

Corso di Dinamica e Modellistica degli Inquinanti – Anno 2018

Esercitazione n.1: rappresentazione e analisi di dati ambientali

Obiettivo dell'esercitazione

Utilizzare tecniche di rappresentazione grafica e tecniche per l'analisi dei dati utili per acquisire informazioni sul processo descritto dai dati.

Vengono fornite serie di dati ambientali che descrivono diversi fenomeni (velocità del vento, mm di pioggia caduta, temperatura, concentrazione di inquinanti monitorati in ambiente urbano).

Si chiede di elaborare graficamente/statisticamente i dati per estrarre conoscenza riguardo ai fenomeni da cui le misure hanno avuto origine.

Esecuzione

1. Sintesi di dati ambientali: costruzione di pdf, cpdf, q-plot, probability plot e boxplot

Analizzare come serie univariate i dati di inquinamento e i dati meteo per valutare lo stato di qualità dell'aria in ambiente urbano (files 199xINQ.xlsx): utilizzando la funzione FREQUENZA di excel, costruire la funzione distribuzione di probabilità (pdf) e la funzione cumulata della distribuzione di probabilità (cpdf) per la concentrazione di un inquinante e per la temperatura. Quali variazioni di forma si osservano al variare del numero delle classi utilizzate per "aggregare" i dati? Quali differenze di forma ci sono tra le distribuzioni delle due tipologie di dati? Utilizzando le funzioni di Excel per il calcolo dei percentili della distribuzione, costruire i boxplot. Utilizzando la funzione di ordinamento dei dati e la probabilità incrementale data dalla Cunnane position costruire i q-plot. Utilizzando la funzione di Excel inversa della normale, ricavare i quantili di una distribuzione avente stessa media e varianza dei dati e distribuzione normale. Confrontare la distribuzione dei dati di temperatura e inquinamento con quelli di una distribuzione Normale. Se la distribuzione è diversa dalla normale, è possibile identificare una trasformazione dei dati che avvicini la distribuzione dei dati trasformati a quella normale?

2. Analizzare la variabilità della concentrazione di SO_2 su base mensile: costruire i boxplot per confrontare valore medio e variabilità del dato su base mensile (Suggerimento: definire il valore del mese come il criterio per la selezione del dato e utilizzare la funzione filtro avanzato per costruire i sottogruppi di dati mensili da elaborare statisticamente).

3. Costruzione/comparazione di distribuzioni di probabilità doppia

Utilizzare i dati di intensità e direzione di provenienza del vento misurato dalla centralina posta in ambiente urbano per costruire la distribuzione di probabilità doppia (Suggerimento per calcolo con Excel: 1. definire i settori di provenienza del vento limitando inferiormente e superiormente il valore della variabile direzione; 2. utilizzare la funzione filtro avanzato per selezionare i dati secondo il settore di provenienza del vento e valutare la probabilità marginale di avere vento da quel settore; 3. costruire per ogni sottoinsieme di dati la distribuzione di probabilità in funzione dell'intensità del vento – probabilità condizionata – e riscalarla in funzione della probabilità marginale di avere vento da quella direzione per ottenere la distribuzione doppia; 4. costruire la cumulata della distribuzione di probabilità doppia per ogni settore di provenienza; 5. plottare questi valori come polar plot.) Verificare il risultato contro quanto prodotto con WRPLOT (<https://www.weblakes.com/products/wrplot/index.html>).

Utilizzare WRPLOT per confrontare i dati di vento misurati nelle stazioni anemometriche di Fagagna e Gemona (file fagagna.xls e gemona.xls scaricabili dal sito <http://158.110.32.35/CLASS/dinamica-inquinanti.html>) per valutare la variabilità locale del vento. L'analisi può essere condotta confrontando le rose dei venti; confrontando le pdf dell'intensità del vento relative alle 2 centraline; calcolando i parametri statistici (robusti e non) per i due set di dati; costruendo i box-plot; confrontando i q-plot.

4. Analisi di dati multivariati

Analizzare i dati di concentrazione di inquinante (es. SO_2 o CO_2) simultaneamente monitorato in corrispondenza di 4 centraline urbane per identificare se è possibile ridurre il numero di centraline di monitoraggio: costruire gli scatterplot e valutare la matrice di correlazione tra le variabili (utilizzare la funzione CORRELAZIONE di Excel); esistono relazioni di chiara

dipendenza funzionale tra le concentrazioni monitorate alla stazione a e b? È possibile sviluppare un modello multivariato ($c_a = f(c_b)$) per “stimare” la concentrazione misurata nella stazione a in funzione di quella misurata nella stazione b? (estrarre la linea di tendenza dallo scatterplot dei dati). Con riferimento ai dati di inquinamento di una singola stazione, costruire scatterplot e matrice di correlazione per valutare se (i) la concentrazione di qualche inquinante può essere considerata “dipendente” dalla concentrazione di un altro inquinante monitorato; (ii) la concentrazione di qualche inquinante mostra una variazione sistematica rispetto a variabili meteorologiche monitorate (temperatura, vento, pioggia). Se è possibile identificare una dipendenza funzionale tra coppie di variabili del set multivariato, valutare la bontà di un modello predittivo utilizzando diverse misure di errore. Quale misura di errore potrebbe essere più adatta per valutare la bontà del modello?

5. Analisi di serie temporali

Utilizzare i dati di pioggia e di temperatura del file di inquinamento urbano (timeseries93-95.xlsx) per valutare la “differenza” nella dinamica di variazione delle due variabili meteorologiche (valutare l’andamento del grafico di autocorrelazione).

6. Analisi di serie temporali

analizzare l’evoluzione dello stato di qualità dell’aria in ambiente urbano (file timeseries93-95.xlsx): (i) utilizzare la media mobile per identificare il trend dei dati; che effetto ha l’ampiezza della finestra di filtraggio sul segnale risultante?; (ii) detrarre il trend ed analizzare la serie temporale de-trendizzata alla ricerca di componenti periodiche: costruire il correlogramma per la serie de-trendizzata e per la serie differenziata; (iii) valutare se la componente residua del segnale (segnale originario meno trend meno componente periodica) ha una distribuzione casuale (normale) dell’errore; (iv) provare a estrarre ulteriori componenti deterministiche del segnale (variabilità associata al giorno della settimana e all’ora del giorno) (Suggerimento: utilizzare le opzioni di filtro avanzato per selezionare il sottoinsieme di dati associato al giorno della settimana o all’ora del giorno e trovarne la media; identificare una funzione interpolante che dia rappresentazione analitica alla variazione delle medie così ottenuta).

7. Analisi di serie temporali

Utilizzare i dati di variazione di livello misurati per due sezioni lungo il corso di un fiume (file arno0.xls) per sviluppare un modello “predittivo” per l’evoluzione della portata: analizzare la cross-correlazione tra i dati, individuando il ritardo di propagazione del segnale; sviluppare un modello auto-regressivo per la previsione della variazione di livello; valutare l’errore del modello predittivo (vedi presentazione sull’analisi di serie temporali).

Tutte le analisi proposte possono essere realizzate utilizzando funzioni standard di Excel. è disponibile in versione free-trial (30 giorni) un programma aggiuntivo che semplifica le analisi su Excel (<https://www.xlstat.com/en/>).