Big Data Management and Analysis in Physics Research

Tommaso Redaelli – 830442

Time Series Analysis

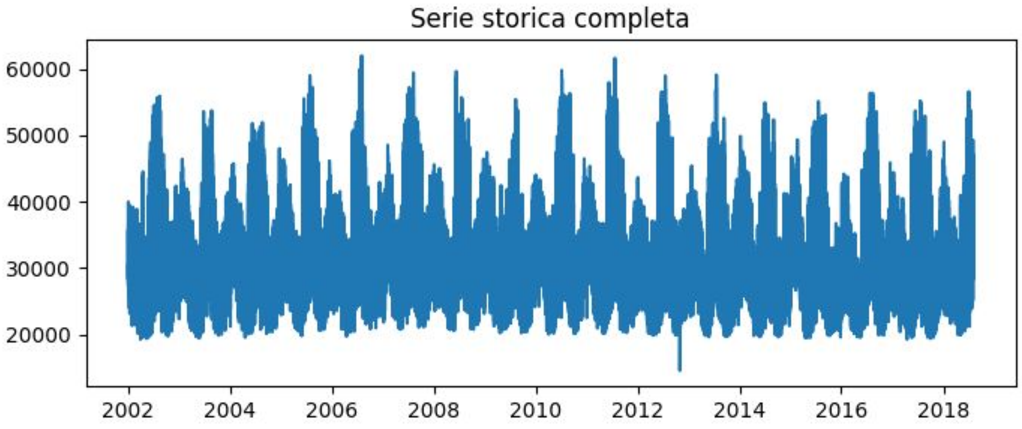
1. Dataset
2. Decomposizione
3. Regressione trend
4. SARIMA model

Dataset — Hourly Energy Consumption

Dimension:

- 145366 records
- 2 colonne

	Datetime	Value
Datetime		
2002-01-01 01:00:00	2002-01-01 01:00:00	30393.0
2002-01-01 02:00:00	2002-01-01 02:00:00	29265.0
2002-01-01 03:00:00	2002-01-01 03:00:00	28357.0
2002-01-01 04:00:00	2002-01-01 04:00:00	27899.0
2002-01-01 05:00:00	2002-01-01 05:00:00	28057.0
...	...	...
2018-08-02 20:00:00	2018-08-02 20:00:00	44057.0
2018-08-02 21:00:00	2018-08-02 21:00:00	43256.0
2018-08-02 22:00:00	2018-08-02 22:00:00	41552.0
2018-08-02 23:00:00	2018-08-02 23:00:00	38500.0
2018-08-03 00:00:00	2018-08-03 00:00:00	35486.0



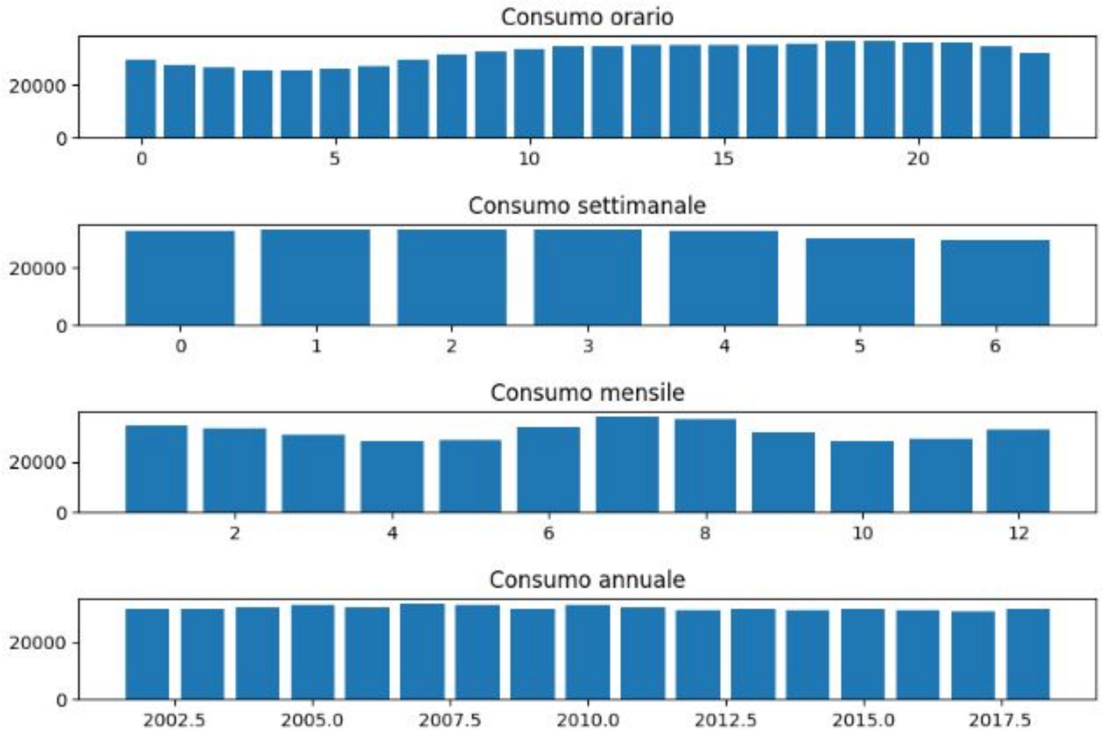
Descrizione periodicità

Media consumo per ogni ora

Media consumo per giorno della settimana

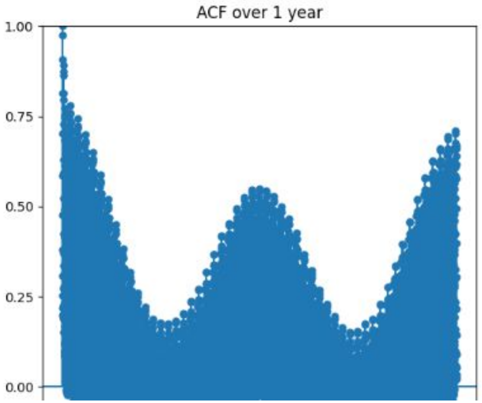
Media consumo per ogni mese

Media consumo per ogni anno

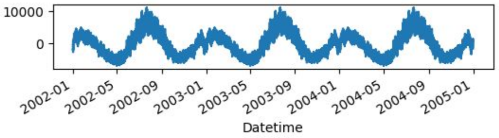


Autocorrelazione e stagionalità

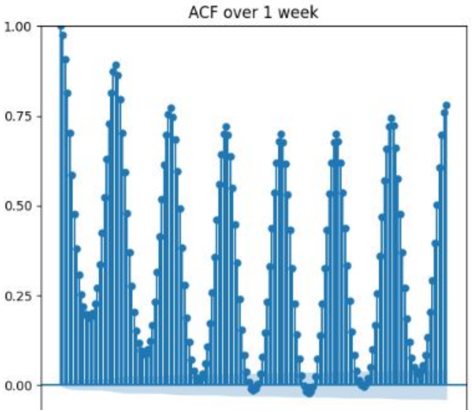
*Lag size = 24 * 365*



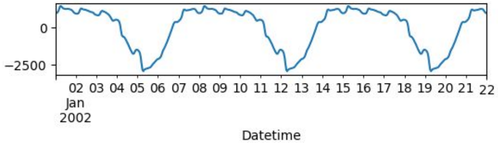
Stagionalità annuale



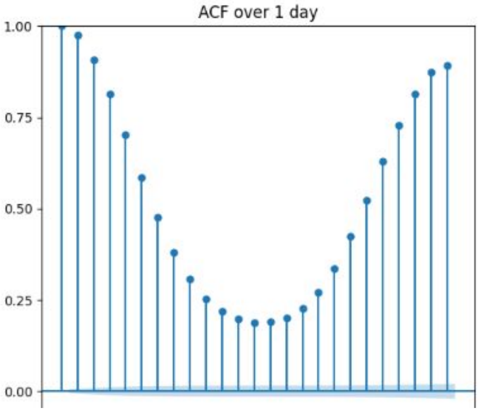
*Lag size = 24 * 7*



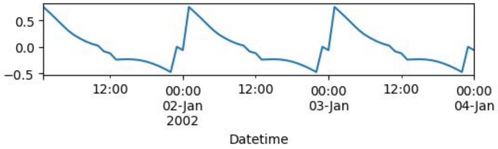
Stagionalità settimanale



Lag size = 24



Stagionalità giornaliera



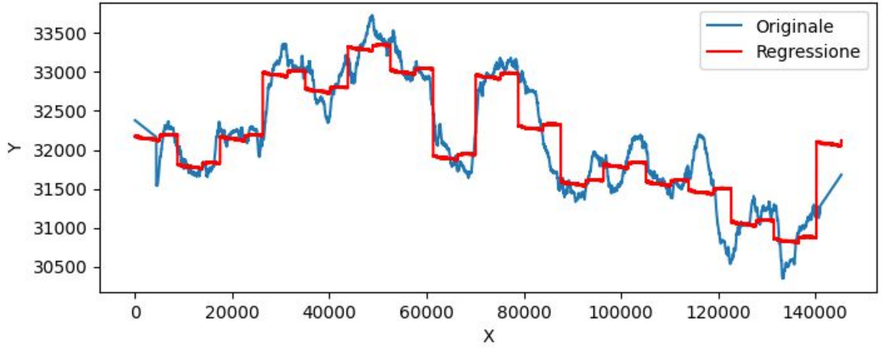
Regressione lineare sul trend

Costruzione variabili esplicative →
- Day of the year $\in \{ 1, \ldots, 365 \}$
- Month $\in \{ 1, \ldots, 12 \}$
- Day of the week $\in \{ 1, \ldots, 7 \}$
- Hour $\in \{ 1, \ldots, 24 \}$

Categorical to dummies → 426 variabili binarie → Riduzione PCA a 50 variabili → Varianza spiegata pari a 74.08 %

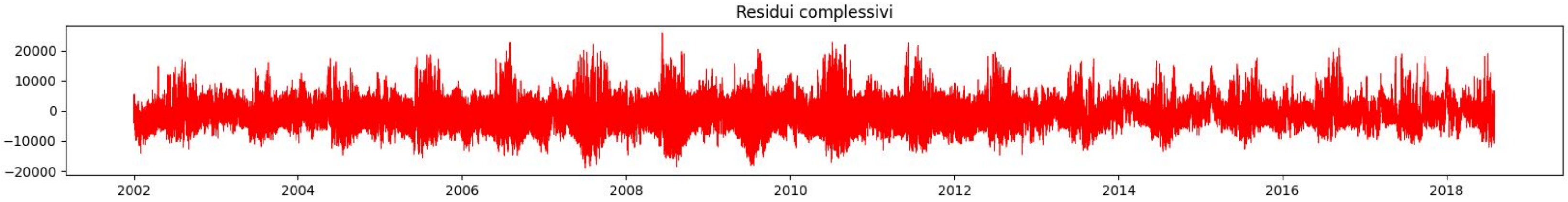
R-squared:	0.850
Adj. R-squared:	0.850
F-statistic:	1.647e+04
Prob (F-statistic):	0.00
Log-Likelihood:	-1.0324e+06
AIC:	2.065e+06
BIC:	2.065e+06

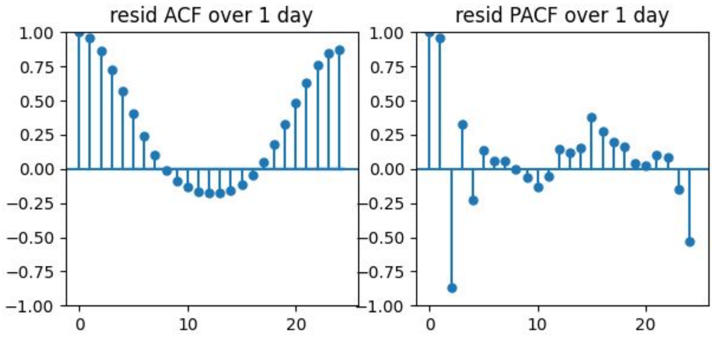
38 su 50 var sono significative per α = 0.05

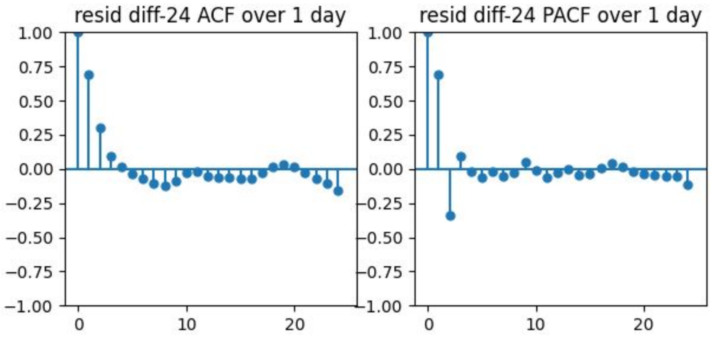


Modello SARIMA — *Definizione*

$\hat{Y}$ = *predizione trend + stag. annuale + stag. settimanale + stag. giornaliera +* ***SARIMA sui residui***





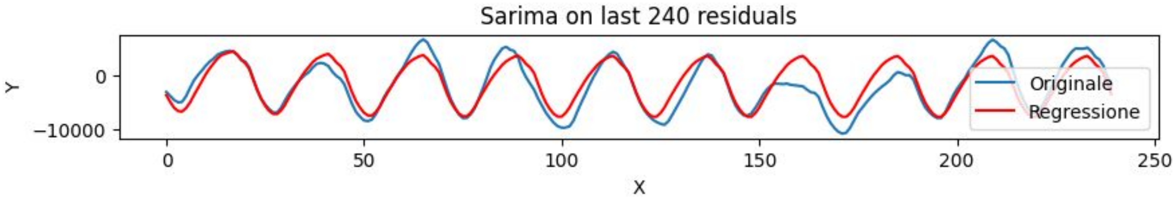


Order: ***p*** *= 2;* ***d*** *= 0;* ***q*** *= 2*

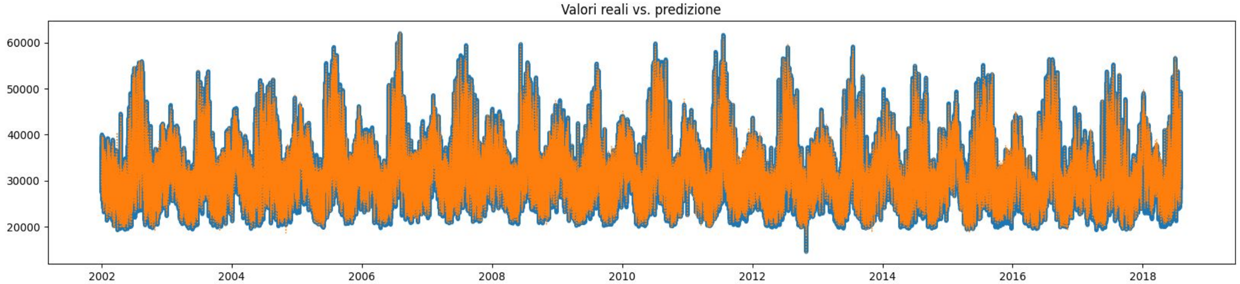
Seasonal order: ***P*** *= 1;* ***D*** *= 1;* ***Q*** *= 1 [24]*

Modello SARIMA — *Risultati*

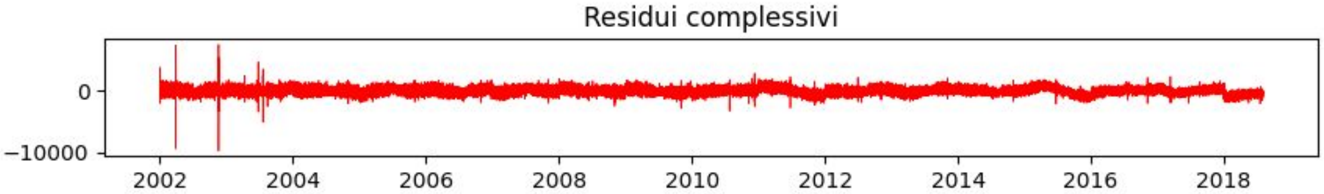
Log Likelihood	-1036723.555
AIC	2073461.110
BIC	2073530.318
HQIC	2073481.766



Valore reale vs. Ŷ



Residui su Ŷ
RMSE: 422



Modello SARIMA — *Previsione*