

Ingegneria Chimica Ambientale

Analisi di serie temporali

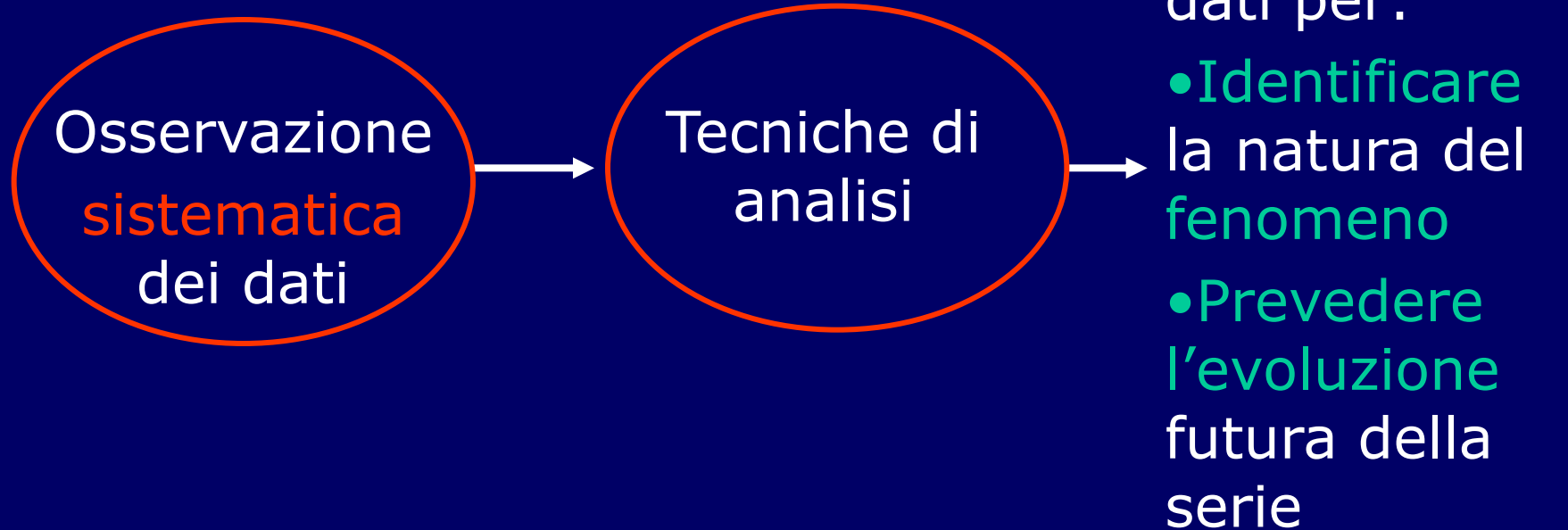
Marina Campolo

Centro Interdipartimentale di Fluidodinamica e Idraulica

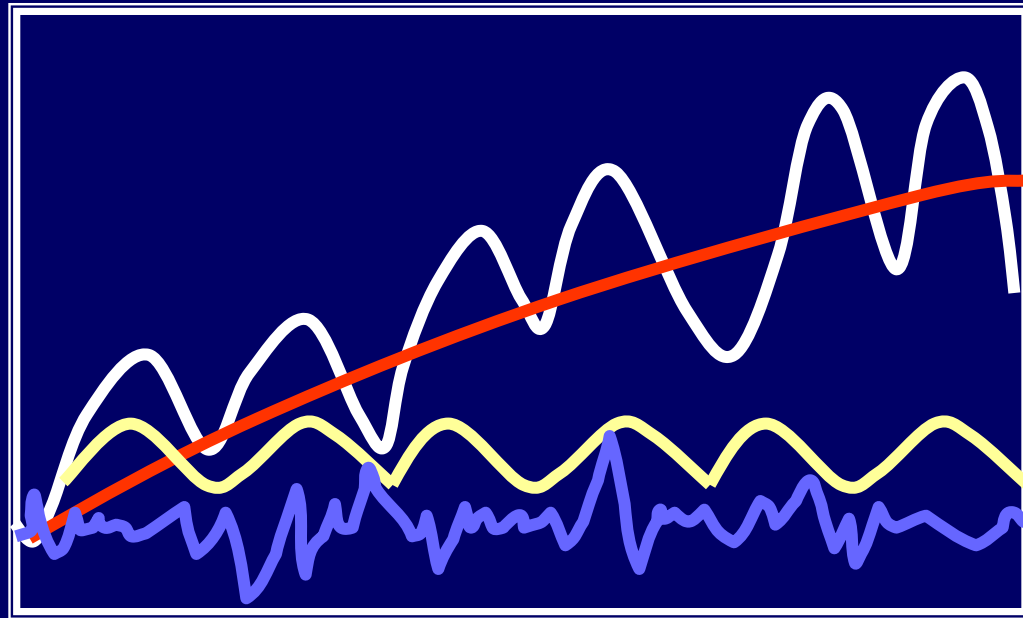
Università di Udine

Obiettivi dell'analisi

Serie temporale: sequenza di misure con ordine prestabilito (intervalli regolari di tempo)



Componenti di una serie temporale



Trend: variazione nel lungo termine

Comp. Stagionale: periodica nel periodo osservato

Comp. Irregolare: residuo random (non deterministico)

OBJ: Separare le componenti "deterministiche" (a lungo e a breve termine) più facilmente modellabili dal residuo

Scomposizione della serie

Modello additivo:

$$O_t = T_t + S_t + I_t$$

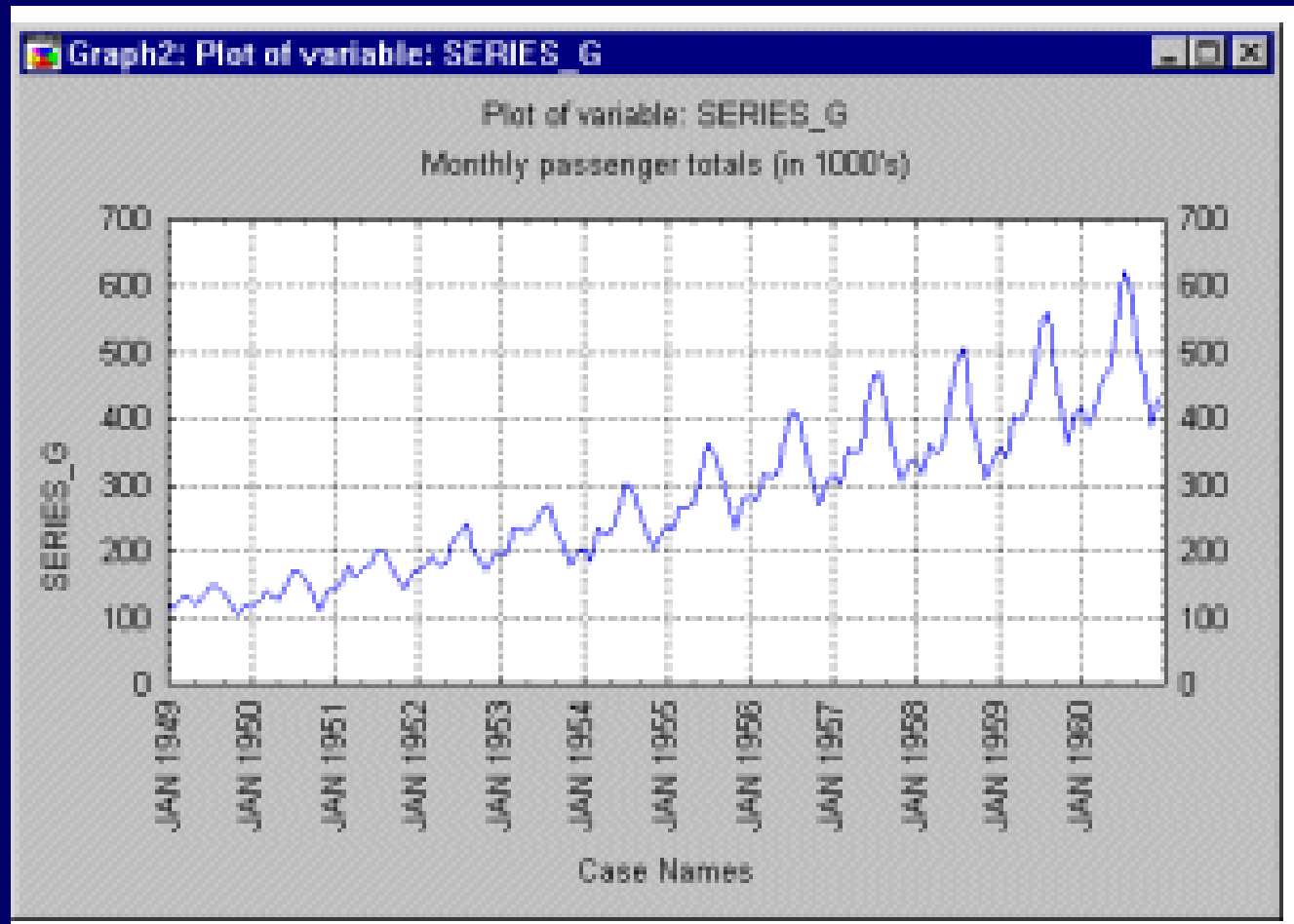
Modello moltiplicativo:

$$O_t = T_t \times S_t \times I_t$$

Modello misto:

$$O_t = T_t + T_t \times (S_t - 1) + T_t \times (I_t - 1)$$

Scomposizione della serie G



Serie del n° passeggeri trasportati in aereo 1949-1960

Che decomposizione usare?

Modello additivo: ampiezza variazioni stagionali e irregolari indipendenti dal trend

Modello moltiplicativo: ampiezza variazioni stagionali e irregolari dipendenti dal trend

Procedura:

1. esaminare un grafico della serie temporale
2. Provare i modelli di decomposizione
3. Scegliere quello che dà componente stagionale più stabile

Individuazione del trend

1. Smoothing: media locale di dati per eliminare le fluttuazioni random (alta frequenza e breve periodo) e periodiche (frequenza più bassa, periodo sottomultiplo del periodo di osservazione)

Media mobile (moving average):

$$O(t_n) = [O(t_{n-m}) + \dots + O(T_{n+m})] / (2m+1)$$

Exponentially weighted smoothing:

$$O(t_n) = [O(t_{n-m}) \exp(-m) + \dots + O(T_{n+m}) \exp(-m)] / (2m+1)$$

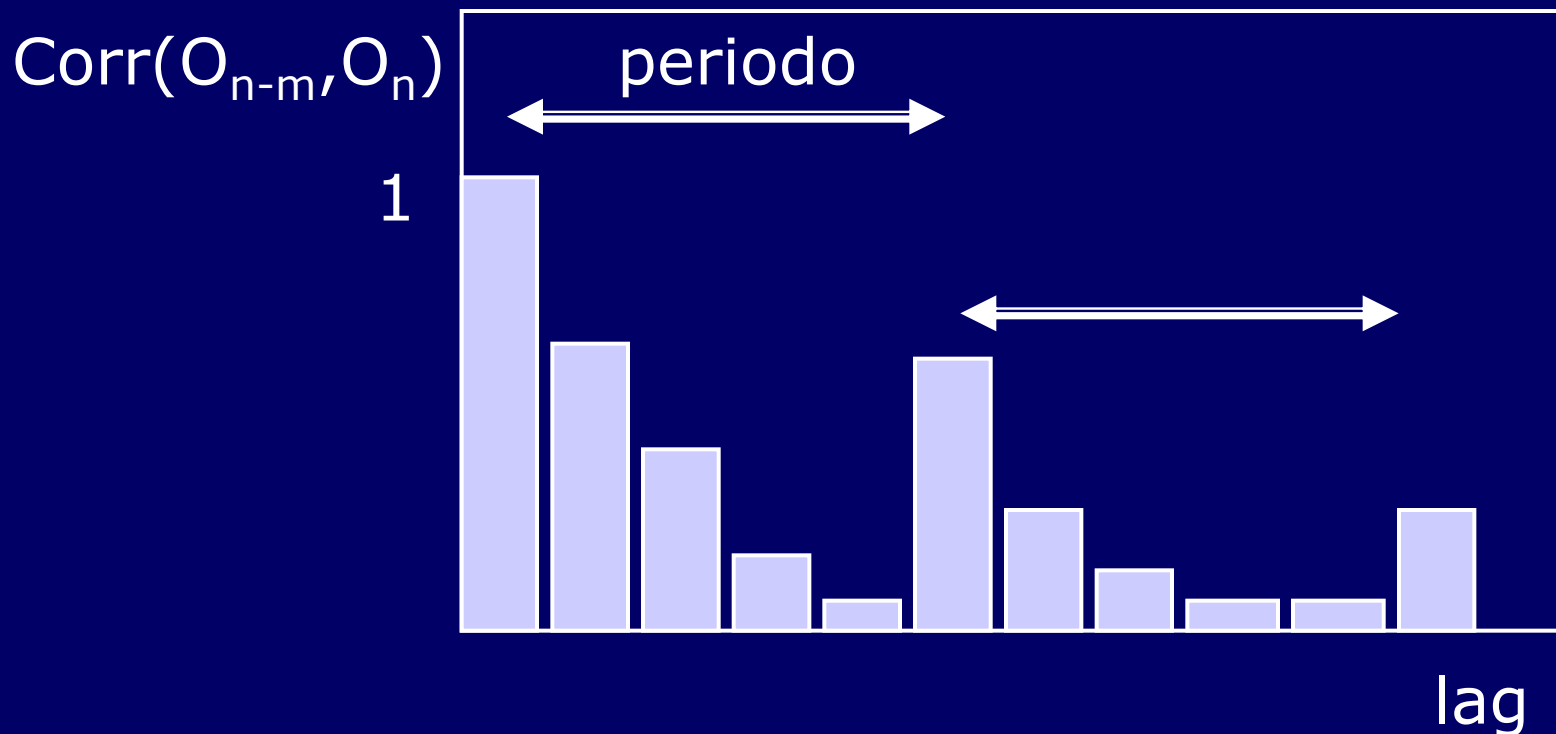
2. Interpolazione (function fitting): con funzioni lineari, logaritmiche, esponenziali o polinomi ...

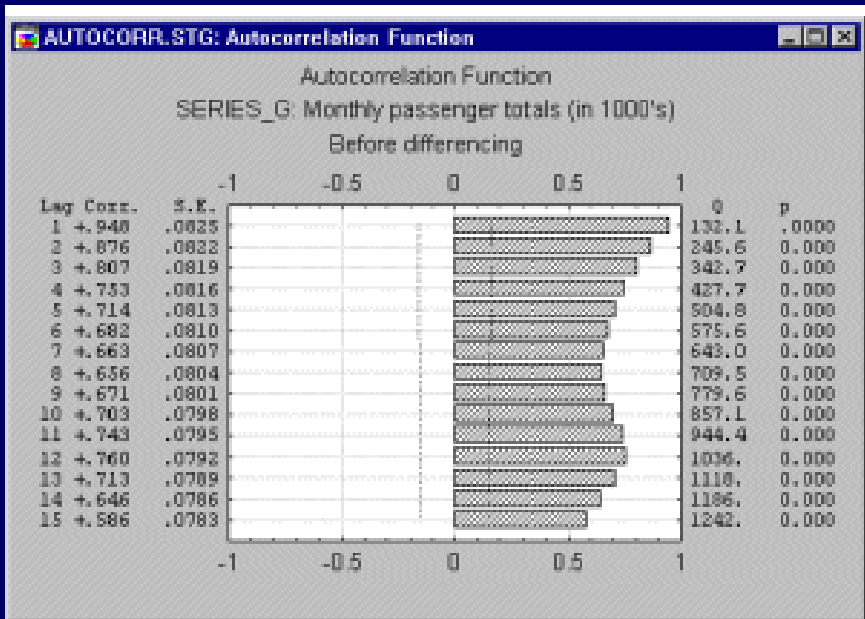
Componente stagionale

1. Analisi di correlazione:

dipendenza tra $O(t_{n-m})$ e $O(t_n)$ -> periodo m (lag)

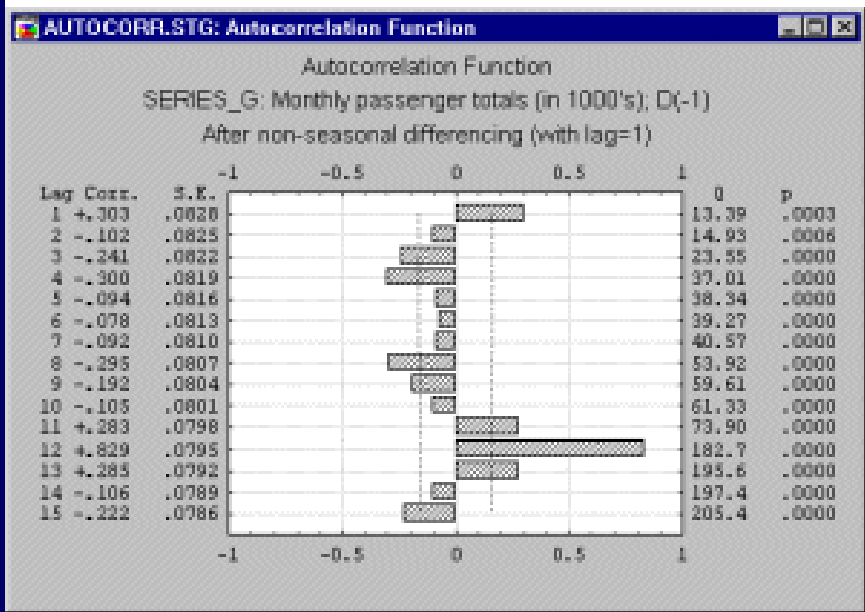
Correlogramma





Correlogramma per la serie G

Serie originale



Serie differenziata

${}^1\Delta$

Componente stagionale

2. Rimozione della componente stagionale (differenziazione della serie)

$$O_t \rightarrow \Delta O_t = O_n - O_{n-m}$$

Vantaggi: serie trasformata è stazionaria, è modellabile con modelli tipo ARIMA (Auto Regressive Integrated Moving Average)

Processi autoregressivi (Box-Jenkins 1976)

AR: Valori della serie prevedibili a partire da valori precedenti (combinazione lineare)

$$x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_n x_{t-n} + \text{err}$$

a_0, a_1, a_2, a_n coefficienti di autoregressione

err = errore random

Vincolo di stazionarietà:

$$x_t = a_0 + a_1 x_{t-1} + \text{err}$$

Limitato $\leftrightarrow -1 < a_1 < 1$

Processi a media mobile (Box-Jenkins 1976)

MA: Valori della serie dipendenti da valori precedenti di errore (combinazione lineare)

$$x_t = \mu + \text{err}_t + b_1 \text{err}_{t-1} + \dots + b_n \text{err}_{t-n}$$

b_1, b_2, \dots, b_n parametri di media mobile

μ = parte costante

Dualità MA – AR per serie stazionarie

Processi ARIMA (Box-Jenkins 1976)

ARIMA(p,d,q)

p = n° parametri di auto-regressione

d = ordine di differenziazione della serie

q = n° parametri di media mobile

ARIMA(0,1,2)

$${}^1\Delta x_t = \mu + \text{err}_t + b_1 \text{err}_{t-1} + b_2 \text{err}_{t-2}$$

${}^1\Delta x_t$ = serie temporale differenziata 1 volta

Processi ARIMA (Box-Jenkins 1976)

1. Controllo stazionarietà serie: media, varianza e autocorrelazione stabili nel tempo

Differenziazione $^d\Delta$

2. Identificazione del modello: valori p, q (<2)

3. Stima dei parametri: valori $a_1, \dots, a_p, b_1, \dots, b_q$
(minimizzazione dei residui)

4. Previsione della serie differenziata

5. Ricostruzione della serie non differenziata

6. Valutazione delle prestazioni del modello

Processi SARIMA (Box-Jenkins 1976)

Seasonal Auto Regressive Integral Moving Average

1. Identificazione lag stagionale e trasformazione della serie $s\Delta$
2. Identificazione lag per avere stazionarietà $d\Delta$
3. Stima dei parametri
4. Previsione della serie differenziata $d\Delta$
5. Ricostruzione della serie de-stagionalizzata
6. Ricostruzione della serie originale
7. Valutazione prestazioni del modello

Simple exponential smoothing

Semplice modello per previsione a 1 passo in avanti

$$x_t = b + \text{err}_t$$

b = costante, cambia lentamente nel tempo

err_t = errore

Stima di b attraverso media mobile:

$$S_t = \alpha x_t + (1 - \alpha) S_{t-1} \quad 0 < \alpha < 1$$

S_t = smoothed value

Valore di α da minimizzazione dell'errore (MSE) sulla previsione 1 passo in avanti

Misure statistiche di errore

Errore medio: $ME = E(x_t - S_t)$

Errore medio assoluto: $MAE = E(|x_t - S_t|)$

Somma degli errori quadrati (SSE):
 $SSE = E(|x_t - S_t|^2)$

Errore percentuale medio: $APE = E((x_t - S_t)/x_t)$

Esercizi

Elaborazione statistica di dati di inquinamento

dati di inquinamento del comune di Udine (4 Centraline)
1 P.le Osoppo; 2 P.le XXVI Luglio; 3 via Manzoni; 5 via Cairoli.
il numero nel nome della variabile indica la centralina

Le variabili misurate sono:

Evar giorno dal 1/1/1960

Doy day of year

Lst local standard time

year,month,day, anno mese giorno

Dow day of week

P precipitazione

Ta temperatura dell'aria

H umidità relativa

Wd wind direction

Ws10m wind speed at 10 m height

Pa atmospheric pressure
Rs solar radiation
Sel solar elevation angle
S1 Sulphur dioxide
C1 carbon monoxide
N1 nitrogen monoxide
Nb1 nitrogen dioxide
Nx1 nitrogen oxides
O3 ozono
Hcnm idrocarburi non metanici
Ch4 metano

rssun ore di sole

dati di inquinamento del comune di Udine (4 Centraline)

1 P.le Osoppo

2 P.le XXVI Luglio

3 via Manzoni

5 via Cairoli

il numero nel nome della variabile indica la centralina

Procedura

- Fare grafico della serie temporale
- Individuare il trend della serie (moving average con finestre di diversa ampiezza)
- Rimuovere il trend
- Individuare la componente stagionale (correlogramma)
- Rimuovere la componente stagionale per individuare il residuo
- Calcolare contributo del trend e della parte stagionale sul segnale totale
- Provare a fittare un modello AR per la previsione della serie dei residui