

Ingegneria Chimica Ambientale

---

# Rappresentazione grafica dei dati

Marina Campolo

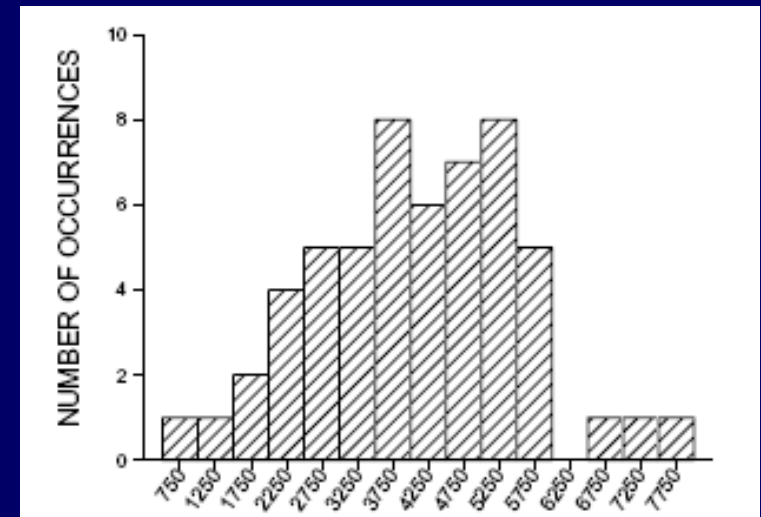
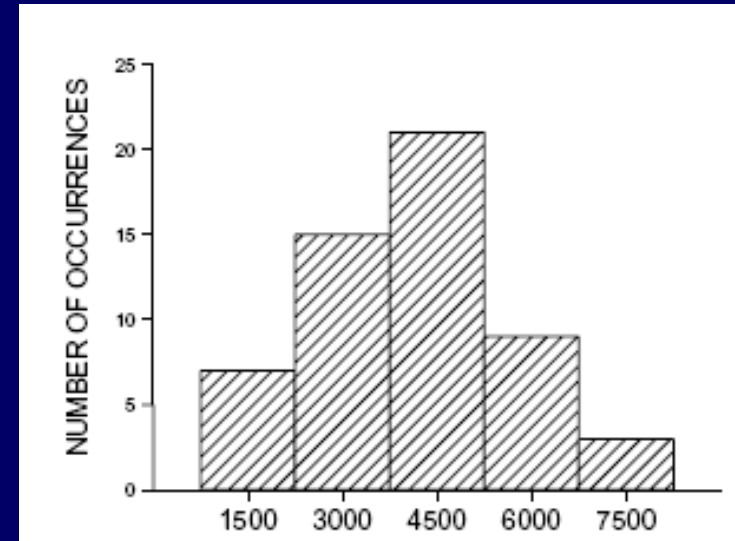
Centro Interdipartimentale di Fluidodinamica e Idraulica

Università di Udine

---

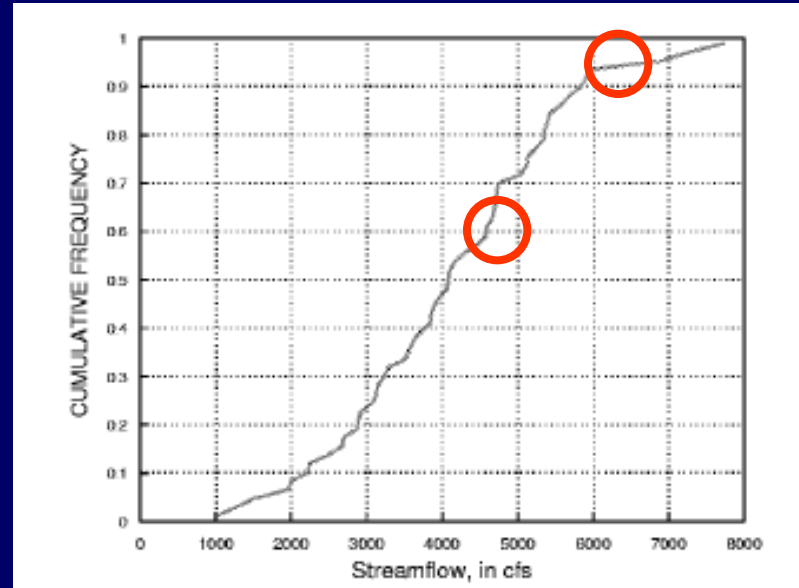
# Istogrammi

Scelta delle classi,  
condiziona la  
rappresentazione



# Quantile plot: cpdf

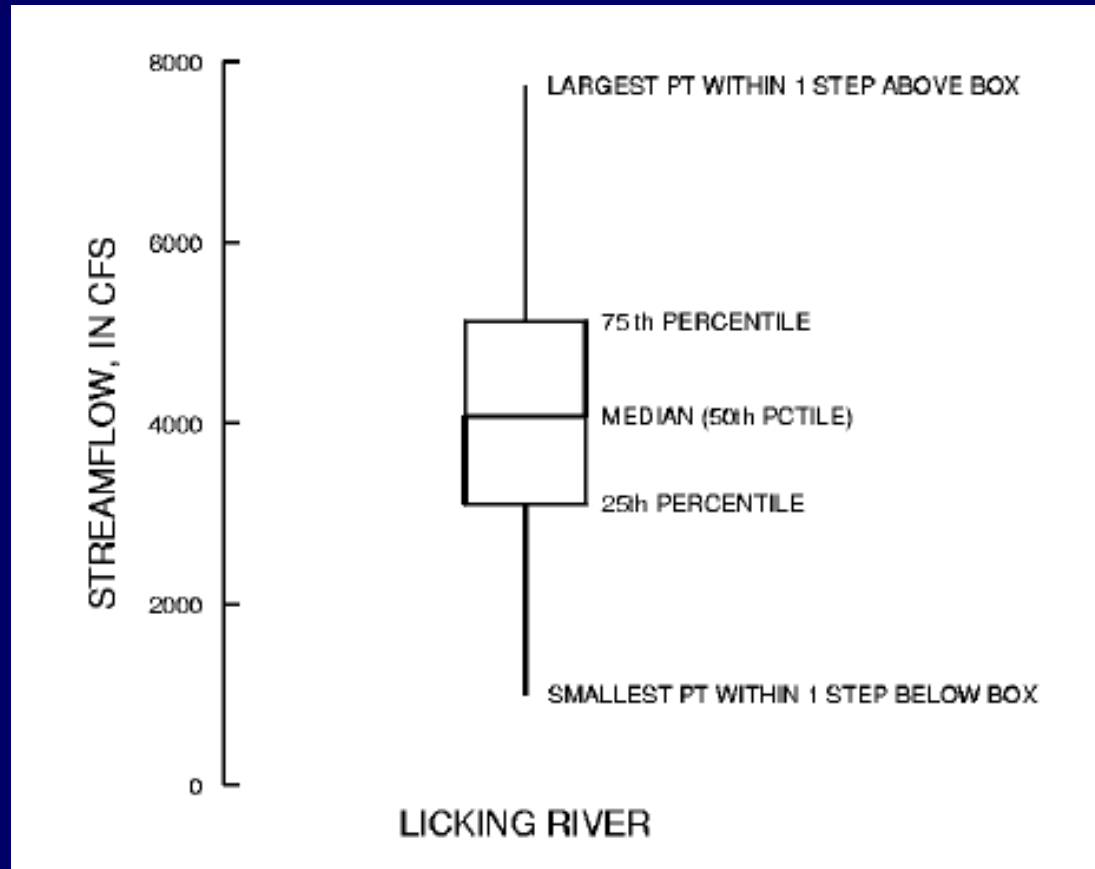
1. Non servono classi arbitrarie
2. Tutti i dati sono rappresentati
3. Ogni punto ha posizione specifica



Cunnane plotting  
position

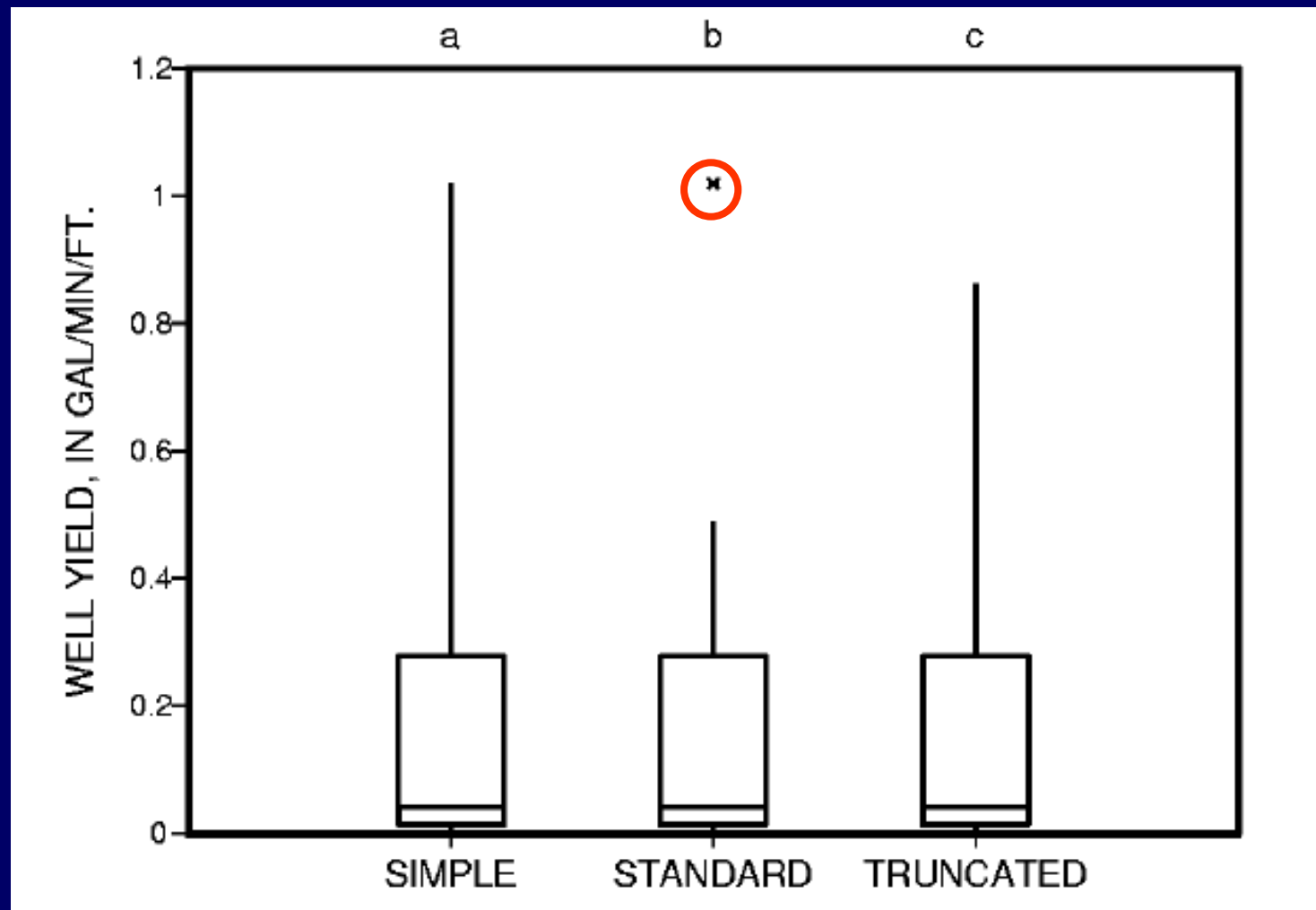
$$p_i = (i-0.4)/(n+0.2)$$

# Box plots

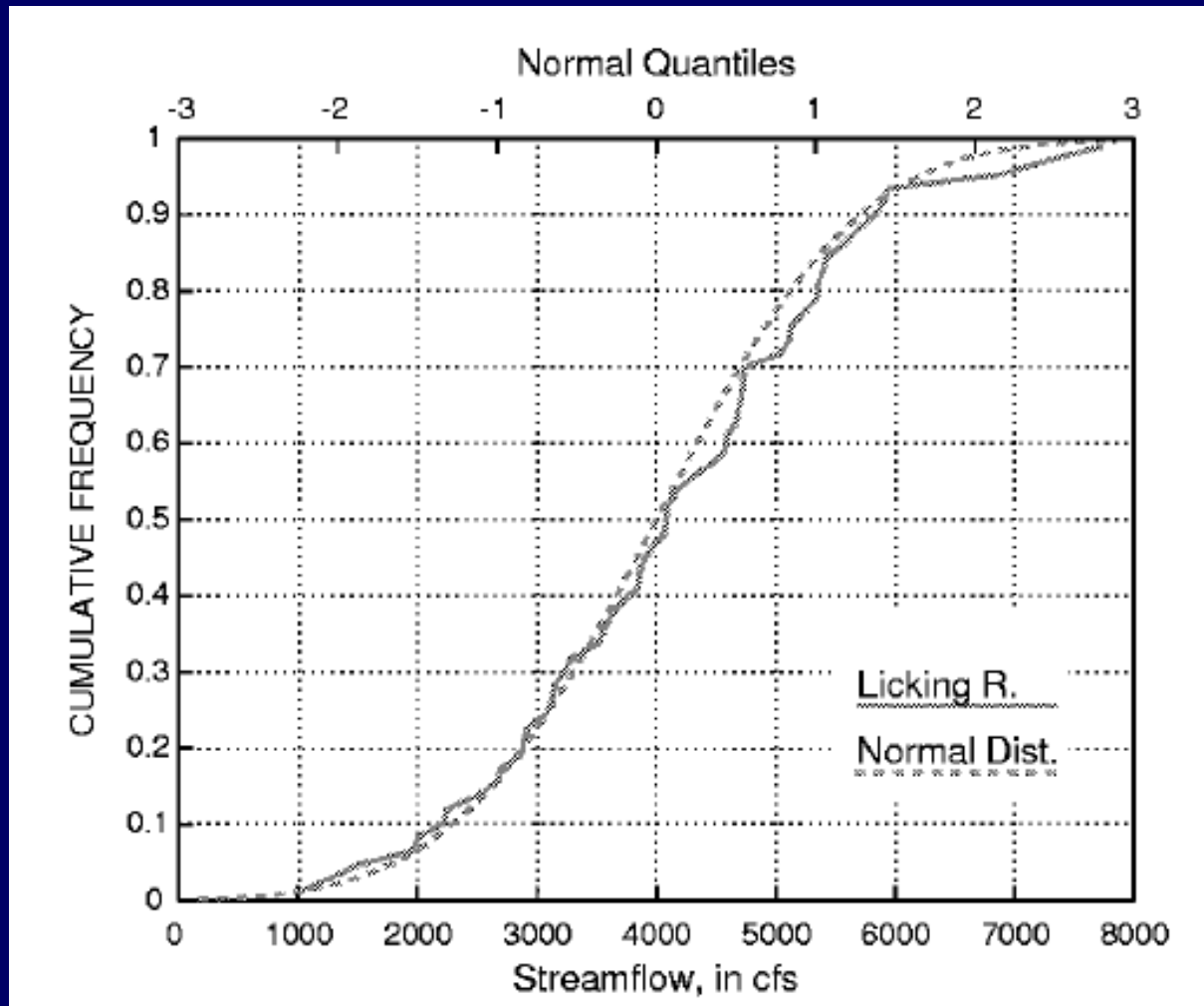


Utali per confronto di distribuzioni

# Tipi di Box plots

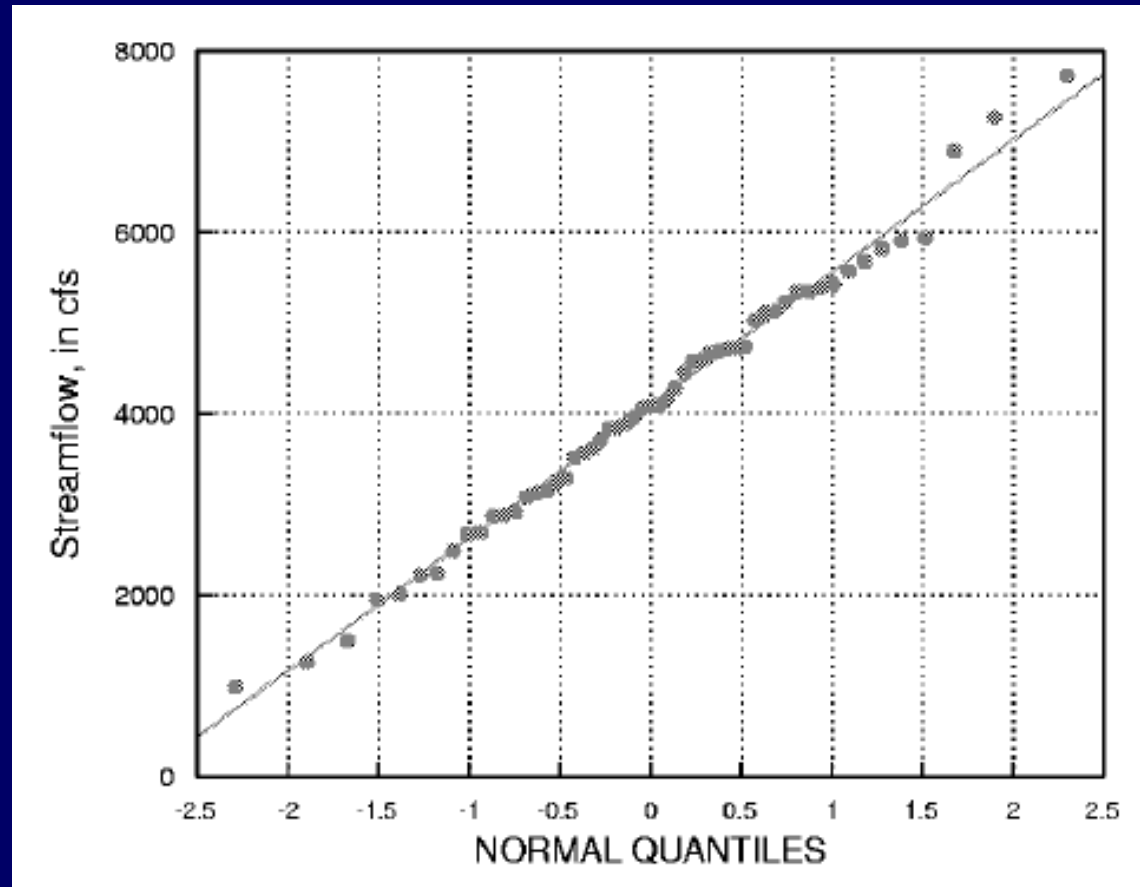


# Probability plots

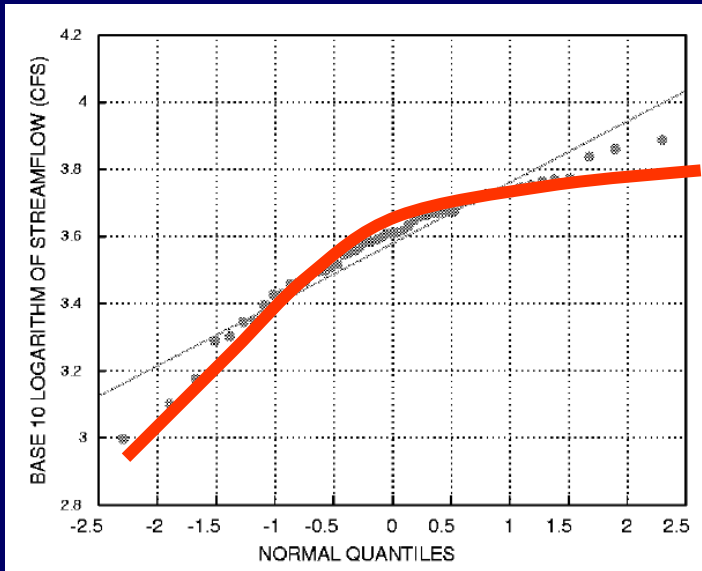


Utili per confronto con distribuzioni di riferimento

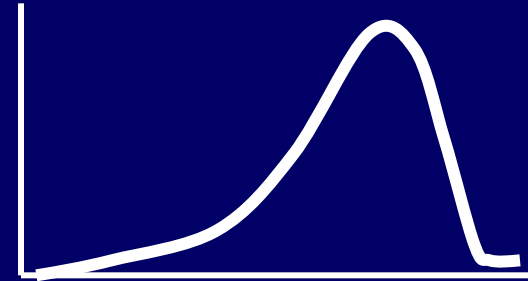
# Probability plots



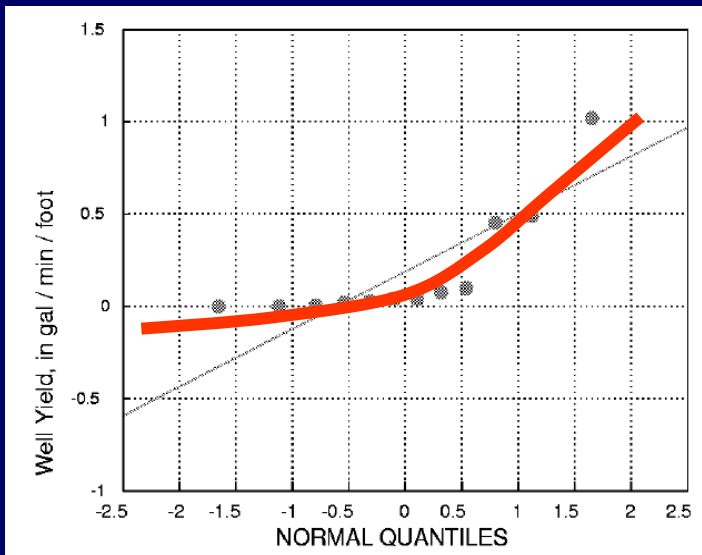
# Skewness



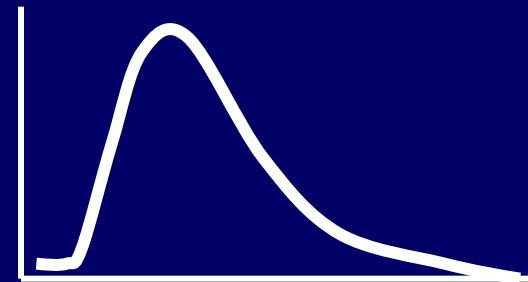
Asimmetria negativa



Trasformazione  $x^\alpha$ ,  $\alpha > 1$



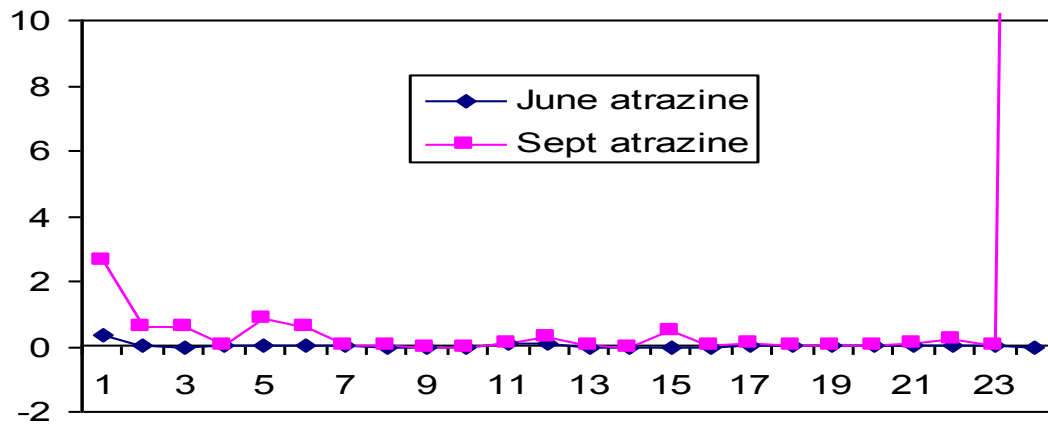
Asimmetria positiva



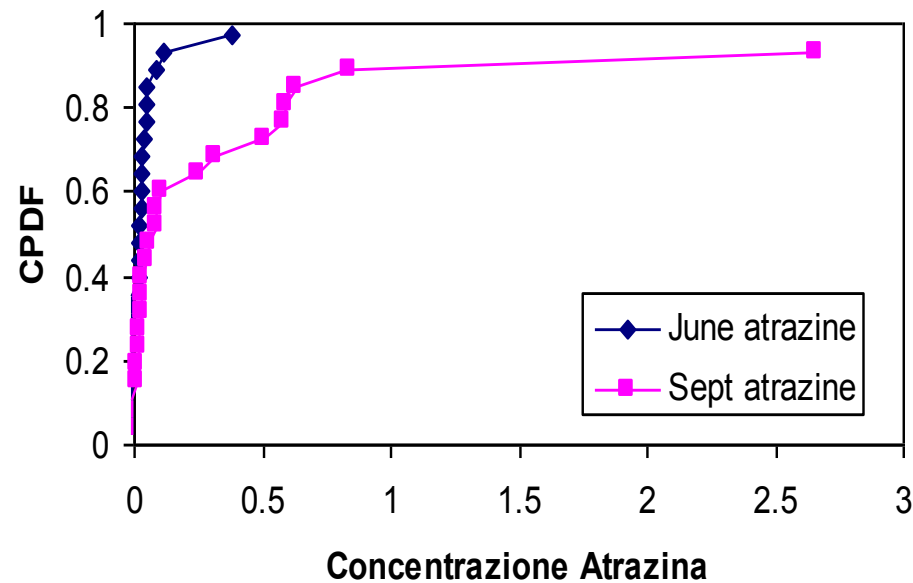
Trasformazione  $\ln(x)$ ,  $x^\alpha$ ,  $\alpha < 1$



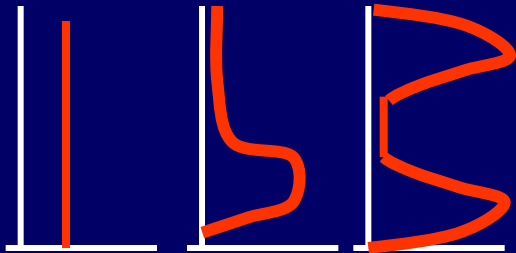
# Confronto dati



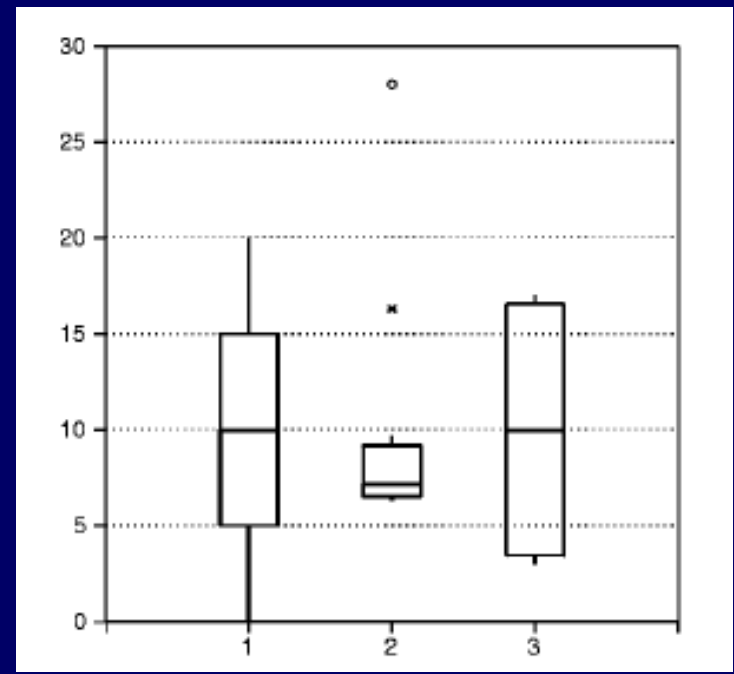
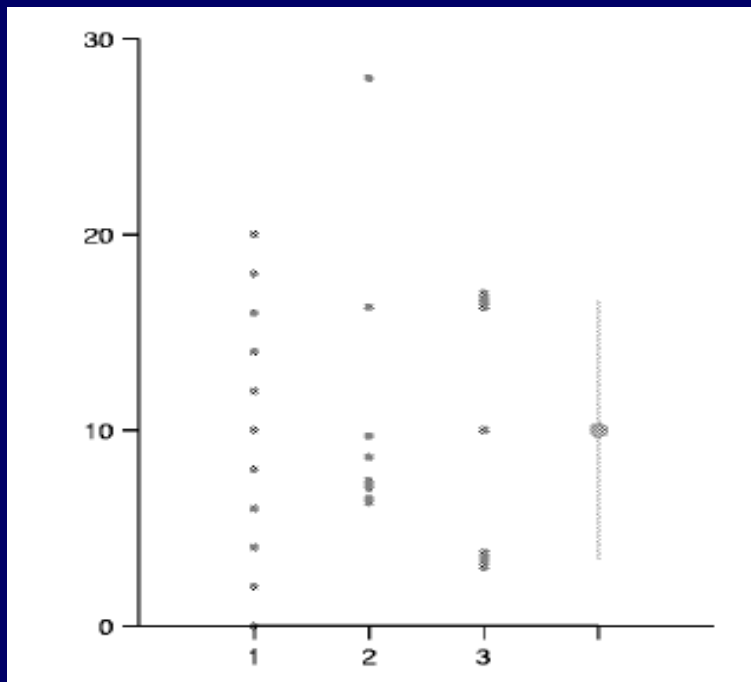
## CPDF



# Box plots per confronto dati



Pdf



Ingegneria Chimica Ambientale

---

# **Analisi multivariata**

Marina Campolo

Centro Interdipartimentale di Fluidodinamica e Idraulica

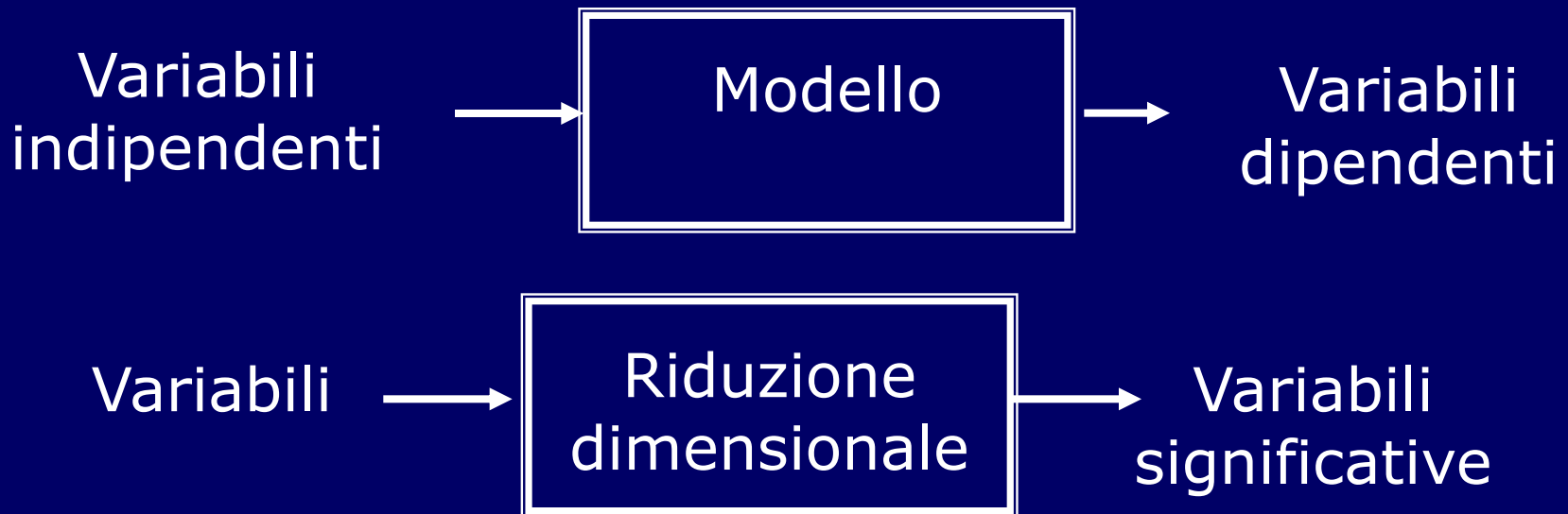
Università di Udine

---

# Obiettivi dell'analisi

---

- Serie multiple:**
- utilizzare dati di una serie come variabili predittive per l'evoluzione di un'altra serie
  - Individuare la combinazione di variabili che meglio descrive il set di dati (min dimensione - massima varianza)



# Serie multi-variate di dati

---

Variabili

Osservazioni

$X_1$

$X_2$

$X_3$

$X_4$

$O_1$

$O_2$

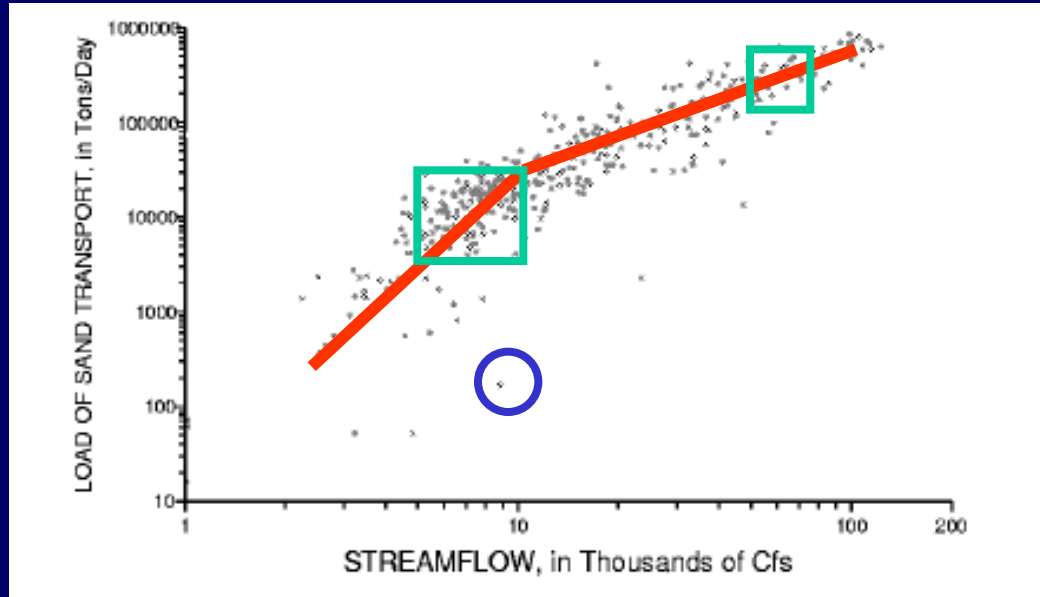
$O_{\dots}$

$O_4$


Matrice dei dati

# Rappresentare relazioni tra dati

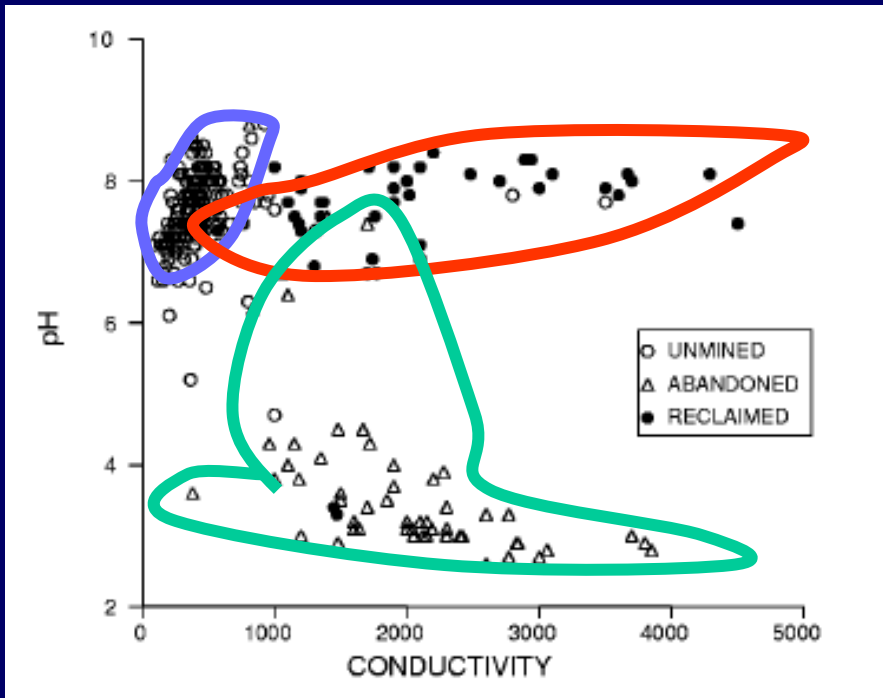
## Scatterplot per coppie di variabili



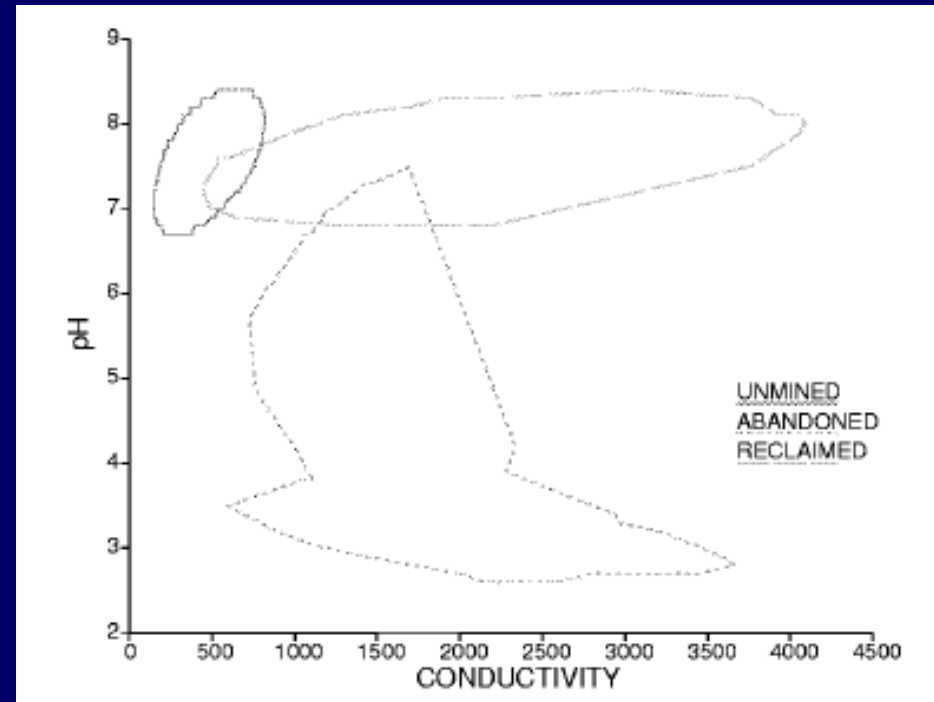
1. Relazioni di linearità (o proporzionalità inversa)
2. Variabilità dei dati in diverse zone dello spazio delle variabili
3. Punti isolati

# Clustering dei dati

Scatterplot (x,y)

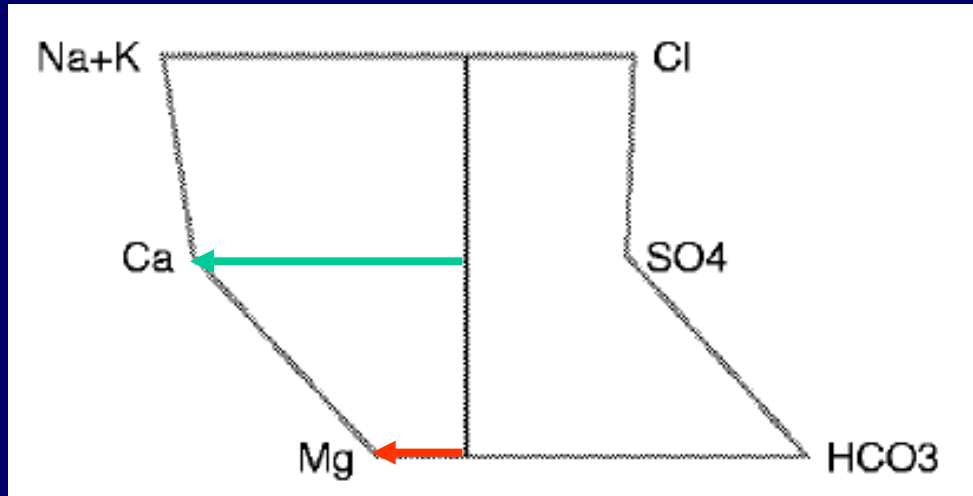


Polar smooth: zona contenente 50% dei dati intorno alla mediana (x,y)



Visual discriminant analysis per gruppi di dati

# Dati multivariati

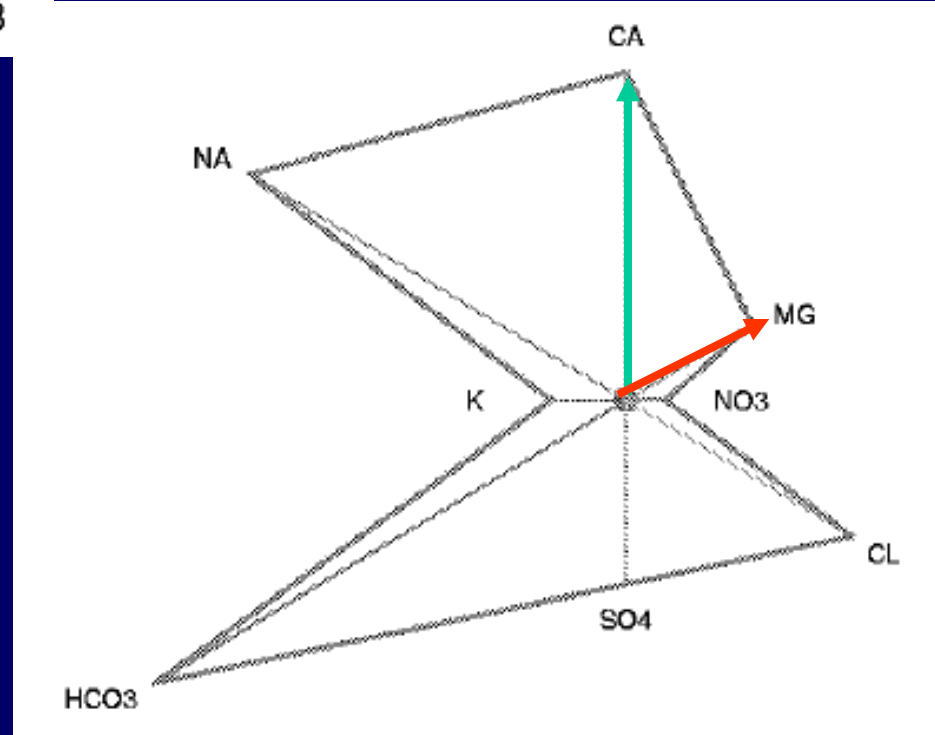


Stiff diagram



Star diagram

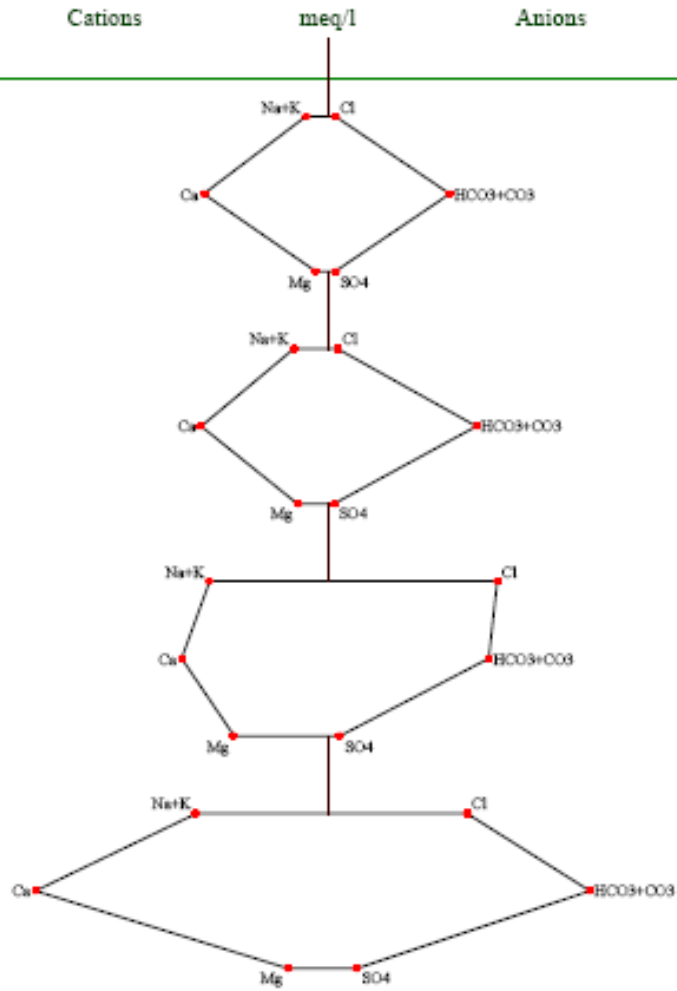
← mediana (o altro percentile)  
←





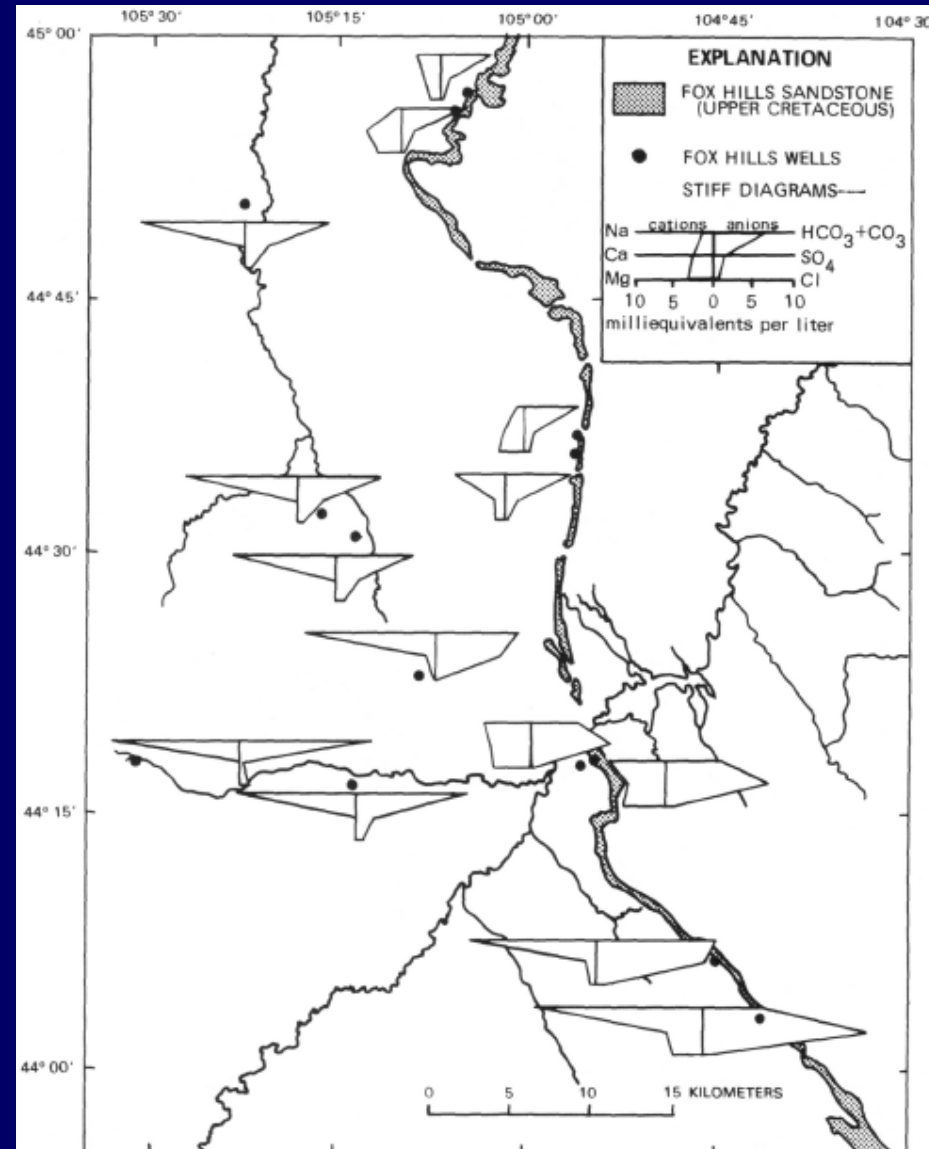
# Stiff Diagram

Floridan Aquifer: WSBH-Ridgeland-Bluflon-Hilton Head

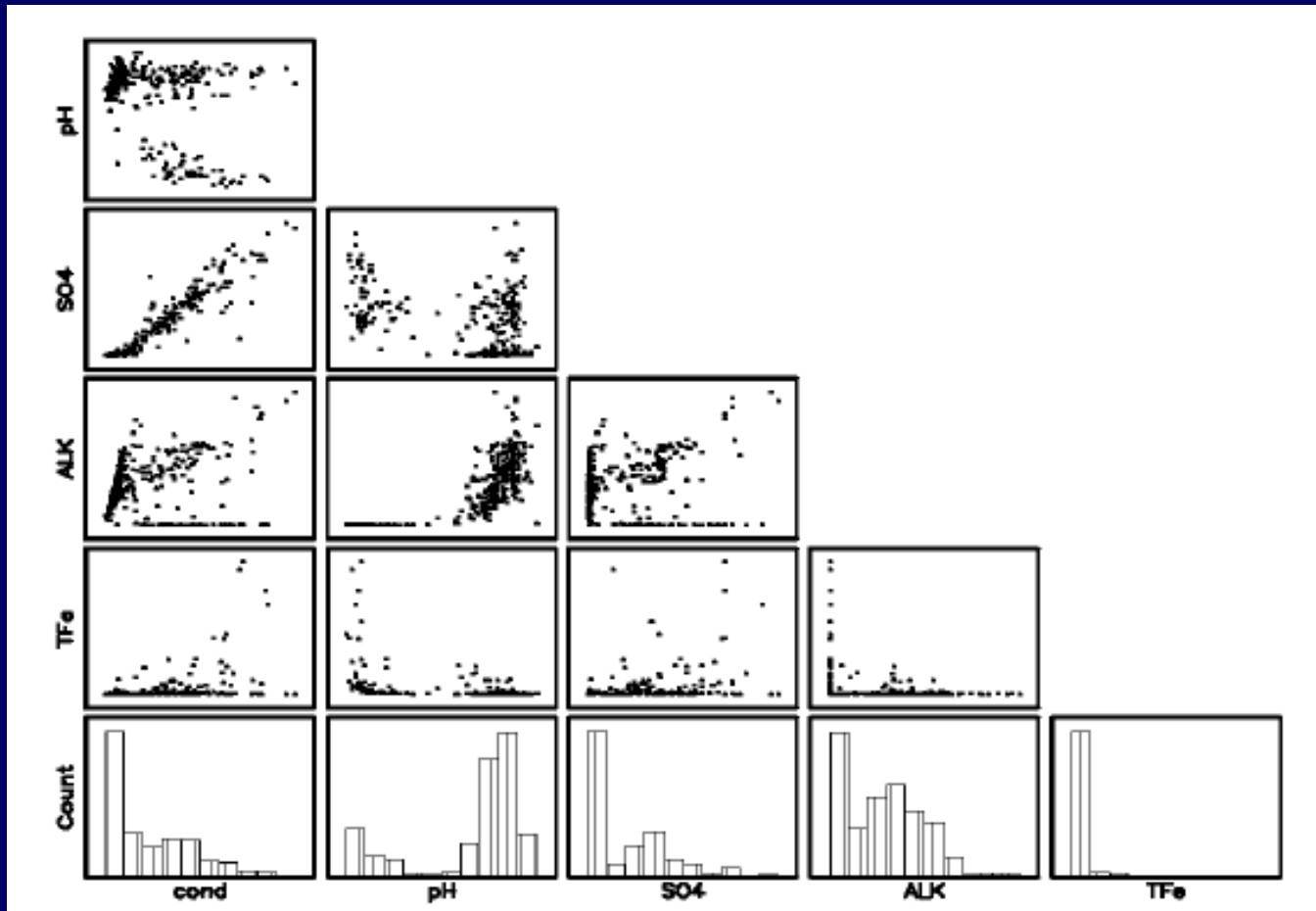


← Evoluzione qualità acque lungo il corso

Rappresentazione in mappa geografica di variazioni osservate →



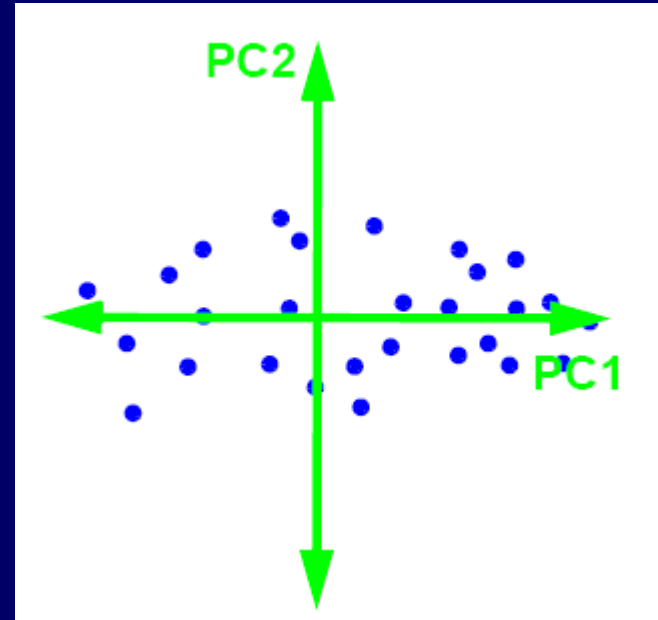
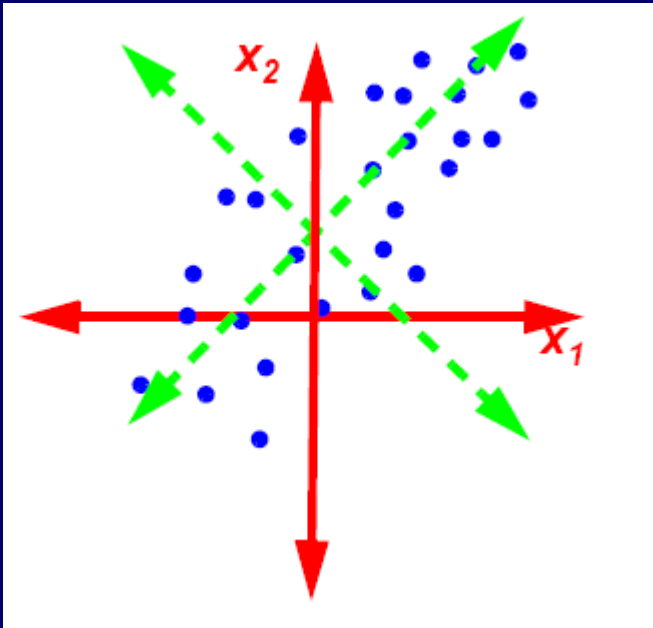
# Relazioni tra più di 2 dati



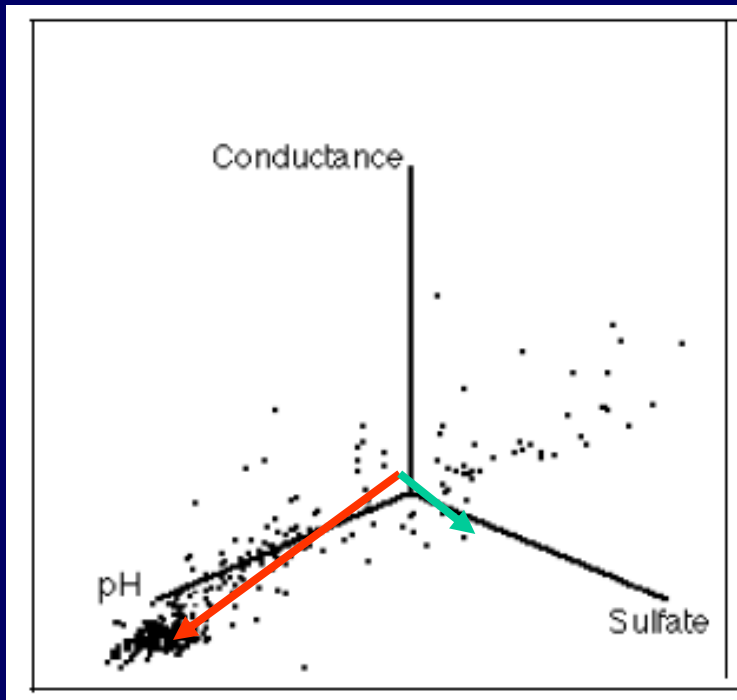
Analisi binaria → Riduzione della dimensionalità dello spazio campionario

# Principal component analysis (PCA)

---



OBJ: Individuare nello spazio delle variabili, la combinazione lineare che produce la rappresentazione più vantaggiosa (numero di coordinate ridotto, massima differenziazione dei dati) → feature extraction



$PCA1 = a \text{ Ph} + b \text{ Cond} + c \text{ Sulfate}$

→ 80% della variazione dei dati

$PCA2 = d \text{ Ph} + e \text{ Cond} + f \text{ Sulfate}$

→ 10% della variazione dei dati

Componenti  
principali

# Principal component analysis

---

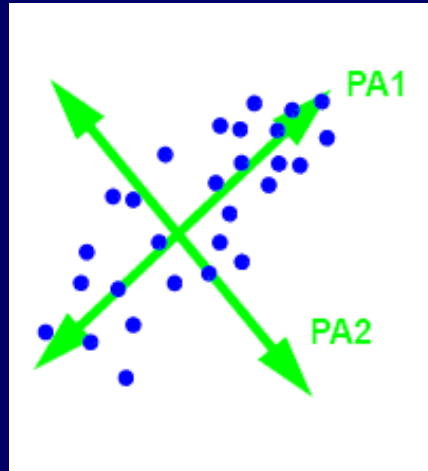
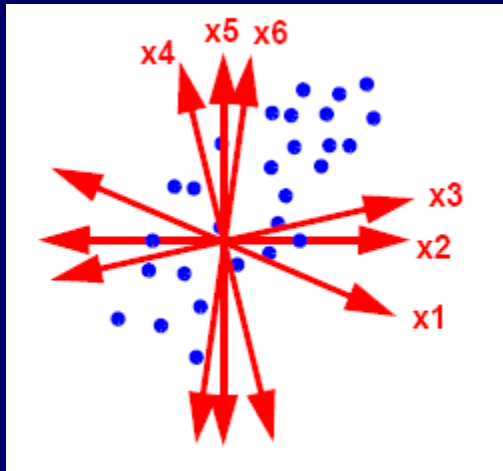
**Approccio algebrico:** Ricerca autovalori/autovettori della matrice della varianza/covarianza dei dati

**Autovalori:** Varianza dei dati

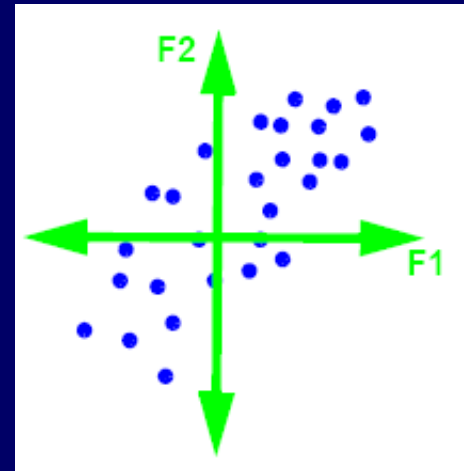
**Autovettori:** Direzioni lungo le quali si allineano i dati

Rappresentazione semplificata dello spazio dei dati (riduzione dimensionale, aggregazione "astratta")

# Factor analysis



PCA



FA

Varianza totale = Varianza comune ad altre variabili +  
Varianza specifica

Procedura: Principal component analysis

Allineamento degli assi principali nella  
direzione che rappresenta la varianza  
comune