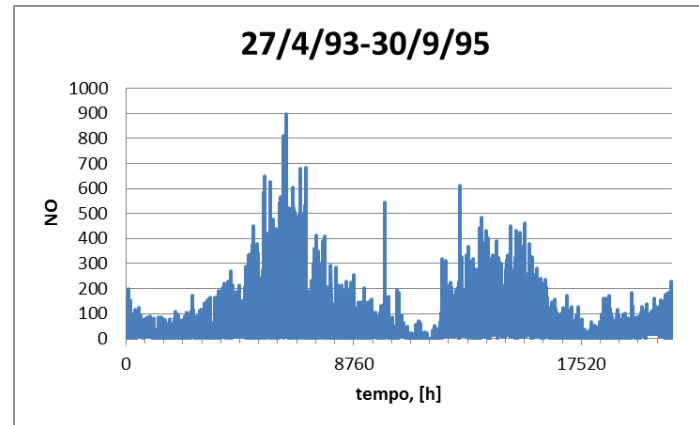
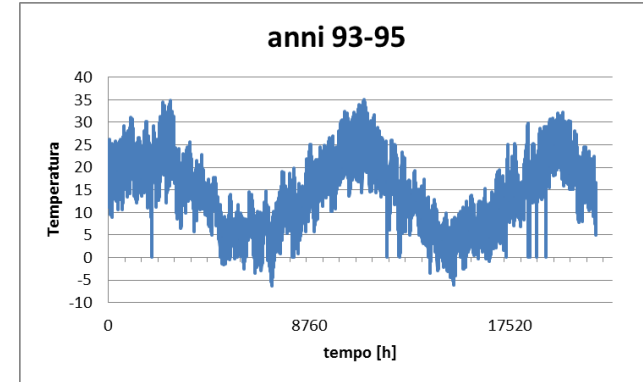
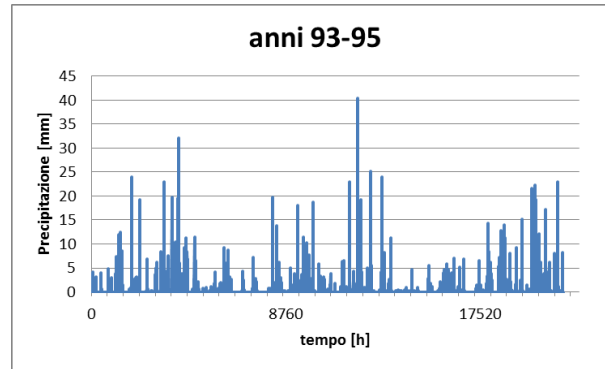


## Analisi serie temporali

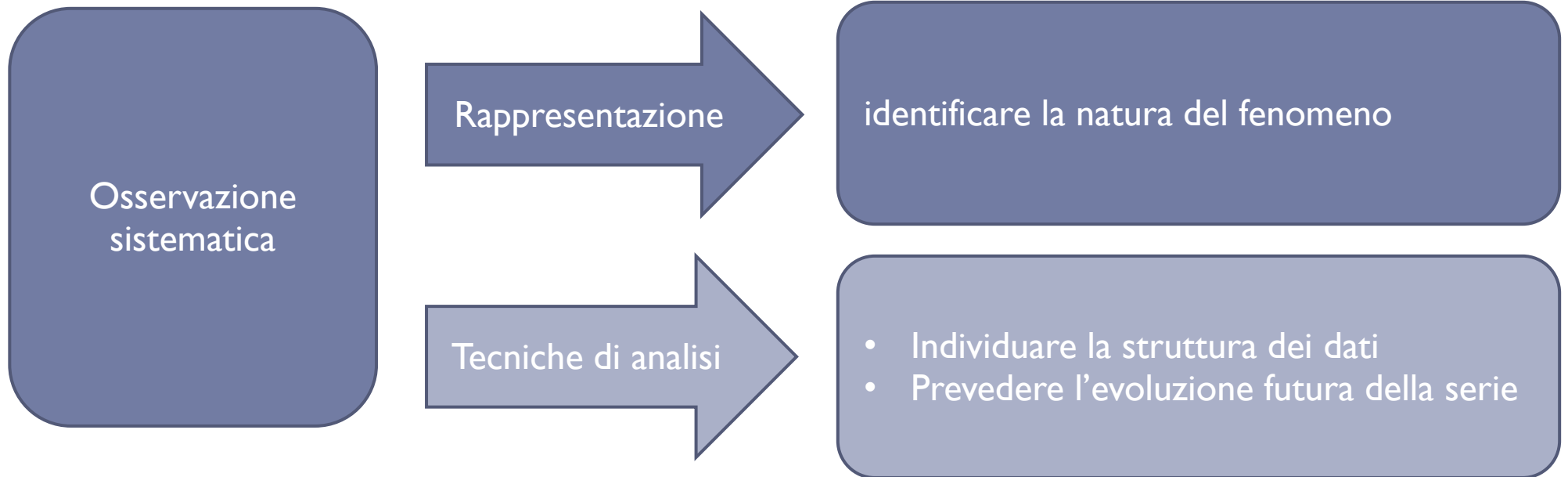
A.A. 2017/2018

Marina Campolo



## Obiettivo dell'analisi

- ▶ Serie temporale: sequenza di misure acquisite ad intervalli regolari di tempo
- ▶ L'ordine conta!

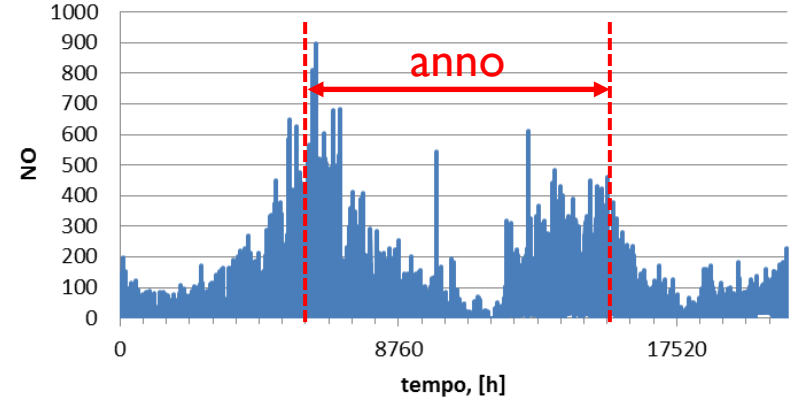


## Scala temporale "significativa"

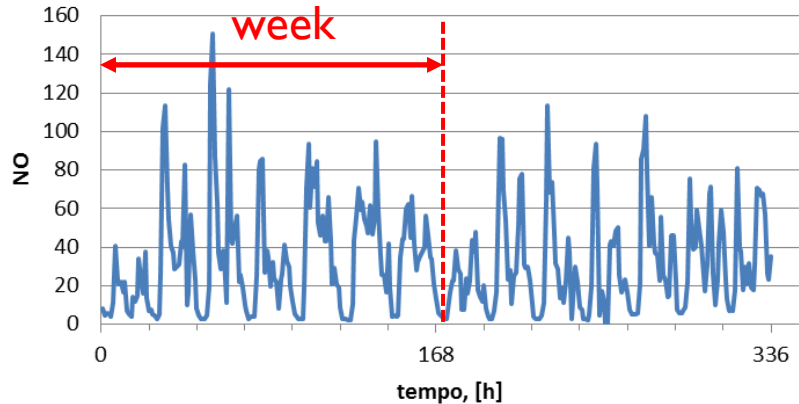
Variabili ambientali influenzate dalla stagionalità

→ grafici a diversa risoluzione per identificare scale significative (annuale, settimanale, giornaliera)

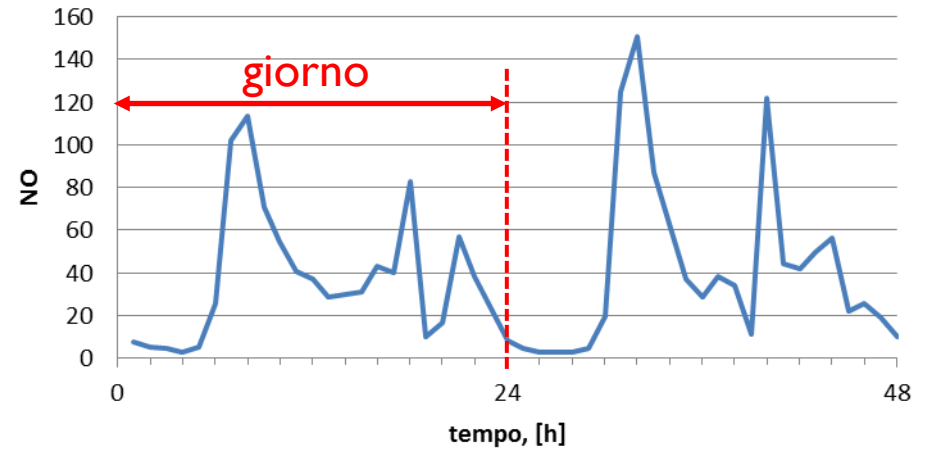
27/4/93-30/9/95



2/5/93-16/5/93



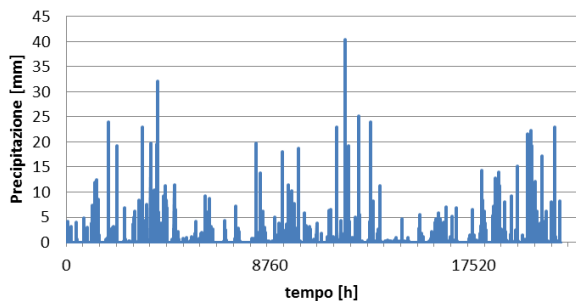
4/5/93-5/5/93



# Caratteristiche dei dati

## precipitazione

anni 93-95

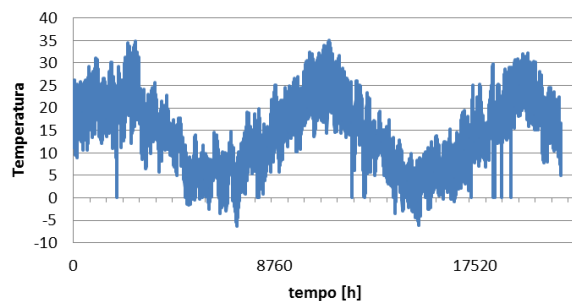


Valori positivi

Serie discontinua

## temperatura

anni 93-95

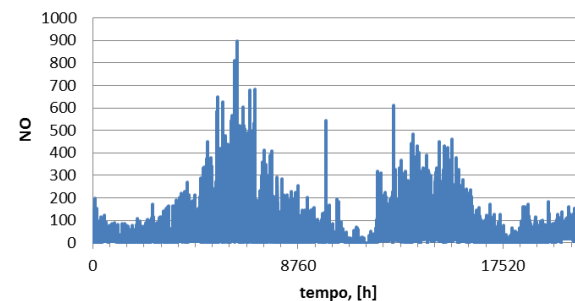


Valori positivi e negativi

Serie continua con fluttuazioni di varia frequenza

## inquinante

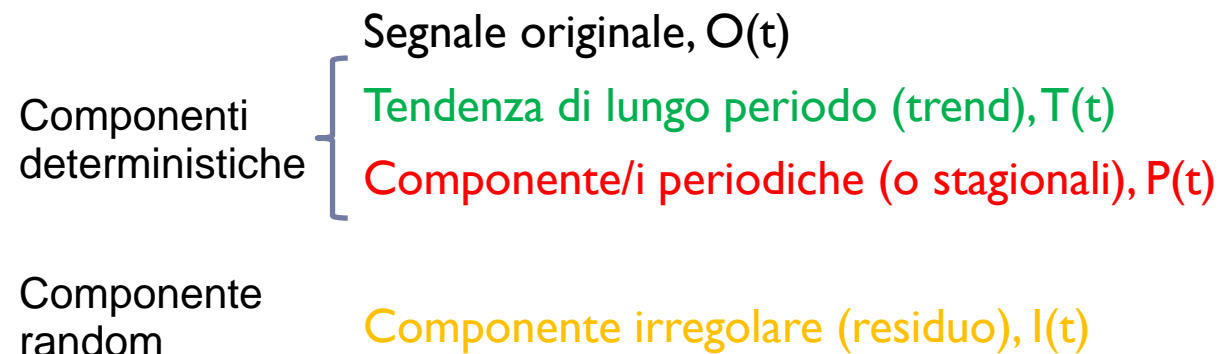
27/4/93-30/9/95



Valori positivi

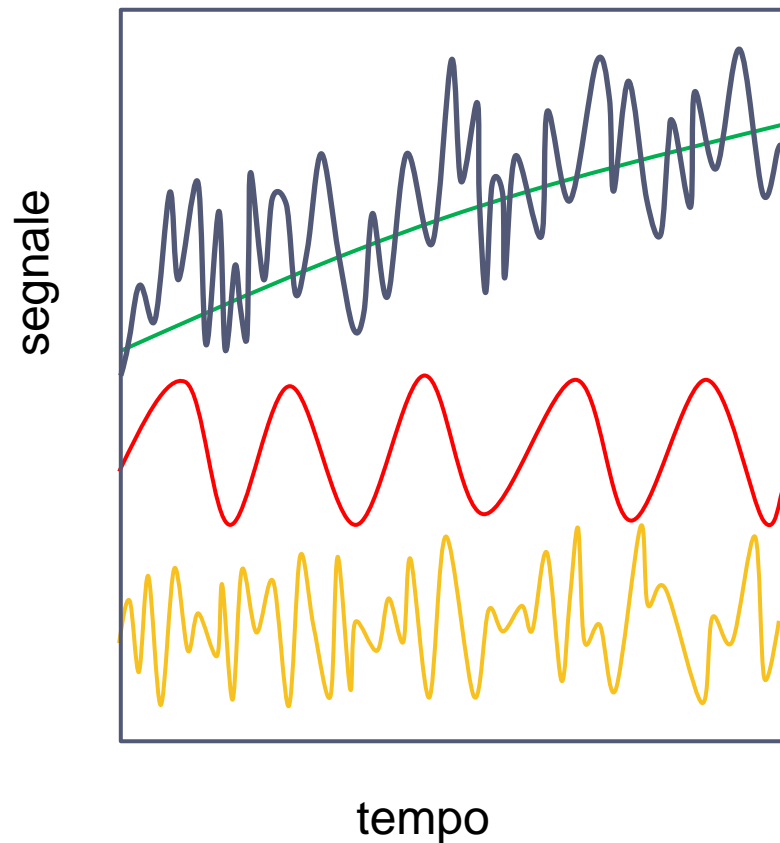
Serie continua con fluttuazioni di varia frequenza

# Elementi della struttura di un segnale



Scomporre un segnale nelle sue componenti deterministiche  $\rightarrow$  sviluppare modelli interpretativi/predittivi

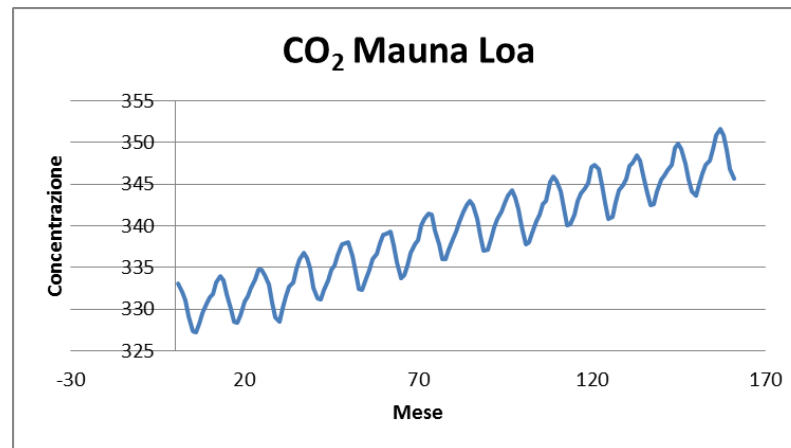
Se residuo= variabile random  $\rightarrow$  decomposizione «ottima»



# Elementi della struttura di un segnale

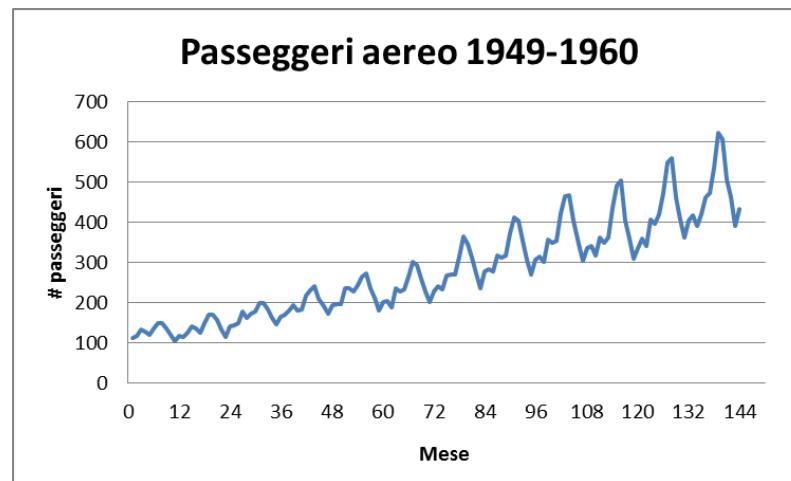
Modello additivo

$$O(t) = T(t) + P(t) + I(t)$$

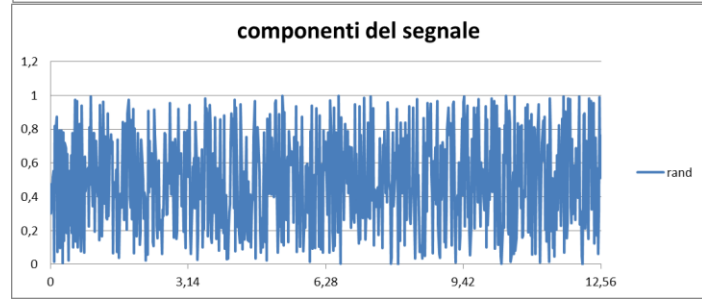
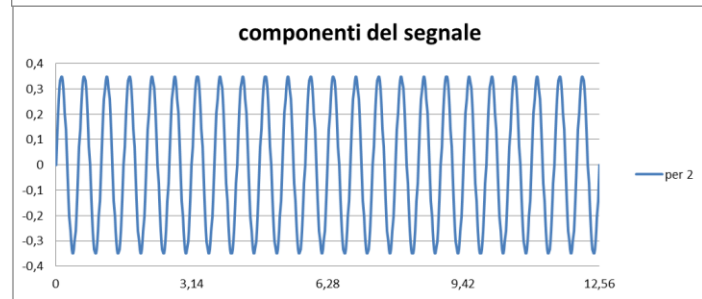
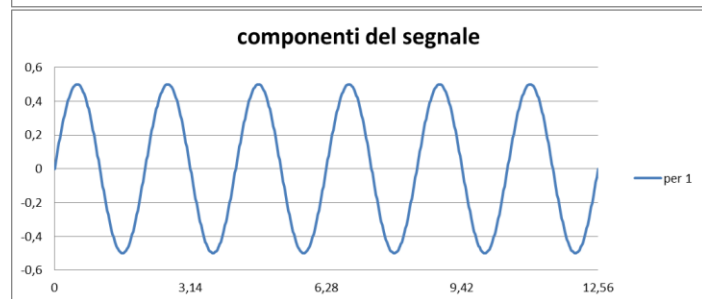
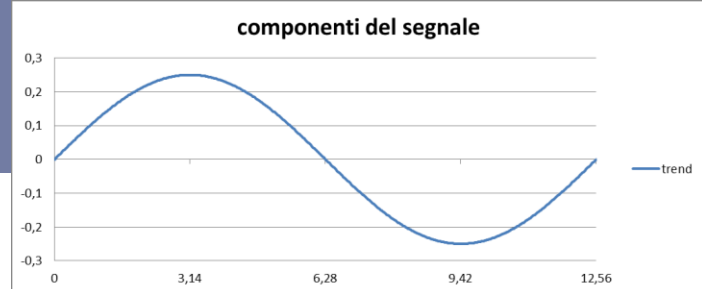
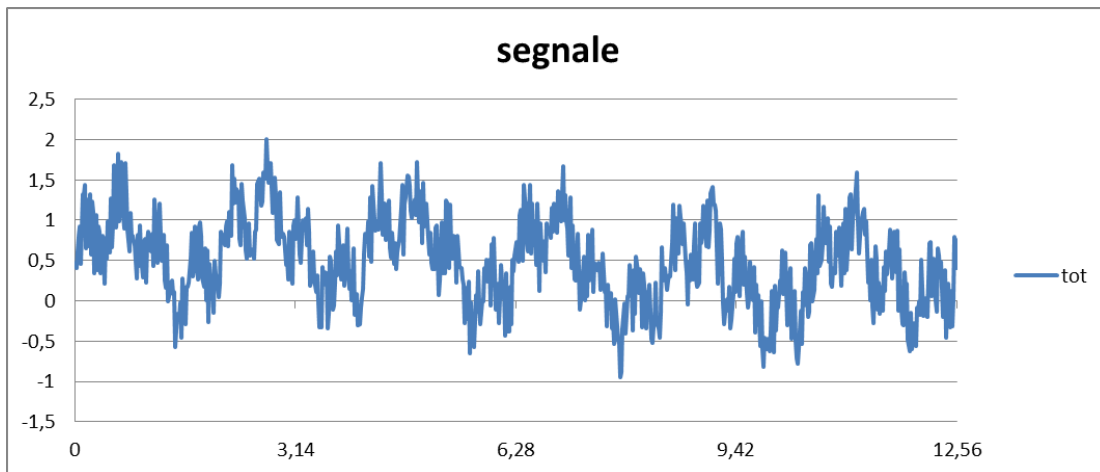


Modello moltiplicativo

$$O(t) = T(t)P(t)I(t)$$



# Elementi della struttura di un segnale



$$y(x) = 0.25 \sin x + 0.5 \sin 3x + 0.35 \sin 12x + \text{rand}()$$

trend

componenti periodiche

rumore

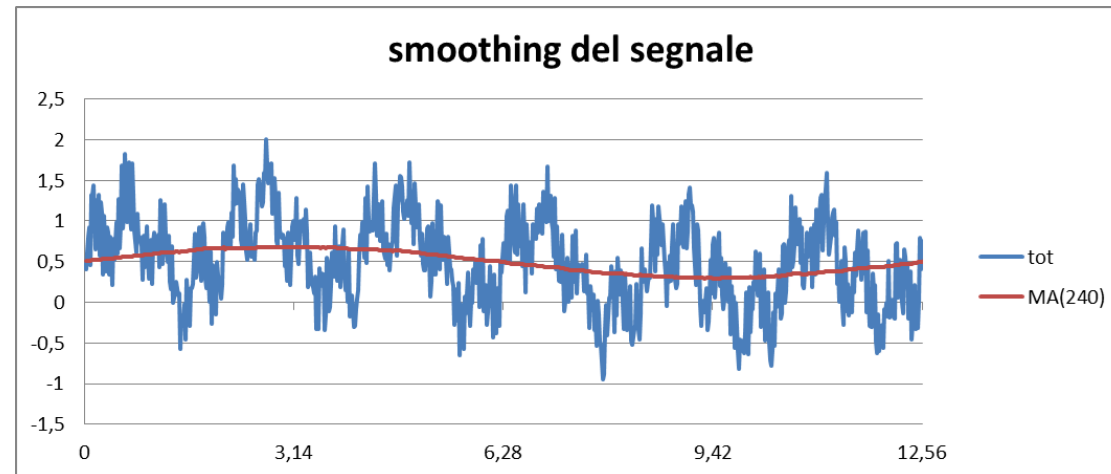
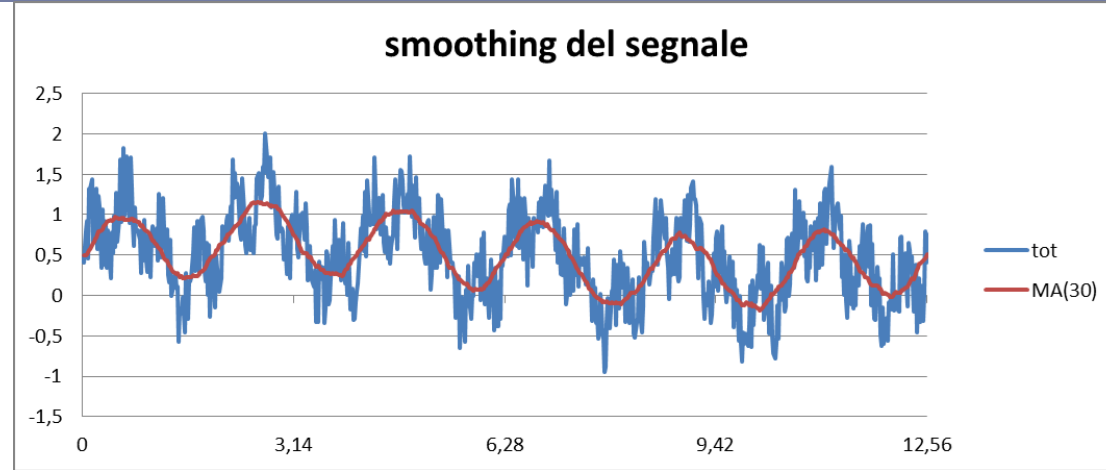
# Estrazione del trend

Variazione di lungo periodo, indipendente dalle variazioni di piccolo periodo (alta frequenza)

► Moving average

Scelta del «periodo» di riferimento determina il «trend» estratto

$$y_f(x, T) = \frac{1}{T} \int_{-T/2}^{+T/2} y(x) dx$$



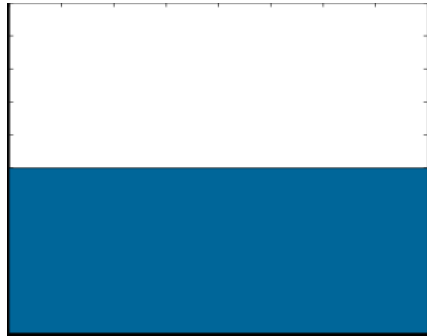


# Estrazione del trend

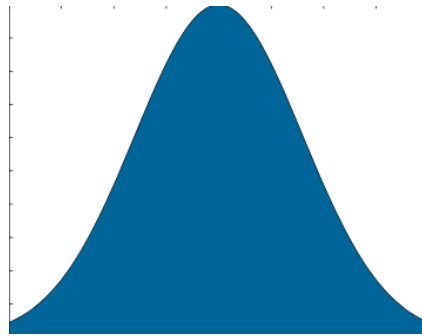
## ► Exponential smoothing

$$y_{exp,smooth}(x, T) = \int_{-T/2}^{+T/2} w(\tau)y(x - \tau)d\tau$$

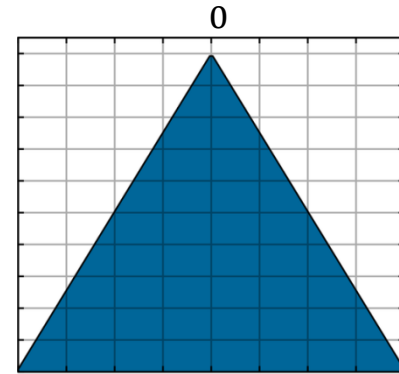
$W(\tau)$



Costante



Gaussiana



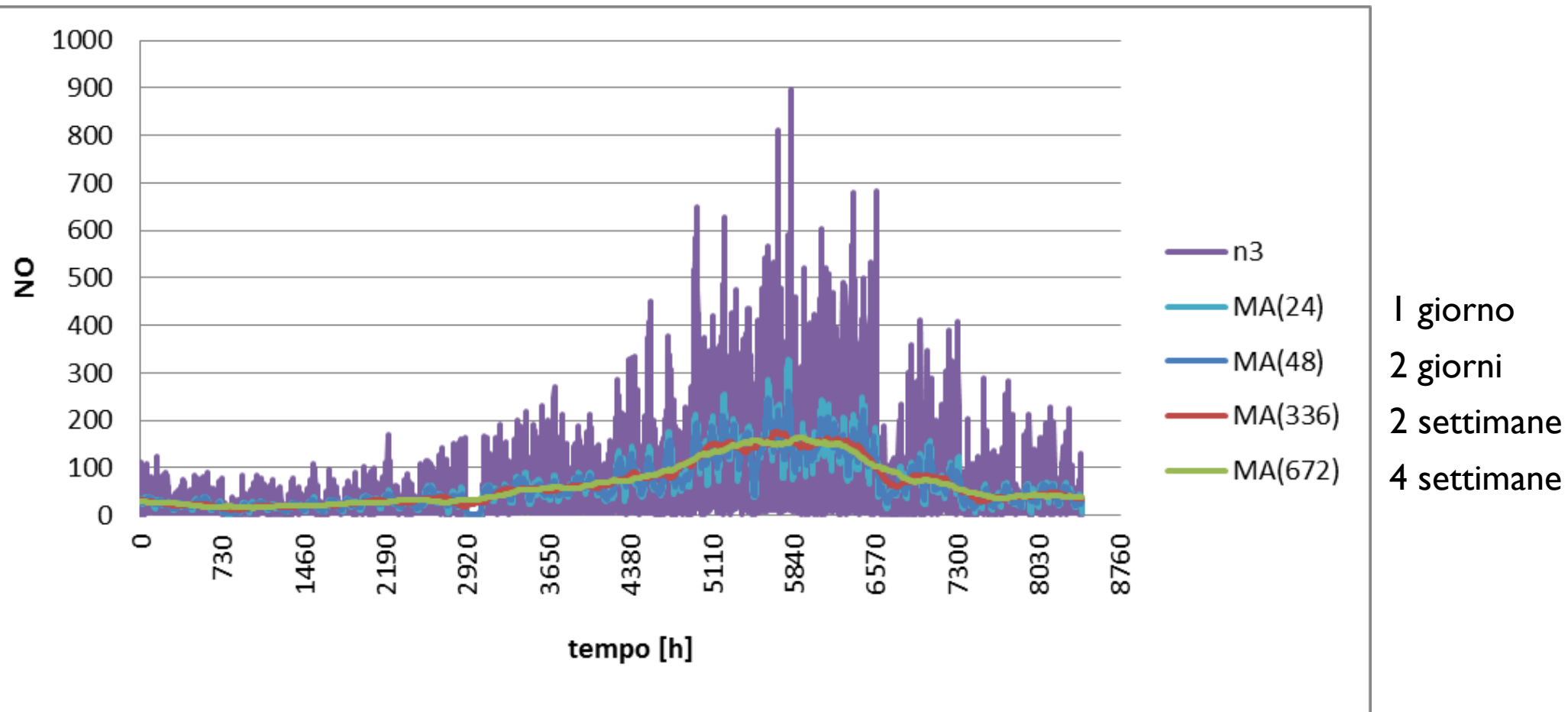
Triangolare

$W(\tau)$  funzione peso (integrale unitario) definita su  $T$

## ► Function fitting

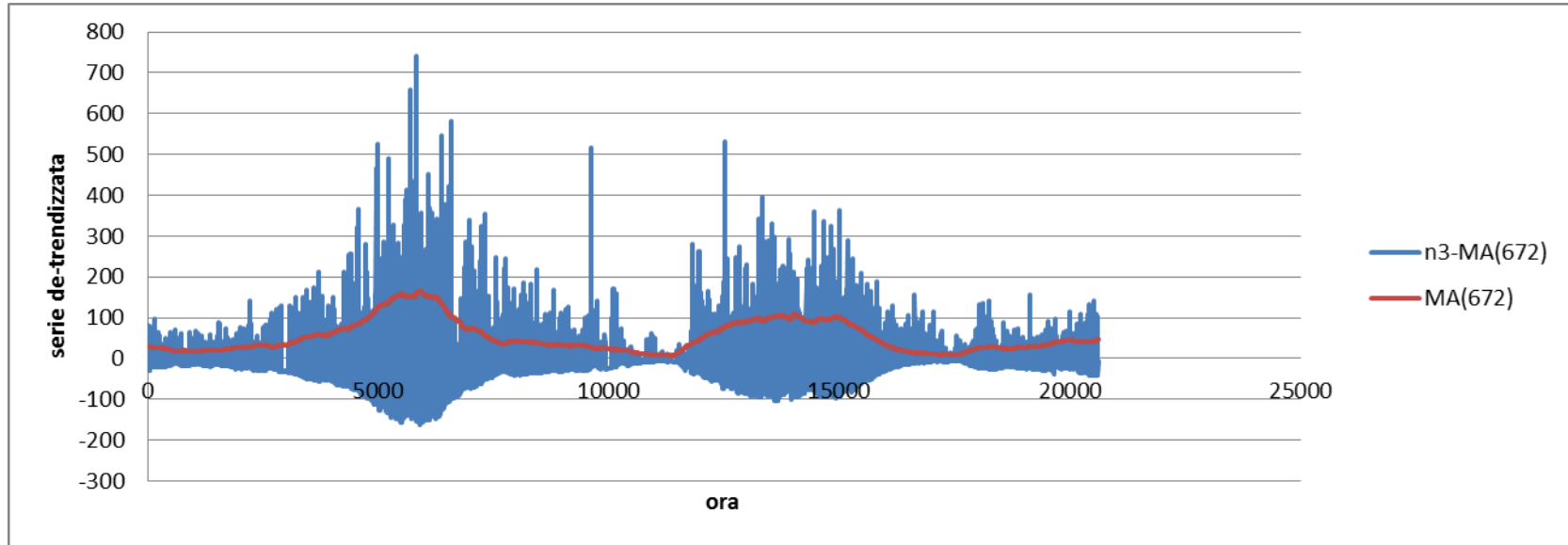
$$y_{fit}(x) = f(x)$$

# Trend concentrazione NO su base annua



# Modello additivo o moltiplicativo?

Analisi della serie de-trendizzata:  $O'(t) = O(t) - T(t)$

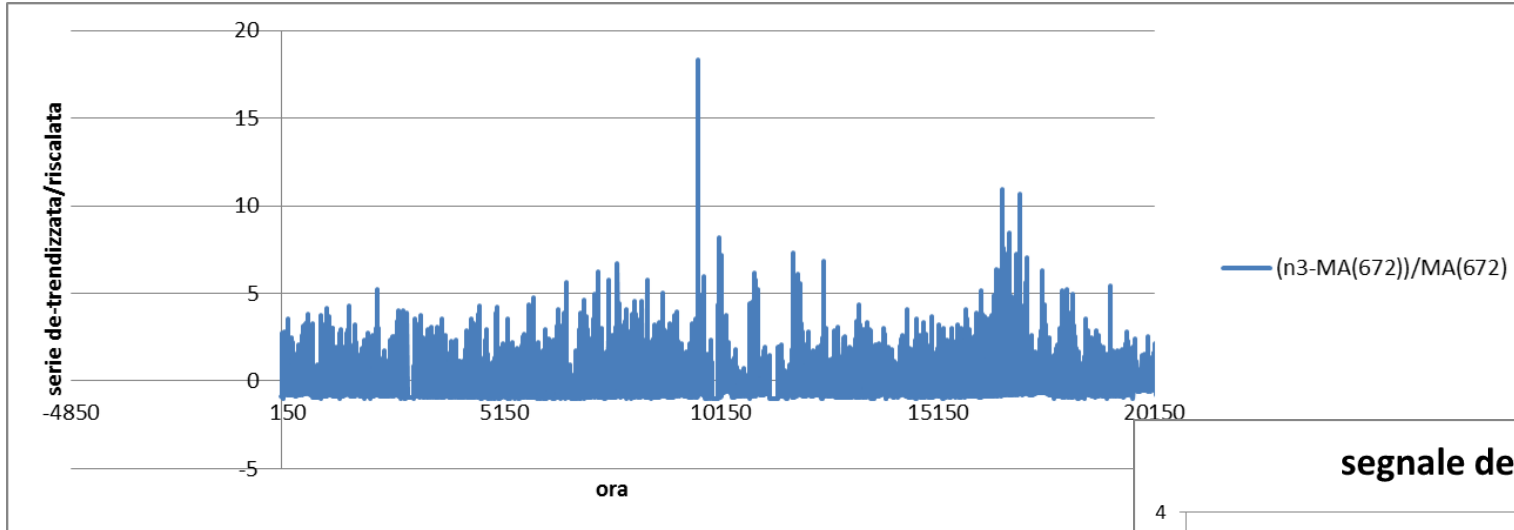


Se l'ampiezza delle fluttuazioni è indipendente dal trend, decomposizione additiva

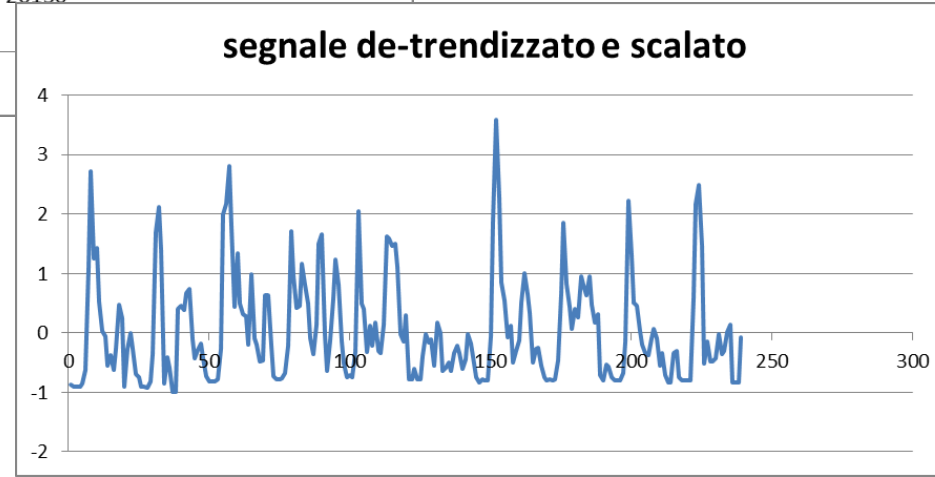
Se l'ampiezza delle fluttuazioni è proporzionale al trend, decomposizione moltiplicativa

# Modello additivo o moltiplicativo?

Serie de-trendizzata e normalizzata rispetto al trend:  $O'(t) = (O(t) - T(t)) / T(t)$



Fluttuazioni periodiche evidenti nella serie de-trendizzata e normalizzata su piccola scala

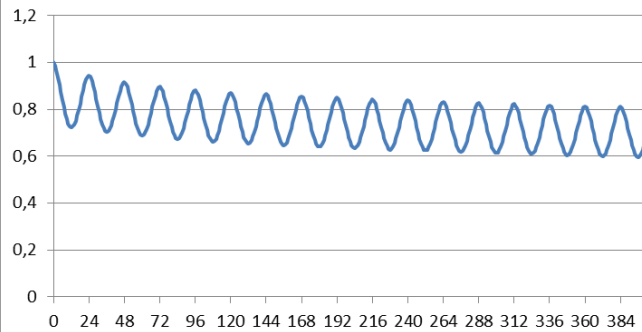


# Identificazione periodo e componente periodica del segnale

Cross-correlazione :  $O'(t)$  vs  $O'(t-\tau)$  per diversi valori del ritardo  $\tau$

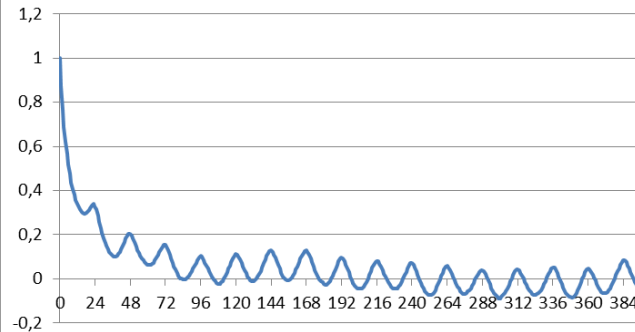
$$\text{correlazione}(x(t), x(t - \tau)) = \sum_{t=0}^{\tau} \frac{(x(t) - \mu_x)}{\sigma_x} \frac{(x(t - \tau) - \mu_x)}{\sigma_x} \in [-1,1]$$

correlogramma T



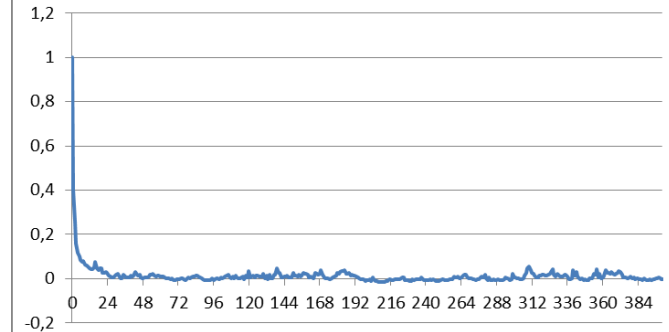
Serie persistente, componente periodica a 24h

correlogramma velocità vento



Serie meno persistente, componente periodica a 24h

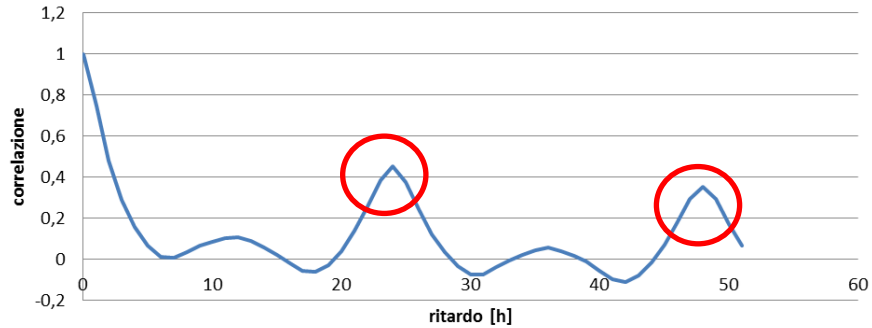
correlogramma precipitazione



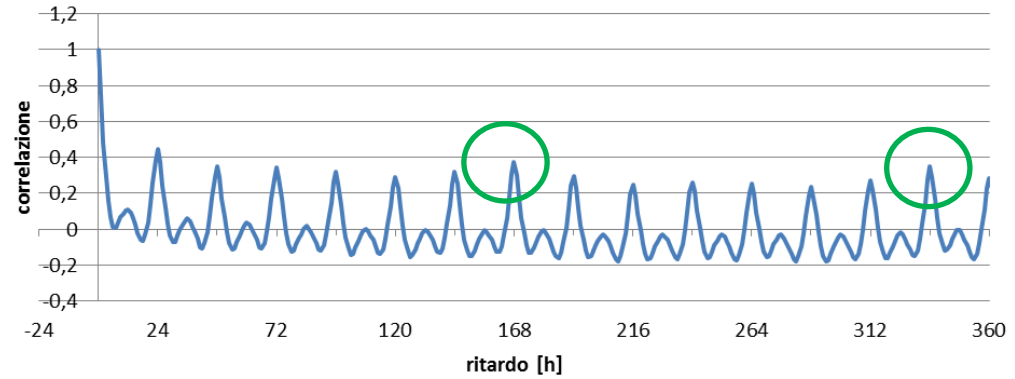
Serie discontinua, irregolare

# Concentrazione NO

correlogramma



correlogramma

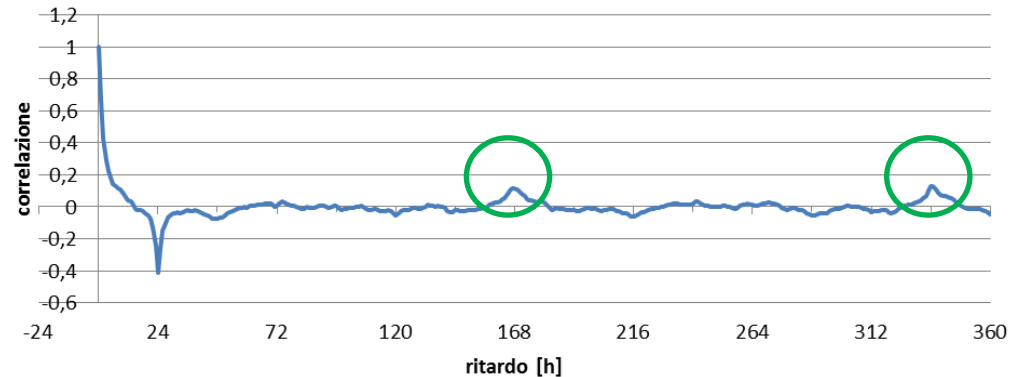


Serie differenziata

$$\Delta^{24}O = o(t) - O(t-24)$$

Rimozione di periodicità note

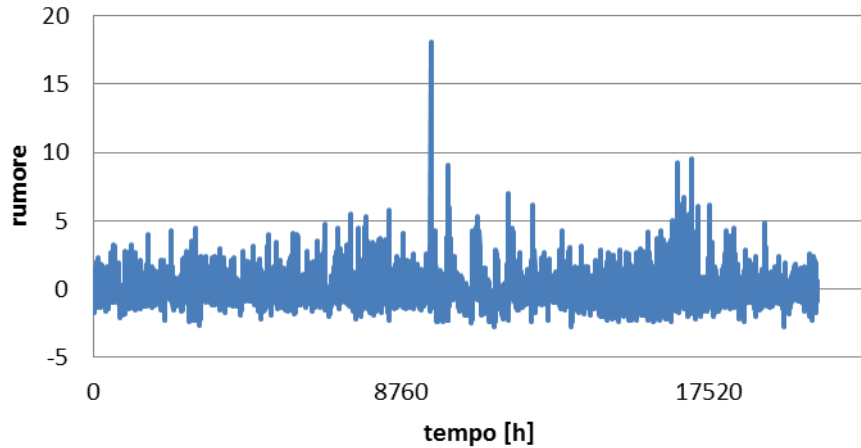
correlogramma serie differenziata (24h)



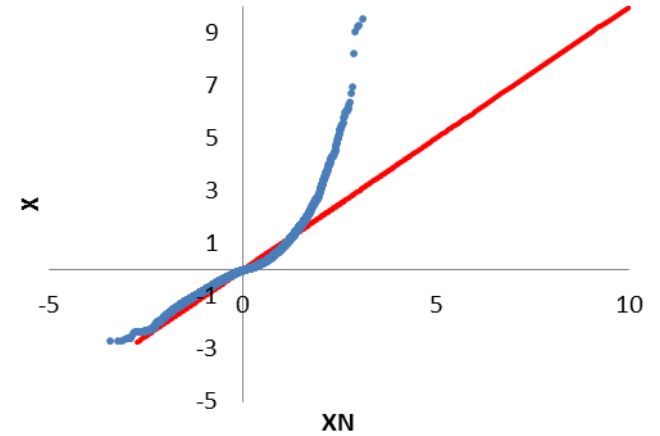


# Componente irregolare

**componente irregolare**



**distribuzione residui**



Distribuzione normale dei residui → solo componente casuale non spiegata

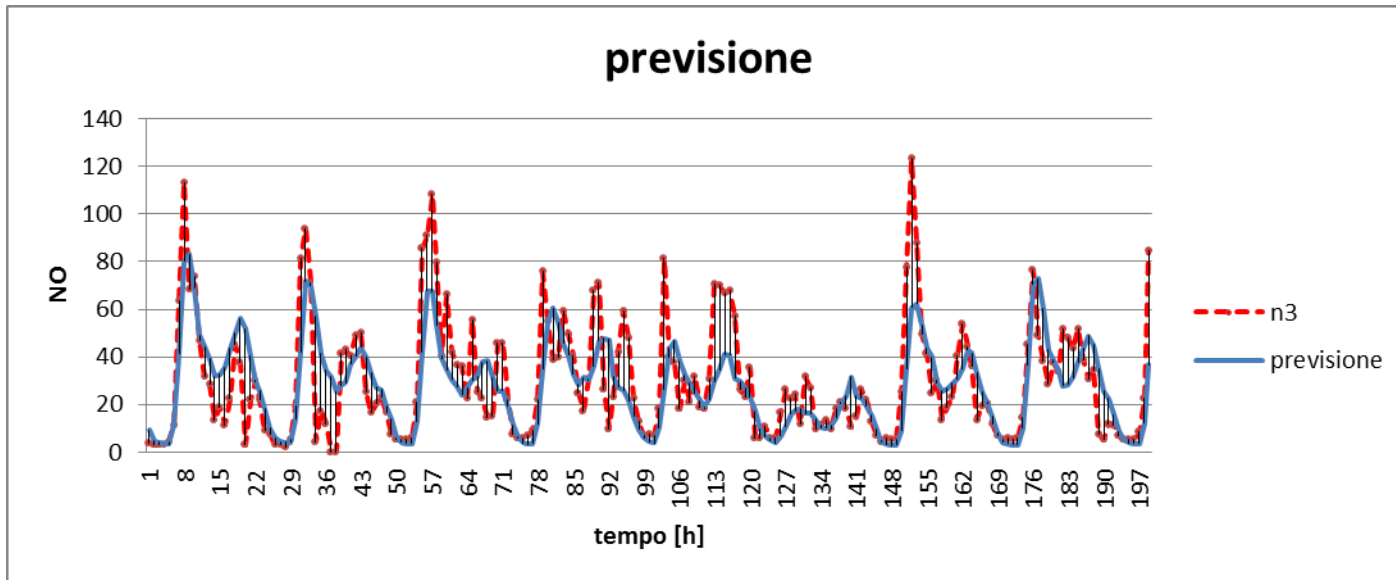
Distribuzione NON normale → esistono componenti «deterministiche» del segnale spiegabili con altri fattori (vento, precipitazione, temperatura?)



# Modello previsionale

Sulla base dell'analisi del segnale (identificazione trend, valutazione della componente periodica) risulta

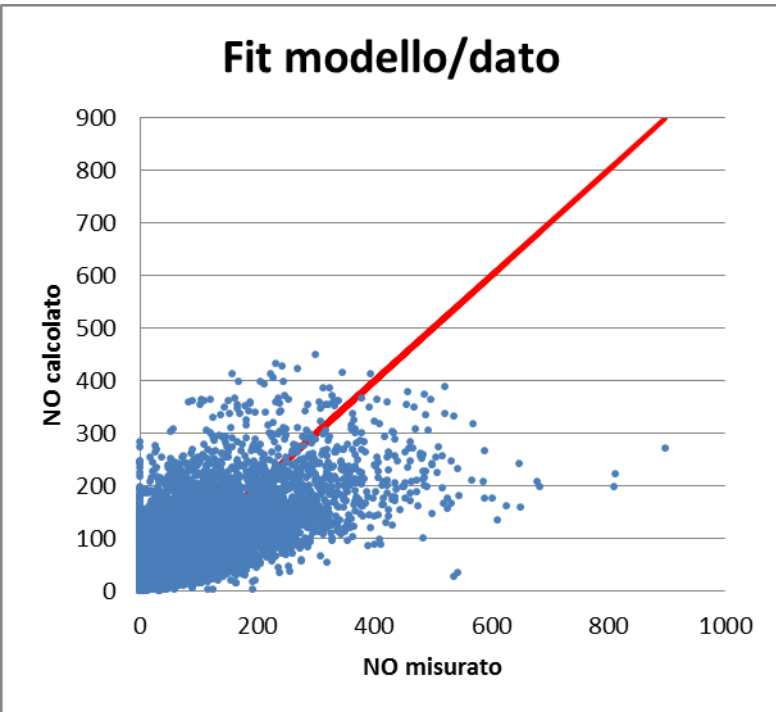
$$M(t) = T(t) + T(t) \cdot P(t)$$



L'errore del modello è

$$\text{Err}(t) = O(t) - M(t)$$

# Modello previsionale: bontà del modello



La bontà del modello è rappresentabile graficamente come scatterplot tra valore previsto  $M(t)$  vs valore osservato,  $O(t)$

Linea rossa= fit perfetto

Dispersione dei punti, mancanza di accordo

E' quantificabile numericamente con diverse misure

Mean Error (ME) 
$$ME = \frac{1}{N} \sum_{k=1}^N O(t) - M(t)$$

Mean Absolute Error (MAE) 
$$MAE = \frac{1}{N} \sum_{k=1}^N |O(t) - M(t)|$$

Mean Square Error (MSE) 
$$MSE = \frac{1}{N} \sum_{k=1}^N [O(t) - M(t)]^2$$

Average Percent Error (APE) 
$$APE = \frac{1}{N} \sum_{k=1}^N \frac{O(t) - M(t)}{O(t)}$$