

# Corso di Dinamica e Modellistica degli Inquinanti – Anno 2014

## 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 che i dati descrivono.

Vengono fornite diverse serie ambientali (dati di vento, dati di pioggia, dati di temperatura, variazione di livello lungo il corso di un fiume, valori di concentrazione di inquinanti monitorati in ambiente urbano) che descrivono diversi fenomeni ambientali.

Si chiede di verificare quale/quanta conoscenza sia possibile estrarre dai dati in termini di migliore comprensione del fenomeno.

### Esecuzione

1. Costruzione di pdf, cpdf, q-plot, probability plot e boxplot  
Analizzare la variabilità delle variabili meteo e di inquinamento tipicamente monitorate per valutare lo stato di qualità dell'aria in ambiente urbano (files 199xINQ.xlsx): costruire la funzione distribuzione di probabilità (pdf) e la funzione cumulata della distribuzione di probabilità (cpdf) per la temperatura e per la concentrazione di un inquinante. 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? Costruire i boxplot e i q-plot, 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. Costruzione/comparazione di distribuzioni di probabilità doppia  
Utilizzare i dati di vento misurati in tre centraline vicine (file dativento2007.xls) per valutare la variabilità locale dello stesso. L'analisi può essere condotta confrontando le rose dei venti; confrontando le pdf dell'intensità del vento relative alle 3 centraline; calcolando i parametri statistici (robusti e non) per i tre set di dati; costruendo i box-plot; confrontando i q-plot.
3. 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; 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? 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?
4. Analisi di serie temporali  
Utilizzare i dati di pioggia e di temperatura (dativento2007.xls) per valutare la "differenza" nella dinamica di variazione delle due variabili meteorologiche. Considerando le due serie come serie temporali, calcolare l'autocorrelazione dei segnali.
5. Analisi di serie temporali  
analizzare l'evoluzione dello stato di qualità dell'aria in ambiente urbano (files 199xINQ.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 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.

6. 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).