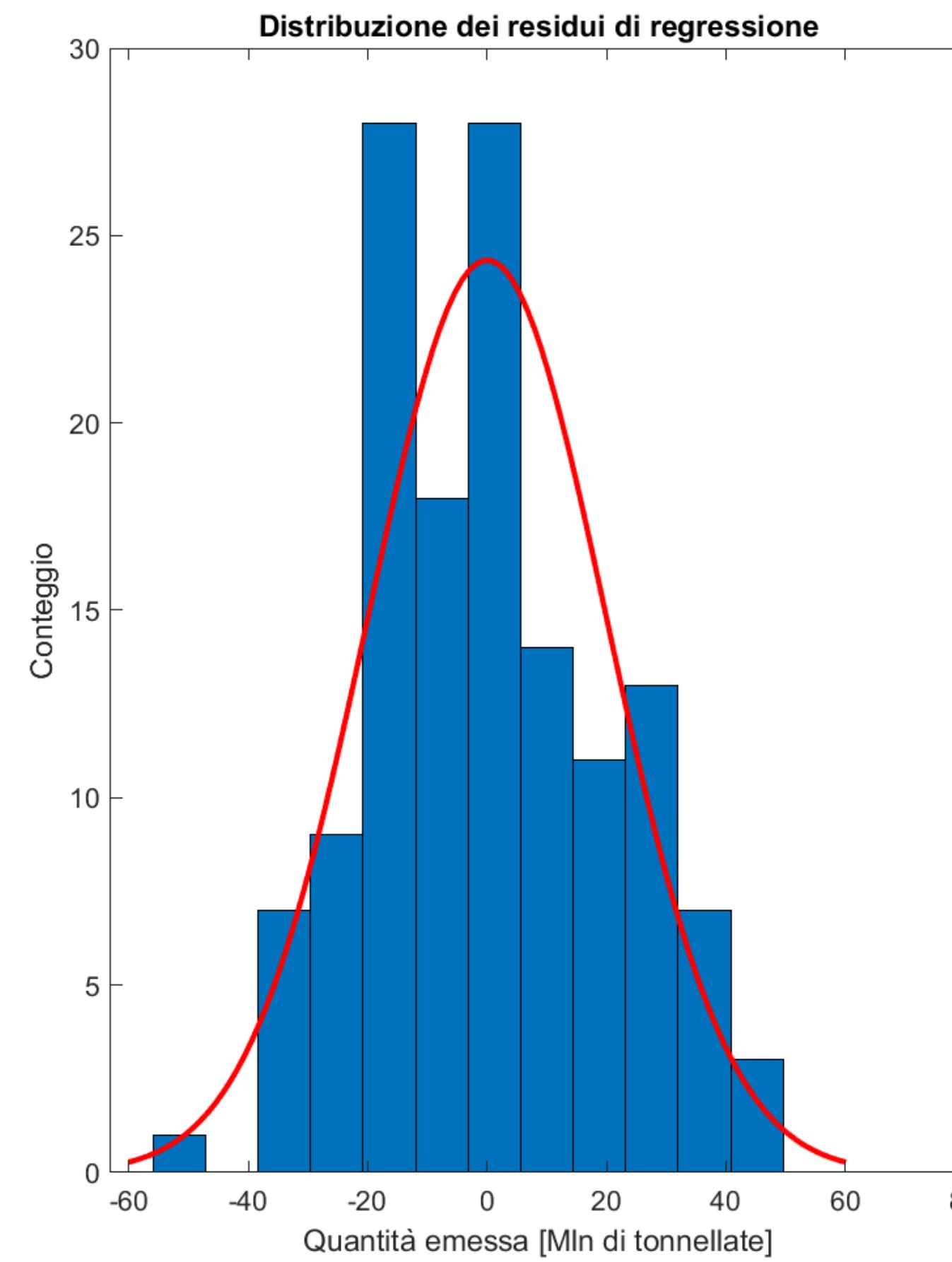
1.6 Regressione multipla con predittori singoli:

- Produzione di energia eolica, di biomasse e di carbone
- Consumo combustibili fossili nel settore dei trasporti e consumo del carbone totale



$R^2 \rightarrow 0.731$

1.7 Analisi dei residui

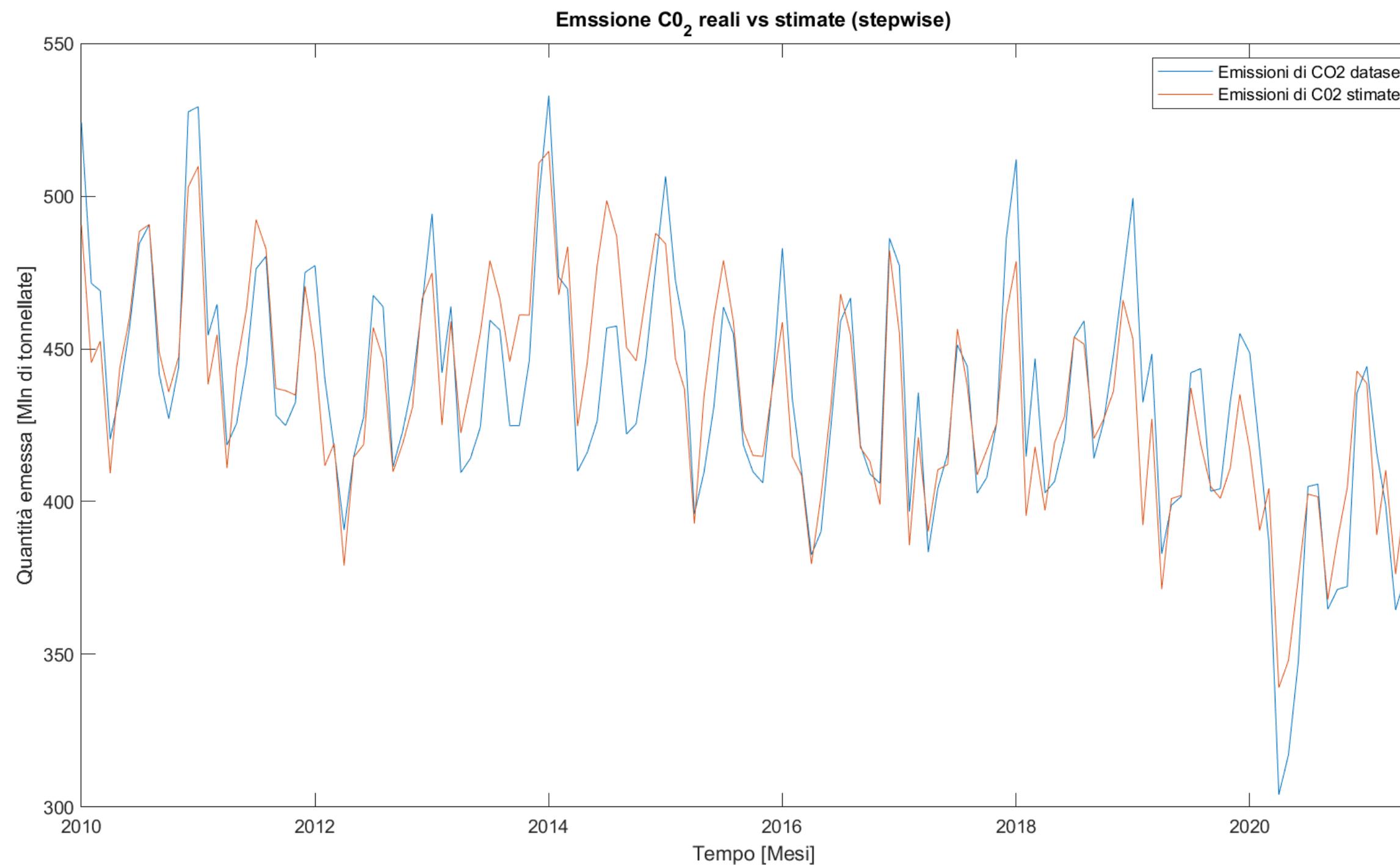


SK = 0.2624
K = 2.6842

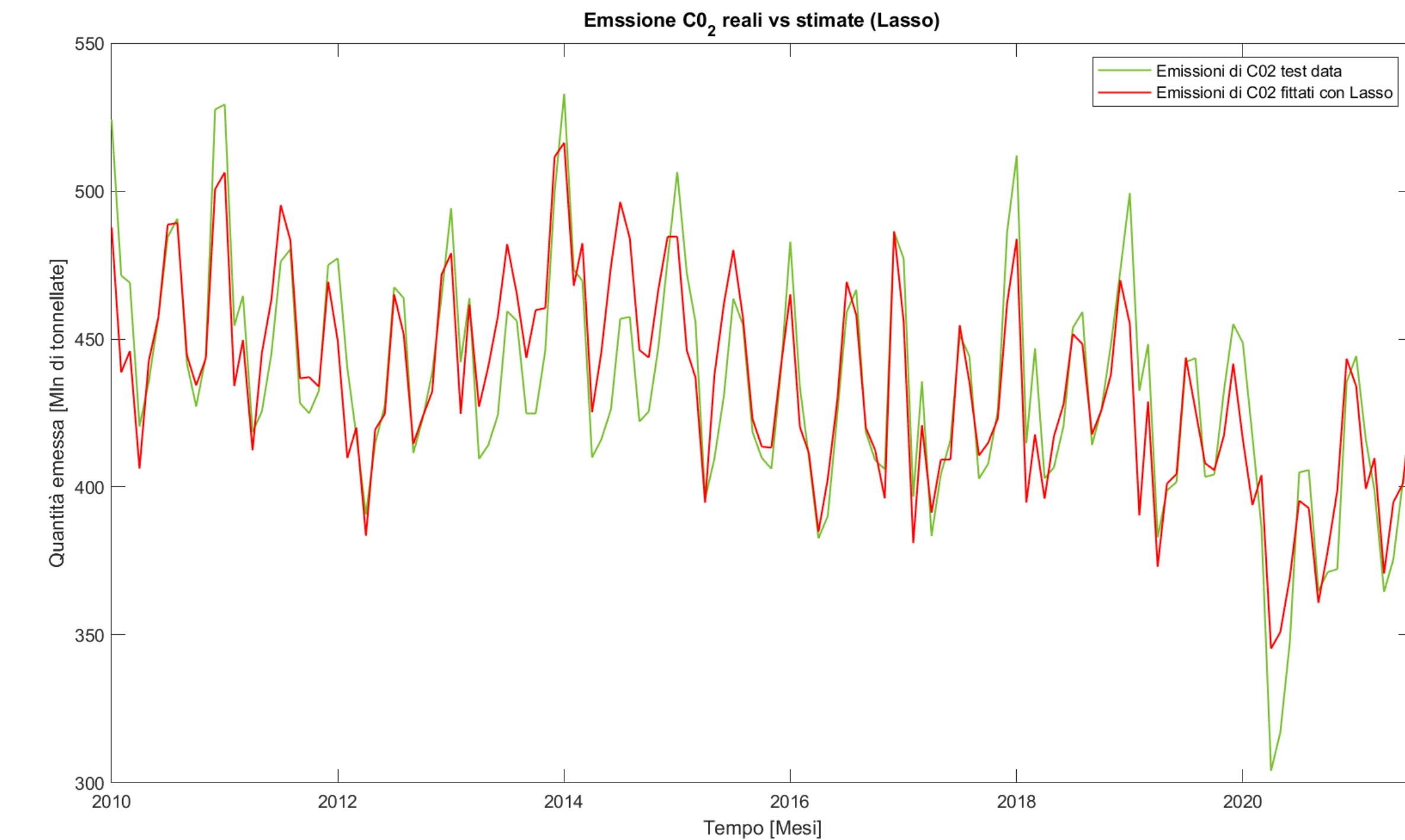
- L'ipotesi di normalità dei residui è confermata dal test di Lilliefors al 5% e da quello Bera-Jarque con livello di significatività dell'1% e del 5%

1.8 Model selection: Stepwise e Lasso

- Variabili da valutare: tutte le colonne del dataset relative a produzioni e consumi singoli



$R^2 \rightarrow 0.785$



$R^2 \rightarrow 0.794$

1.9 Confronto

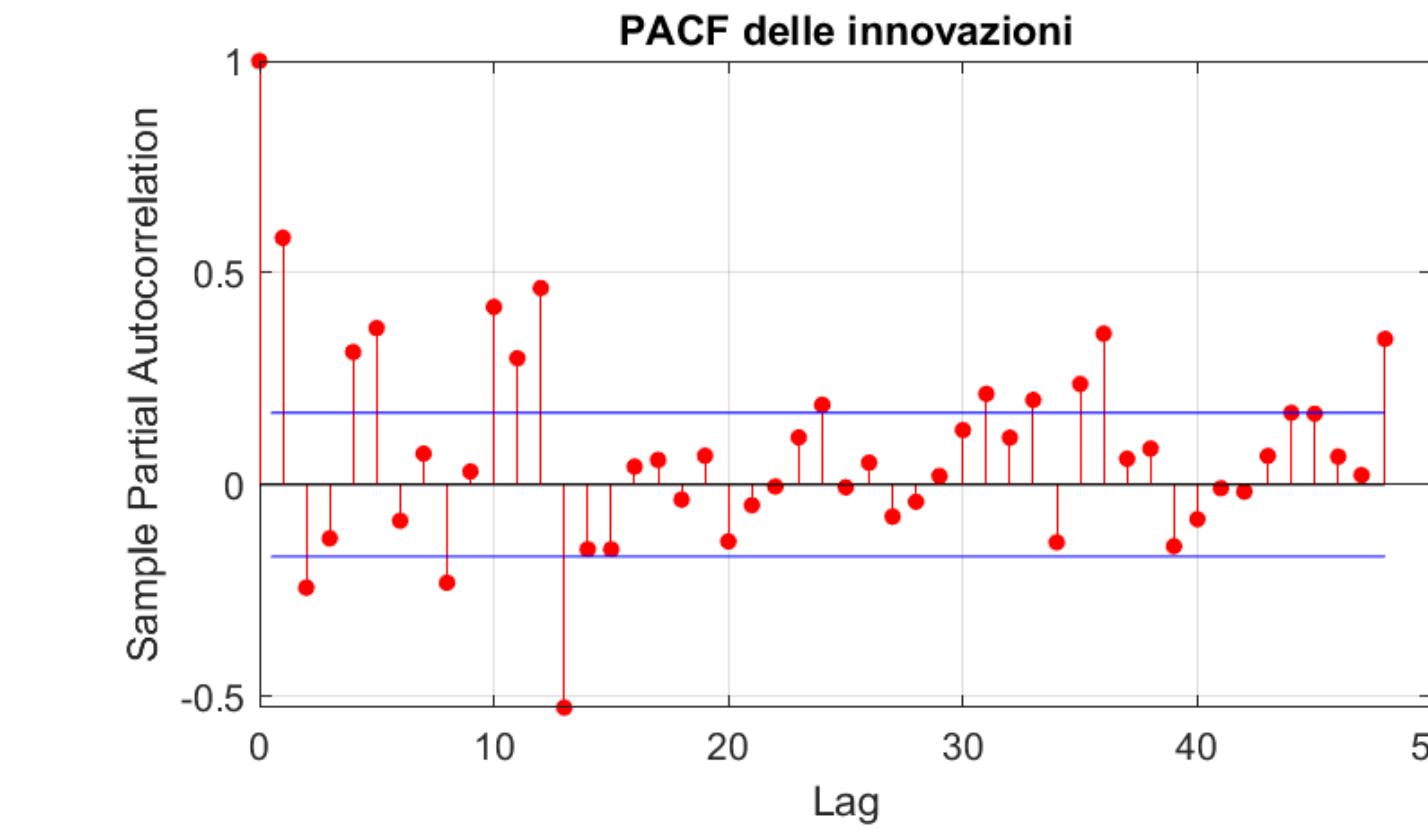
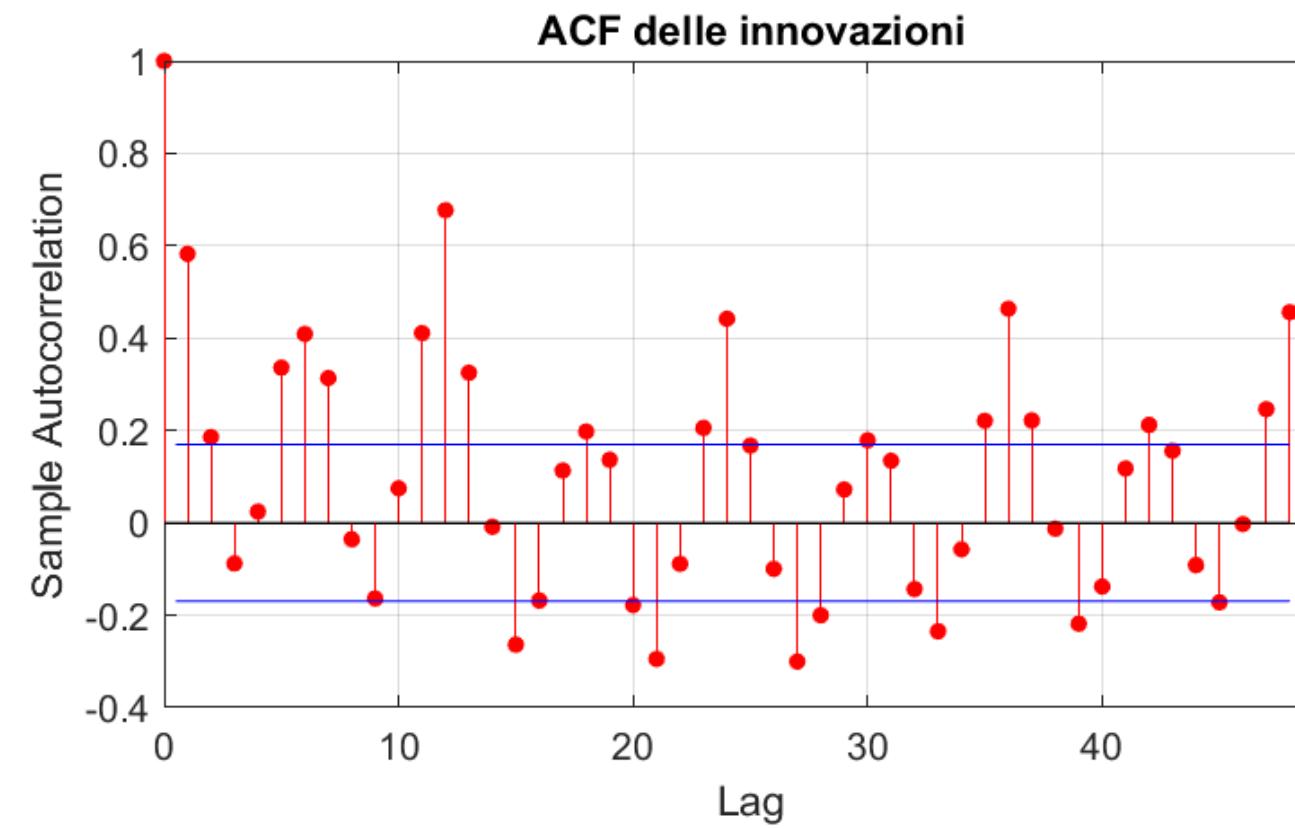
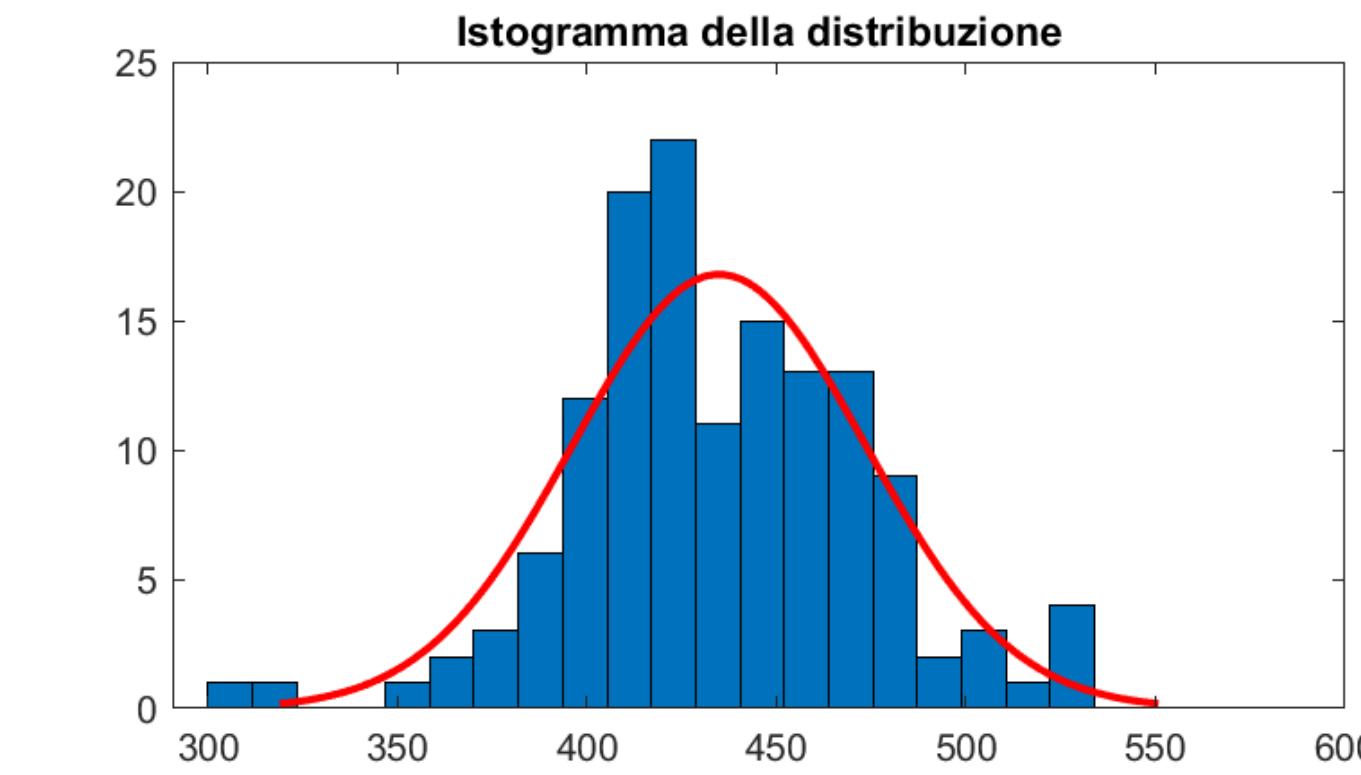
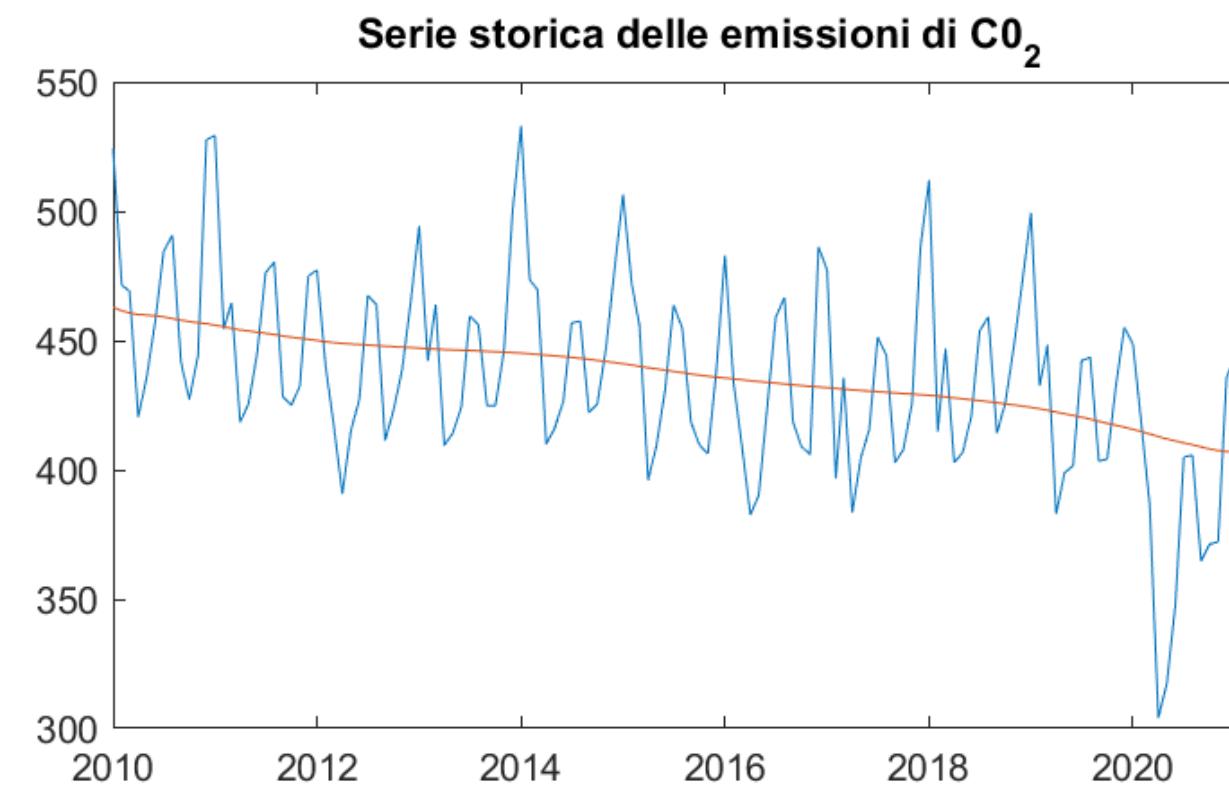
	R ²	sk	k	pv B-J ($\alpha = 5\%$)	pv Lilliefors ($\alpha = 5\%$)
Reg. semplice	0.537	0.5687	3.0127	0.0284	1.0000e – 03
Reg. mult. (aggr.)	0.961	0.2170	2.8723	0.5	0.5
Reg. mult. (sing.)	0.731	0.2624	2.6842	0.2678	0.2249
Stepwise (sing.)	0.785	–	–	–	–
Lasso (sing.)	0.794	–	–	–	–

- Il modello che produce una stima migliore delle emissioni di CO₂, cioè quello con R² maggiore, si rivela essere quello ottenuto dalla regressione multipla con regressori aggregati.
- Per quanto riguarda i modelli costruiti con le variabili singole il migliore si rivela essere quello ottenuto per mezzo del lasso



Previsione delle emissioni di CO₂ negli USA

2.1 Valutazione della serie storica

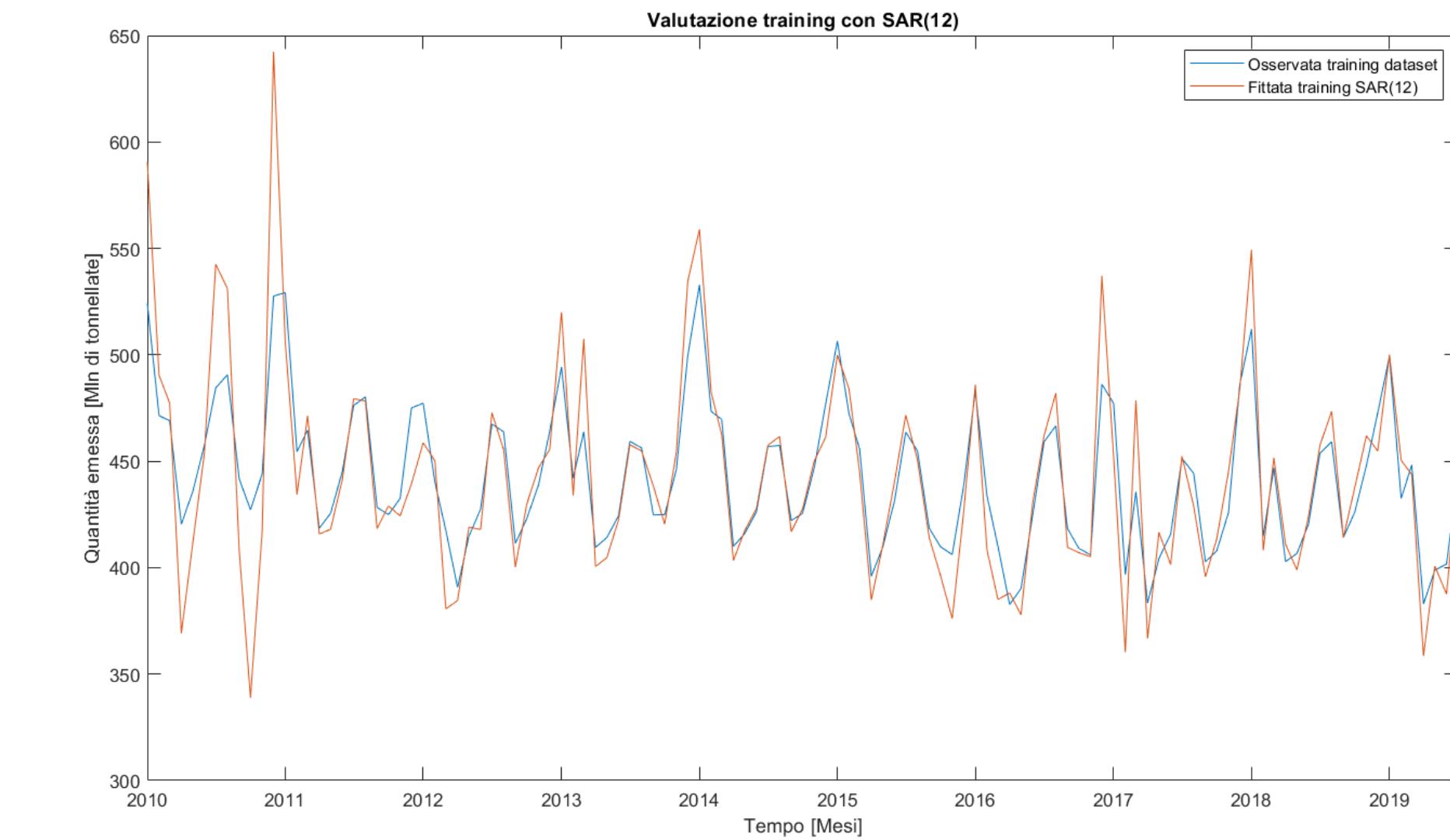


- Il test Augmented Dickey Fuller sui ritardi da 0 a 6 permette di rifiutare l'ipotesi nulla e quindi di dire che la serie è stazionaria con trend decrescente
- il test di Ljung-Box permette di rifiutare l'ipotesi nulla e quindi di dire che la serie è significativamente autocorrelata ai ritardi 6,12,18,24,30,36,48
- il test di Durbin-Watson permette di dire che la serie è autocorrelata di ritardo 1

2.2 Stima dei modelli - con stagionalità

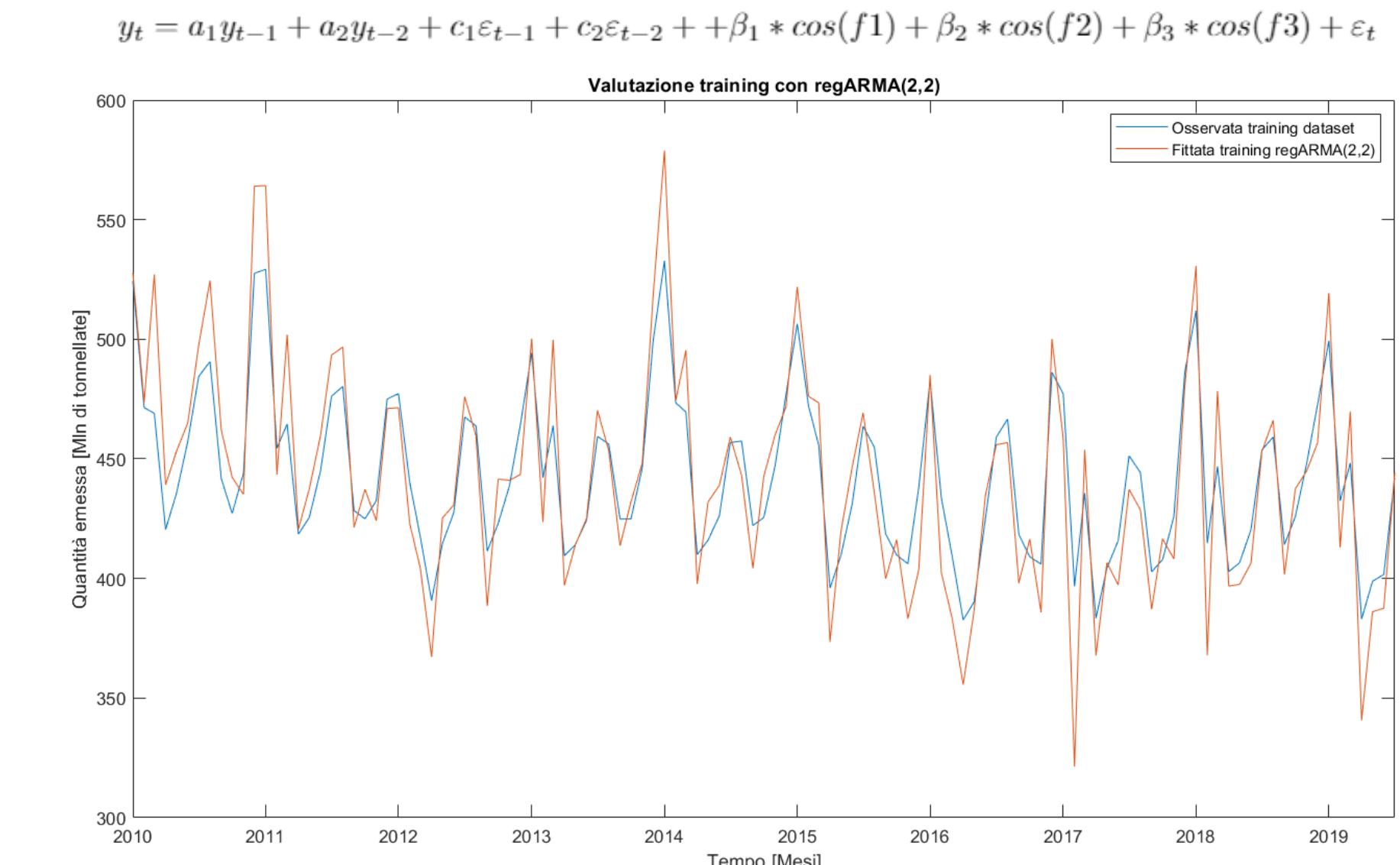
- 85% LEARNING SET (Gennaio 2010 - Giugno 2019)
- 15% TEST SET (Luglio 2019 - Luglio 2021)

ARIMA(1,0,0) Model with Seasonal AR(12) (Gaussian Distribution)			
Effective Sample Size: 115			
Number of Estimated Parameters: 4			
LogLikelihood: -473.941			
AIC: 955.882			
BIC: 966.862			
RMSE training → 23.08			
Value	StandardError	TStatistic	PValue
Constant	23.473	9.2856	2.5279
AR{1}	0.52337	0.075464	6.9354
SAR{12}	0.88377	0.041682	21.203
Variance	222.42	22.378	9.9393



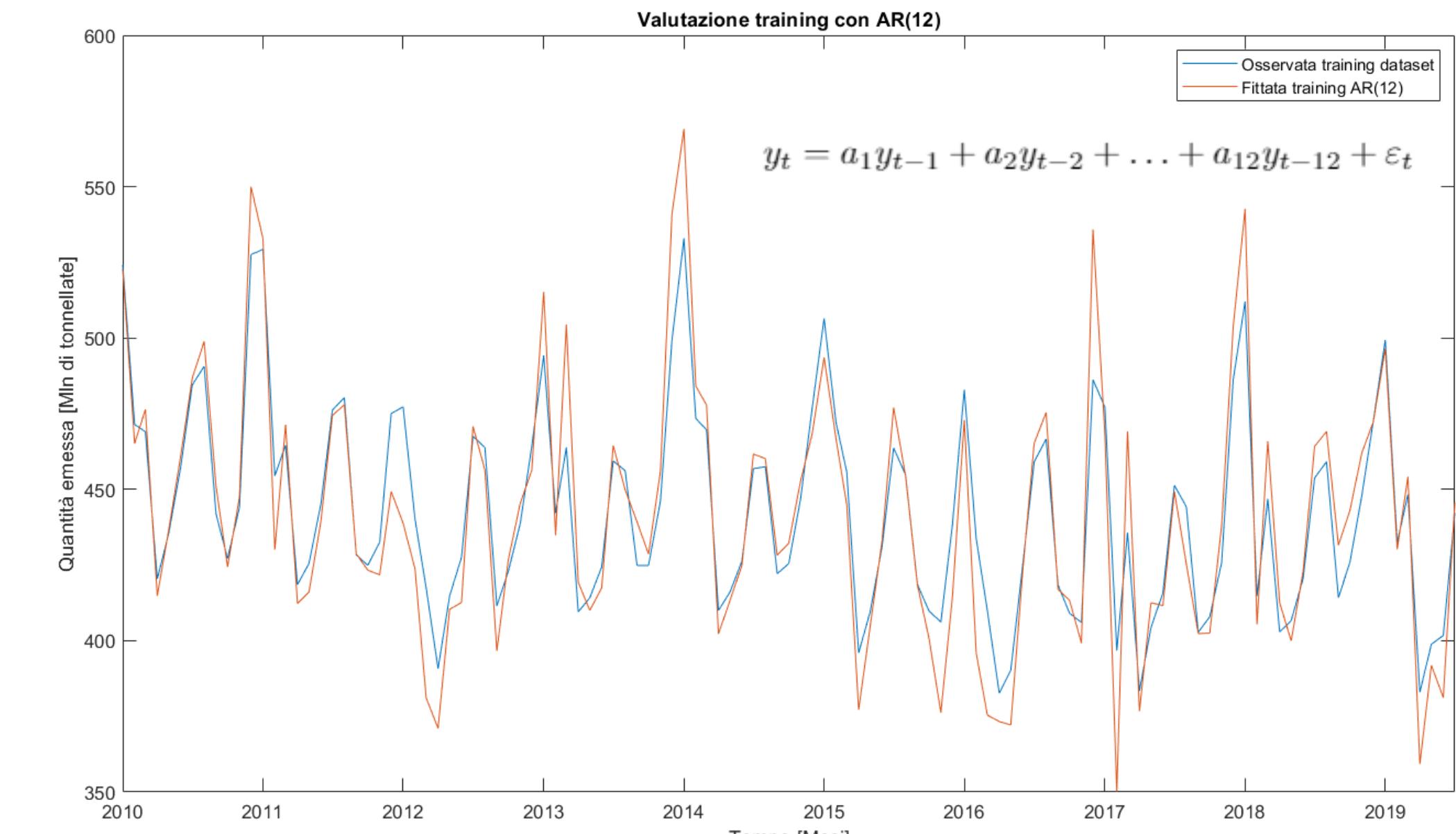
Regression with ARMA(2,2) Error Model (Gaussian Distribution):			
	Value	StandardError	TStatistic
Intercept	442.92	2.7272	162.41
AR{1}	0.98356	0.052947	18.576
AR{2}	-0.96128	0.033078	-29.061
MA{1}	-0.71327	0.053964	-13.217
MA{2}	0.90587	0.056197	16.119
Beta{1}	18.475	4.6889	3.9402
Beta{2}	-3.3657	77.879	-0.043217
Beta{3}	10.119	2.1338	4.7421
Variance	386.17	47.689	8.0978

RMSE training → 19.65

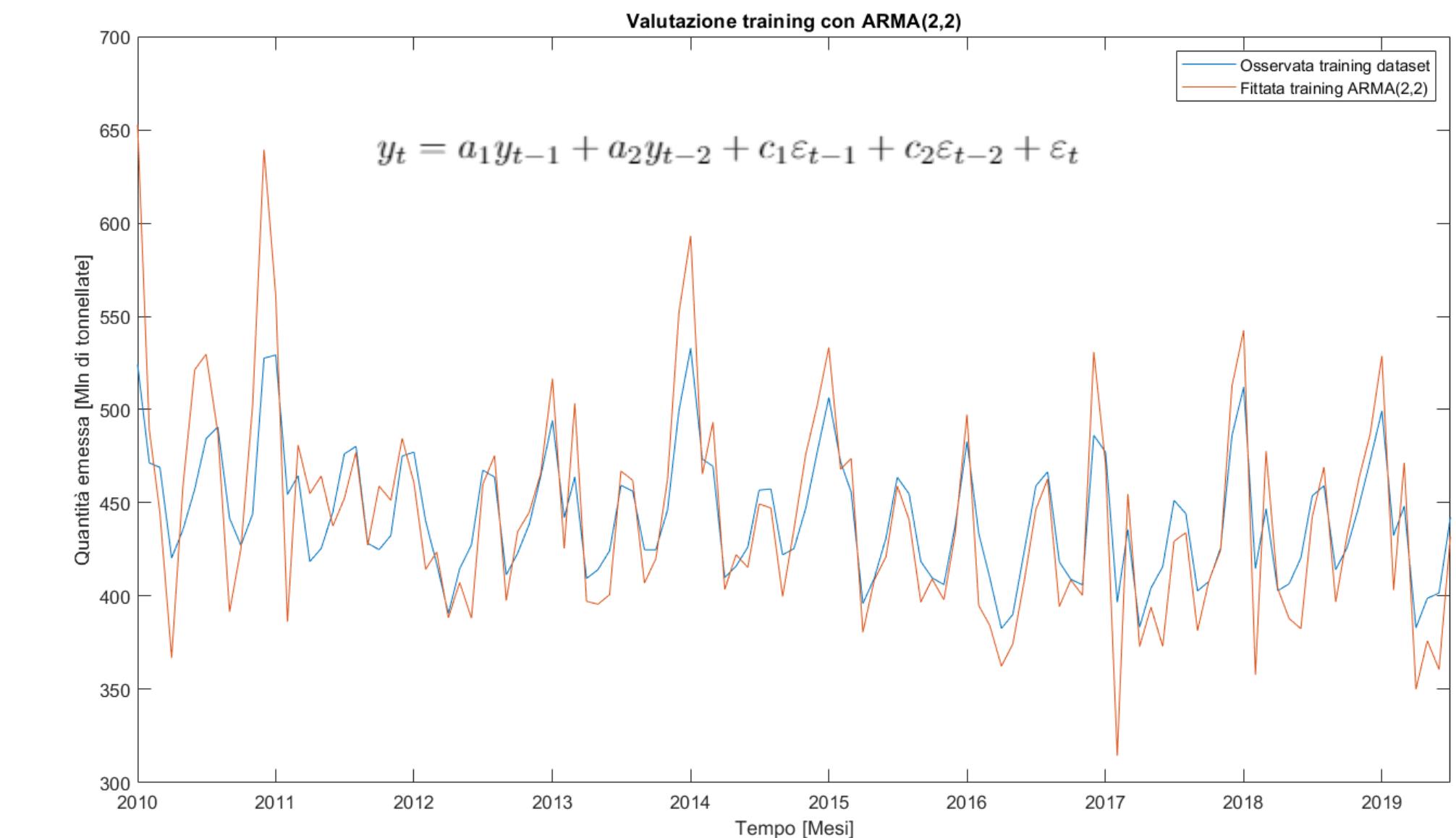


2.2 Stima dei modelli - senza stagionalità

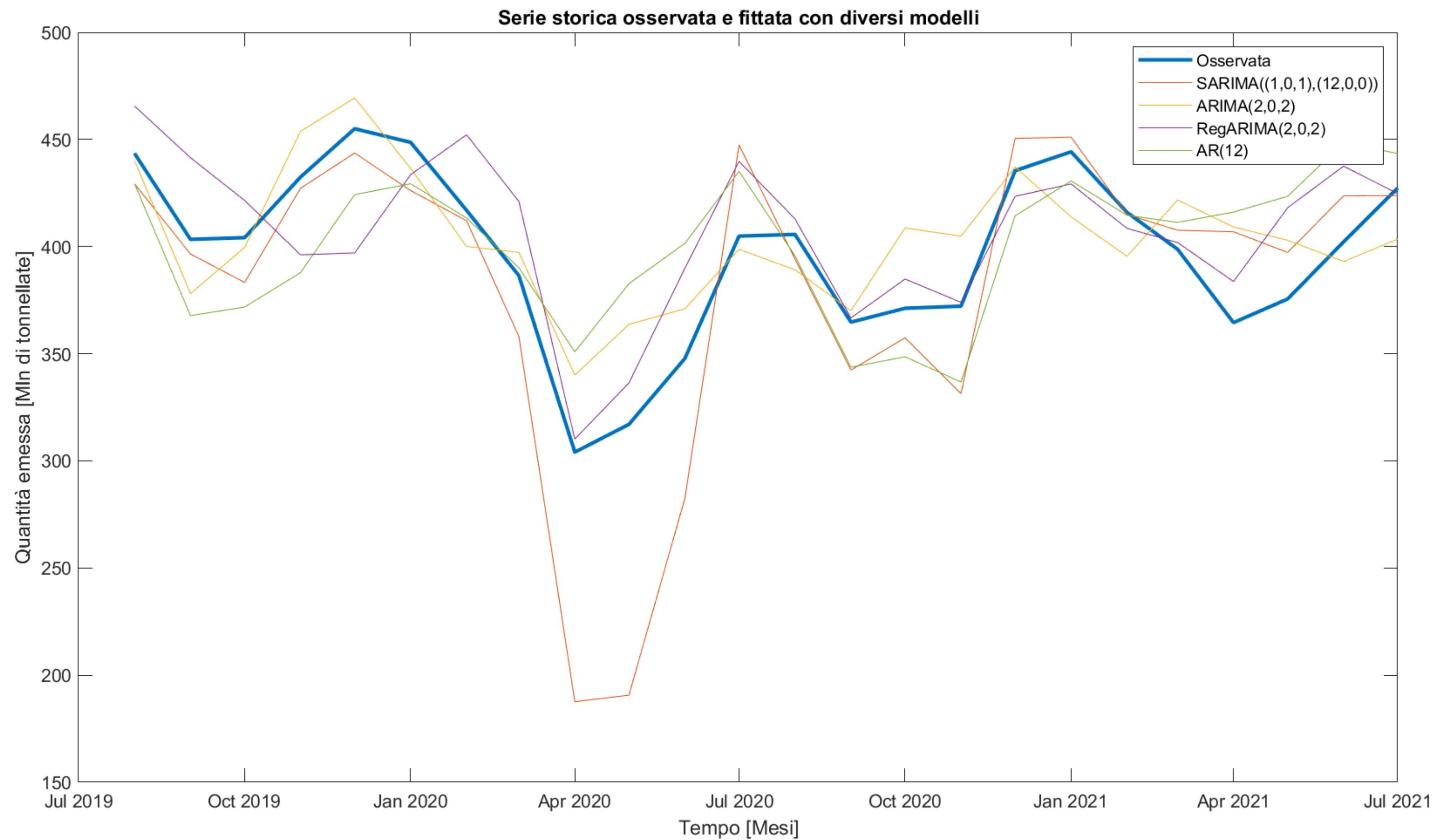
ARIMA(12,0,0) Model (Gaussian Distribution)				
Effective Sample Size: 115 Number of Estimated Parameters: 14 LogLikelihood: -482.741 AIC: 993.482 BIC: 1031.91				
RMSE training → 15.86				
Value	StandardError	TStatistic	PValue	
Constant	35.248	55.128	0.6394	0.52257
AR{1}	0.23349	0.07708	3.0292	0.0024523
AR{2}	0.0033727	0.08631	0.039076	0.96883
AR{3}	-0.040055	0.094621	-0.42332	0.67206
AR{4}	0.052005	0.11952	0.4351	0.66349
AR{5}	0.045041	0.12757	0.35306	0.72404
AR{6}	-0.08504	0.10688	-0.79569	0.42621
AR{7}	0.11521	0.10888	1.0581	0.29001
AR{8}	-0.088496	0.10584	-0.83615	0.40307
AR{9}	-0.13347	0.089312	-1.4944	0.13507
AR{10}	0.11149	0.078287	1.4241	0.15441
AR{11}	0.0022594	0.081469	0.027734	0.97787
AR{12}	0.70115	0.071708	9.7778	1.4017e-22
Variance	259.2	33.113	7.828	4.9576e-15



ARIMA(2,0,2) Model (Gaussian Distribution)				
Effective Sample Size: 115 Number of Estimated Parameters: 6 LogLikelihood: -529.839 AIC: 1071.68 BIC: 1088.15				
RMSE training → 30.33				
Value	StandardError	TStatistic	PValue	
Constant	422.57	15.782	26.775	6.323e-158
AR{1}	0.98688	0.033894	29.117	2.2123e-186
AR{2}	-0.94196	0.022871	-41.185	0
MA{1}	-0.7019	0.052808	-13.292	2.5932e-40
MA{2}	0.86977	0.050836	17.109	1.2654e-65
Variance	587.99	70.548	8.3346	7.7765e-17



2.3 Confronto sul test set

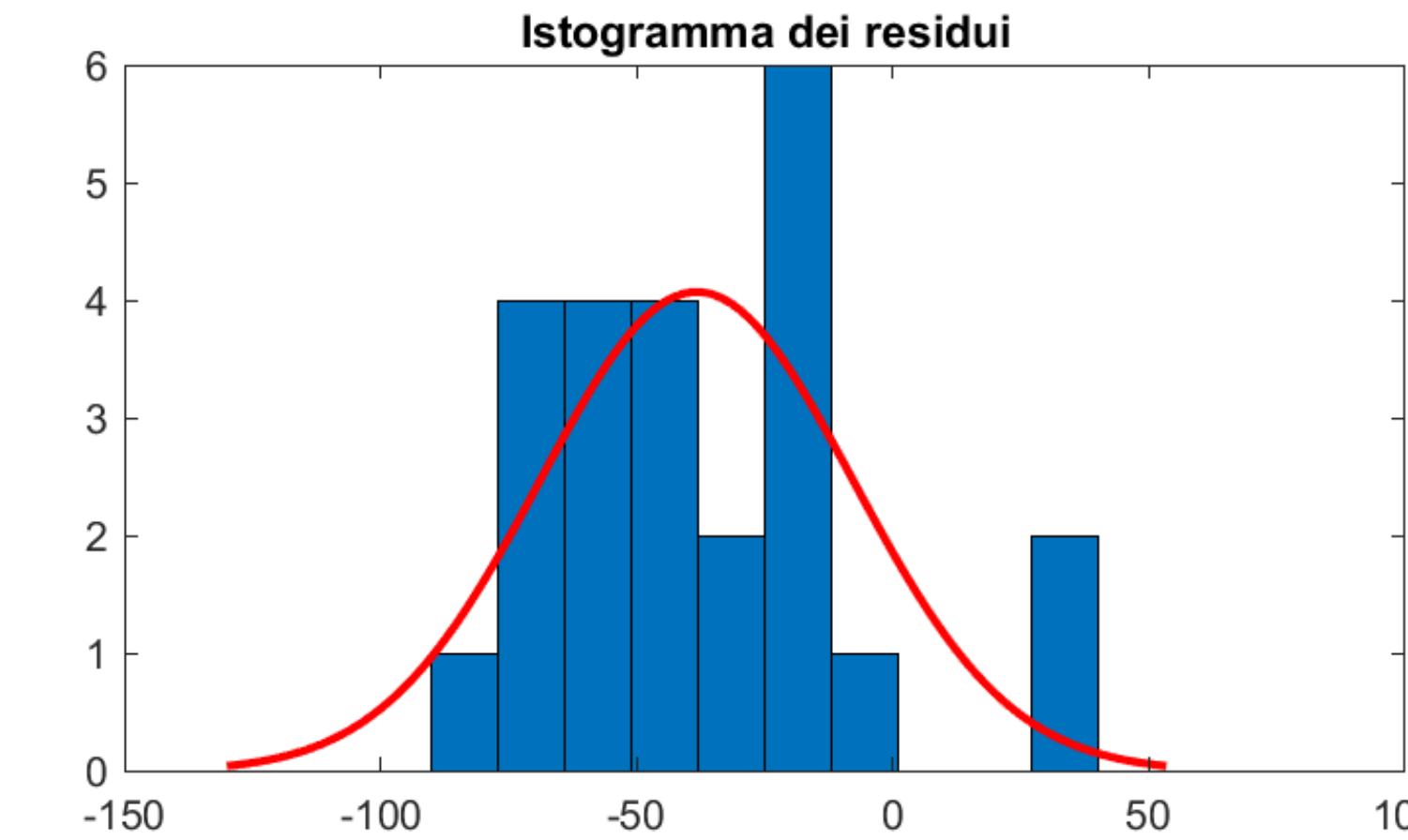
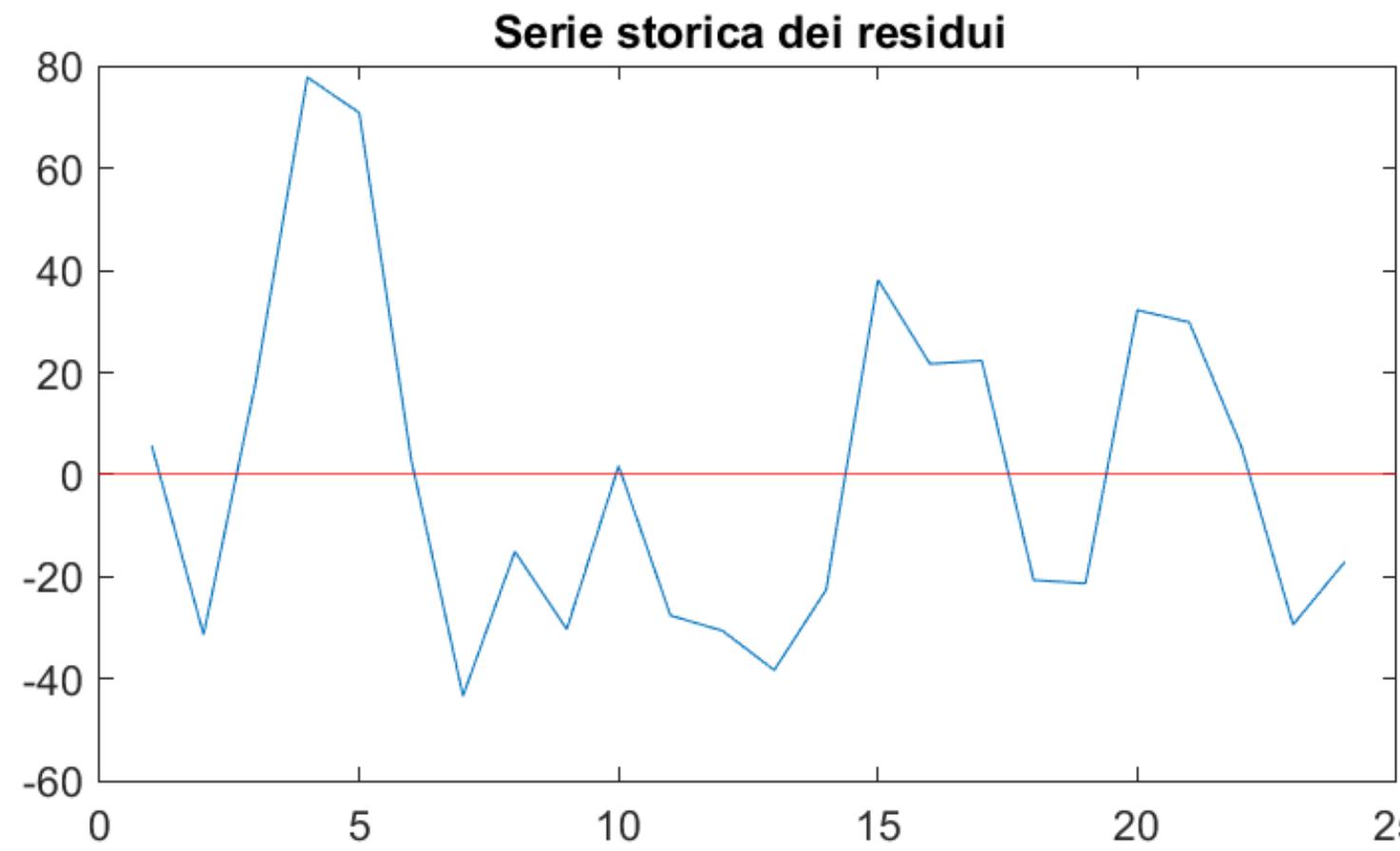


2.4 Valutazione modello ottimo

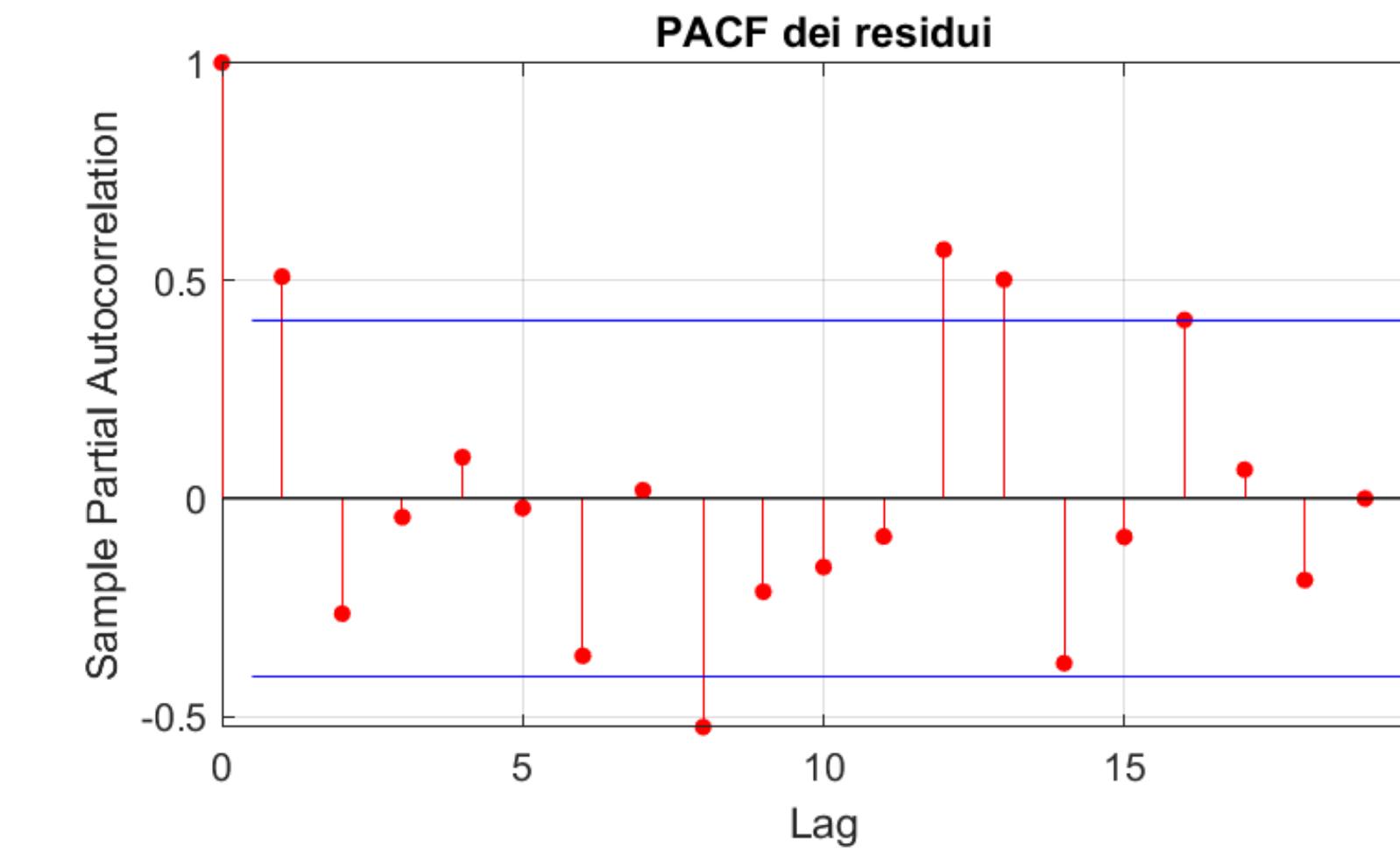
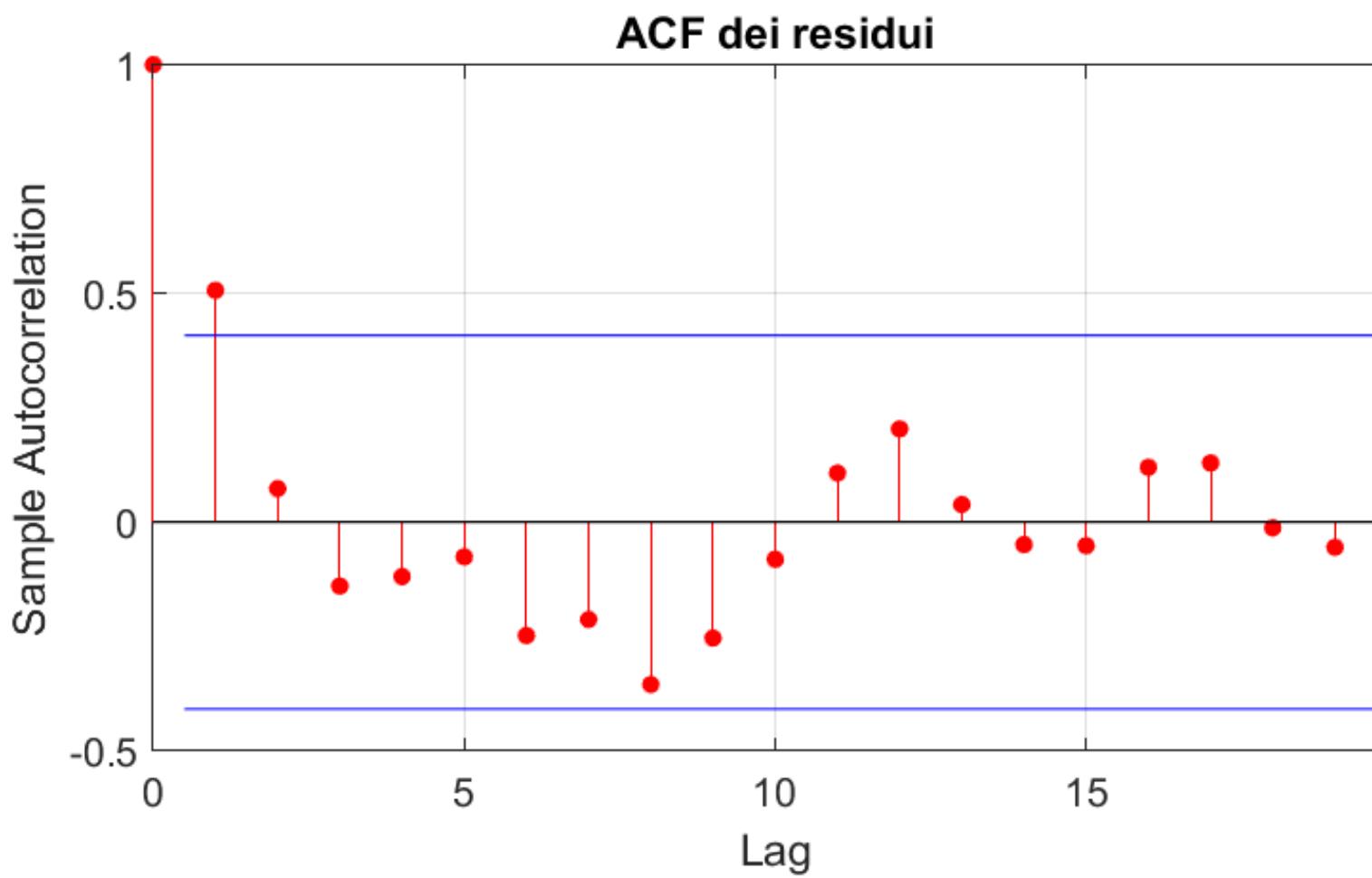
	Valutazioni migliore fitting sul test set			
	AR (12)	SAR (12)	ARMA (2,0,2)	RegARIMA (2,0,2)
RMSE	33.31	42.52	24.13	26.70

Modello ottimo a RMSE minore

2.5 Analisi dei residui di ARIMA (2,0,2)



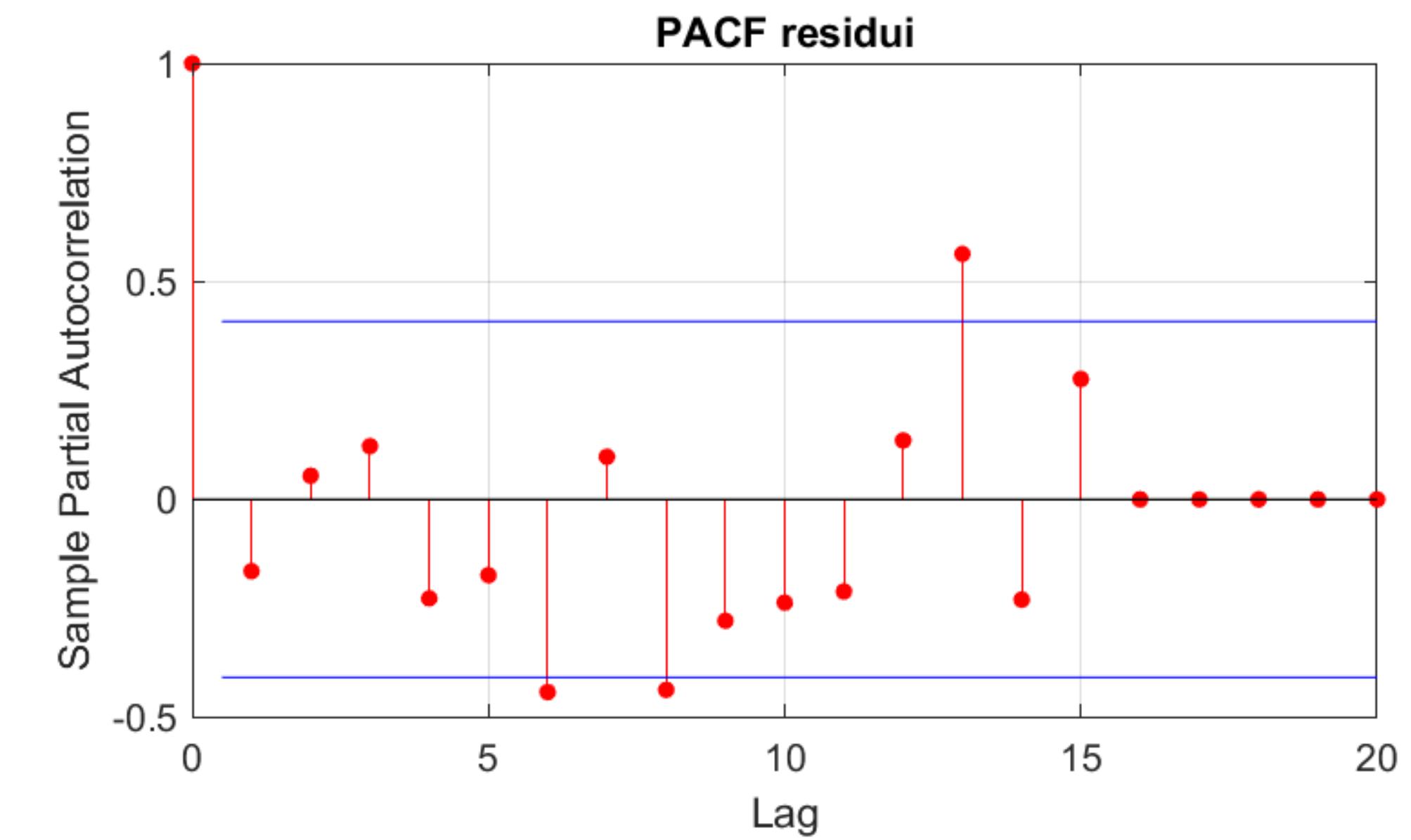
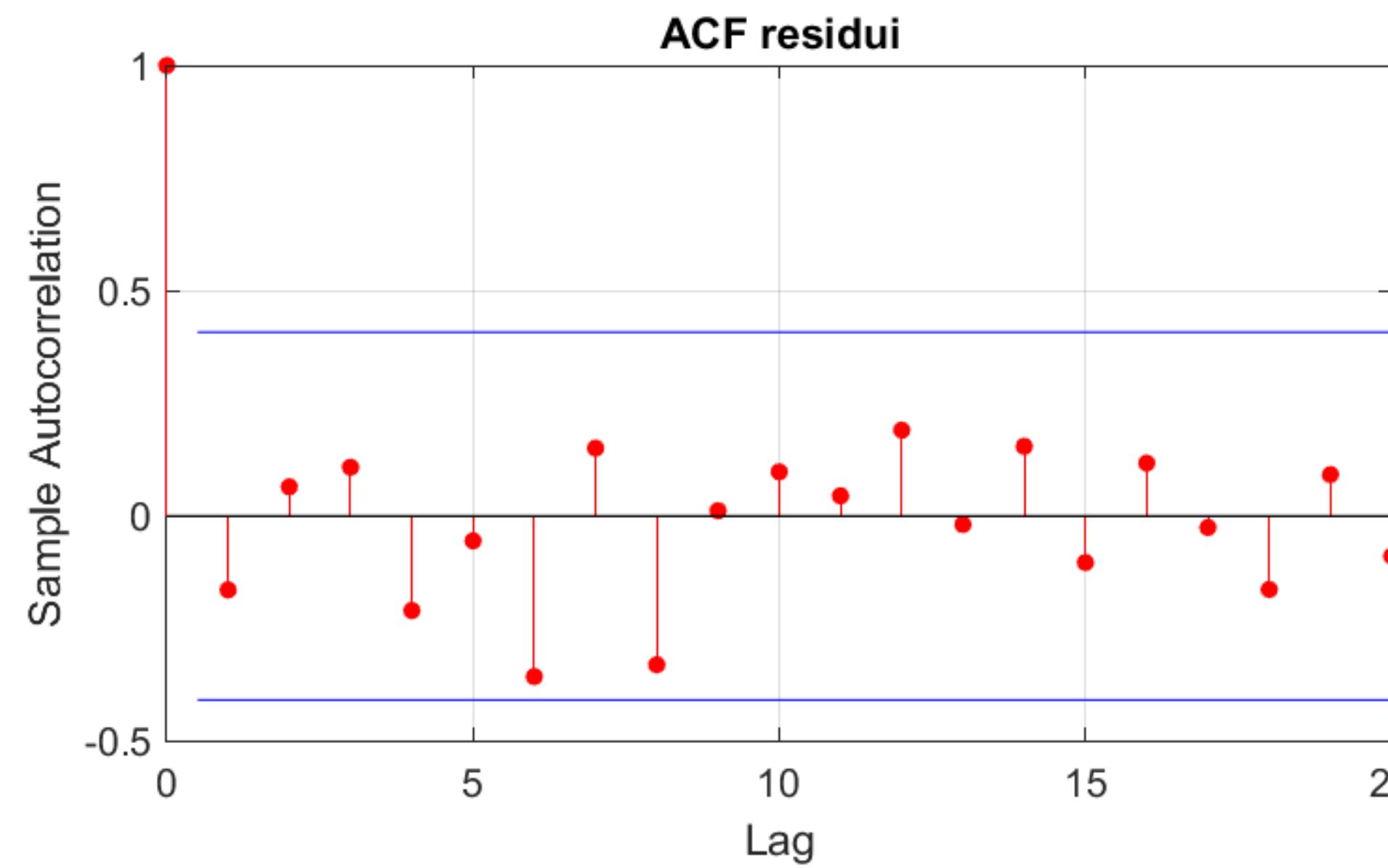
- Il test di Bera-Jarque al 5% permette di accettare l'ipotesi nulla: residui normali —> pvalue 15%
- Il test di Dickey Fuller permette di rifiutare l'ipotesi nulla e quindi di dire che i residui sono stazionari—> pvalue 0,63%



- Il test di Ljung Box permette di rifiutare l'ipotesi nulla: i residui sono autocorrelati nei ritardi 1,8,12,13
- Residui presentano bassa ACF ma buona PACF

2.6 Modellazione dei residui di ARIMA (2,0,2)

- Applicazione di ARIMA (3,2,1)



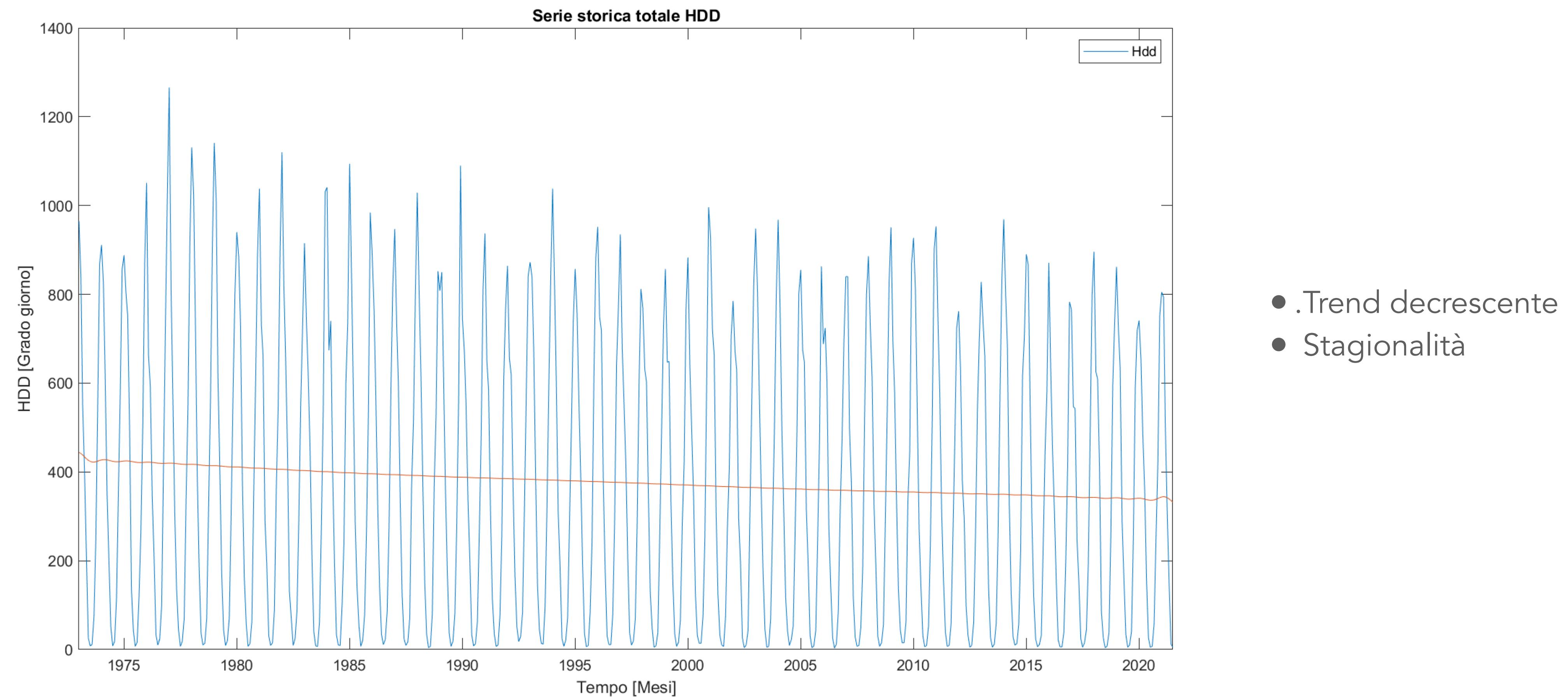
- ACF eliminata
- PACF non più statisticamente significativa, valutata con Ljung-Box test ai ritardi 6,8,12

```
>> [h,pValue,stat,cValue] = lbqtest(res_k,'lags',[6,8,12])  
h =  
1x3 logical array  
0 0 0  
pValue =  
0.3143 0.1445 0.2656
```

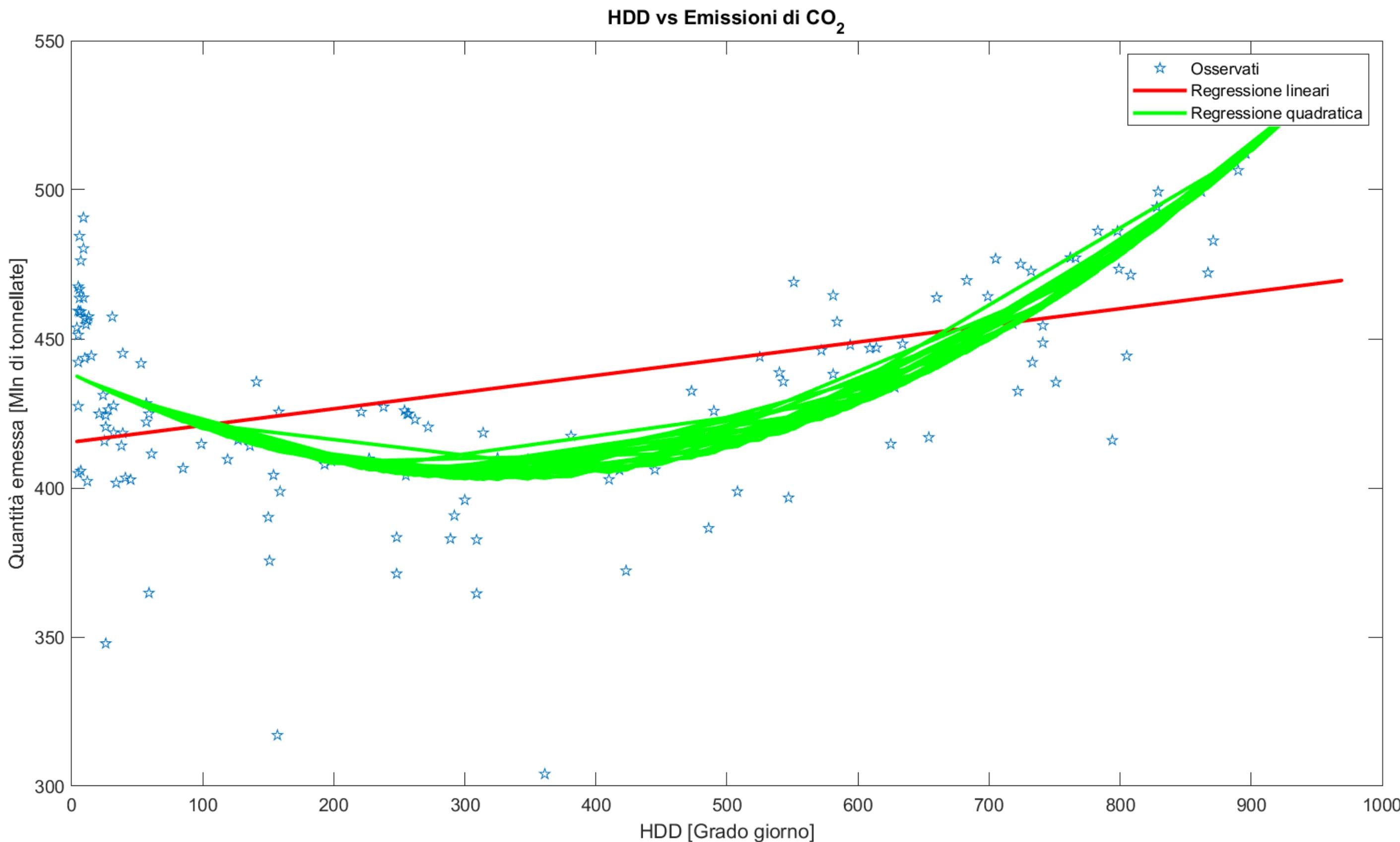


Relazione tra emissioni di CO₂ e HDD negli USA

Andamento della variabile predittiva HDD utilizzata per le successive analisi



3.1 Regressione statica



Regressione lineare semplice

CORRELAZIONE 0.44857

R² → 0.201

Regressione lineare multipla

CORRELAZIONE 0.5951

Estimated Coefficients:

	Estimate	SE	tStat	pValue
(Intercept)	438.33	3.8858	112.8	3.0803e-136
x1	-0.21769	0.026165	-8.3201	8.1938e-14
x2	0.00033403	3.0774e-05	10.854	3.7205e-20

Number of observations: 139, Error degrees of freedom: 136

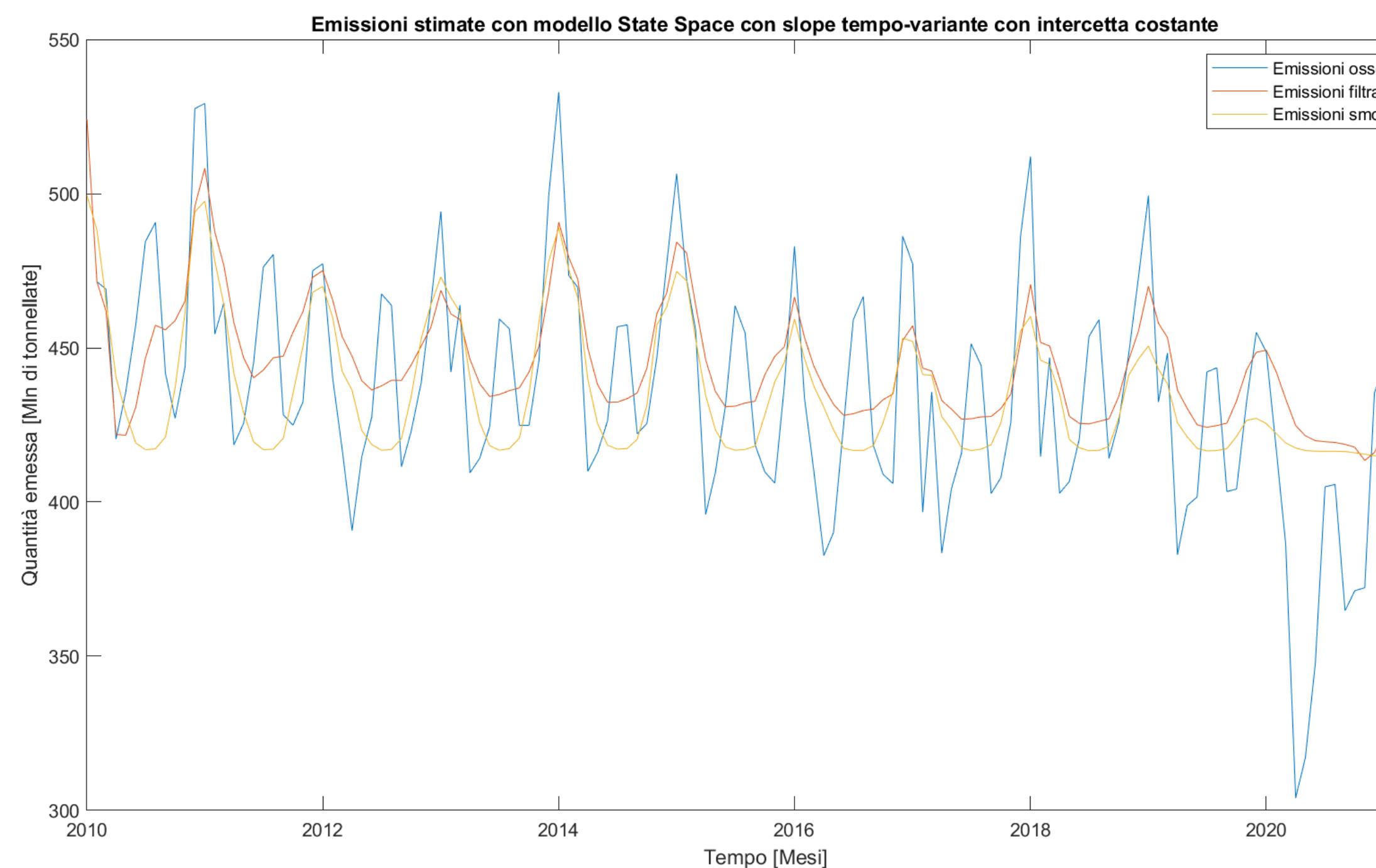
Root Mean Squared Error: 25.5

R-squared: 0.572, Adjusted R-Squared: 0.566

F-statistic vs. constant model: 90.9, p-value = 8.68e-26

3.2 Regressione dinamica

- Coefficiente angolare (β) tempo-variante e intercetta (α) costante



AIC 1420,31
BIC 1426,17
 R^2 filtrata $\rightarrow 0,3958$
 R^2 smussata $\rightarrow 0,3576$

```
tVarPar (params, x, lm.Coefficients.Estimate(1), lm.Coefficients.Estimate(2))
```

↓ ↓ ↓

Valore iniziale di alfa statico x(1) Valore iniziale di beta statico x(2)

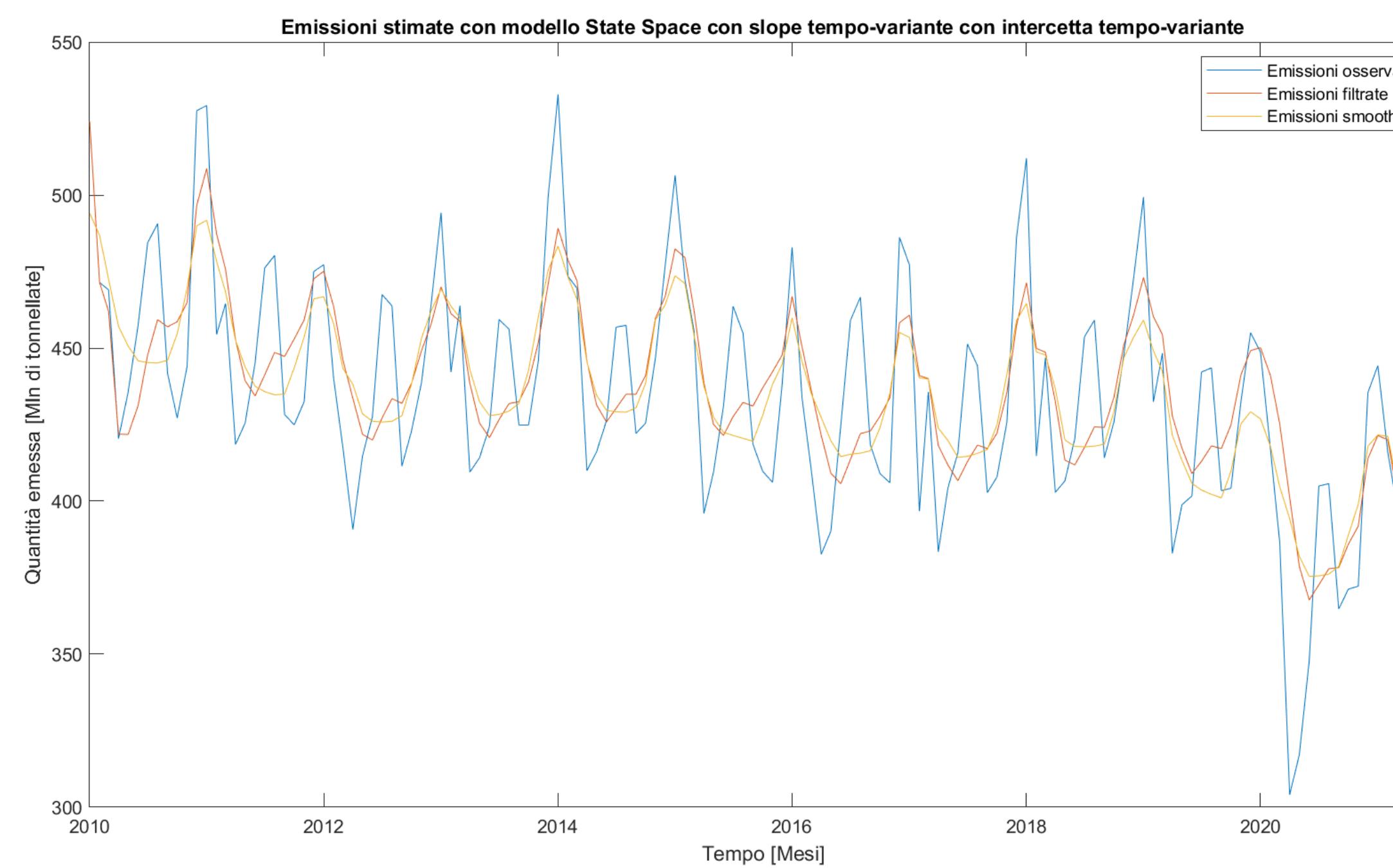
[0.10,log(var(lm.Residuals.Raw))]
Vettore di due elementi contenente il valore iniziale del logaritmo della varianza del coefficiente (es. 0,10) e il valore iniziale del logaritmo della varianza dell'errore statico

Method: Maximum likelihood (fminunc)
Sample size: 139
Logarithmic likelihood: -708.153
Akaike info criterion: 1420.31
Bayesian info criterion: 1426.17

	Coeff	Std Err	t Stat	Prob
c(1)	-10.29361	0.93546	-11.00377	0
c(2)	6.92038	0.11459	60.39372	0
	Final State	Std Dev	t Stat	Prob
x(1)	416.45722	4.08505	101.94662	0
x(2)	-0.00491	0.02205	-0.22284	0.82366

3.2 Regressione dinamica

- Coefficiente angolare (β) e intercetta (α) tempo-varianti



AIC 1399,13
BIC 1407,93
 R^2 filtrata $\rightarrow 0,5763$
 R^2 smussata $\rightarrow 0,5213$

```
tVarPar (params, x, lm.Coefficients.Estimate(1), lm.Coefficients.Estimate(2))
```

↓

Valore iniziale di alfa statico $x(1)$

↓

Valore iniziale di beta statico $x(2)$

[0.10, 0.10, log(var(lm.Residuals.Raw))]

Vettore di tre elementi contenente il valore iniziale del logaritmo della varianza dell'intercetta, il valore iniziale del logaritmo della varianza del coefficiente, e il valore iniziale del logaritmo della varianza dell'errore statico.

Method: Maximum likelihood (fminunc)

Sample size: 139

Logarithmic likelihood: -696.563

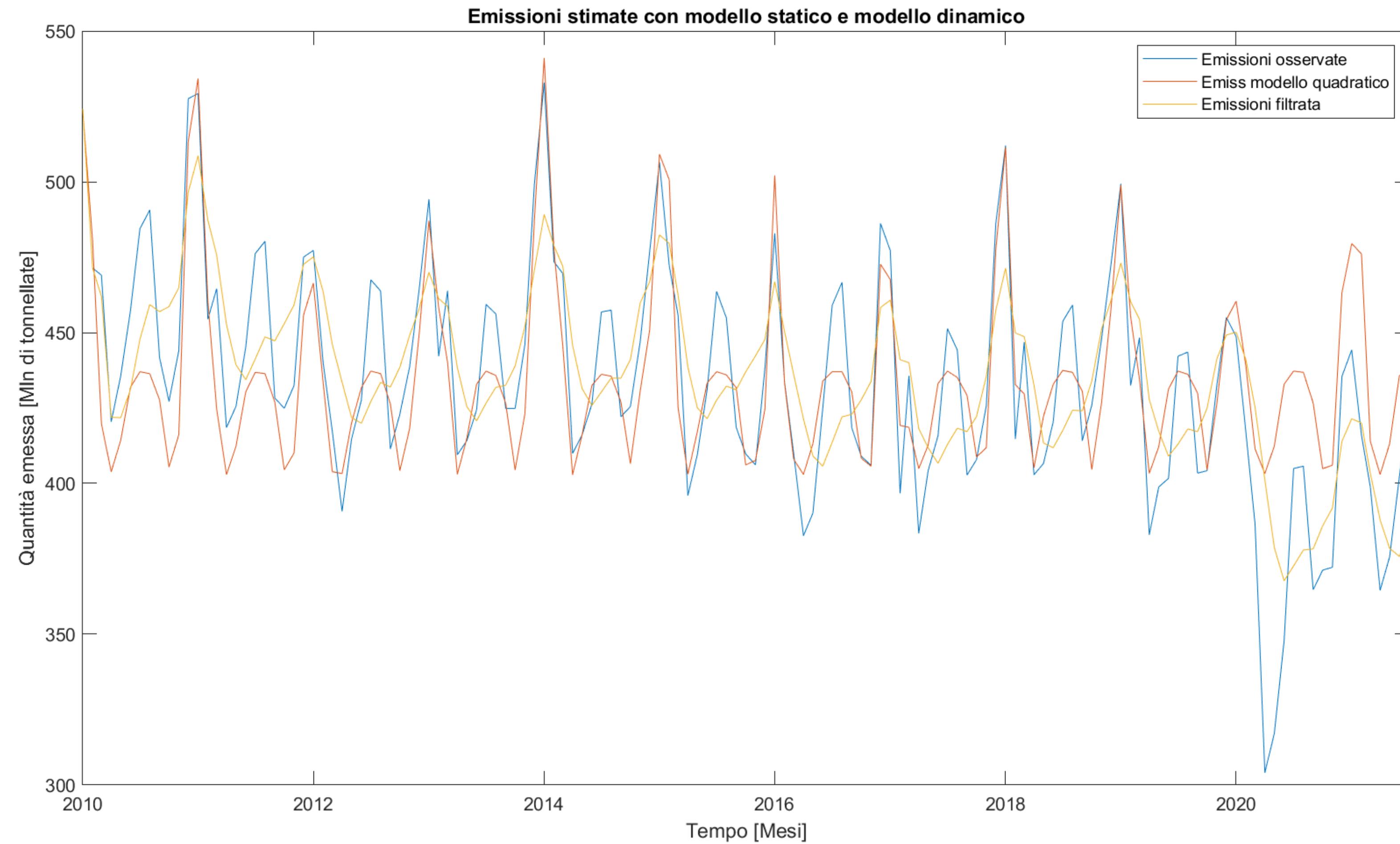
Akaike info criterion: 1399.13

Bayesian info criterion: 1407.93

	Coeff	Std Err	t Stat	Prob
<hr/>				
c (1)	3.10601	0.50506	6.14974	0
c (2)	-25.82321	1801.37878	-0.01434	0.98856
c (3)	6.66820	0.16307	40.89208	0
<hr/>				
	Final State	Std Dev	t Stat	Prob
x (1)	383.47635	11.28410	33.98378	0
x (2)	0.05414	0.00812	6.66321	0

3.3 Confronto fra modelli e valutazione ottimo

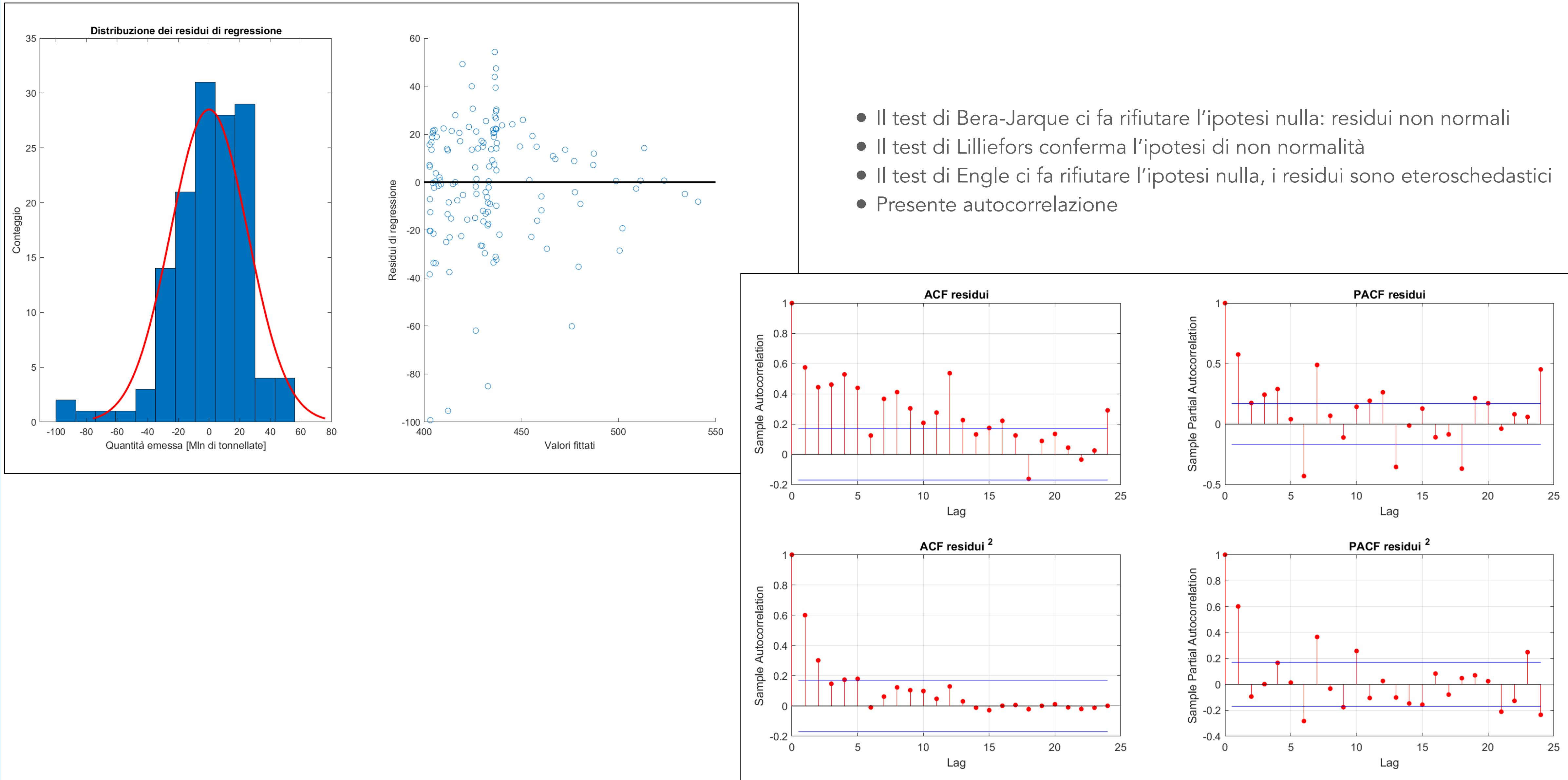
- Migliore dinamico (parametri variabili filtrato)
- Migliore statico (quadratico)



$R^2 \rightarrow 0.57$

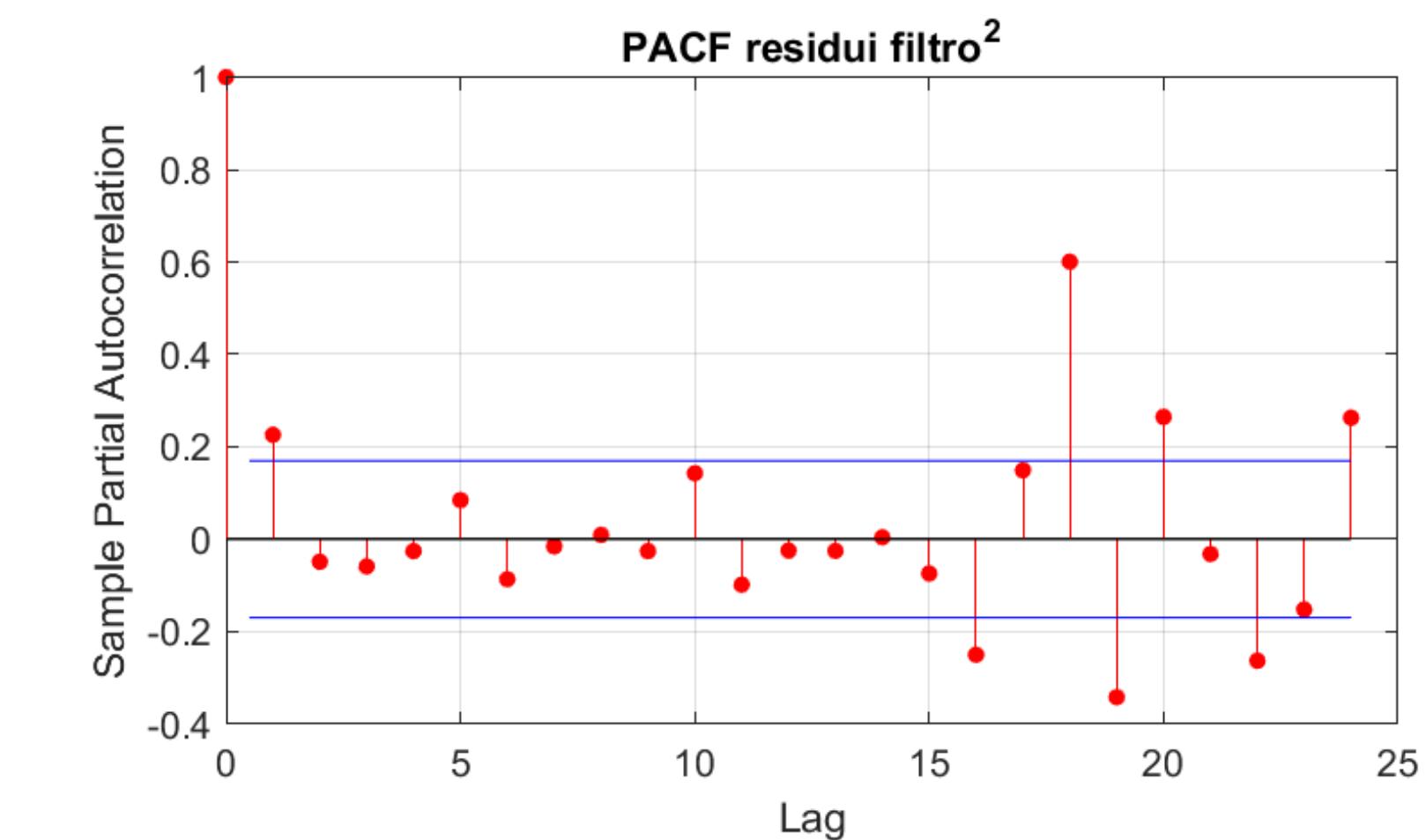
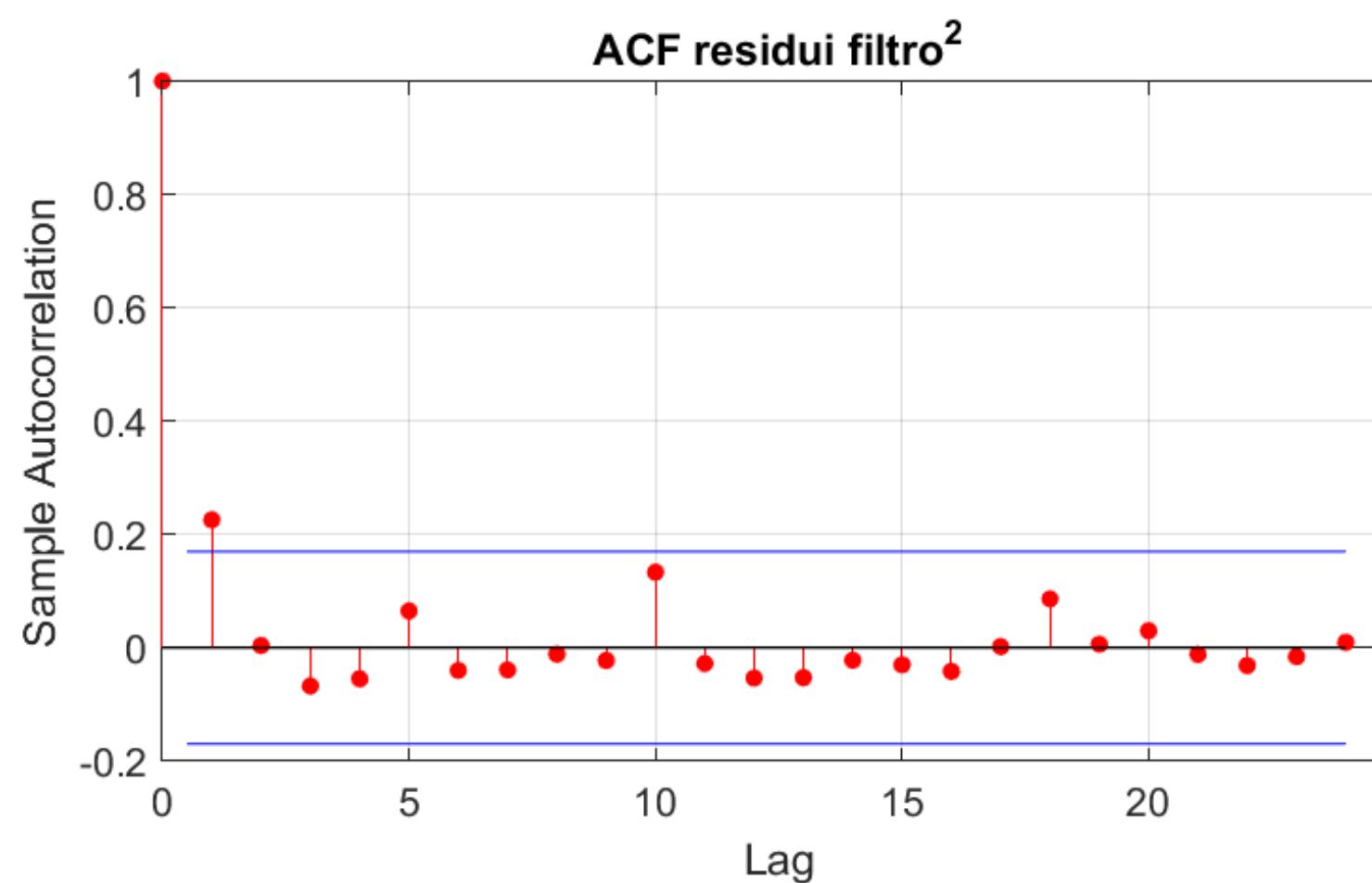
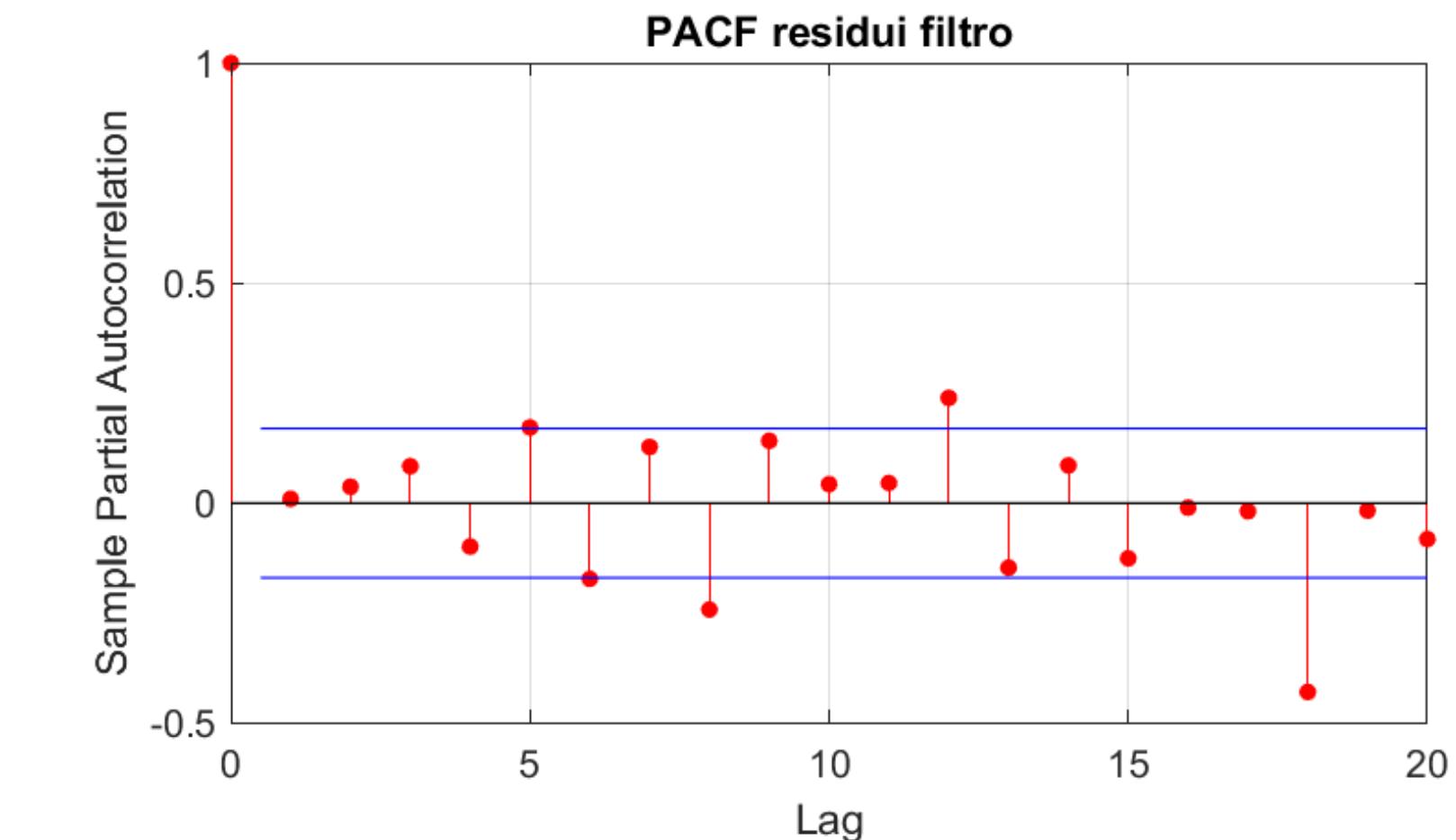
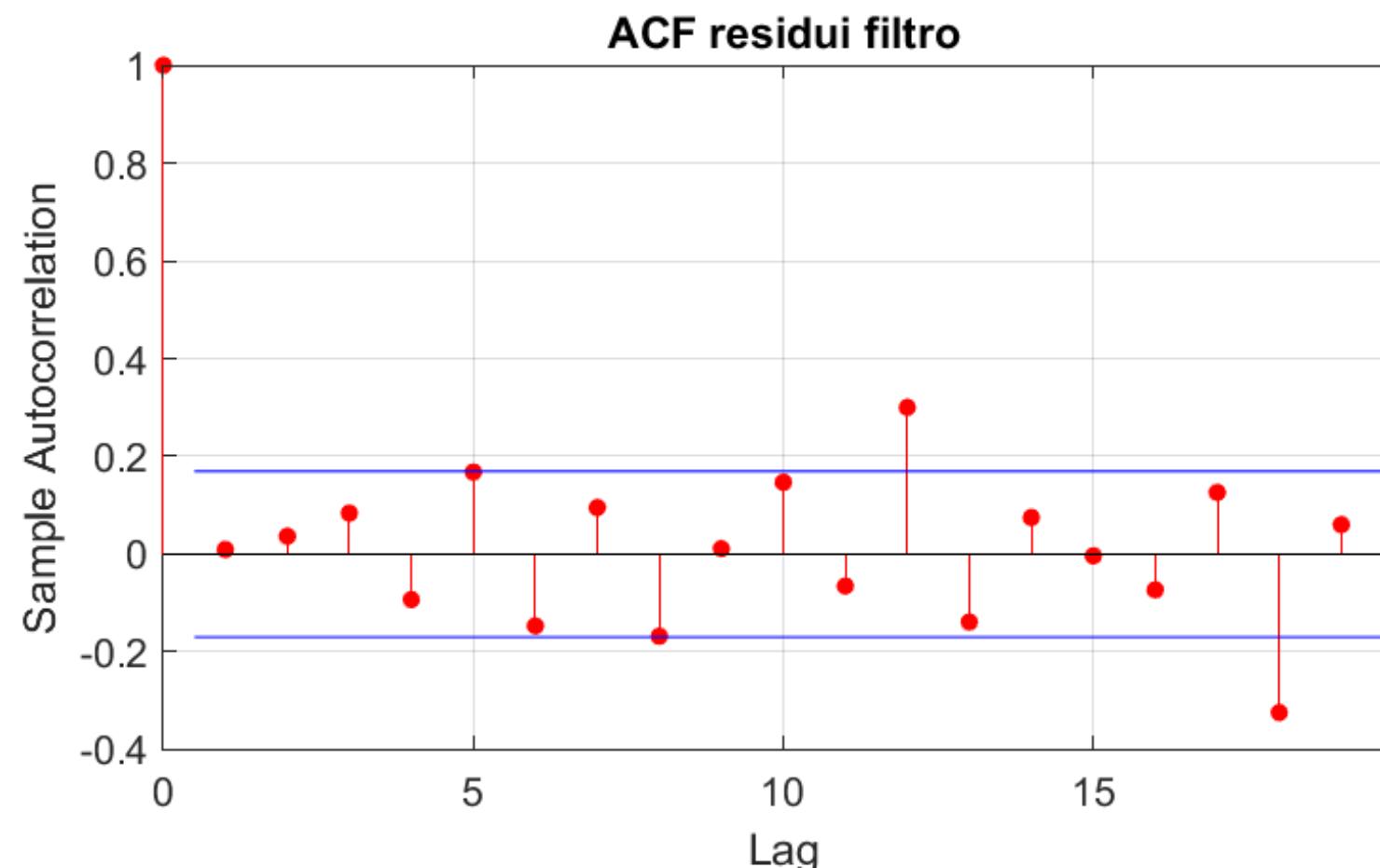
3.4 Analisi dei residui miglior modello

- Statico (quadratico)



3.5 Modellazione dei residui

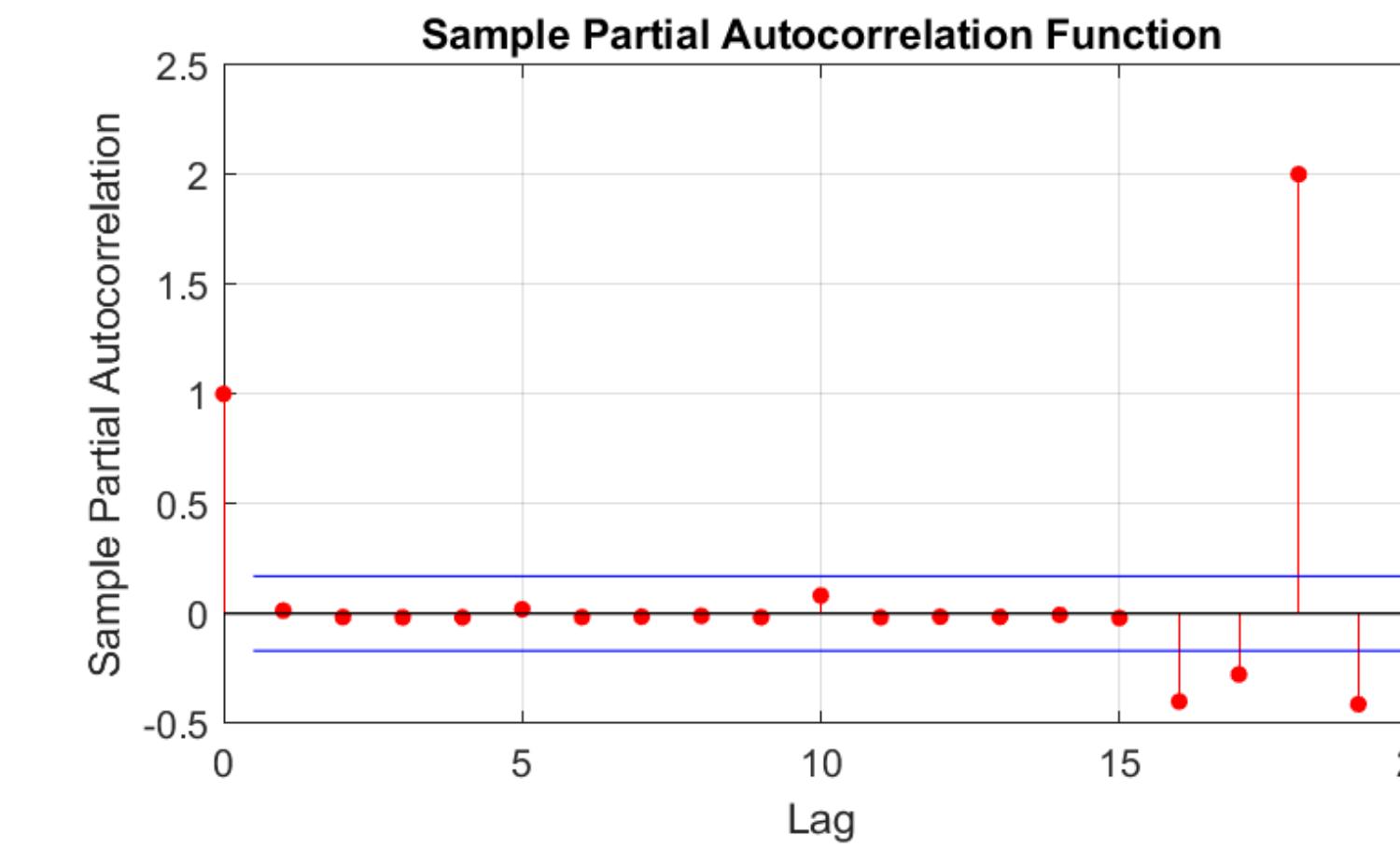
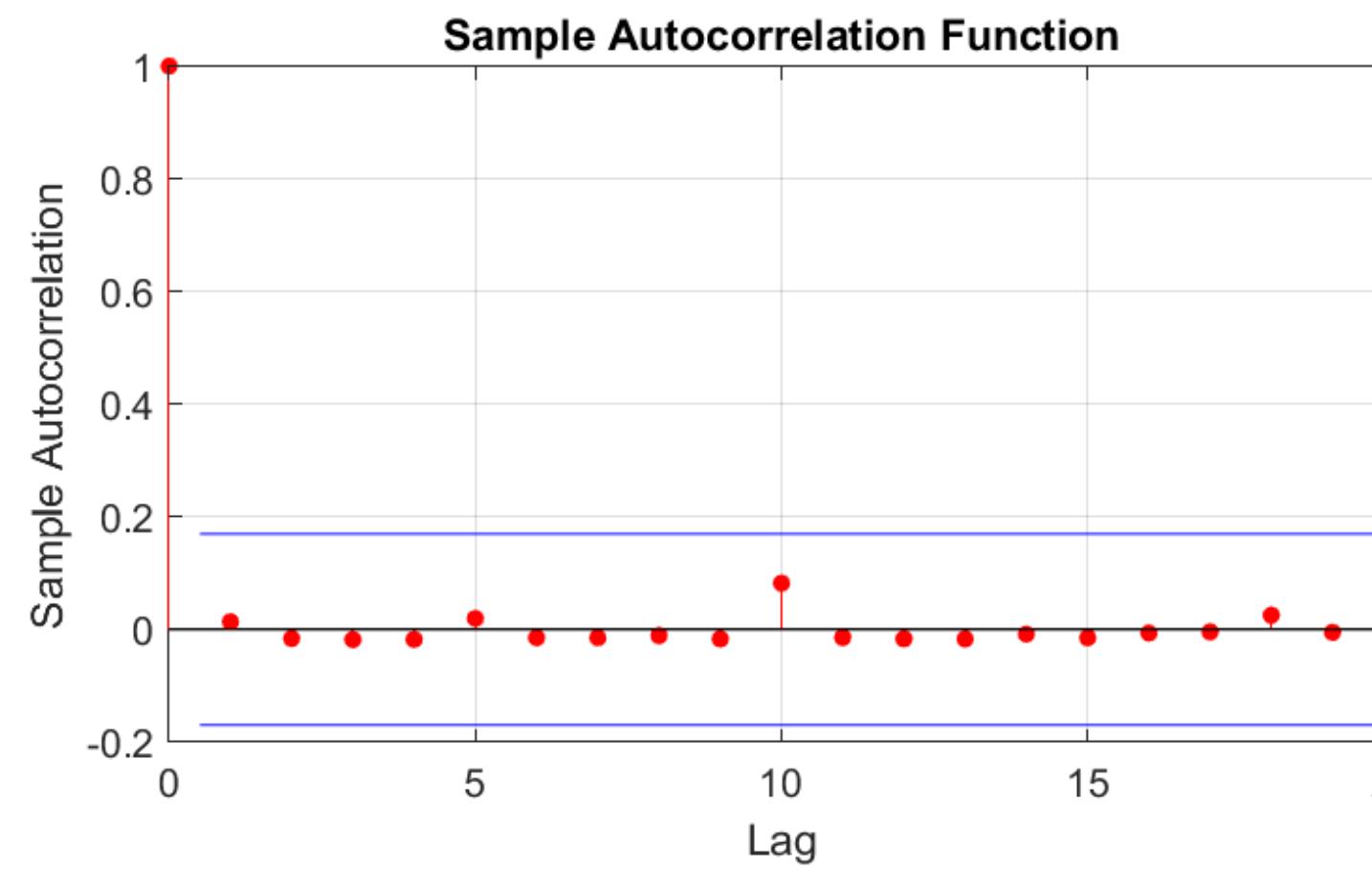
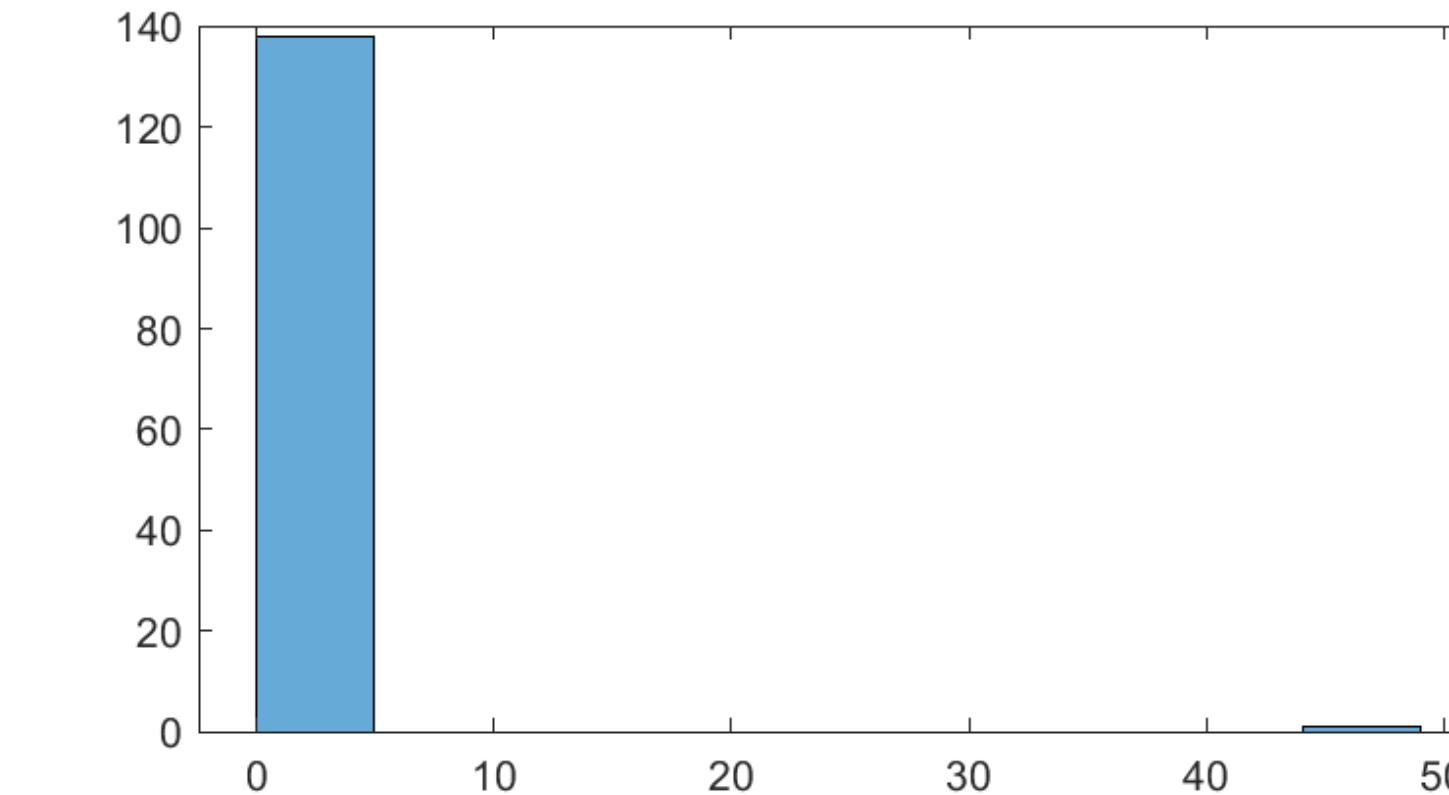
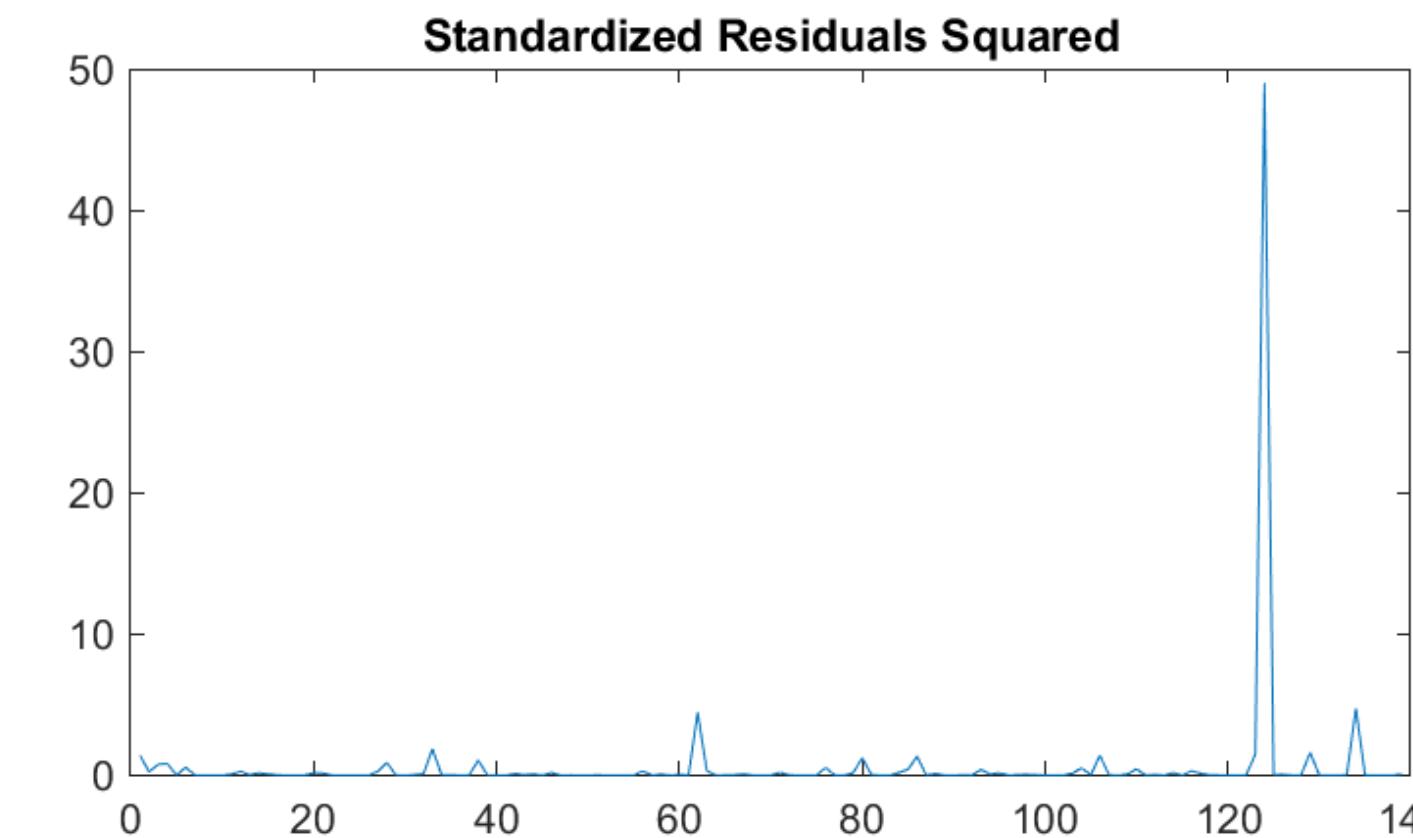
- Applicazione di ARIMA (3,0,5) finalizzato a rimuovere autocorrelazione



- ACF e PACF parzialmente ridotte

3.5 Modellazione dei residui

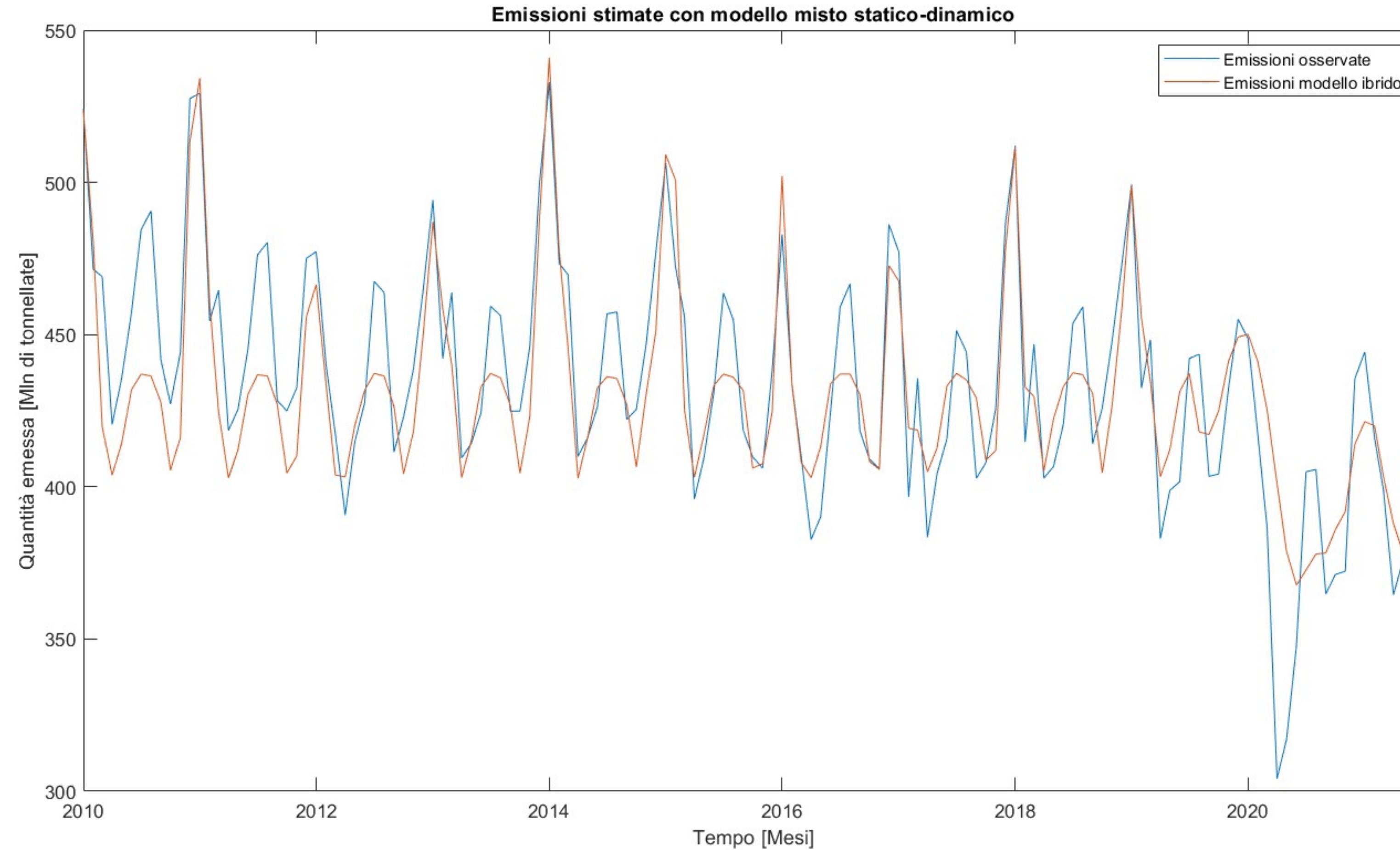
- Applicazione di GARCH (0,2) finalizzato a rimuovere autocorrelazione dai residui al quadrato



- I residui standardizzati al quadrato non presentano più autocorrelazione

3.6 Modello ibrido

- Modello statico per i primi 9 anni e modello dinamico per gli ultimi 2



$R^2 \rightarrow 0.68$

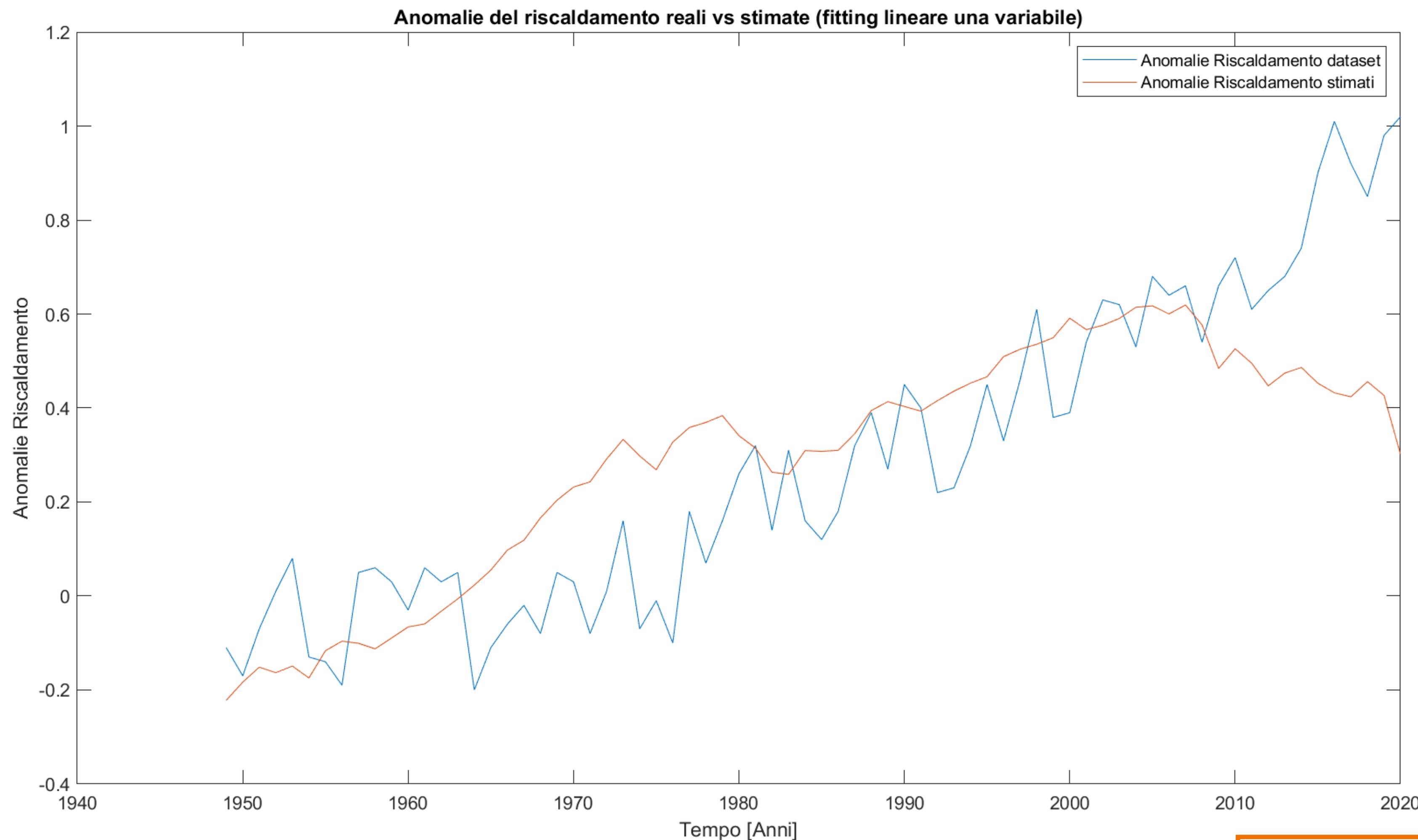


Stima delle anomalie delle temperature globali

Passaggio a dataset annuale

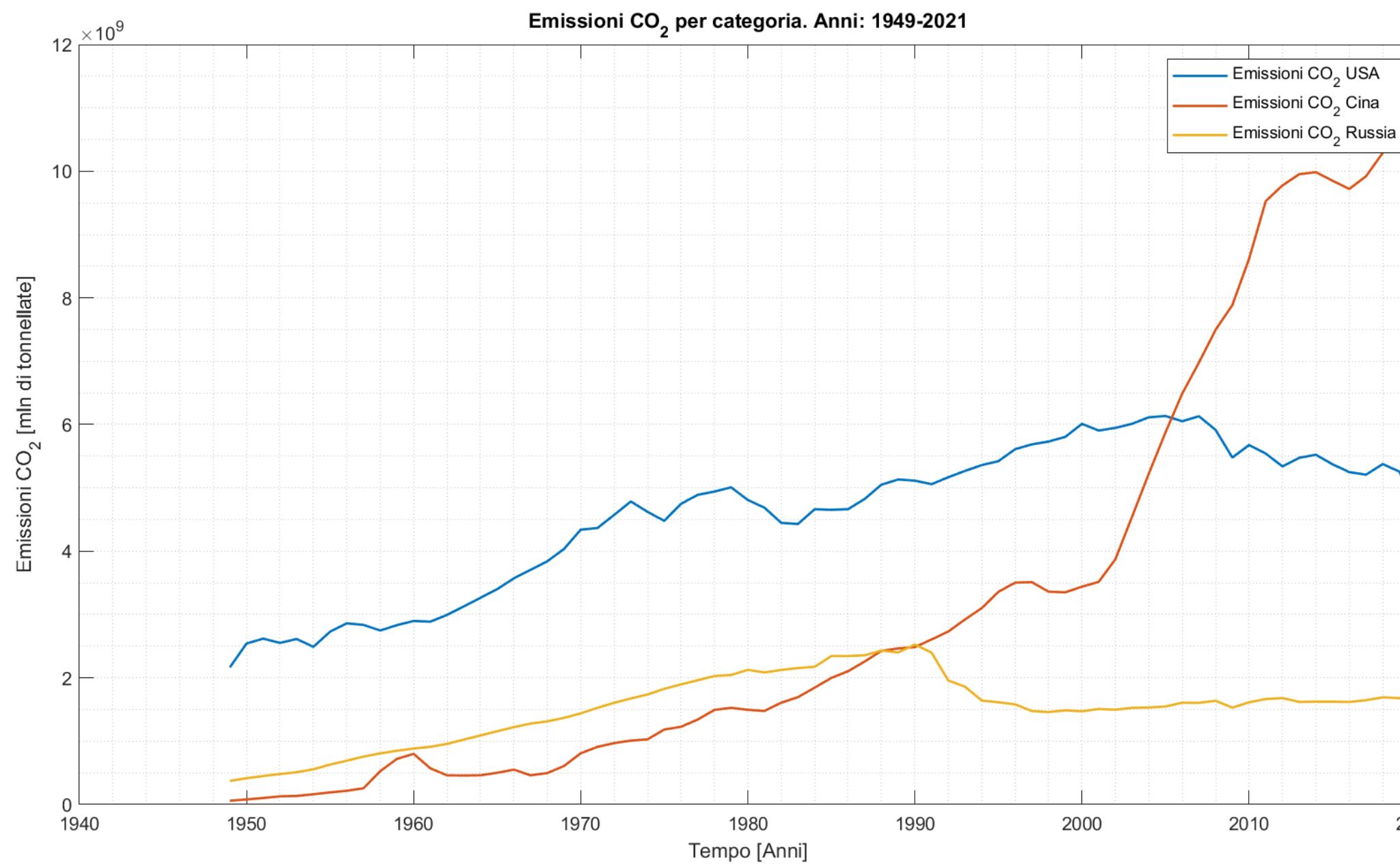
1949 - 2021

4.1 Regressione semplice fra anomalie e emissioni di CO₂ degli USA



CORRELAZIONE 0.74
 $R^2 \rightarrow 0.562$

4.2 Regressione multipla fra anomalie e emissioni di CO₂ di USA, Cina e Russia



Estimated Coefficients:

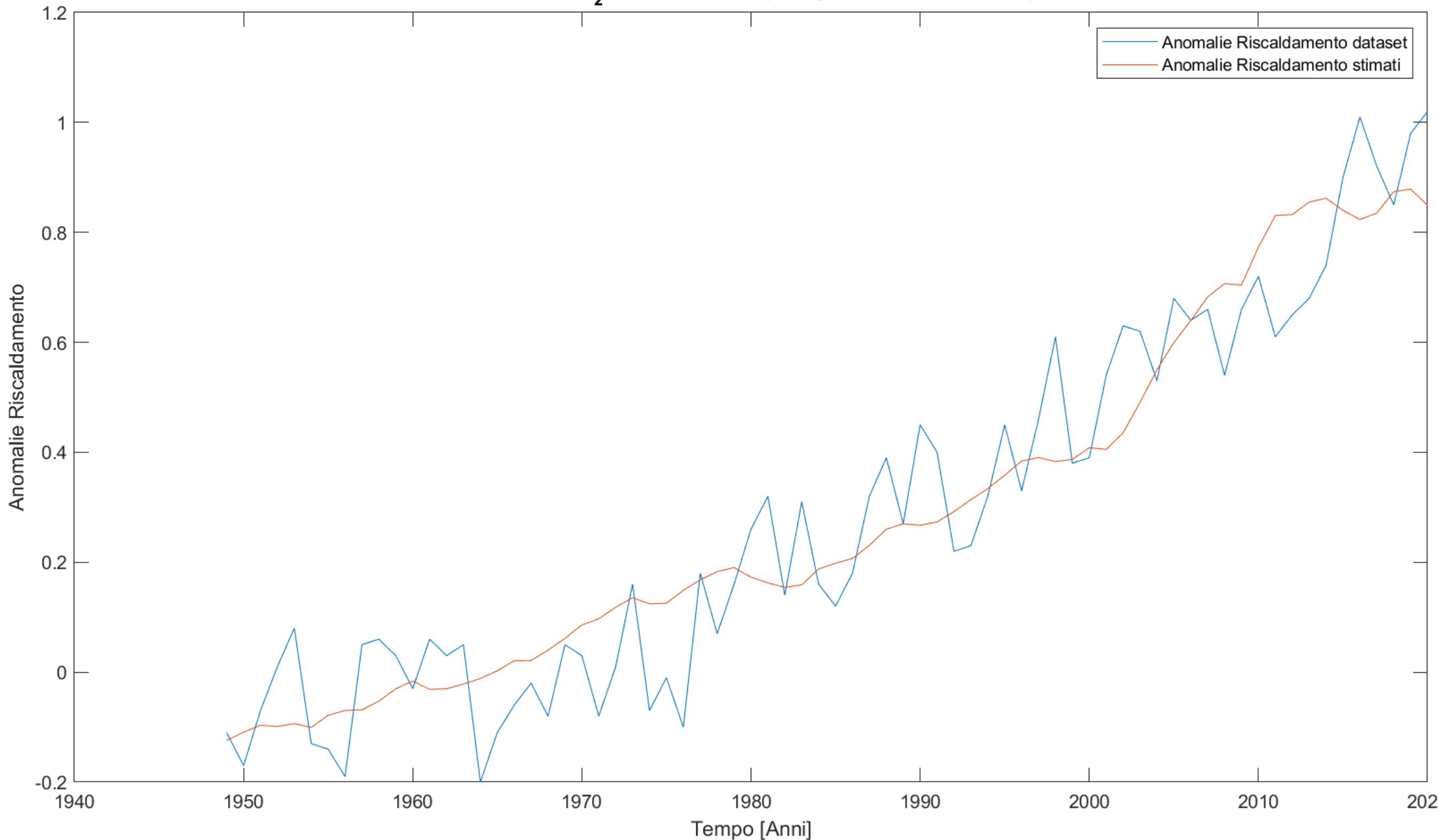
	Estimate	SE	tstat	pValue
(Intercept)	-0.29878	0.063714	-4.6894	1.3654e-05
TotalEnergyCO2EmissionsUSA	7.7038e-11	2.1994e-11	3.5027	0.00081879
TotalEnergyCO2EmissionChina	7.4244e-11	5.6321e-12	13.182	2.1034e-20
TotalEnergyCO2EmissionRussia	-4.3702e-12	3.7112e-11	-0.11776	0.90661

Number of observations: 72, Error degrees of freedom: 68

Root Mean Squared Error: 0.115

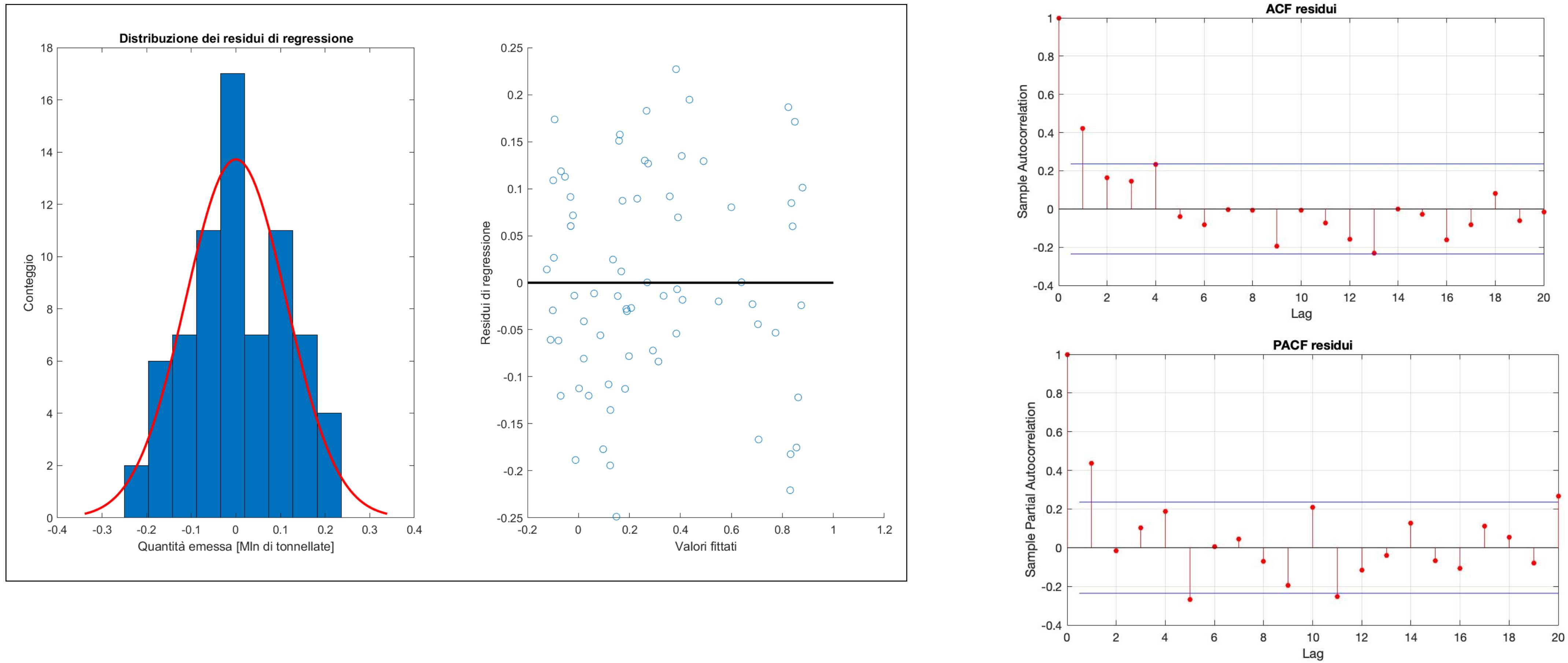
R-squared: 0.886, Adjusted R-Squared: 0.881

F-statistic vs. constant model: 177, p-value = 4.72e-32



R² → 0.886

4.3 Analisi dei residui miglior modello



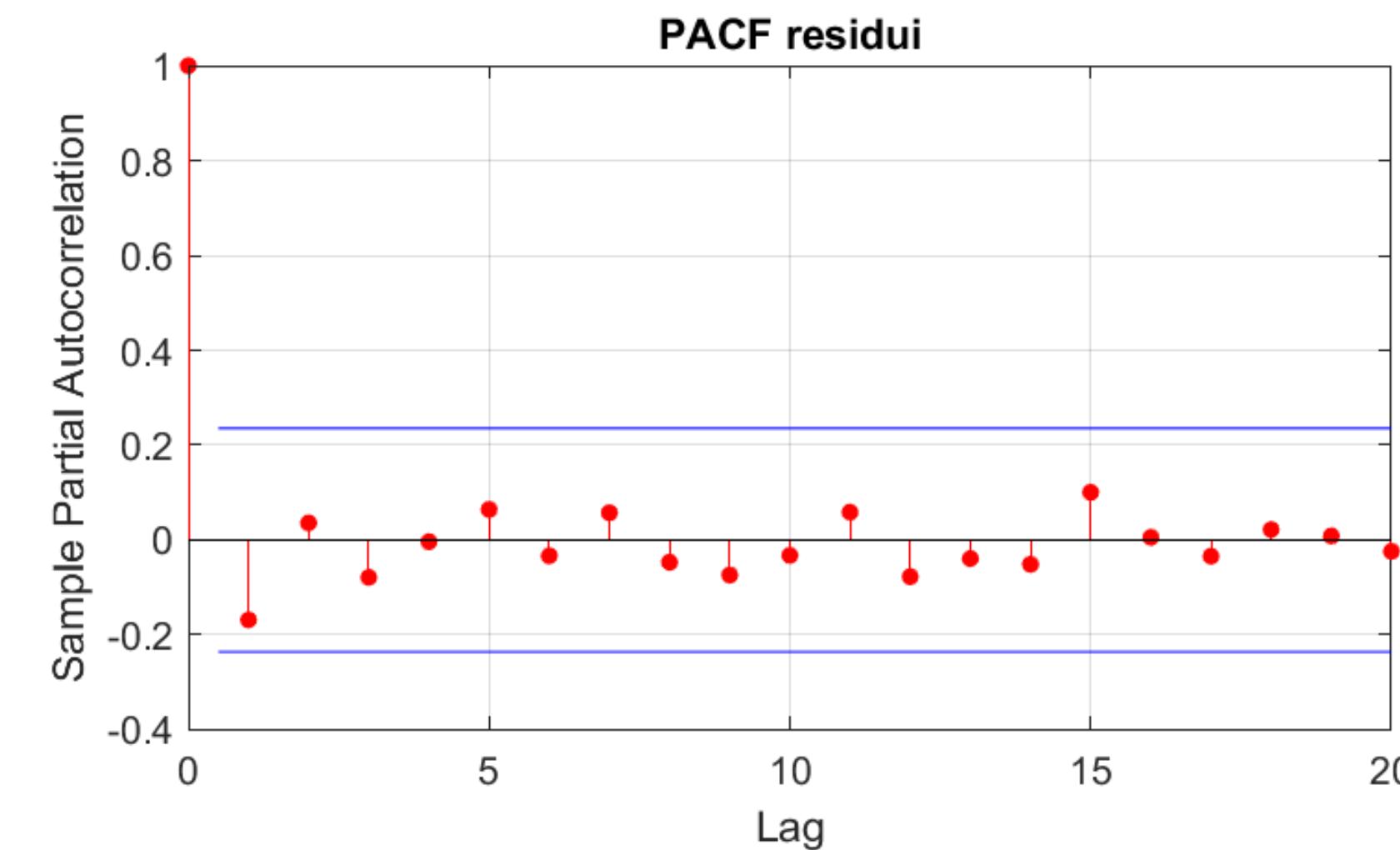
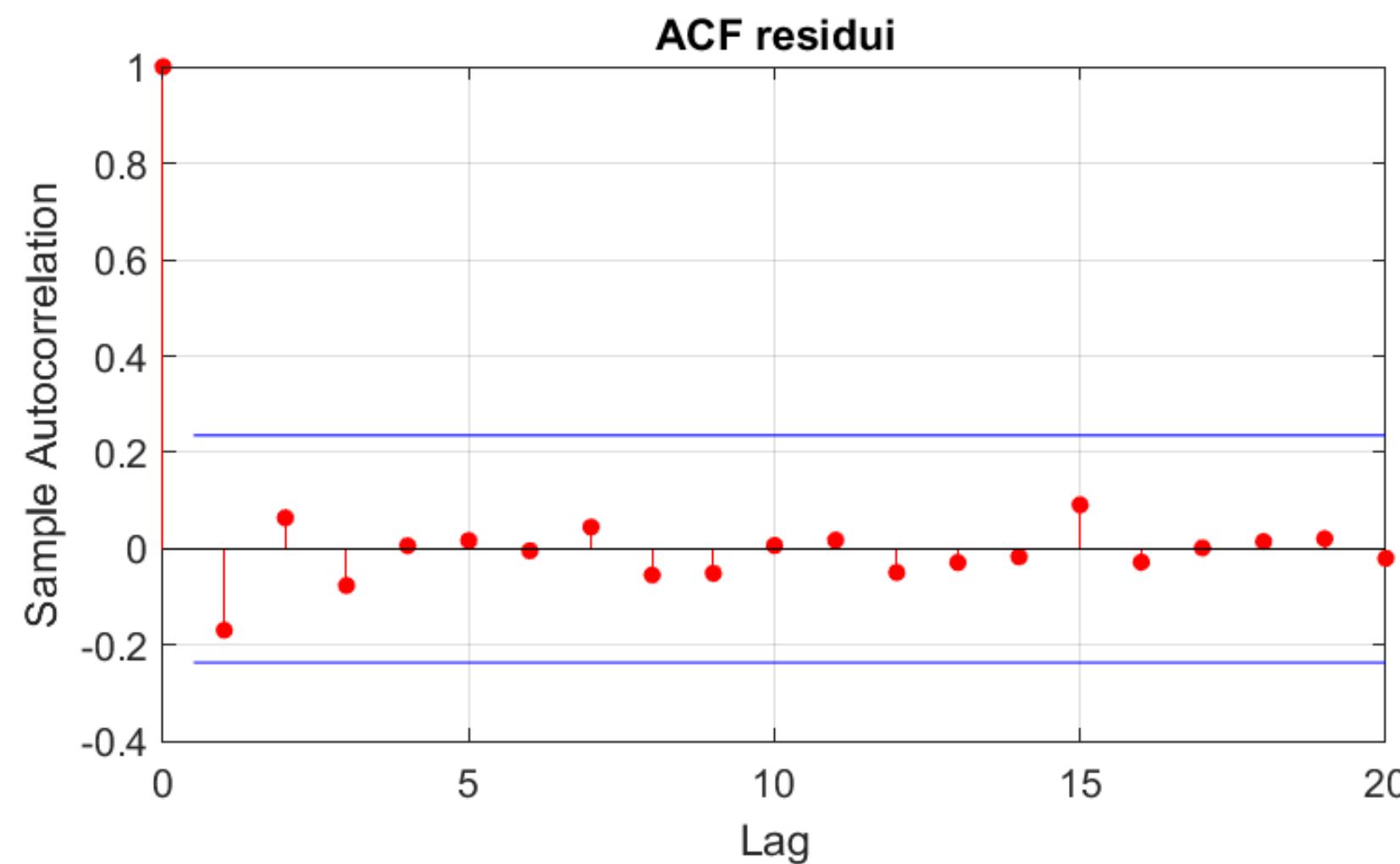
- T-test ci fa accettare l'ipotesi nulla, la media dei residui è 0
- L'ipotesi di normalità dei residui è confermata dal test di Lilliefors al 5% e da quello Bera-Jarque con livello di significatività dell'1% e del 5%
- Il test di Engle ci fa accettare H_0 , i residui risultano omoschedastici
- I residui risultano poco autocorrelati

4.4 Modellazione dei residui

- Applicazione di ARMA (2,1) finalizzato a rimuovere autocorrelazione

$$\text{Equazione} \rightarrow y_t = 5 + \underline{\alpha_1 y_{t-1}} + \underline{\alpha_2 y_{t-2}} + \varepsilon_t + \underline{\theta_1 \varepsilon_{t-1}}$$

↓ ↓ ↓
0.1 0.2 0.5



- Ora i residui rispettano le ipotesi di omoschedasticità, non correlazione, media nulla e stazionarietà, si tratta di residui white noise

IMPOSTAZIONE LAVORO

DOMANDE: inizialmente ci eravamo posti domande diverse, in corso d'opera ci siamo trovati a doverle rivedere in quanto i dati che avevamo a disposizione non erano sufficienti per dare risposte adeguate.

SCRIPT: siamo riusciti a collaborare e a stare al passo con il lavoro svolto servendoci del software GitHub; ognuno ha provato a risolvere i vari quesiti e insieme li abbiamo argomentati.

REPORT: ci siamo suddivisi le parti da rielaborare per la relazione finale; insieme le abbiamo compattate servendoci di Latex.