






agenzia regionale per la protezione ambientale

SCUOLA DI ALTA FORMAZIONE AMBIENTALE

GIORNATA DI STUDIO  
«LA CARATTERIZZAZIONE CHIMICA DEL PARTICOLATO ATMOSFERICO»  
V edizione nazionale

Agenzie ambientali  
**AssoARPA**  
Emilia Romagna Toscana Umbria Marche Lazio Toscana Puglia Lombardia Veneto Friuli Venezia Giulia Abruzzo Molise Basilicata Sardegna Valle d'Aosta Piemonte Liguria Sicilia Trentino-South Tyrol

**ARPA** LOMBARDIA  
Agenzia Regionale per la Protezione dell'Ambiente

**arpae**  
agenzia prevenzione ambiente energia emilia-romagna

## L'ANALISI MULTIVARIATA NELLA CARATTERIZZAZIONE CHIMICA DEL PM

Exploratory Data Analysis (PCA etc.)  
e metodi di classificazione (SIMCA etc.)

Dott. Andrea Mistaro  
ARPA FVG  
Laboratorio Acque Marino - Costiere e Qualità dell'Aria  
Trieste

Indice:

- 1) Tecniche AMV
- 2) Esempi ARPA FVG

Terni, 21 novembre 2022 dott. Andrea Mistaro, SOS Laboratorio AMCQA (Trieste) 1



*«Ci tocca gestire argomenti del sapere molto complessi»*  
(Giorgio Assennato, ex-direttore di ARPA Puglia)

**1° problema: grande numero variabili (n)**

*“...the world is generally multivariate”*  
(Edward Tufte)

**2° problema: grande numero campioni (m)**

*«Imbalance (...) between the technical capacity to **generate** lots of good data and the human capacity to **interpret** all these data»*  
(Riccardo Leardi, UNI GE)

→ ***(m x n)...***

***«too much data = too little information»***  
(Harald Martens, Norwegian Univ. of Science & Technology)

Necessità di basare le conclusioni su **ds grandi** (*m* campioni x *n* parametri)  
Difficoltà ad estrarre informazioni interessanti latenti in **ds grandi**

dott. Andrea Mistaro, SOS Laboratorio AMCQA (Trieste) 2

arpa FVG

CHEMOMETRICS: *“the chemical discipline that uses mathematical and statistical methods to (...) provide maximum information by analysing chemical data”*

## Exploratory Data Analysis (EDA)

“The amount of data retrieved may be so large as to preclude analysis by means other than EDA”

```

graph LR
    A[CAMPIONE] -- "Analisi Chimica" --> B[DATA SET]
    B -- "Exploratory Data Analysis (EDA)" --> C[INFORMAZIONE]
    C --> D[Visualizzazione «semplice» di problemi complessi]
  
```

Visualizzazione «semplice» di problemi complessi

dott. Andrea Mistarò, SOS Laboratorio AMCQA (Trieste) 3

## Exploratory Data Analysis

[\[https://infometrix.com\]](https://infometrix.com) arpa FVG

*PREMESSA:*

- **Patterns of association** exist in many data sets...
- ...but the **relationships between samples [and variables]** can be difficult to **discover** when the data matrix exceeds three or more features.
- **Exploratory Data Analysis can reveal hidden patterns in complex data...**

*... COME LO FACCIAMO:*

- ...by reducing the information to a more comprehensible form.
- Exploratory algorithms such as **Principal Component Analysis (PCA)** and **Hierarchical Cluster Analysis (HCA)** are designed to **reduce large complex data sets into a series of optimized and interpretable views.**

*...PERCHÉ LO FACCIAMO:*

- These views **emphasize the (natural?) groupings in the data** and show **which variables most strongly influence those patterns.**
- Such a chemometric analysis **can expose possible outliers and indicate whether there are patterns or trends** in the data.

dott. Andrea Mistarò, SOS Laboratorio AMCQA (Trieste) 4

arpa **FVG**

## TECNICHE AMV

*«Imperocché la mente umana essendo circonscritta, & di poca capacità, dalla moltitudine si confonde, & resta la lezione senza frutto....»*

[P. Sarpi, «Supplimento dell'Historia degli Uscocchi», 1676]

→ **Approccio tradizionale «riduzionista»:**  
*«data has **traditionally** been analyzed using one or two variables at a time.*

*However (...)*  
*to discover the relationships among all samples and variables **efficiently**, we must process all of the data simultaneously»*

[https://infomatrix.com]

→ **Analisi multivariata: metodi proiettivi**  
*«immense visual capacities of the human eye-brain system»*

[E. Tufte, 2006]

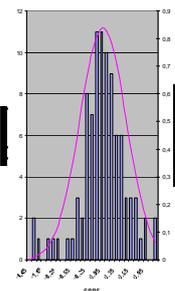
**scopi:**

**“data visualization” + “dimension reduction” → “EXPLORATION ABILITIES”**

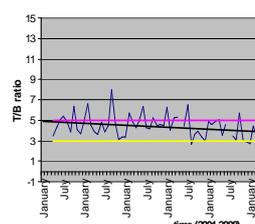
dott. Andrea Mistarò, SOS Laboratorio AMCQA (Trieste) 5

...dalla **STATISTICA CLASSICA**...

arpa **FVG**



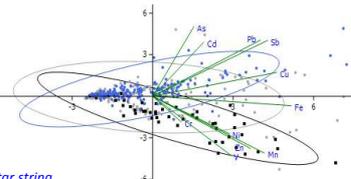
- **Caratteristiche della distribuzione:**  
normale, lognormale, ...
- **Parametri di localizzazione:**  
Medie ( $\mu$ ), Mediana, Moda
- **Parametri di dispersione:**  
Varianza ( $\sigma^2$ ) e Dev. St. ( $\sigma$ )
- **Parametri di asimmetria:**  
Skewness
- **Trend in serie temporali:**  
pendenze (m) e quote (q)



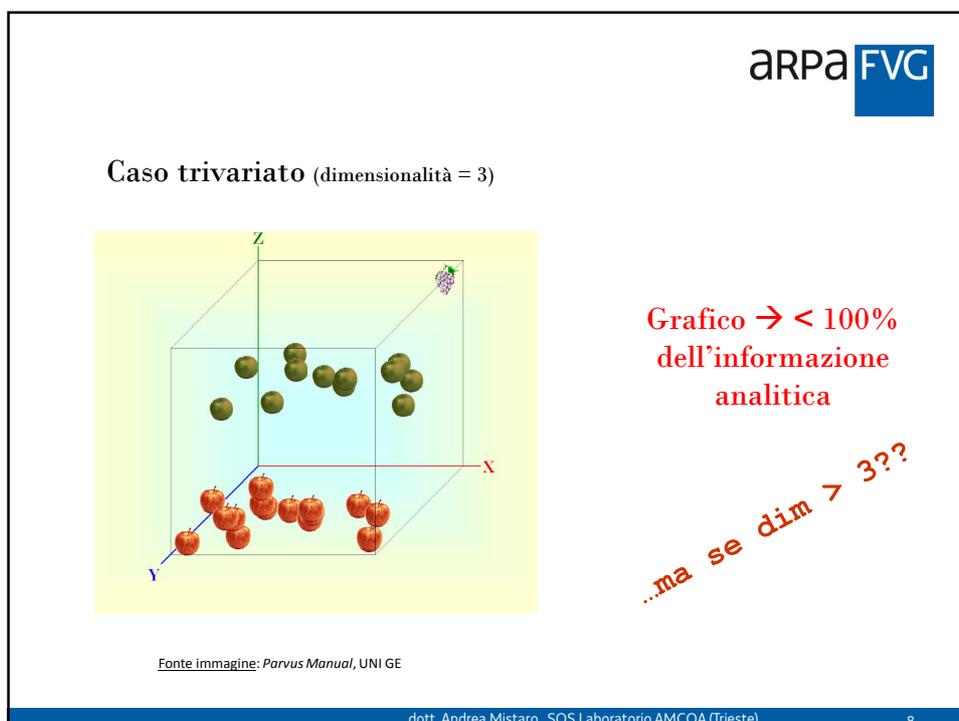
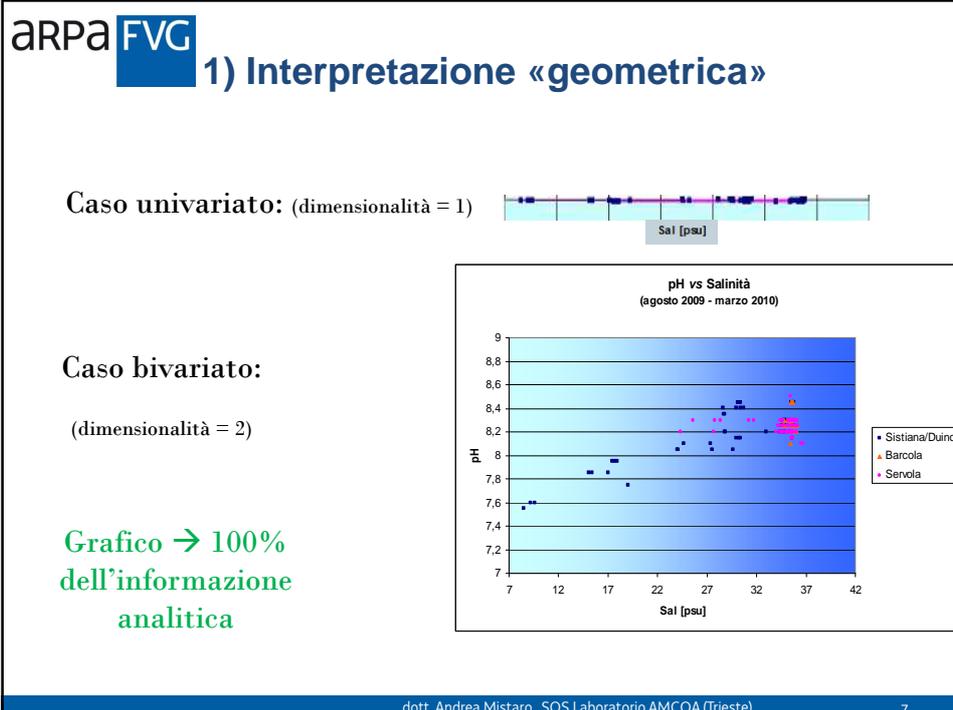
... alla **ANALISI MULTIVARIATA:**

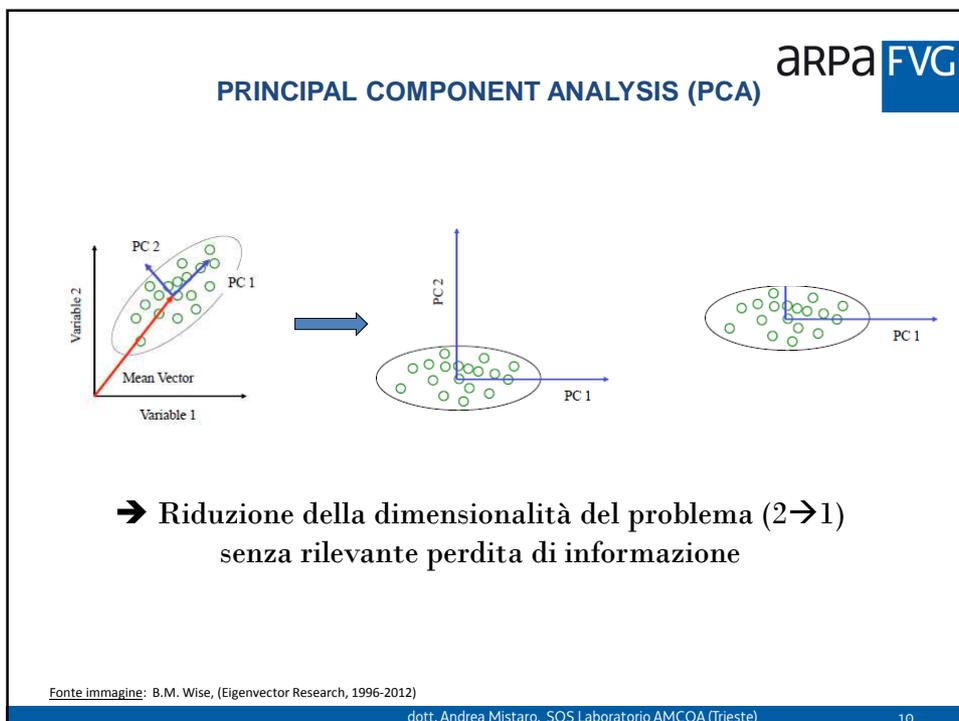
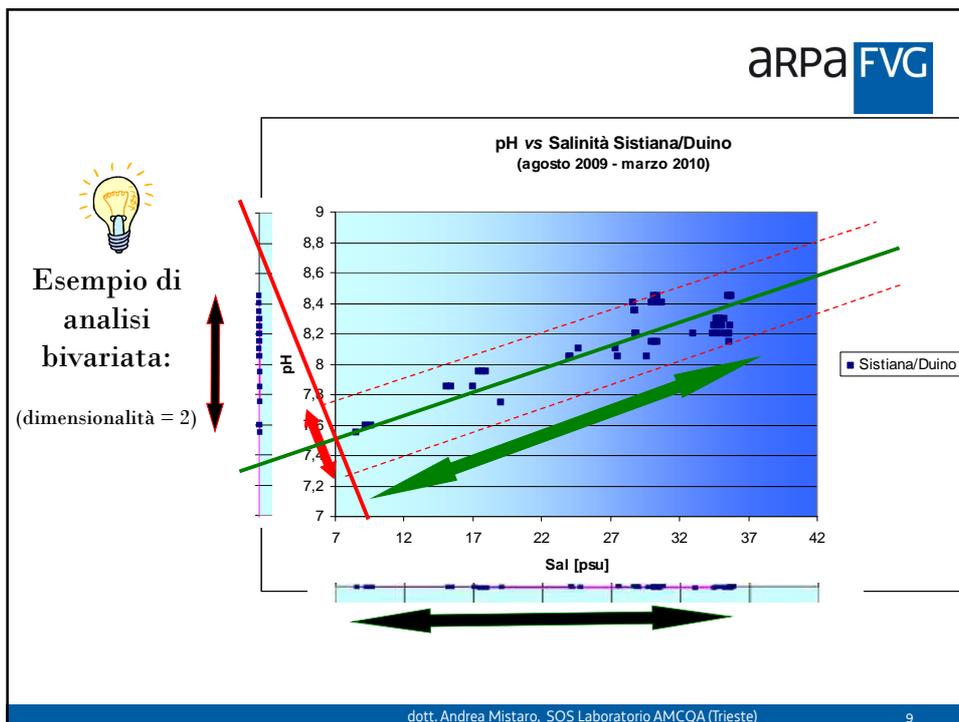
*«A problemi complessi servono risposte complesse, non banali»*  
 (F. Mariottini, Resp. Comunicazione ARPA Umbria)

*«... suonare la chitarra con tutte le corde»*  
 In 1983 Harald Martens (...) began his lecture playing a piece of music with only one guitar string. He said: "This is univariate statistics". Then he played the same piece using all the guitar strings. He said: "... and this is multivariate statistics".  
 M. Forina, Fifty years of Chemometrics, fifty years with Chemometrics



dott. Andrea Mistarò, SOS Laboratorio AMCQA (Trieste) 6

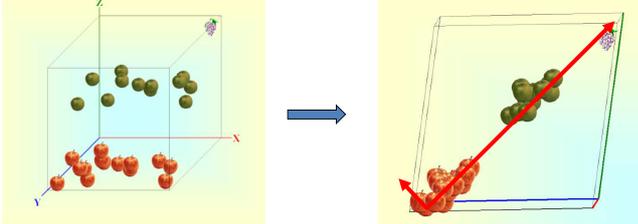




ARPA FVG

PCA

Fonte immagine: Porvus Manual, UNI GE



«All models are wrong  
...but some are useful»  
[G.E.P. Box]



«L'Arte è una menzogna che  
aiuta a capire la realtà»  
[P. Picasso]

**«scomposizione volumetrica»:**  
il pittore può scomporre liberamente il senso del volume secondo accostamenti che nulla hanno a che vedere con la prospettiva, rappresentando sulla tela, simultaneamente, varie prospettive possibili.

«Il cubismo non è (...) un'arte di imitazione, ma di pensiero (...). Il pittore può quindi dare l'apparenza delle tre dimensioni (...). Egli non potrebbe farlo, rendendo semplicemente la realtà-vista»  
(Apollinaire)

➔ **Riduzione della dimensionalità del problema (3→2 (→1?))  
senza rilevante perdita di informazione...**

dott. Andrea Mistaro, SOS Laboratorio AMCQA (Trieste) 11

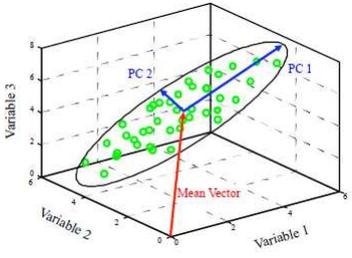
ARPA FVG

«metodi proiettivi»:

*“projection of a multiway signal from the high-dimensional space of objects onto a much lower-dimensional space (usually 2D, i.e. a plane)”*

*«immense visual capacities of the human eye-brain system»*  
[E. Tufte, 2006]

**permette:**  
“data visualization”  
“dimension reduction”  
➔ “EXPLORATION ABILITIES”



1<sup>a</sup>: “Analysis of a Complex of Statistical Variables with Principal Components”, 1933. *Journal of Educational Psychology*

Fonte immagine: B.M. Wise, (Eigenvector Research, 1996-2012)

dott. Andrea Mistaro, SOS Laboratorio AMCQA (Trieste) 12

ARPA FVG

## 2) Interpretazione «algebraica»

$X$  (data, 100% dell'info) =  $T$  (Score matrix, (100 - ε) %)  $P$  (Loading matrix) +  $E$  (Error matrix, ε %)

«componente deterministica» (info sistematica)  
 («Score m.»: relazioni tra campioni)  
 («Loading m.»: relazioni tra variabili)

«componente stocastica» (info residua: noise, errori misura e camp.to, info ridondanti, multicollinearità...)

**algebra matriciale:**  
 (rototraslazione degli assi)  
 (riduzione della dimensionalità)

$$X = TP^T + E$$

dott. Andrea Mistarò, SOS Laboratorio AMCOA (Trieste) 13

ARPA FVG

**Loading matrix** : relazione tra il «nuovo» e il «vecchio» sistema di assi:

$$X = \begin{bmatrix} | & | & & | \\ t_1 & p_1 & + & t_2 & p_2 & + \dots + & t_k & p_k \\ | & | & & | \end{bmatrix} + E$$

$X = TP^T + E$

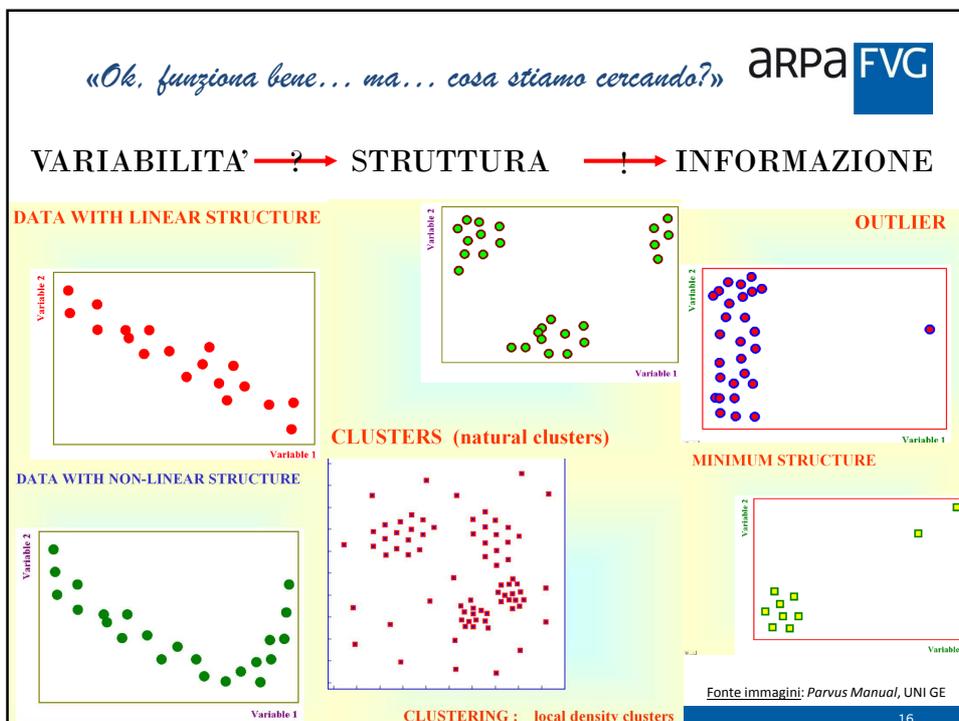
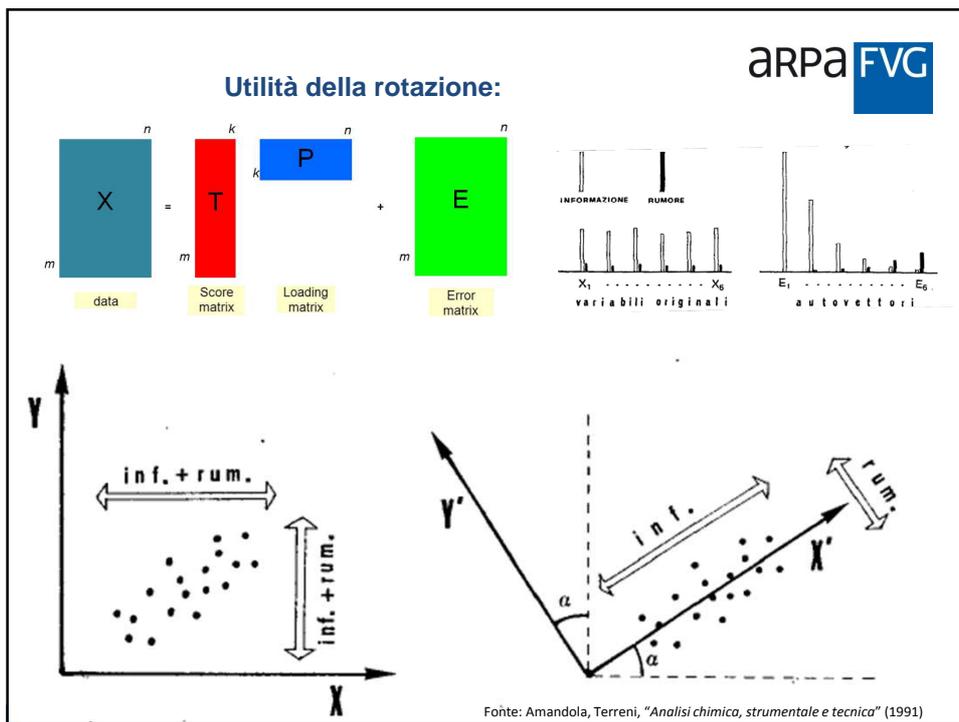
The  $p_i$  are eigenvectors of the covariance matrix of  $X$

**determinazione del nuovo sistema di  $k$  assi:**  
 (matematicamente: problema di autovalori/autovettori e diagonalizzazione della matrice)

**Nuovi assi («var. latenti» o «componenti») = combinazione lineare delle var. orig.:**  
 $PC_i = \alpha_{i,1}v_1 + \alpha_{i,2}v_2 + \alpha_{i,3}v_3 + \dots + \alpha_{i,n}v_n$

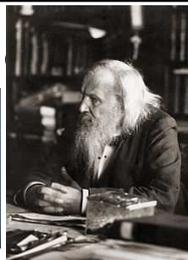
Numerati progressivamente in ordine decrescente di informazione catturata →  
 «componenti principali»:  $PC_1 > PC_2 > PC_3 \dots > PC_k$

dott. Andrea Mistarò, SOS Laboratorio AMCOA (Trieste) 14



**ESEMPIO: se Mendeleev avesse avuto la PCA...**

	Tfus	Teb	dens	STOX = com.	e.n.
H	14	20	90	1	2.2
Li	454	1615	534	1	0.98
Na	371	1156	970	1	0.93
K	337	1032	860	1	0.82
Rb	313	961	1530	1	0.82
Cs	302	944	1870	1	0.79
Be	1550	3243	1800	2	1.57
Mg	924	1380	1741	2	1.31
Ca	1120	1760	1540	2	1
Sr	1042	1657	2600	2	0.95
Ba	1002	2078	3500	2	0.89
F	53.5	85	2	-1	3.98
Cl	172	239	3	-1	3.16
Br	296	332	3100	-1	2.96
I	387	457	4940	-1	2.66
He	0.9	4	0.2	0	0
Ne	25	27	0.8	0	0
Ar	84	87	1.7	0	0
Kr	117	121	3.5	0	0
Xe	161	166	5.5	0	0
Sc	1614	3104	2999	3	1.36
Ti	1943	3560	4510	3	1.54
V	2183	3682	6110	4	1.63
Cr	2130	2945	7190	3	1.86
Mn	1517	2370	7440	2	1.5
Fe	1808	3300	7870	2	1.8
Co	1765	3170	8900	2	1.8
Ni	1726	3005	8900	2	1.8
Cu	1356	2868	8930	2	1.9
Zn	693	1180	7140	2	1.6
Bj	544	1837	9780	3	2.02
Pb	601	2022	11340	2	1.8
Tl	577	1746	11850	3	1.62

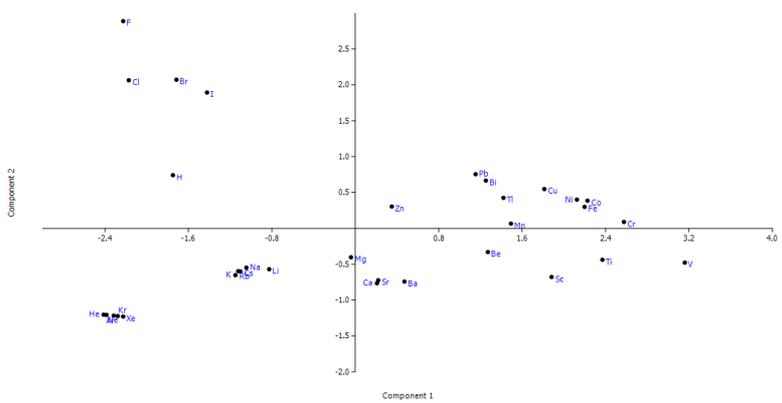


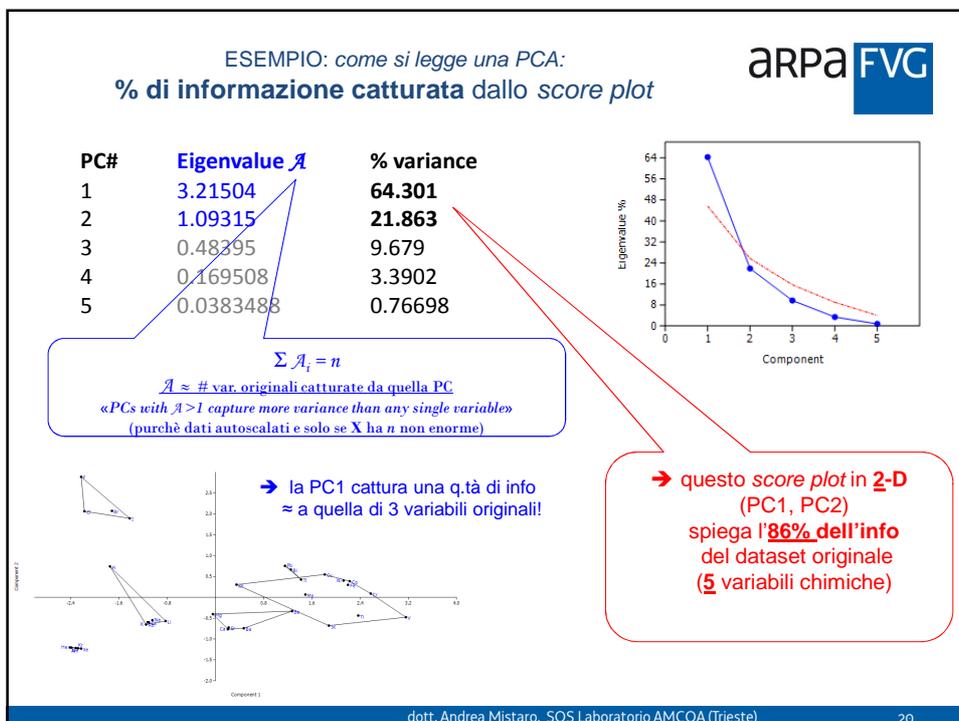
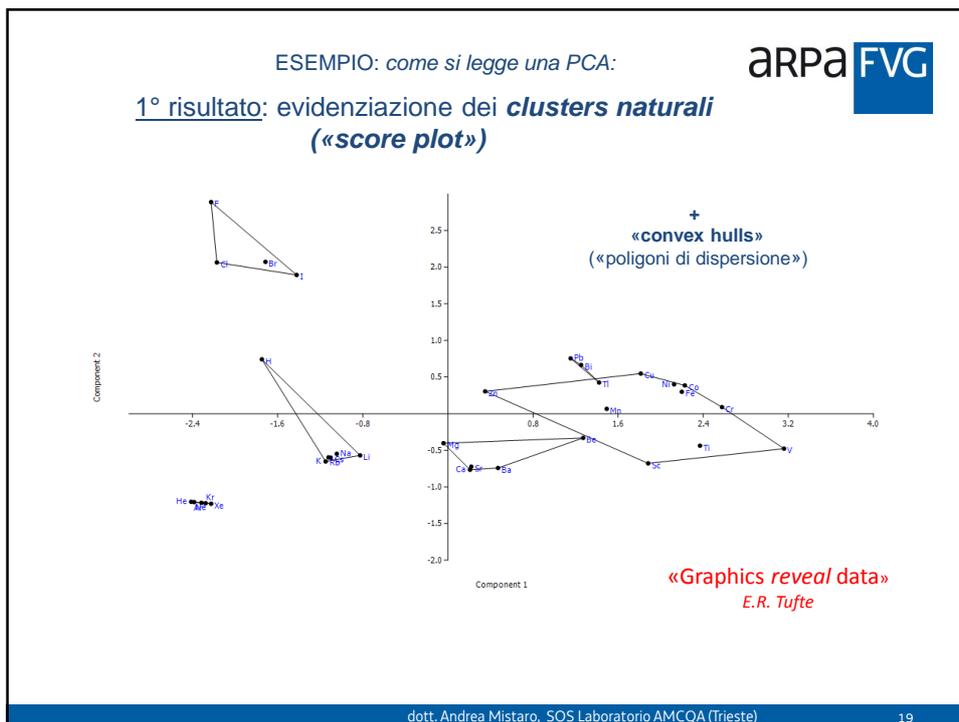
Anordnung der Elemente nach Gruppen und Reihen.

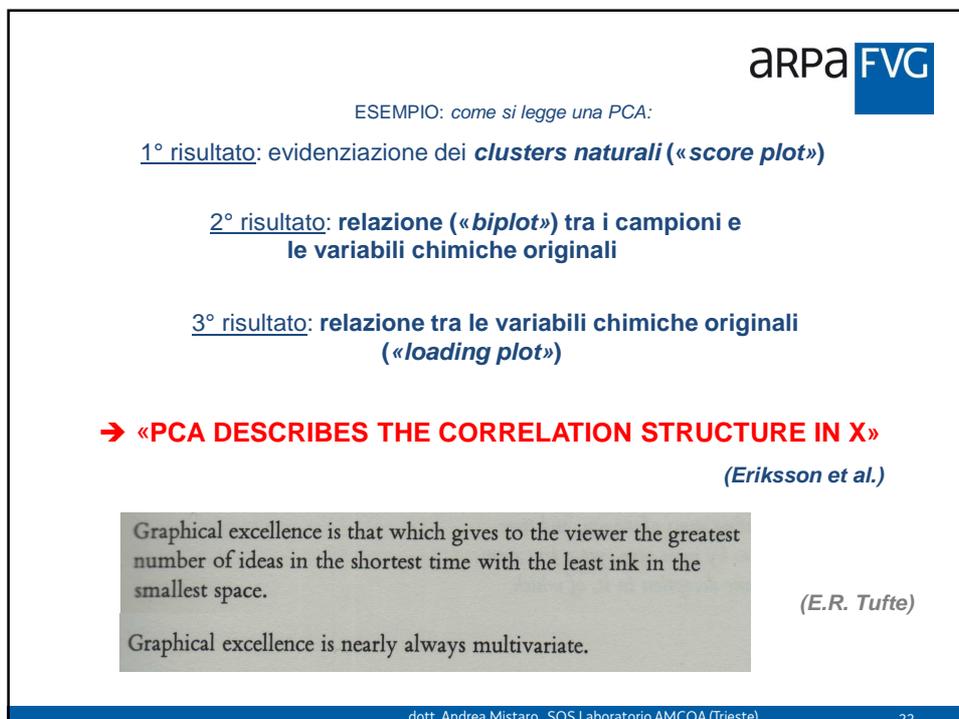
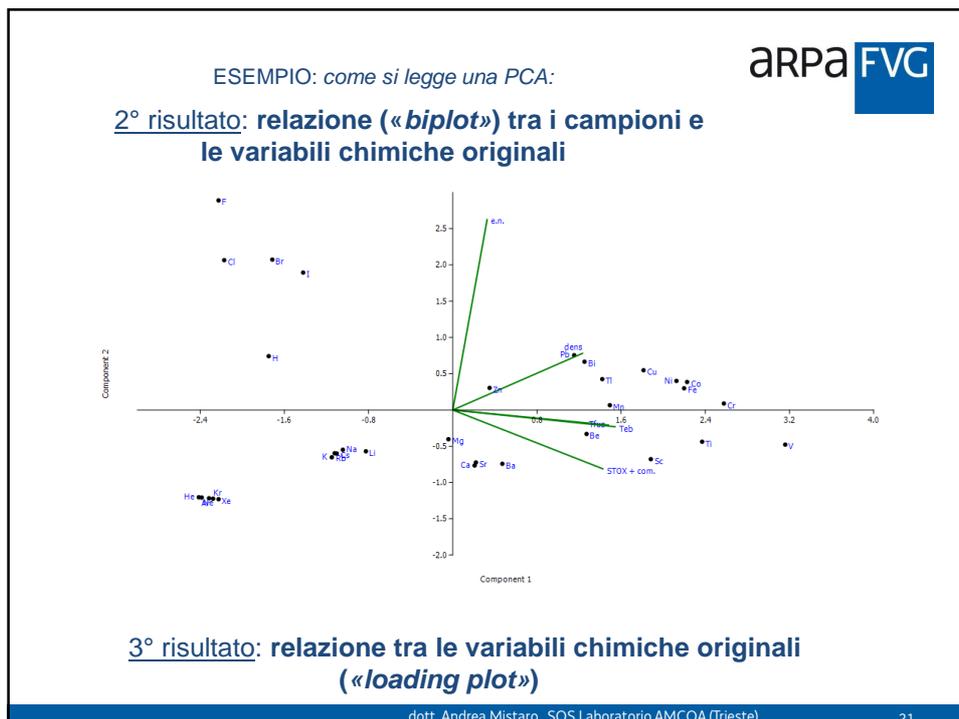
GRUPPE:	I	II	III	IV	V	VI	VII	VIII	IX
Reihe 1	H								
» 2	Li	Be	B	C	N	O	F		
» 3	Na	Mg	Al	Si	P	S	Cl		
» 4	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co
» 5	(Rb)	Zn	Ga	Ge	As	Se	Br		
» 6	Rb	Sr	Y	Zr	Nb	Mo		Ra	Rh
» 7	(Ag)	Cd	In	Sn	Sb	Te	J		
» 8	Cu	Ba	La	Ce	Pr				
» 9									
» 10				Yb		Ta	W		
» 11	(Au)	Hg	Tl	Pb	Bi				
» 12				Tb		U			

tratta da "Grundlagen der Chemie" di Dmitrij I. Mendeleev (St. Petersburg, 1892)

**ESEMPIO: come si legge una PCA: lo «SCORES PLOT»**

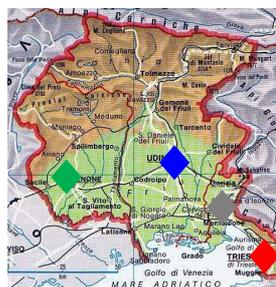






## L'ANALISI MULTIVARIATA NELLA CARATTERIZZAZIONE CHIMICA DEL PM:

Esempi FVG



dott. Andrea Mistaro, SOS Laboratorio AMCQA (Trieste)

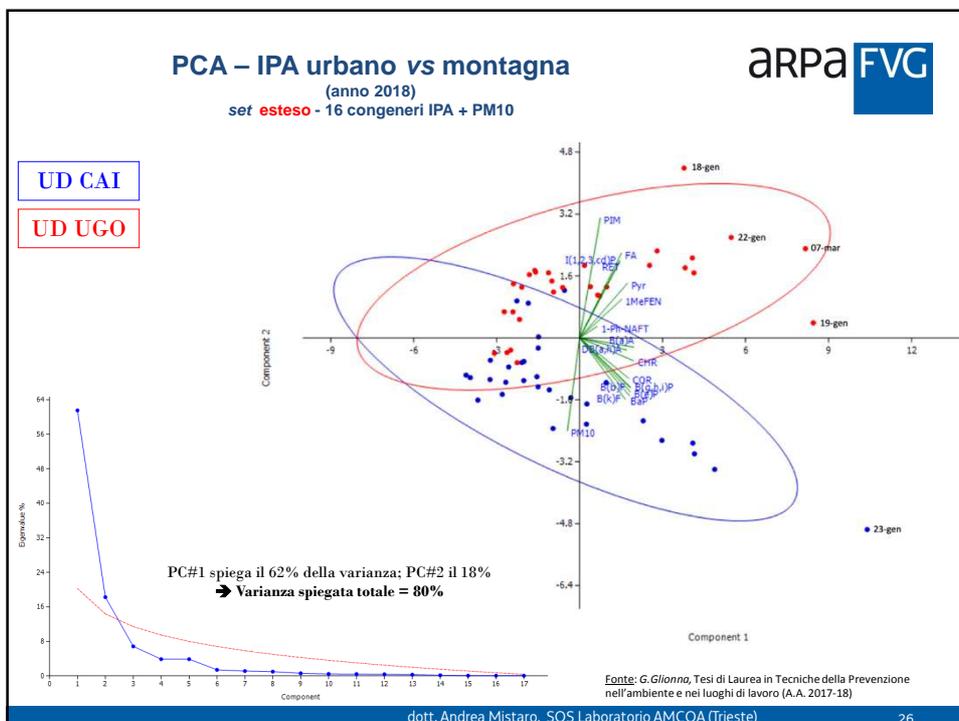
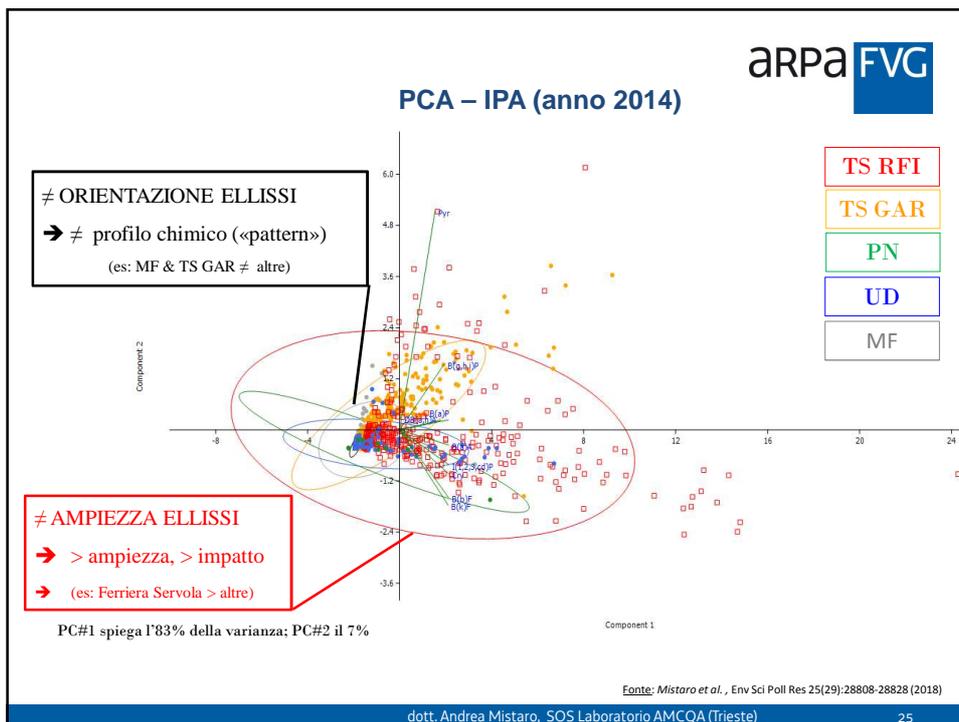
23

### ESEMPI DA DATI ARPA FVG

I  
—  
IPA

dott. Andrea Mistaro, SOS Laboratorio AMCQA (Trieste)

24



**ARPA FVG**

**ESEMPI DA DATI ARPA FVG**

**II**

**METALLI E SEMIMETALLI**

dott. Andrea Mistaro, SOS Laboratorio AMCQA (Trieste) 27

**ARPA FVG**

**Analisi multivariata (PCA)**  
4 → 16 metalli normati

→ ORIENTAZIONE ELLISSI ≠:

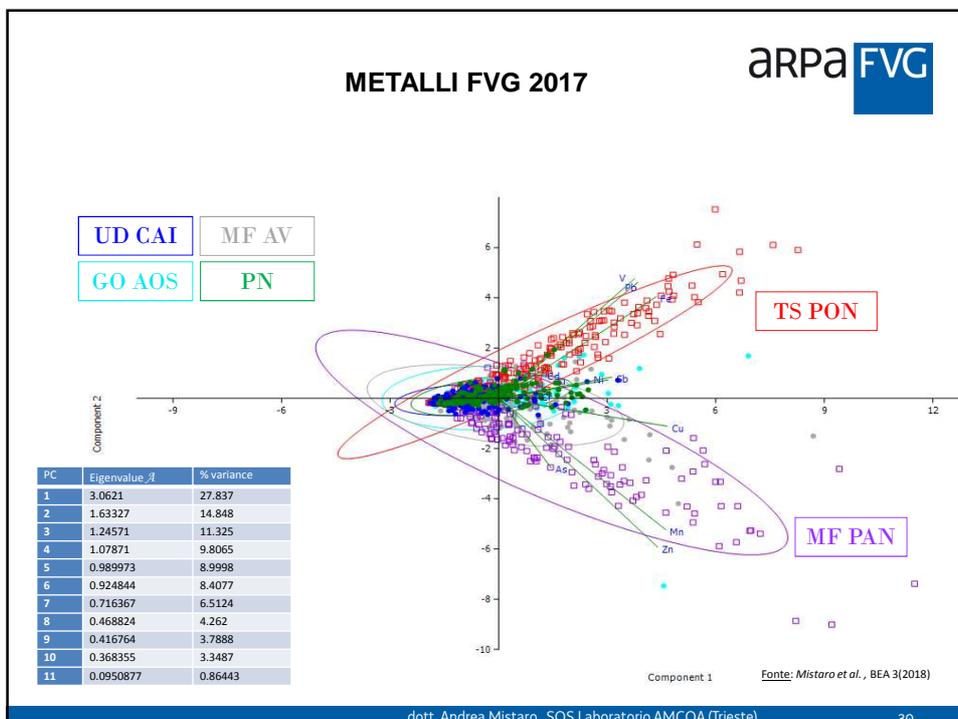
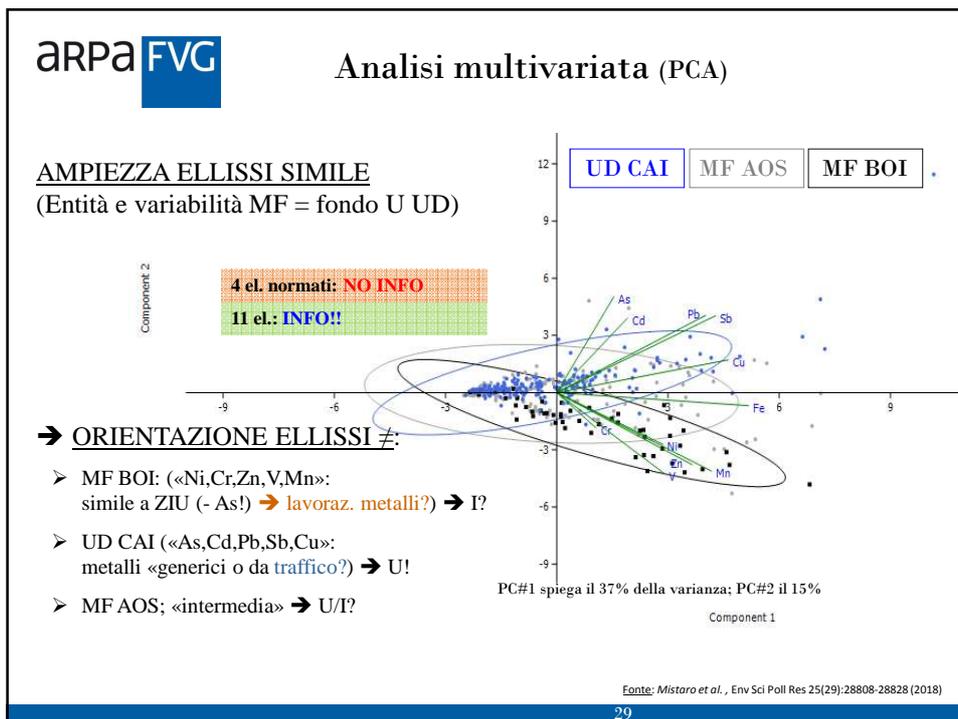
- ZIU («As, Ni»)
- ZIRO («Pb, Cd»)

UD CAI MF  
PN TS CAR

**Variabilità del biplot dominata dalle 2 (!) stazioni c/o le 2 acciaierie UD ZIU e ZIRO (staz. urbane non visibili)** →

Fonte: Mistaro et al., Env Sci Poll Res 25(29):28808-28828 (2018)

dott. Andrea Mistaro, SOS Laboratorio AMCQA (Trieste) 28

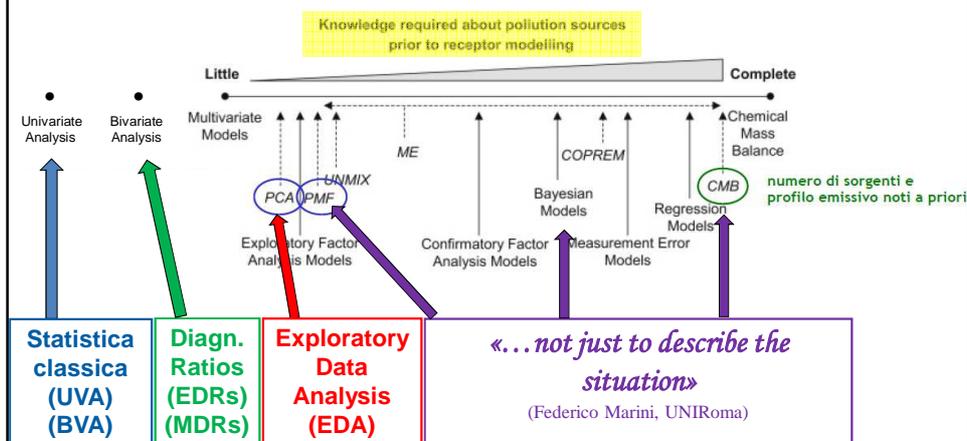


...oltre la PCA

ALTRE TECNICHE DI ANALISI MULTIVARIATA

- PCA: esplorativa; *unsupervised* → scopre clusters naturali
- CA: esplorativa; *unsupervised* → clusterizza i campioni; + similarità
- CLASS: predittiva; *supervised* → classifica i campioni in classi note dalla EDA

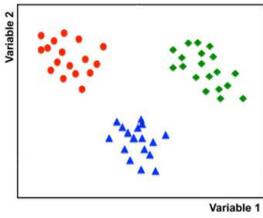
ALTRE TECNICHE AMV



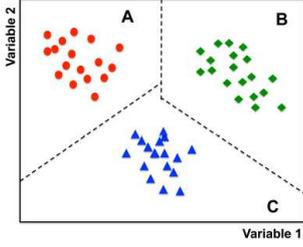
### TECNICHE AMV

1) dalla PCA...      ...alla 2) CLASSIFICAZIONE





Variable 2  
Variable 1



Variable 2  
Variable 1

*"To find a criterion to assign an object (sample) to one class (defined a priori) based on a set of measurements performed on the object itself"*

**DISCRIMINATION** (es. LDA)...  
*Try to define the **boundaries** which separate the classes in the n-D spaces (1936)*

- Ogni oggetto è assegnato ad una classe (perchè classi "aperte")

**...vs MODELING** (es. SIMCA)  
*Each **category** is modeled (1976)*

- Ogni oggetto può essere assegnato a una, classe, a più d'una, o a nessuna (perchè classi "chiuse")

Fonte: F. Marini, "Classification: a gentle introduction", CMA4CH 2016

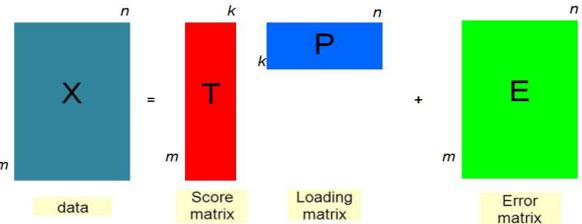
dott. Andrea Mataro, SOS Laboratorio AMCQA (Trieste)

33

### TECNICHE AMV

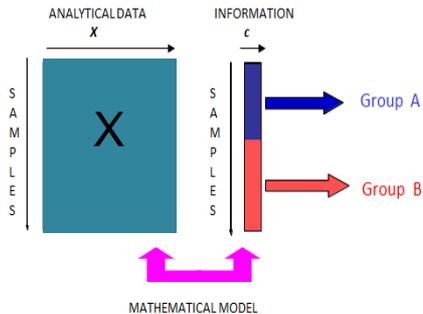
1) dalla PCA...      ...alla 2) CLASSIFICAZIONE





*"Exploratory [unsupervised] methods [PCA] see whether there are groupings"*

*"Supervised methods [classification, class modeling] assign samples into these groups"*



ANALYTICAL DATA      INFORMATION

X      c

S A M P L E S      S A M P L E S

Group A  
Group B

MATHEMATICAL MODEL

Fonte: R. Brereton, "Chemometrics for Pattern Recognition", Wiley (2009)

dott. Andrea Mataro

**2b) CLASSIFICAZIONE – SIMCA**  
 (Soft\* Independent Modeling by Class Analogy)  
 (Svante Wold, 1976)

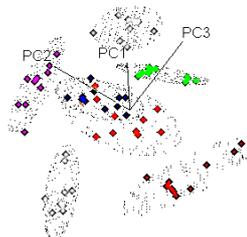
Tecnica non probabilistica, ma basata sulle PC dei dati appartenenti a ciascuna classe



\*nessuna assunzione è necessaria sulle caratteristiche della distribuzione

Specific models of different classes are made by applying PCA on each class. At the end, each class will be modelled by a specific PC-based model which will delimit it in a region of space [a sub-space].

A new individual will be affiliated into class *k* if its distance to corresponding region *k* is lower than those to other regions (classes).

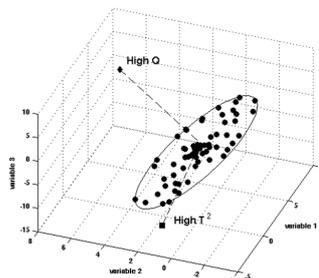
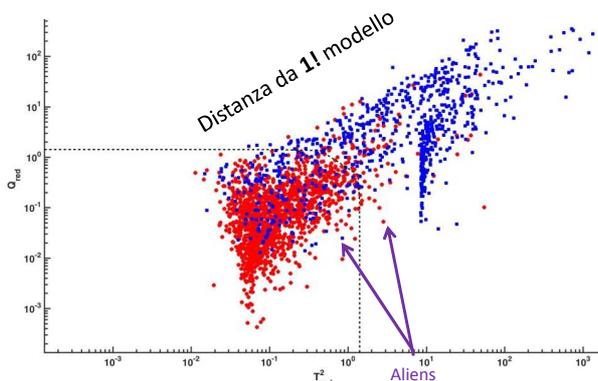


Foods2016, 5(4), 77; doi:10.3390/foods5040077

Fonte: <https://infomatrix.com/chemometrics/methods/>



OUTPUT SIMCA (1):



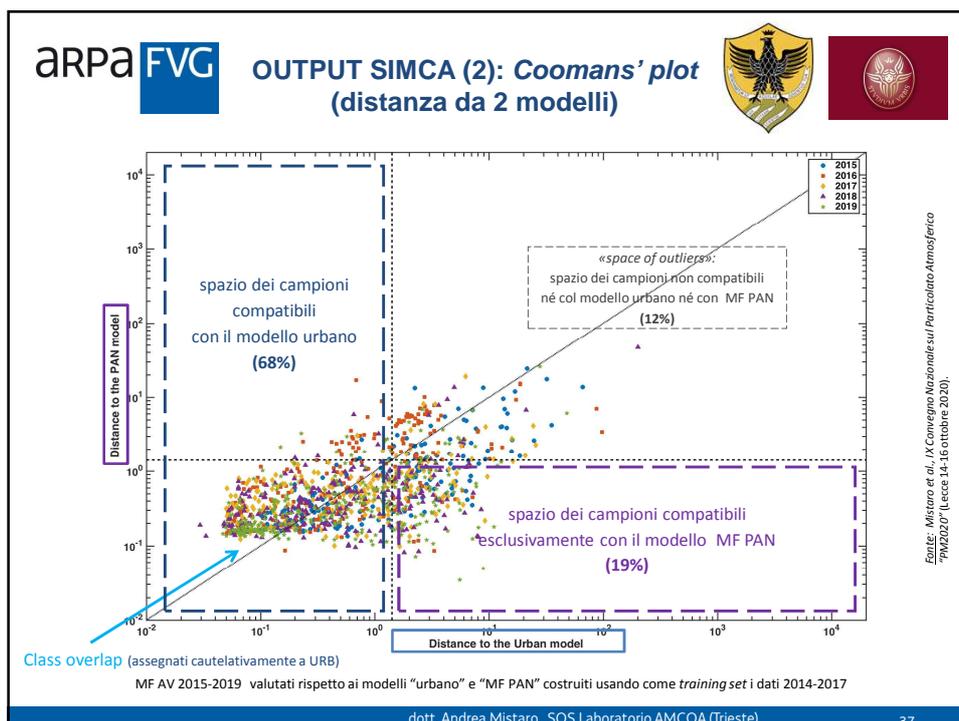
Fonte: F. Marini, "Classification: a gentle introduction", CMA4CH 2016

Class Modeling (SIMCA) del particolato atmosferico della regione Friuli Venezia Giulia in base al contenuto di 11 metalli nel PM10

Andrea Mistaro<sup>1,\*</sup>, Alessandra Biancolillo<sup>2</sup>, Anna Abatangelo<sup>3</sup>, Tazio Asquini<sup>1</sup>, Flavio Moimas<sup>3</sup>, Clorinda Del Bianco<sup>4</sup>, Federico Marini<sup>5</sup>

<sup>1,3,4</sup> ARPA FVG - <sup>2</sup> Università dell'Aquila - <sup>5</sup> Università di Roma La Sapienza,

Fonte: Mistaro et al., IX Convegno Nazionale sul Particolato Atmosferico "PM2020" (Lecce 14-16 ottobre 2020).



### «GRADO DI URBANITÀ» delle stazioni FVG

stazioni valutate rispetto ai modelli costruiti usando come training set i dati 2014-2017

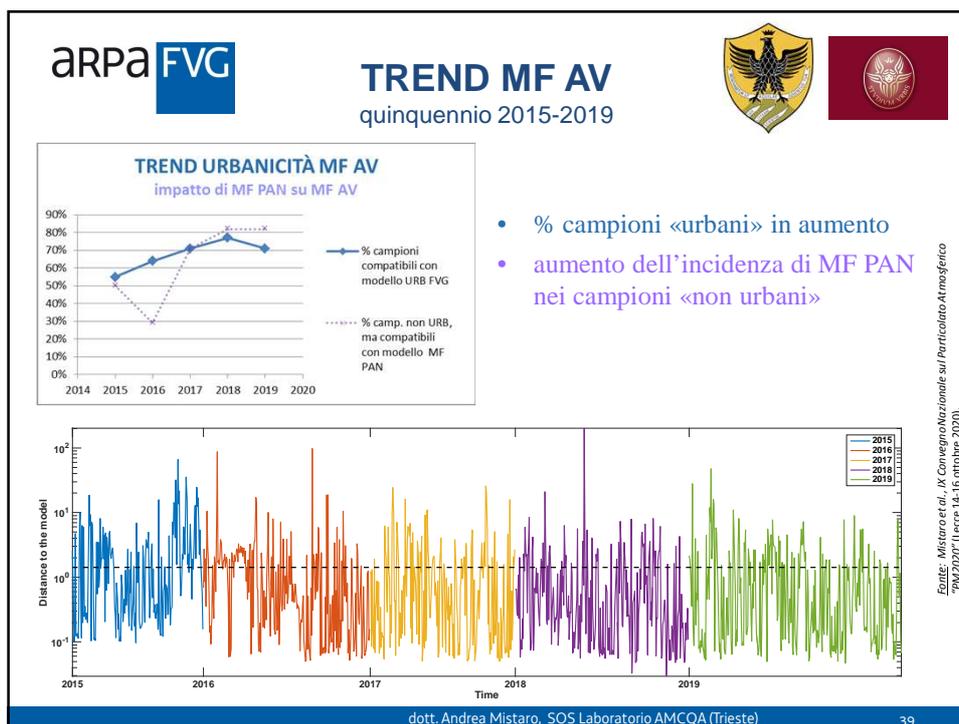
STAZIONE DA CLASSIFICARE:	% di campioni accettati / scartati dal modello "URB FVG 11-var"	% dei campioni scartati dal modello "URB FVG 11-var" considerata compatibile con ciascun modello industriale		
		UD ZIRO	MF PAN	TS PON
MF AOS (11-v.) (anno 2014)	90% / 10%	11%	72%	solo 10 metalli nel pacchetto analitico → modello a 11 variabili non applicabile
MF AV (11-v.) (anni 2015-2019)	68% / 32%	37%	66%	
MF BOI (11-v.) (estate 2014)	51% / 49%	31%	85%	
TS CAR (4-v.) (anni 2014-2017)	69% / 31%	10%	1%	

→ «grado di urbanità» delle stazioni MF: AOS (90%) > AV (68%) > BOI (51%)  
(altre stazioni regionali confermate urbane: GO AOS 86%; PN PNC 81%)

→ i campioni scartati dal modello urbano risultano maggiormente compatibili con la sorgente industriale più vicina: PAN (cantier navale) per le due stazioni MF (48-85%); PON (Ferriera) per la stazione TS CAR (75%)

Fonte: Mistarò et al., IX Convegno Nazionale sul Particolato Atmosferico "PM2020" (Lecce 14-16 ottobre 2020).

dott. Andrea Mistarò, SOS Laboratorio AMCQA (Trieste) 38



arpa FVG

## CONCLUSIONI (metodo)

- **PCA:** permette di *esplorare i dati per distinguere ed individuare rapidamente le peculiarità e la tipologia delle diverse stazioni di misura*
- **Classificazione:** permette di *classificare ciascun campione come compatibile o meno con alcuni modelli di tipologia di inquinamento (urbano vs diversi modelli industriali)*
- **PMF:** permette il **source apportionment completo (individuazione del contributo % al PM10 di ciascuna fonte)**

dott. Andrea Mistaro, SOS Laboratorio AMCQA (Trieste) 40

arpa FVG

GRAZIE PER L'ATTENZIONE  
andrea.mistaro@arpa.fvg.it



«As scientist, we must understand that data are dangerous»  
(R. Brereton)

Trieste, piazza Unità d'Italia e le «Rive»

dott. Andrea Mistaro, SOS Laboratorio AMCQA (Trieste) 41

arpa FVG

## BIBLIOGRAFIA (teoria)

- Richard G. Brereton: *Applied Chemometrics for scientists* (Wiley, 2007)
- Eriksson, Byrne et al.: *Multi- and Megavariate Data Analysis – Basic principles and applications* (3<sup>rd</sup> rev. ed., Umetrics 2013 )
- Barry M. Wise: *Chemometrics I: Principal Components and Exploratory Data Analysis* (Eigenvector Research, 1996-2012)
- Miller J.N., Miller J.C.: *Statistics and chemometrics for analytical chemistry*, (Prentice Hall – Pearson Education, 2010)
- Edward R. Tufte: *The visual display of quantitative information* (Graphic Press, 2001 - 2<sup>nd</sup> edition 5<sup>th</sup> printing, 2007)

**BIBLIOGRAFIA**  
(esempi ARPA FVG)

- A. Mistaro, *L'analisi multivariata nella caratterizzazione chimica del particolato atmosferico: approccio generale e risultati dei casi di studio in Friuli Venezia Giulia*. Bollettino degli esperti ambientali (BEA) (settembre 2018) N.3/2018 (Elpo Ed., Como)
- A. Mistaro et al., *Chemical Characterization of Atmospheric Particulate Matter in Friuli Venezia Giulia (NE Italy) by Exploratory Data Analysis with Multisite and Multivariate Approach*. Environmental Science and Pollution Research 25(29):28808-28828 (Springer 2018)
- G. Glionna, *Valutazione degli idrocarburi policiclici aromatici alchilati di origine terpenica come potenziali traccianti di combustione di legna nel PM10*, Tesi di Laurea in Tecniche della prevenzione nell'ambiente e nei luoghi di lavoro, relatore F. Dugulin, correlatore A. Mistaro (Università degli studi di Trieste, A.A. 2017-18)
- A. Mistaro et al., *Class Modeling (SIMCA) del particolato atmosferico della regione Friuli Venezia Giulia in base al contenuto di 11 metalli nel PM10*, IX Convegno Nazionale sul Particolato Atmosferico "PM2020" (Lecce 2020); Workshop di Chemiometria della Divisione di Chimica Analitica della Società Chimica Italiana" (L'Aquila 2022)