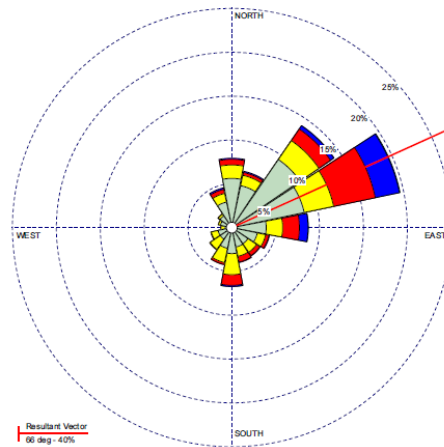
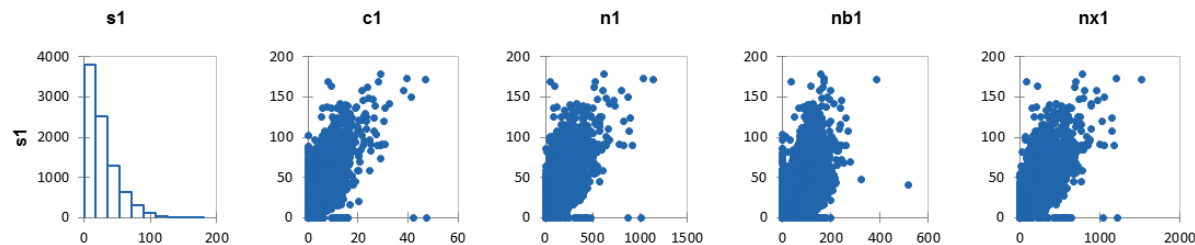


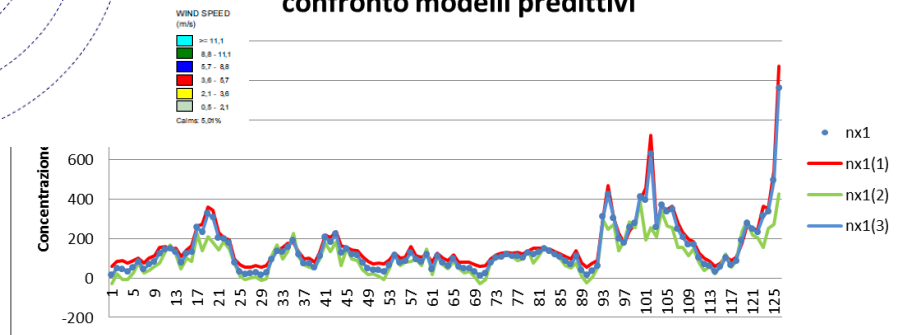
Analisi Multivariata

A.A. 2017/2018

Marina Campolo



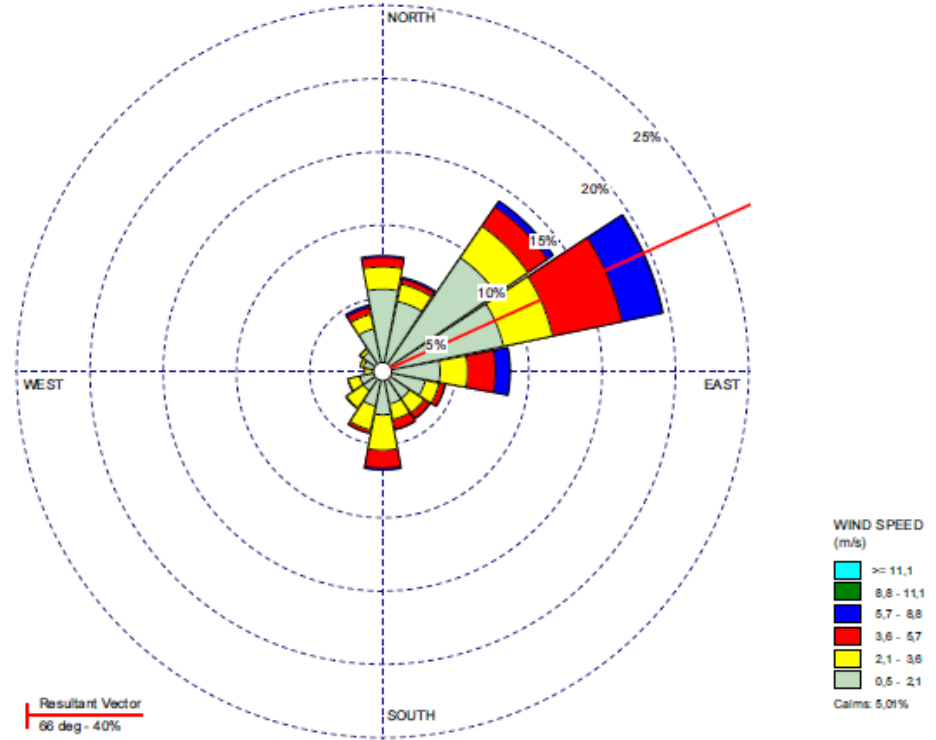
confronto modelli predittivi



Wind rose – Distribuzione doppia

Wind Direction (Blowing From) / Wind Speed (m/s)	0,5 - 2,1	2,1 - 3,6	3,6 - 5,7	5,7 - 8,8	8,8 - 11,1	>= 11,1	Total
348,75-11,25	491	134	56	11	0	0	692
11,25-33,75	427	105	30	12	0	0	574
33,75-56,25	822	228	140	40	1	0	1231
56,25-78,75	736	299	424	246	6	0	1711
78,75-101,25	342	164	167	86	3	0	762
101,25-123,75	259	85	33	6	0	0	383
123,75-146,25	195	103	41	3	0	0	342
146,25-168,75	198	95	58	5	0	0	356
168,75-191,25	259	210	106	12	1	0	588
191,25-213,75	211	145	20	1	0	0	377
213,75-236,25	140	122	4	1	0	0	267
236,25-258,75	140	69	4	1	0	0	214
258,75-281,25	67	41	3	0	0	0	111
281,25-303,75	116	21	2	0	0	0	139
303,75-326,25	143	21	3	1	0	0	168
326,25-348,75	260	88	42	15	0	0	405
Total	4806	1930	1133	440	11	0	8759

Frequency of Calm Winds: 439
Average Wind Speed: 2,17 m/s



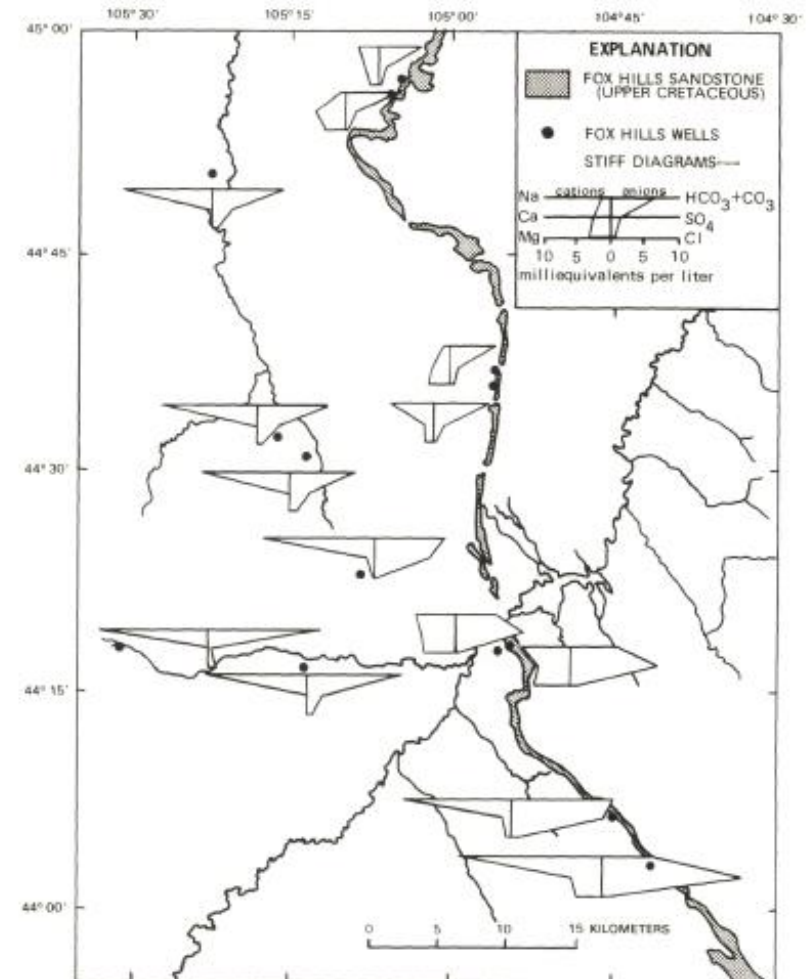
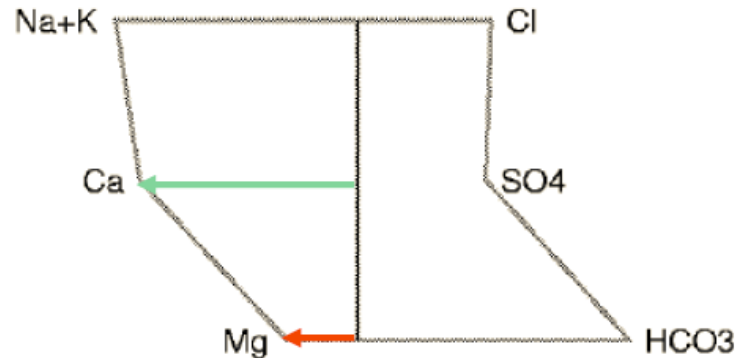
- ▶ Frequenza congiunta
- ▶ Frequenza marginale
- ▶ Numerosità del campione

Rappresentazione grafica di dati multivariati

► Diagrammi rigidi (stiff diagram)

- Posizione= variabile
- Segmento= valore

Forma semplice, confronto visivo tra 2 o più distribuzioni multivariate

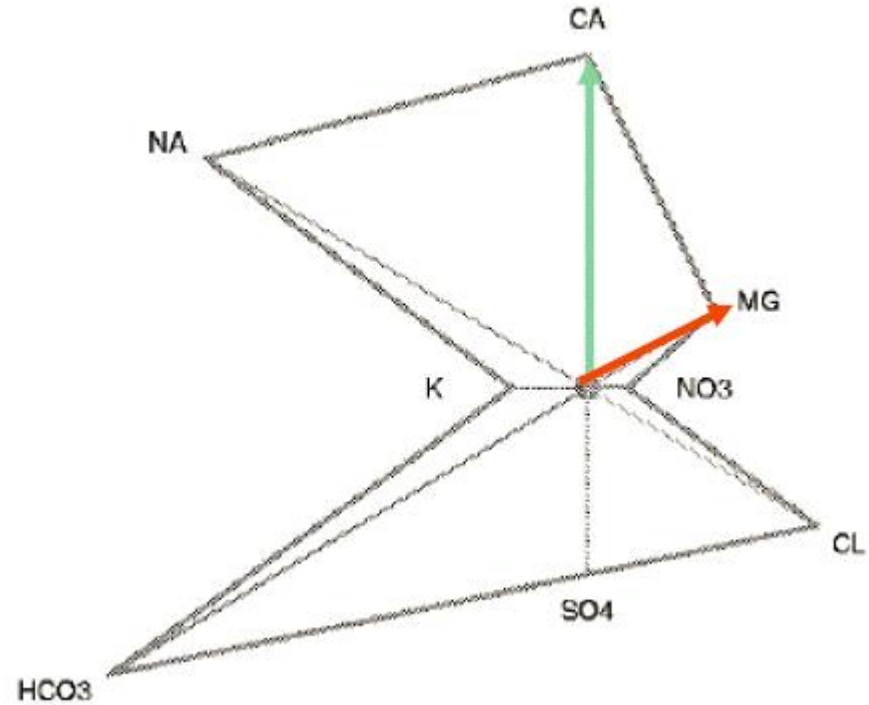
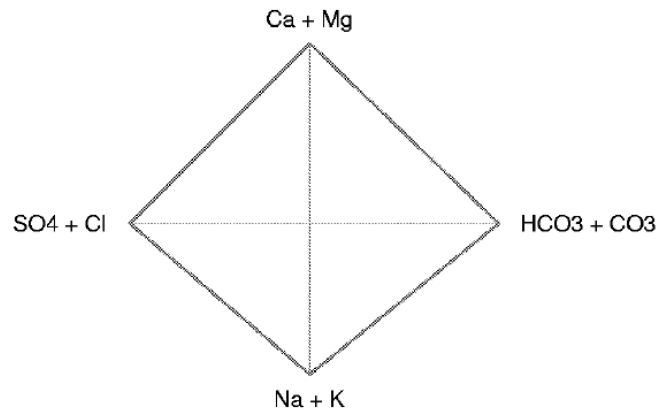


Rappresentazione grafica di dati multivariati

► Diagrammi a stella/ad aquilone (star/kite diagram)

- Posizione angolare= variabile
- Segmento= valore

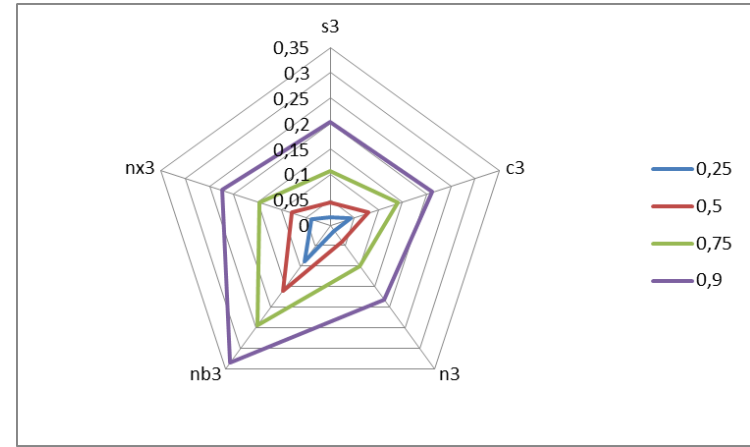
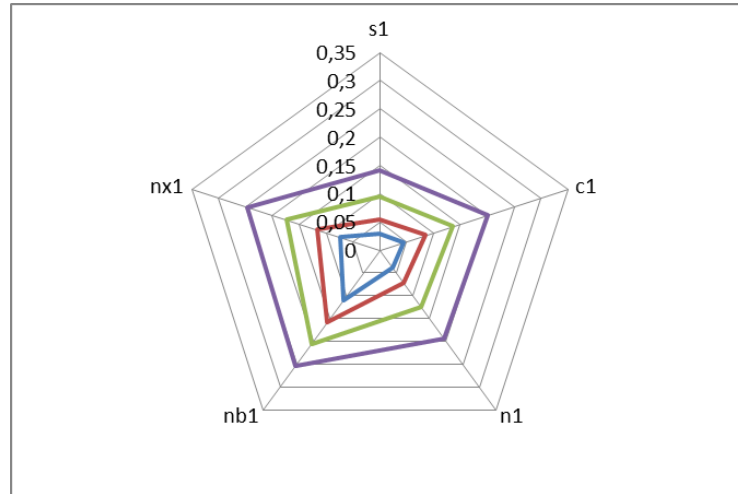
Forme semplici, confronto visivo tra 2 o più distribuzioni multivariate



Rappresentazione grafica di dati multivariati

▶ Diagrammi a stella (star diagram)

- Coordinata radiale=valore percentile
- Coordinata angolare=variabile
- ▶→ Possibilità di confrontare gruppi di dati

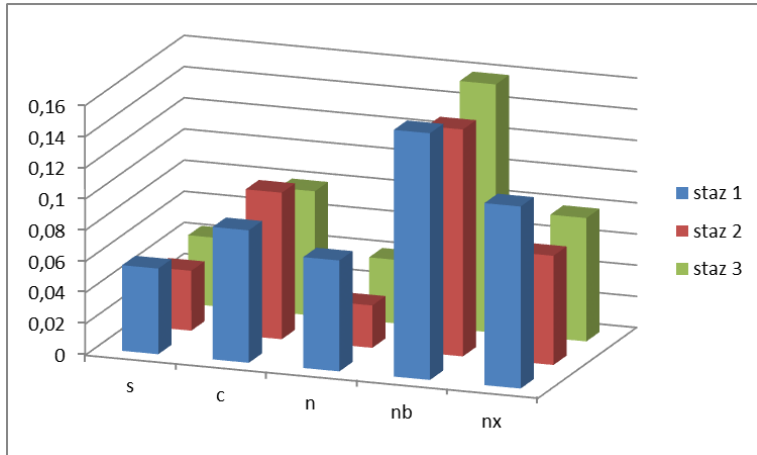


Rappresentazione grafica di dati multivariati

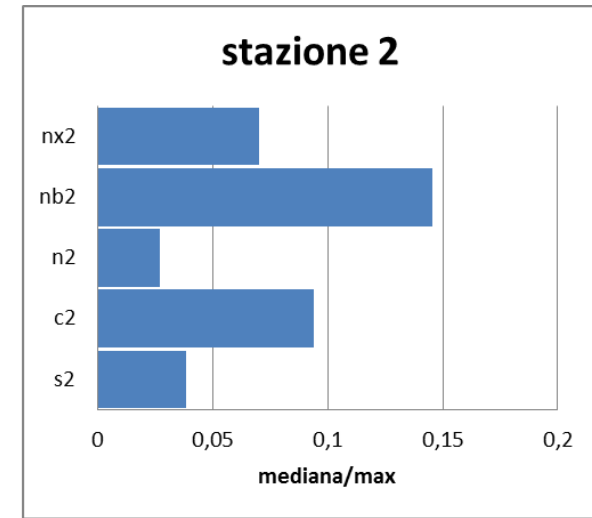
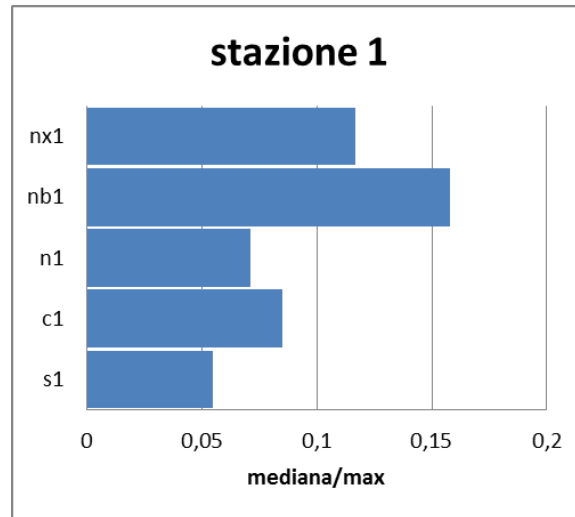
► Diagrammi a linea (line diagram) 2D

- Coordinata 1 = variabile
- Coordinata 2 = valore variabile (quantile)

3D



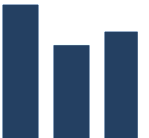

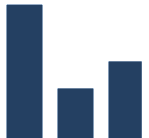


2D



Rappresentazione grafica di dati multivariati

Tabelle di dati con Spark line (linee di tendenza)

- Dato tabulato
- Trend grafico (per immediata leggibilità)

					
	s	c	n	nb	nx
staz 1	0,054924	0,084926	0,071088	0,158097	0,116655
staz 2	0,038404	0,09396	0,027158	0,14545	0,069982
staz 3	0,043982	0,079812	0,040765	0,159327	0,0791

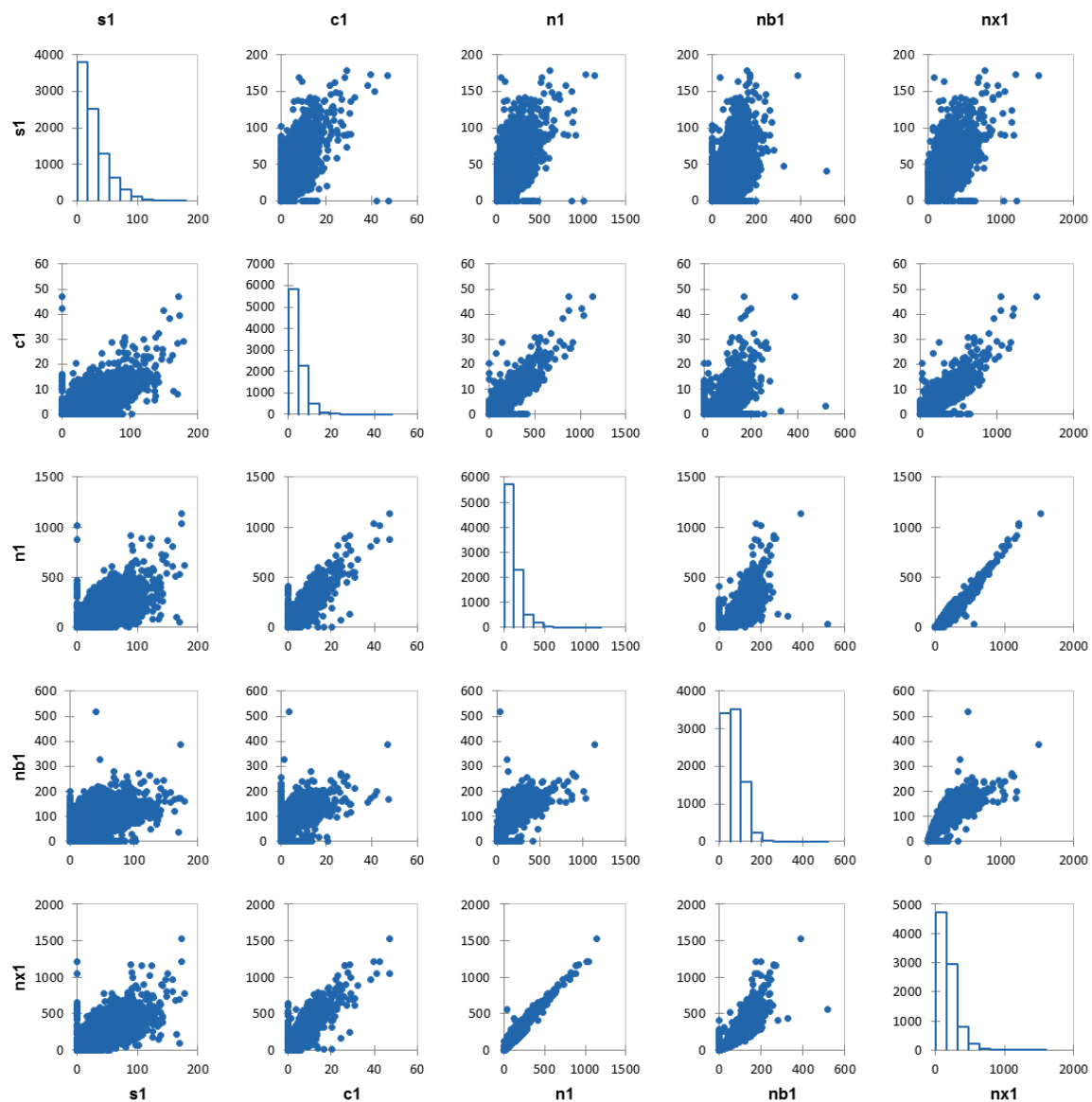
Mediana annua della concentrazione rilevata

Analisi delle dipendenze

- ▶ **OBJ:** evidenziare relazione tra i dati
- Metodo grafico: matrice di scatterplot
- Metodo analitico: tabella di correlazione

▶ Tabella di correlazione

	s1	c1	n1	nb1	nx1
s1	1	0,688	0,601	0,532	0,637
c1	0,688	1	0,821	0,555	0,808
n1	0,601	0,821	1	0,600	0,956
nb1	0,532	0,555	0,600	1	0,808
nx1	0,637	0,808	0,956	0,808	1

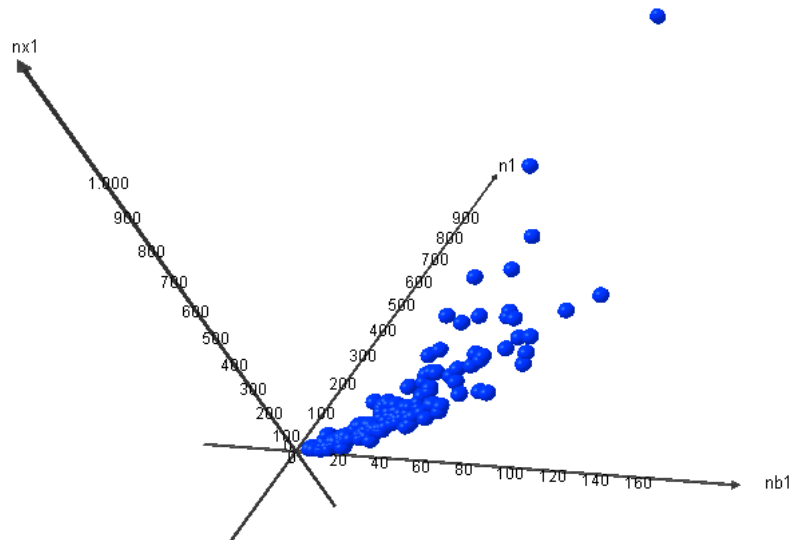


Tipi di relazione tra dati

1. Esistenza di relazione funzionale tra variabili:
allineamento/addensamento di punti in
scatterplot n-dimensionale=curva interpolante

→ Possibilità di sviluppare modelli predittivi

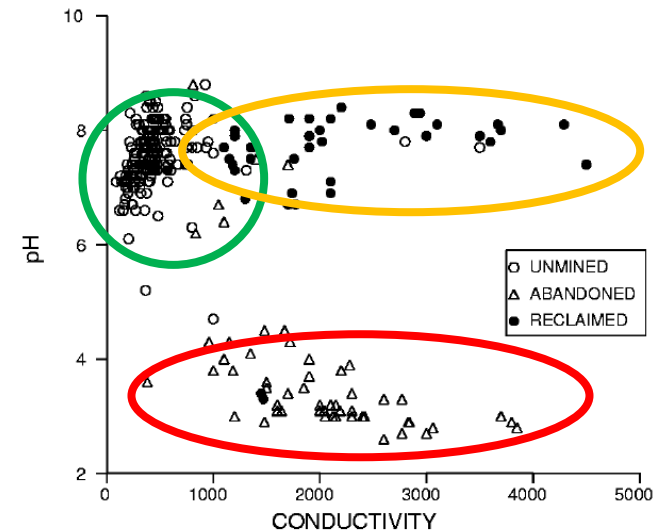
$$Y=f(x_1, x_2)$$



2. Esistenza di gruppi all'interno dei dati

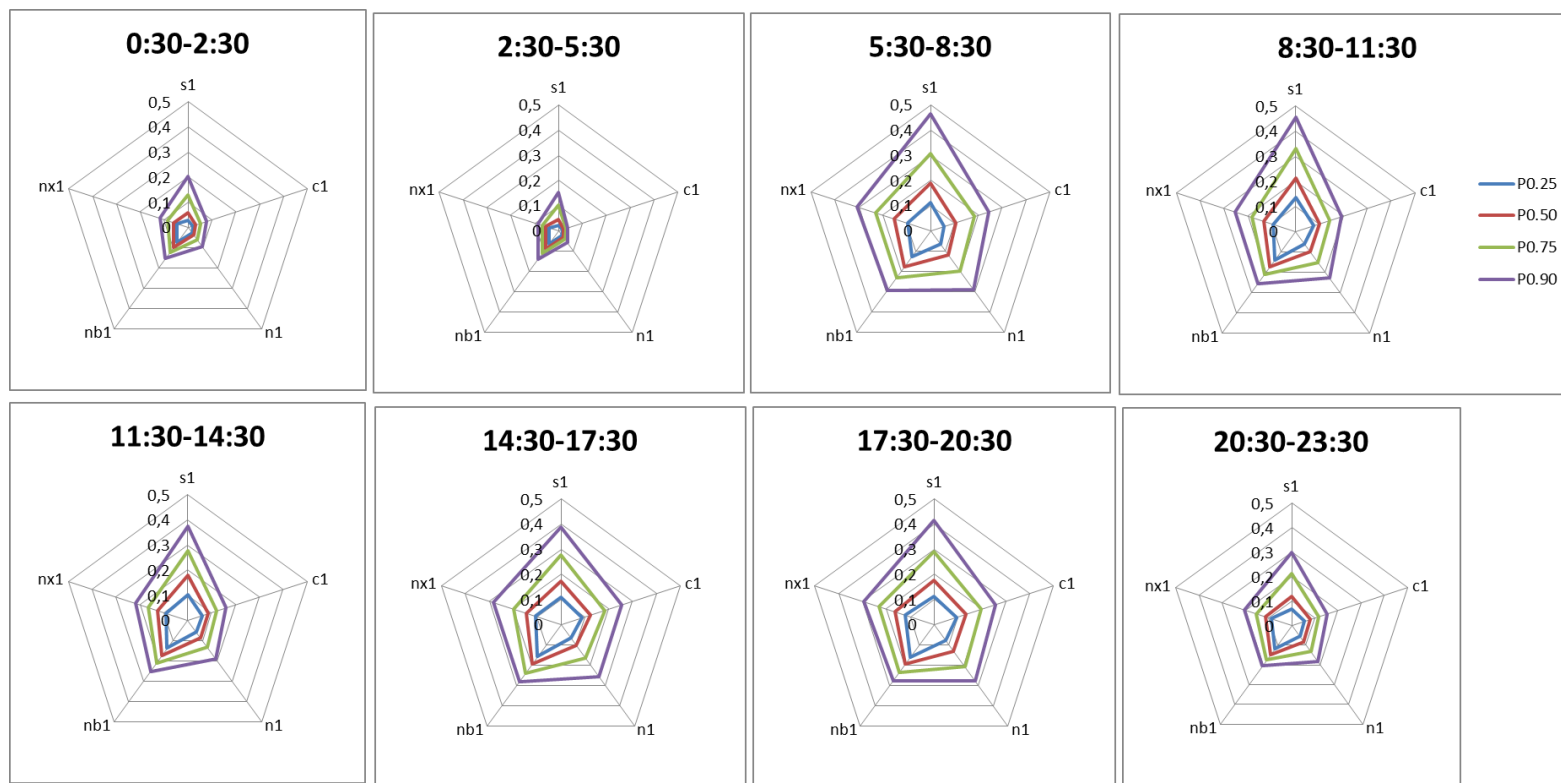
→ possibilità di aggregare/sintetizzare/ridurre la
dimensionalità dei dati

(Principal Component Analysis, Cluster Analysis)



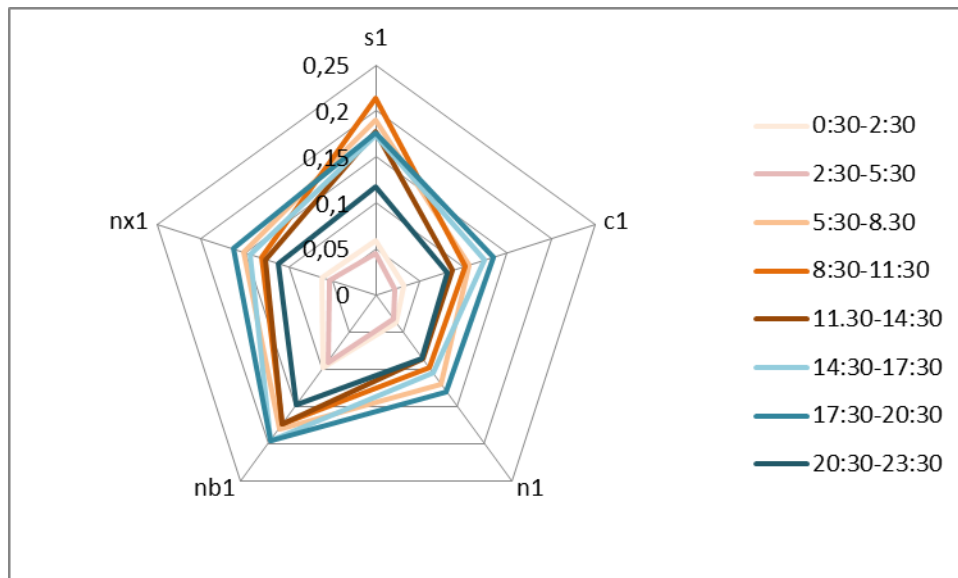
Es: Variazione del livello di qualità dell'aria durante il giorno

Star plot realizzato valutando i quantili della serie multivariata filtrata secondo la finestra temporale indicata: dati normalizzati rispetto al valore massimo del data set



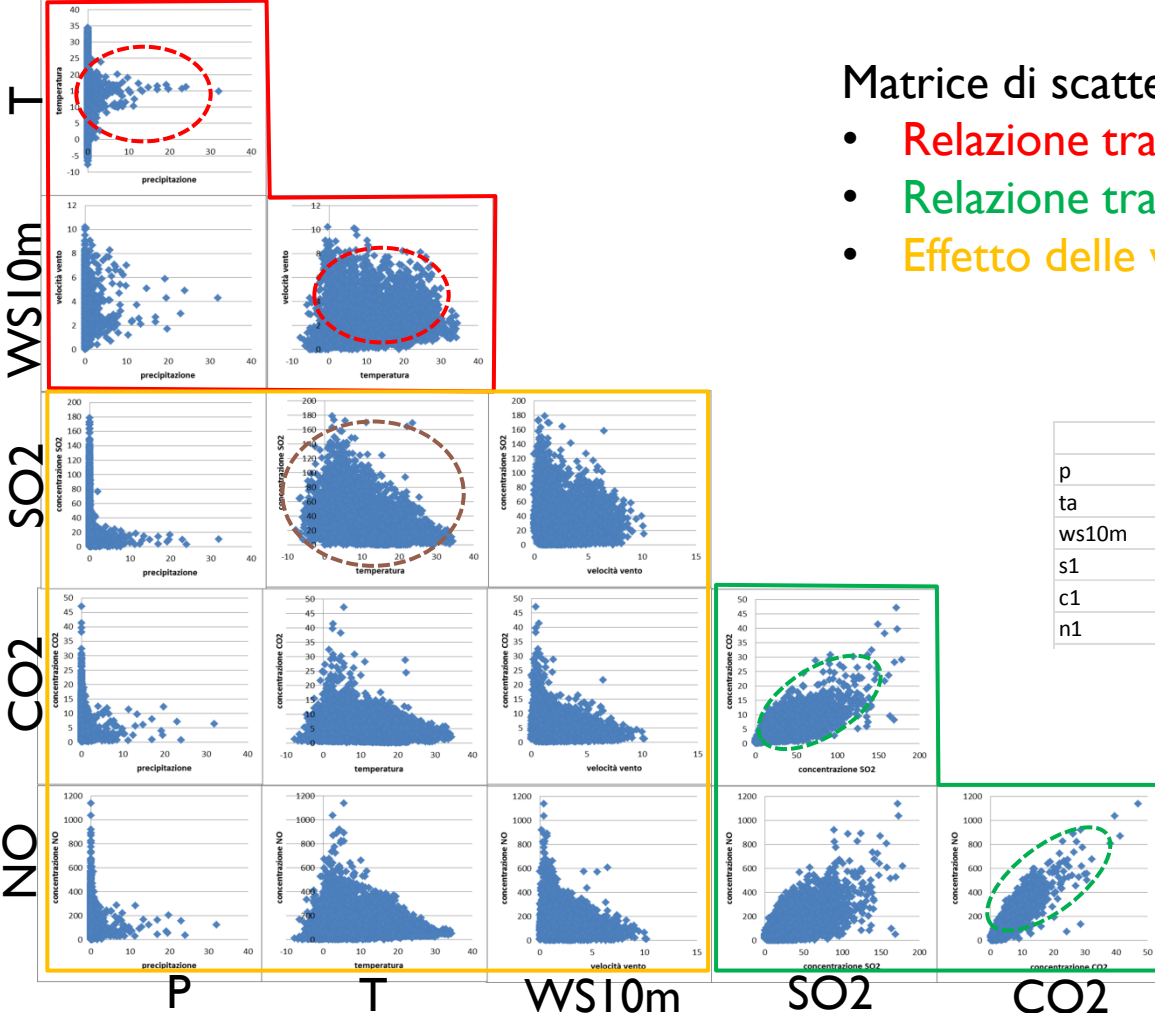
Es: Variazione del livello di qualità dell'aria durante il giorno

Star plot relativo alla variazione temporale del valore medio delle distribuzioni: dati normalizzati rispetto al valore massimo del data set



Il livello di inquinamento è minimo nelle prime ore del giorno, aumenta nelle ore centrali e decresce nuovamente dopo le 20:30

Es: Analisi dipendenze tra i dati



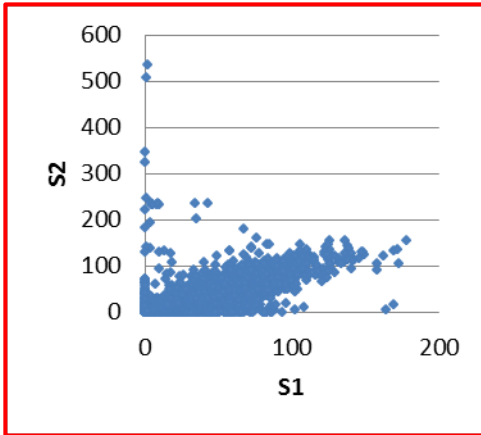
Matrice di scatterplot:

- Relazione tra variabili meteo
- Relazione tra variabili qualità aria
- Effetto delle variabili meteo sulla qualità dell'aria

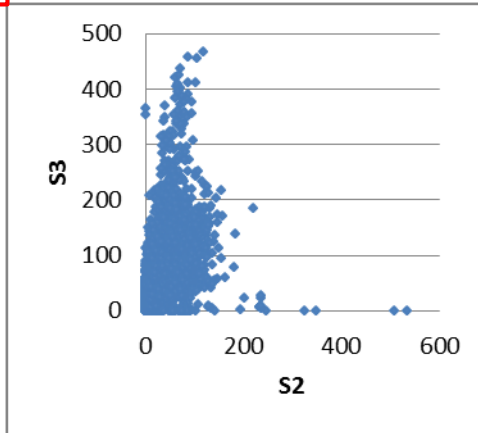
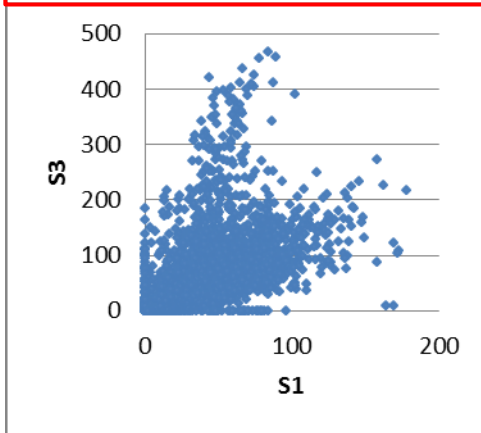
	p	ta	ws10m	s1	c1	n1
p	1,00	0,04	0,12	-0,12	-0,01	-0,02
ta	0,04	1,00	0,06	-0,40	-0,18	-0,20
ws10m	0,12	0,06	1,00	-0,06	-0,19	-0,19
s1	-0,12	-0,40	-0,06	1,00	0,73	0,64
c1	-0,01	-0,18	-0,19	0,73	1,00	0,87
n1	-0,02	-0,20	-0,19	0,64	0,87	1,00

- Relazione di indipendenza?
- Relazione lineare
- Relazione di anti-correlazione?

Es: Analisi dipendenze – qualità dell'aria in stazioni diverse

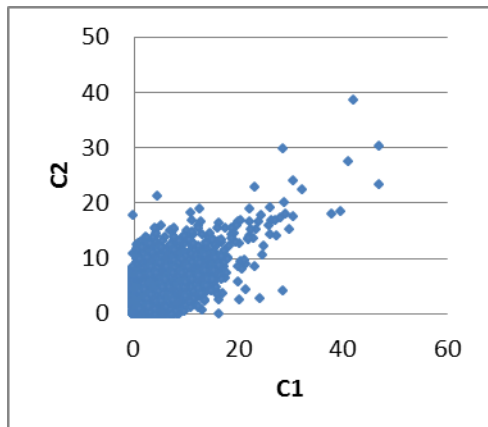


	s1	s2	s3
s1	1	0,733862	0,630583
s2	0,733862	1	0,591838
s3	0,630583	0,591838	1

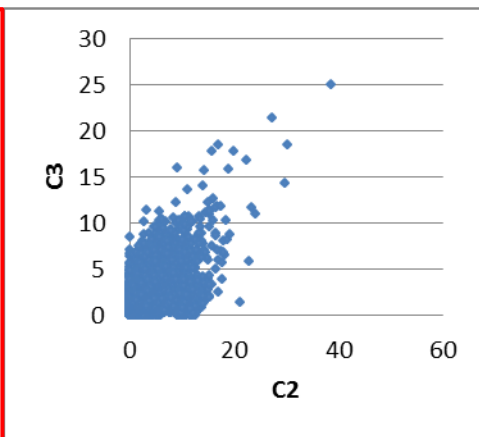
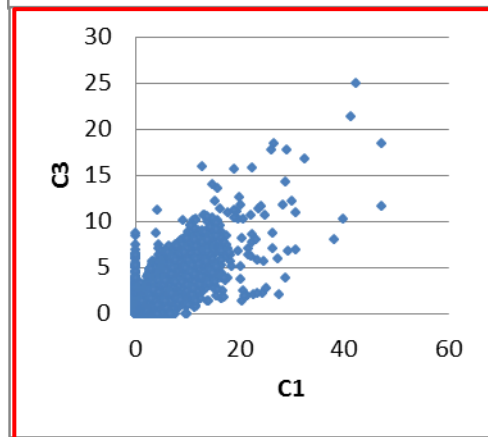


SO2 in stazione 2 **miglior predittore** per SO2 in stazione 1

Es: Analisi dipendenze – qualità dell'aria in stazioni diverse



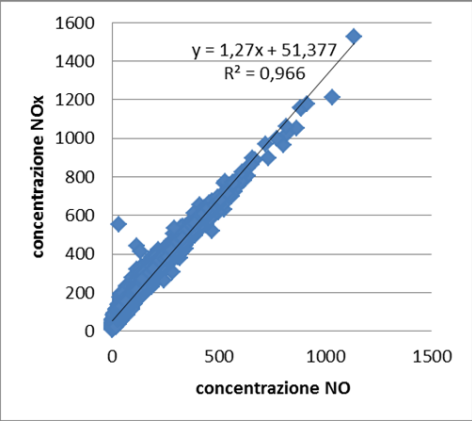
	c1	c2	c3	
c1		1	0,599873	0,759938
c2	0,599873		1	0,519824
c3	0,759938	0,519824		1



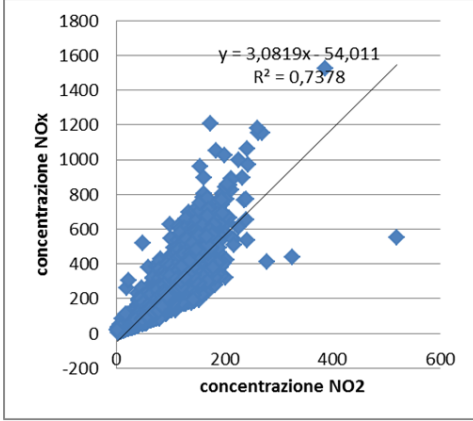
CO2 in stazione 3 **miglior predittore** per
CO2 in stazione 1

Es: Analisi dipendenze – previsione del livello di NOx

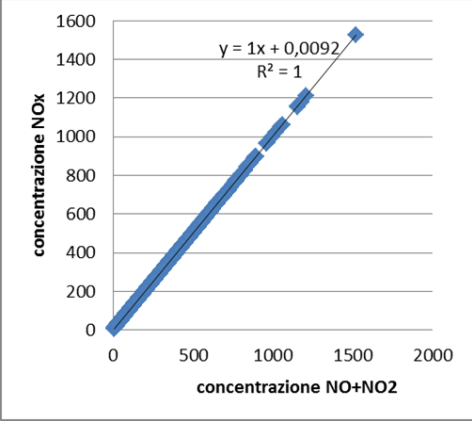
Modello 1: $NO_x = f(NO)$



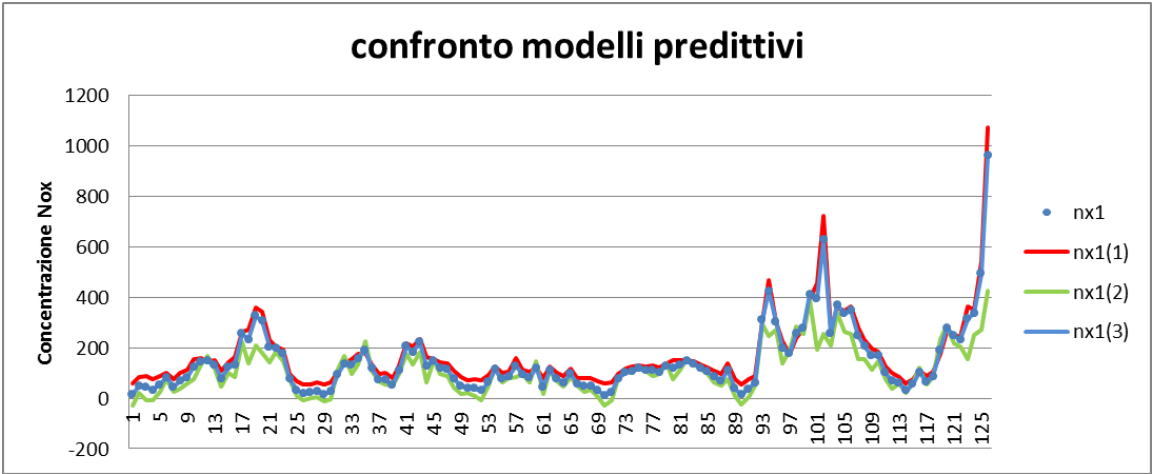
Modello 2: $NO_x = f(NO_2)$



Modello 3: $NO_x = f(NO + NO_2)$



confronto modelli predittivi



previsione NOx

