Climathon Torino

chimica

L'interpretazione dei dati chimico-ambientali

Le potenzialità dell'Analisi Multivariata

Eugenio Alladio*, Silvia Berto, Mery Malandrino, Claudia Barolo, Davide Vione

Dipartimento di Chimica, Università degli Studi di Torino

**ealladio@unito.it*

Climate-KIC is supported by the EIT, a body of European Union

Torino, 20 ottobre 17 – HACKADEMY

DALL'UNIVARIATO AL MULTIVARIATO

UNIVARIATO

1 variabile per volta

MULTIVARIATO

Tutte le variabili simultaneamente

DALL'UNIVARIATO AL MULTIVARIATO - ESEMPIO

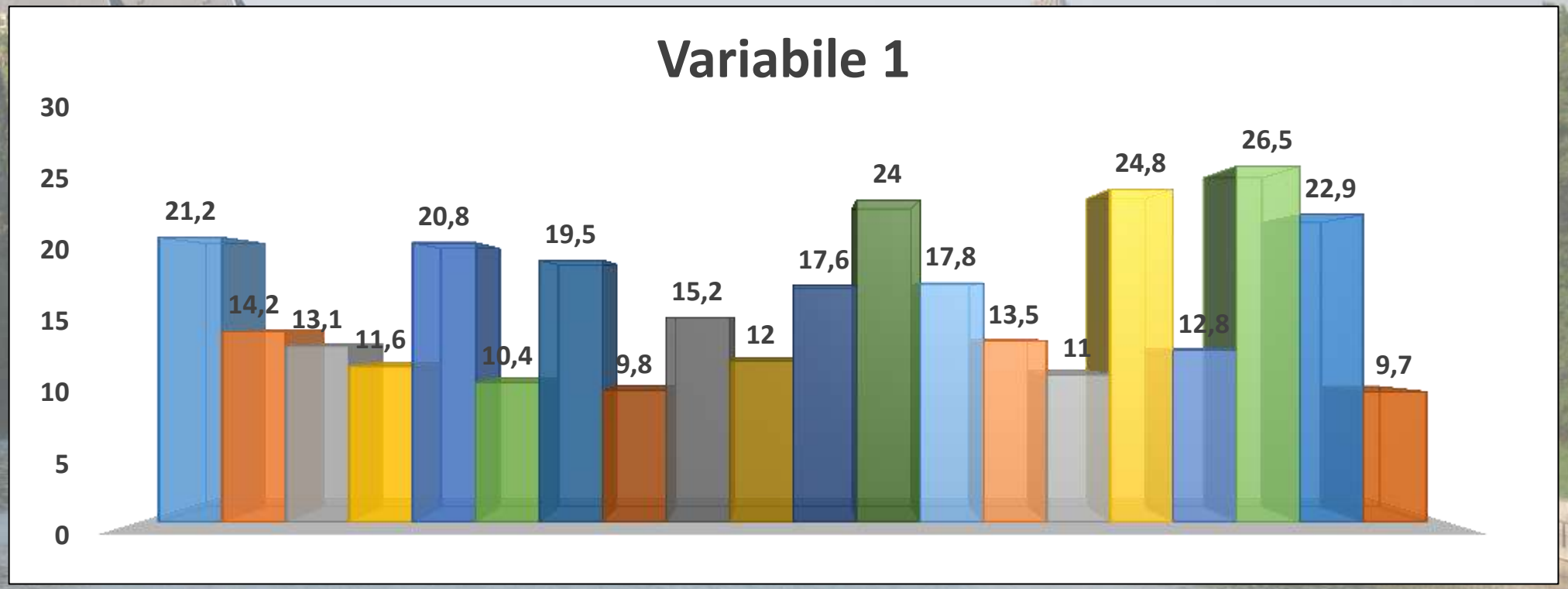
Oggetti	Variabile 1	Variabile 2
1	21.2	32.5
2	16.2	21
3	13.1	21.7
4	11.6	21.3
5	20.8	29.9
6	10.4	20.6
7	19.5	26.8
8	9.8	25.2
9	15.2	31.2
10	12	26
11	17.6	28.5
12	24	30
13	17.8	33.1
14	15	24
15	11	24.2
16	24.8	25.3
17	12.8	23.3
18	26.5	30.6
19	22.9	27.5
20	9.7	22.8

Cosa possiamo dire di questi dati?

IL MONDO UNIVARIATO

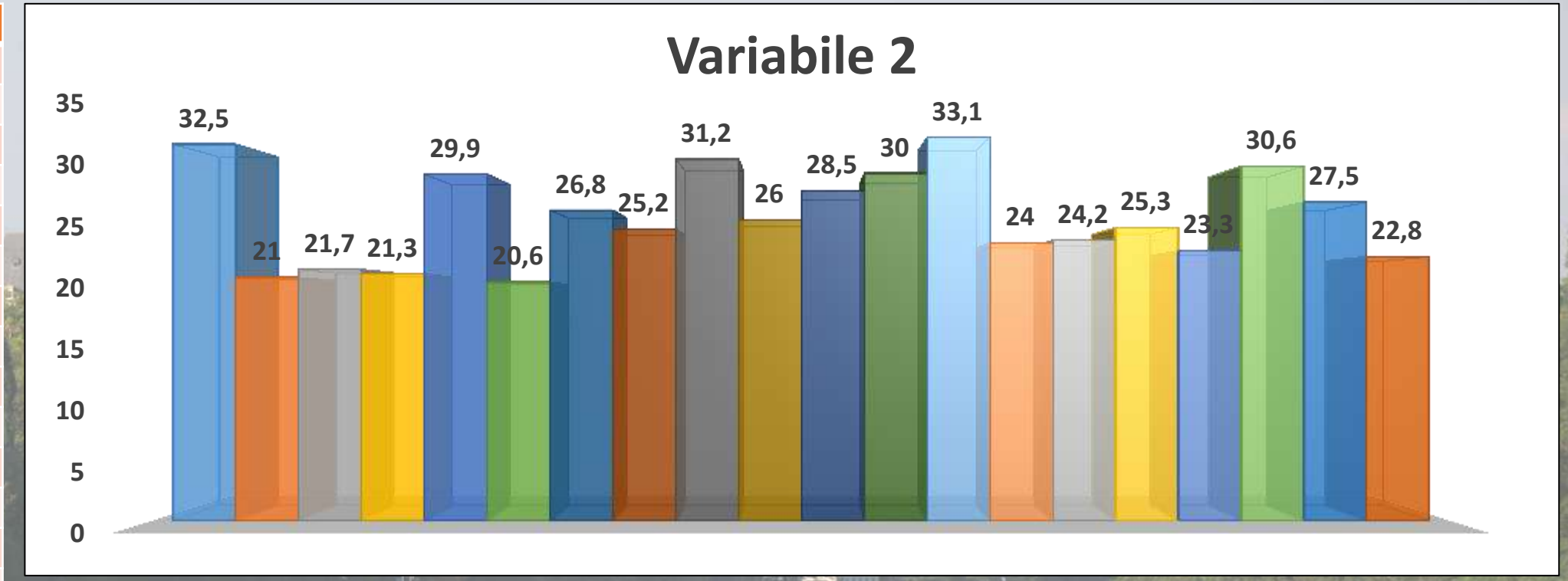
Oggetti	Variabile 1
1	21.2
2	16.2
3	13.1
4	11.6
5	20.8
6	10.4
7	19.5
8	9.8
9	15.2
10	12
11	17.6
12	24
13	17.8
14	15
15	11
16	24.8
17	12.8
18	26.5
19	22.9
20	9.7

1 variabile per volta



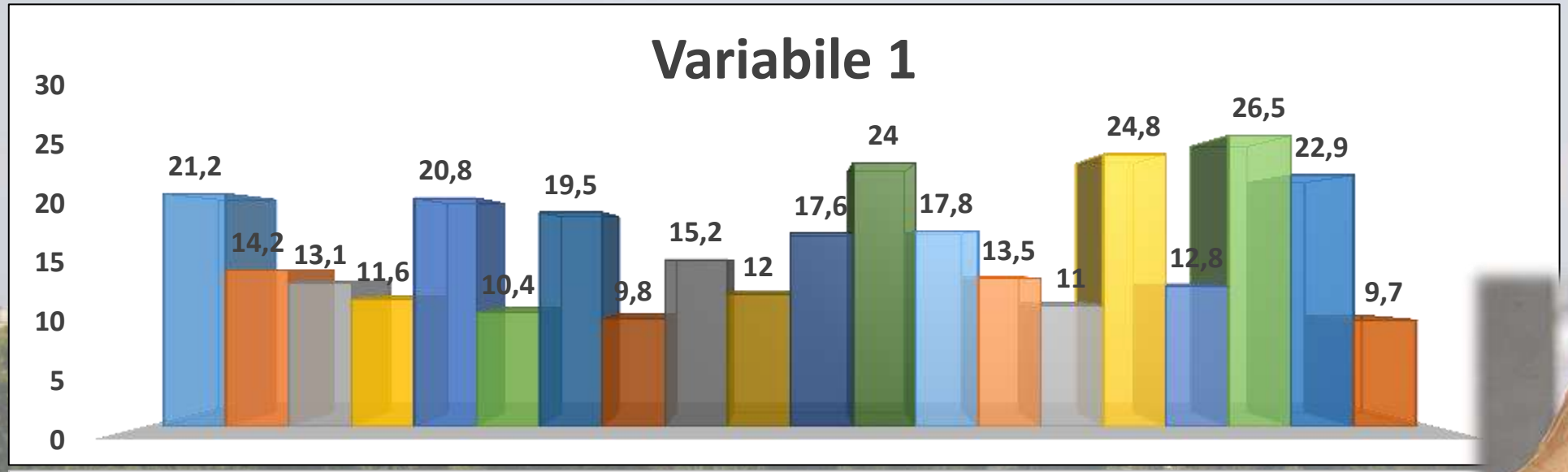
IL MONDO UNIVARIATO

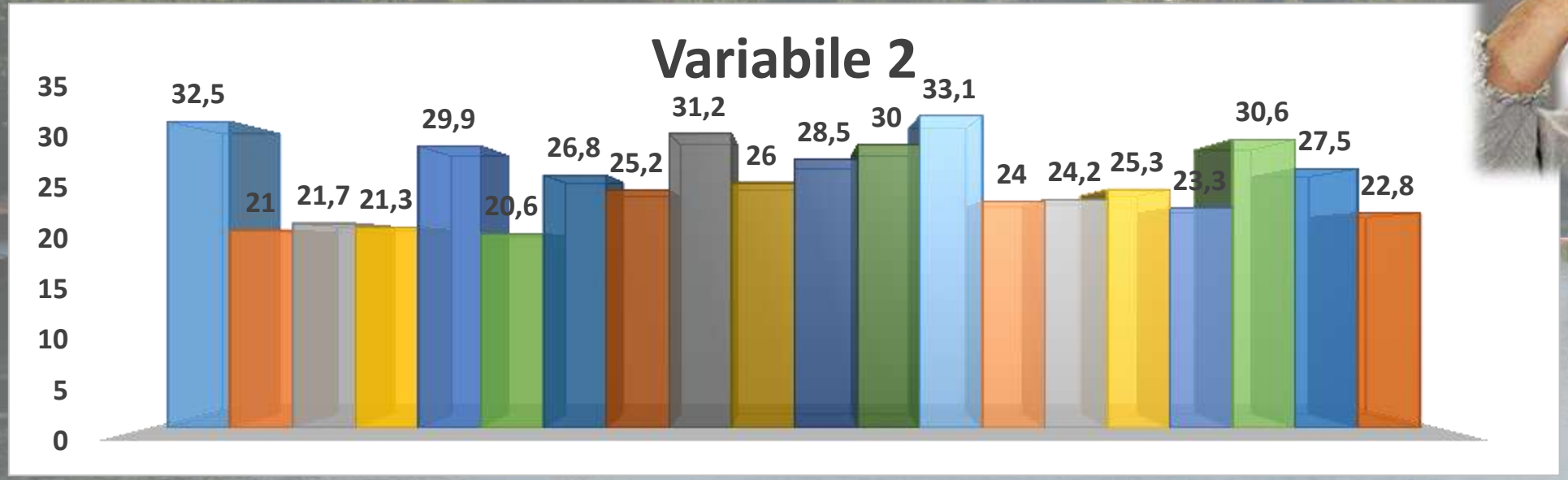
Oggetti	Variabile 2
1	32.5
2	21
3	21.7
4	21.3
5	29.9
6	20.6
7	26.8
8	25.2
9	31.2
10	26
11	28.5
12	30
13	33.1
14	24
15	24.2
16	25.3
17	23.3
18	30.6
19	27.5
20	22.8



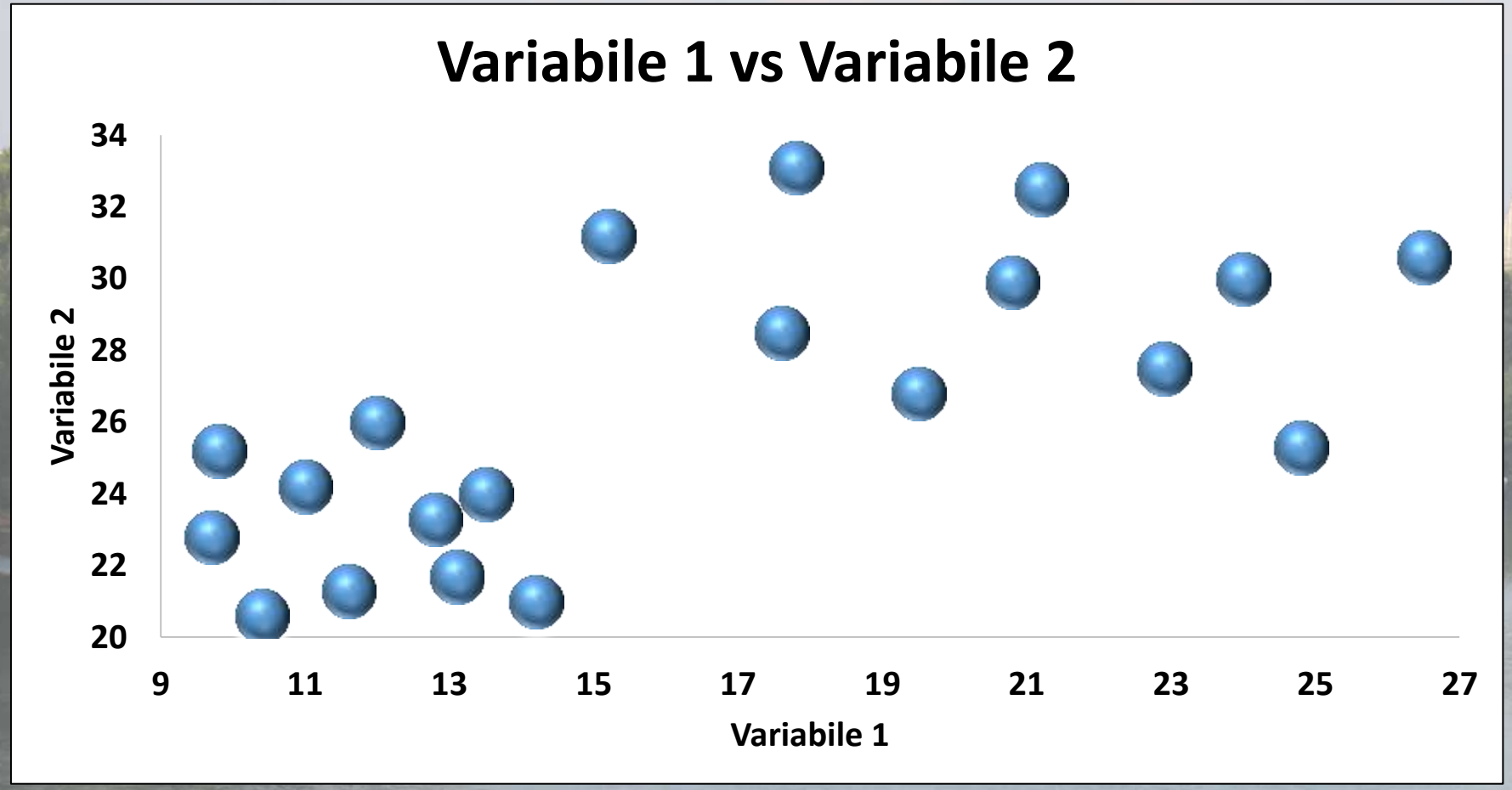
1 variabile per volta

IL MONDO UNIVARIATO





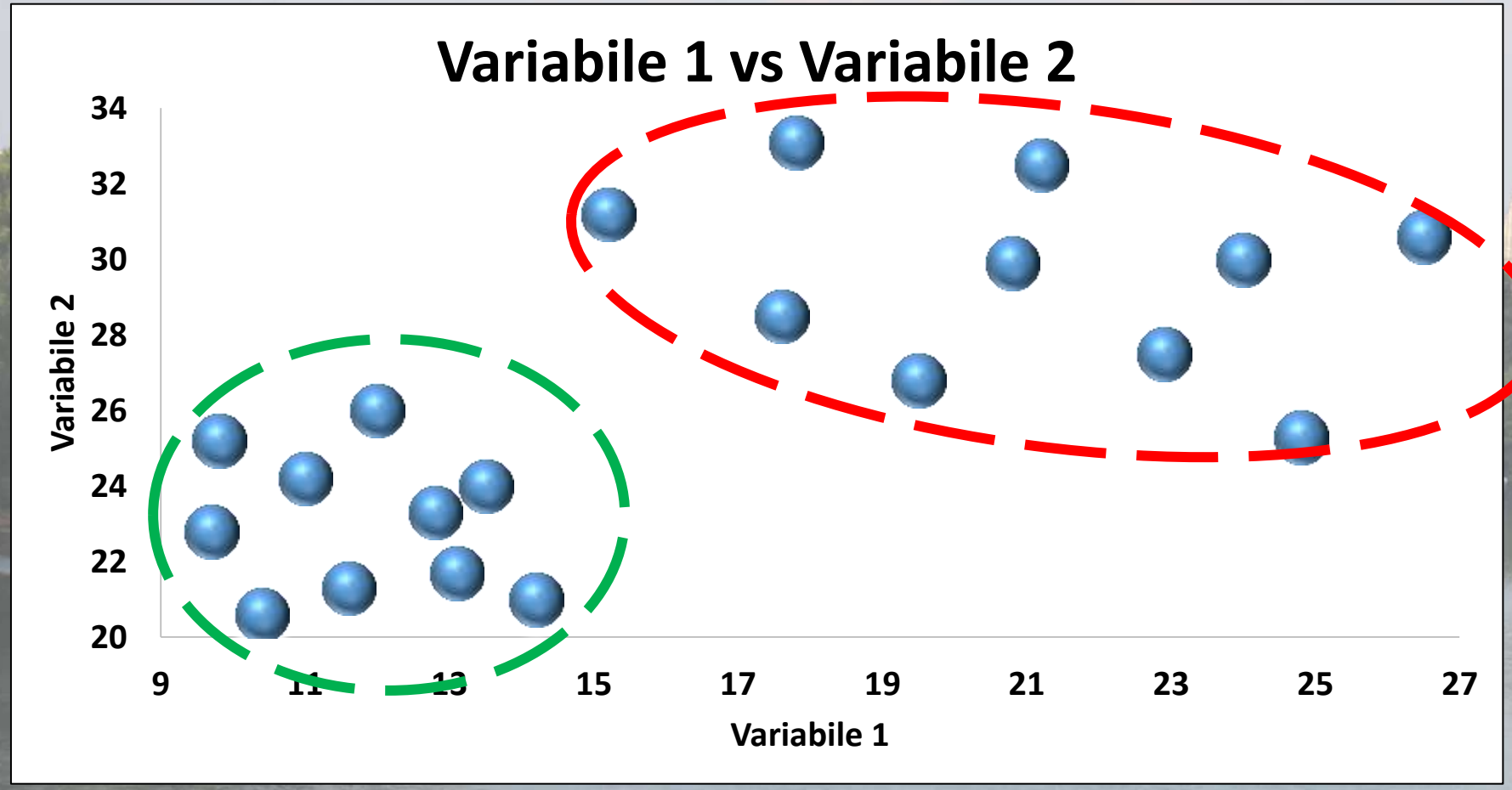
Valutiamo INSIEME le 2 variabili…



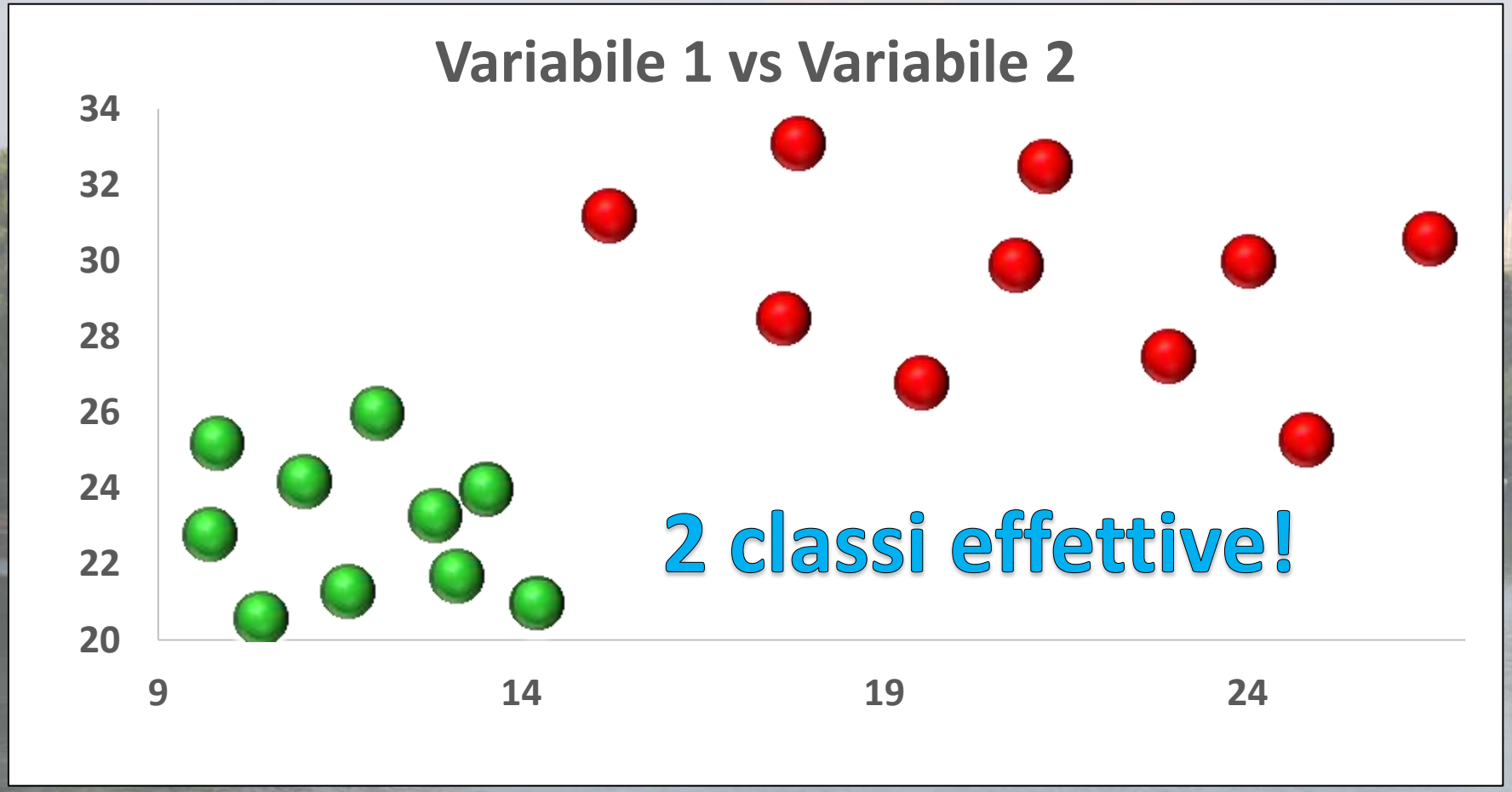
Differenze rispetto a univariato?

IL MONDO MULTIVARIATO

Si vedono due raggruppamenti!

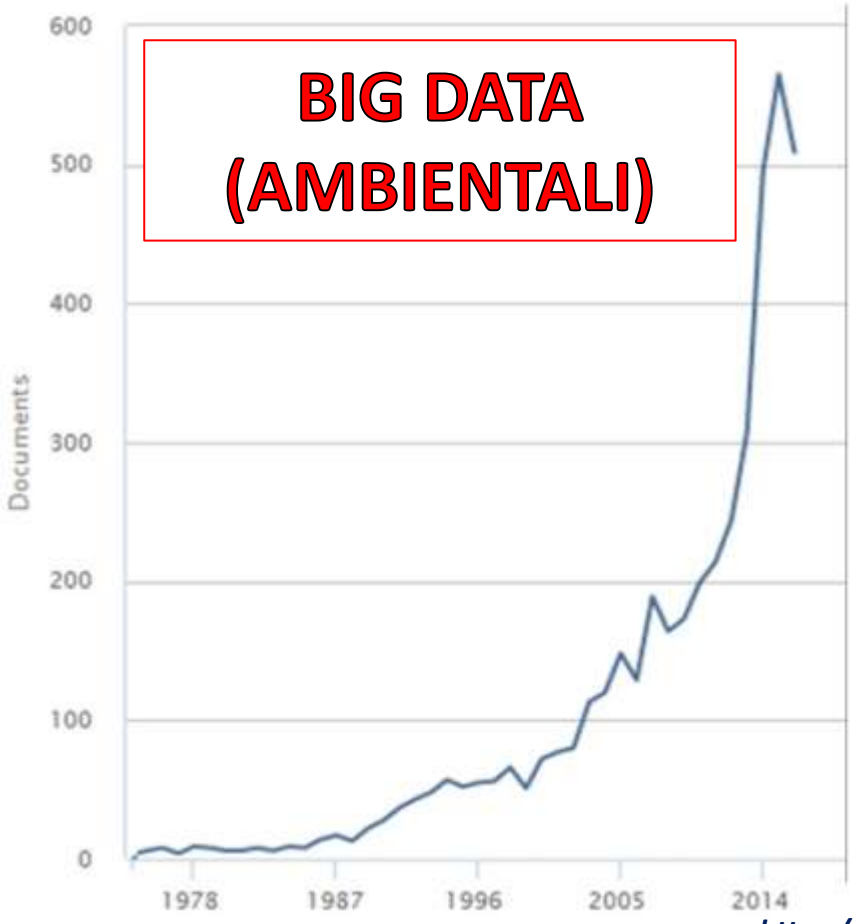


Si vedono due raggruppamenti!



BIG DATA & AMBIENTE... UN CRESCENTE INTERESSE

Nr. di articoli scientifici per anno



https://www.scopus.com

PERCHE' BIG DATA - MULTIVARIATO IN AMBIENTE?

3V-data

Velocità

Volume

Varietà

Continui progressi della strumentazione analitica

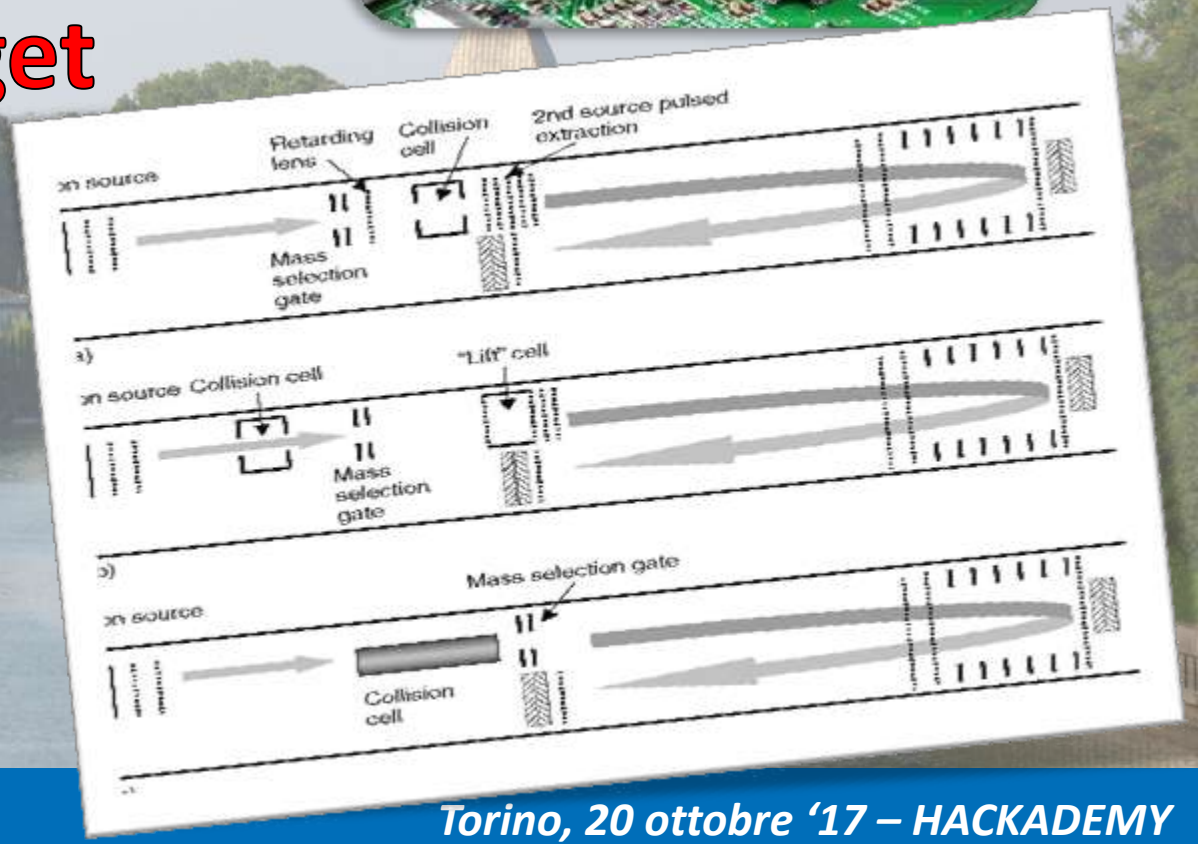
Analisi chimico – ambientali

VELOCITA'

Elettronica veloce

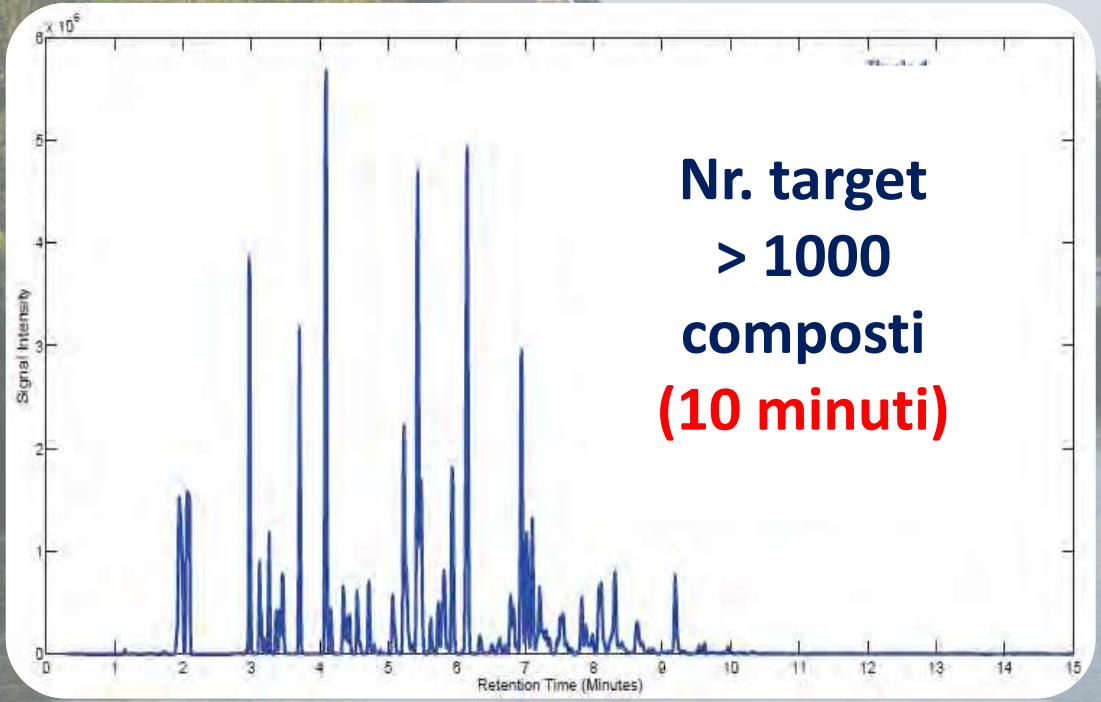
Tecniche analitiche multi-target

- Fast Chromatography
- Multiple Mass Spectrometry $(MS)^n$
- Time-of-Flight
- Orbitrap



VOLUME

Numero di composti



Big Data Storage

VARIETA'

VARIABILI

OGGETTI

2D

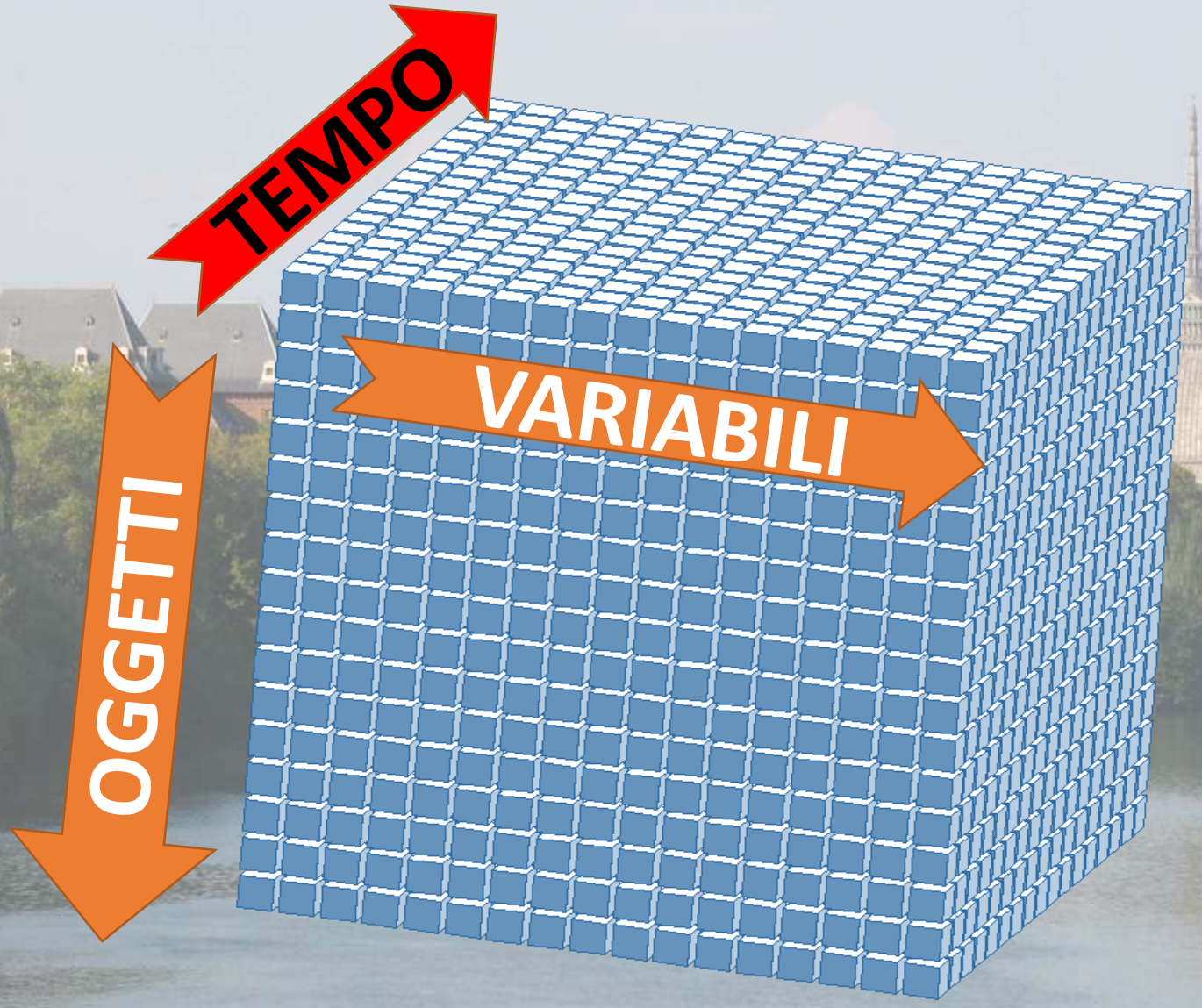
Viewer

Relation: final

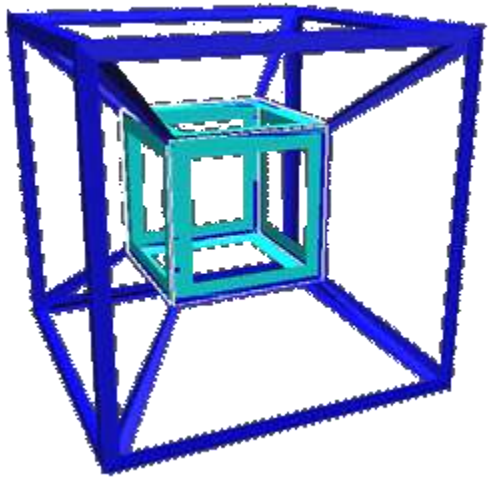
No.	Rating Numeric	Survey Numeric	Prize Numeric	Punishment Numeric	Aspen Numeric	Snowmass Numeric	Breckenridge Numeric	Keystone Numeric	ABasin Numeric	Loveland Nominal	CrestedButte Nominal	Vail Numeric	Silverton Numeric	WinterPark Numeric	Mary Jane Numeric	**Eldora** Numeric
1	0.675	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0		0.0	0.0	0.0
2	1.0	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
3	0.9	20.0	10.0	30.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0
4	0.95	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0
5	0.6	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0
6	0.95	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1-	0.0	0.0	0.0	0.0	0.0	0.0
7	1.0	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0
8	0.8	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	Q	1.0	1.0	1.0	1.0	1.0
9	0.9	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
10	0.85	20.0	10.0	30.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0
11	0.94	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0
12	1.0	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0
13	0.8	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	Q	1.0	1.0	1.0	1.0	1.0
14	1.0	20.0	10.0	30.0	0.0	0.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
15	0.95	20.0	10.0	30.0	1.0	0.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0
16	0.9	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0
17	0.85	20.0	10.0	30.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0
18	0.9	20.0	10.0	30.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	1.0	0.0
19	1.0	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
20	0.675	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0		0.0	0.0	0.0
21	0.575	20.0	1.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0
22	0.925	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
23	0.9	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	0.0	0.0	0.0	0.0
24	0.6	20.0	10.0	30.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0

Undo OK Cancel

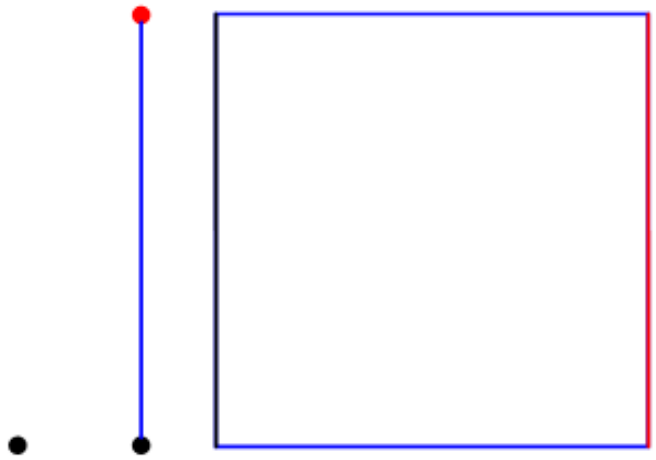
VARIETA'



VARIETA'

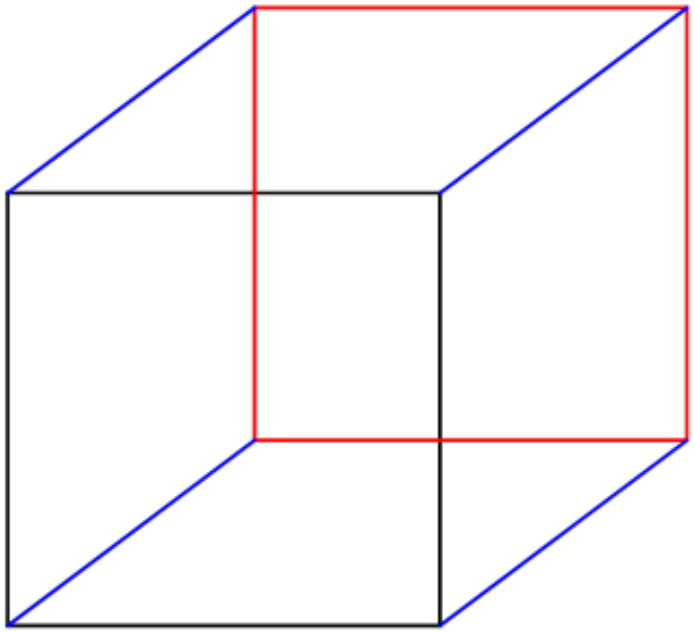


1. Oggetti
2. Variabili
3. Tempo di campionamento
4. **Luogo di campionamento**

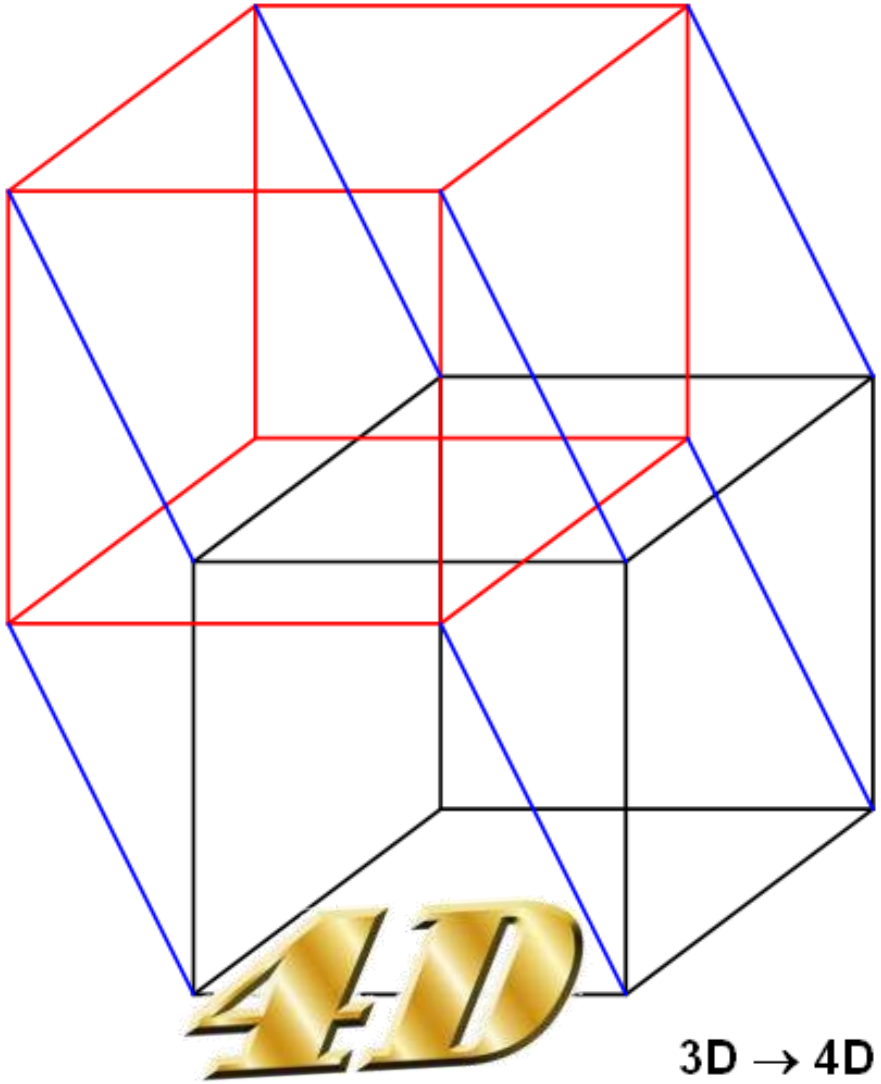


0D → 1D

1D → 2D

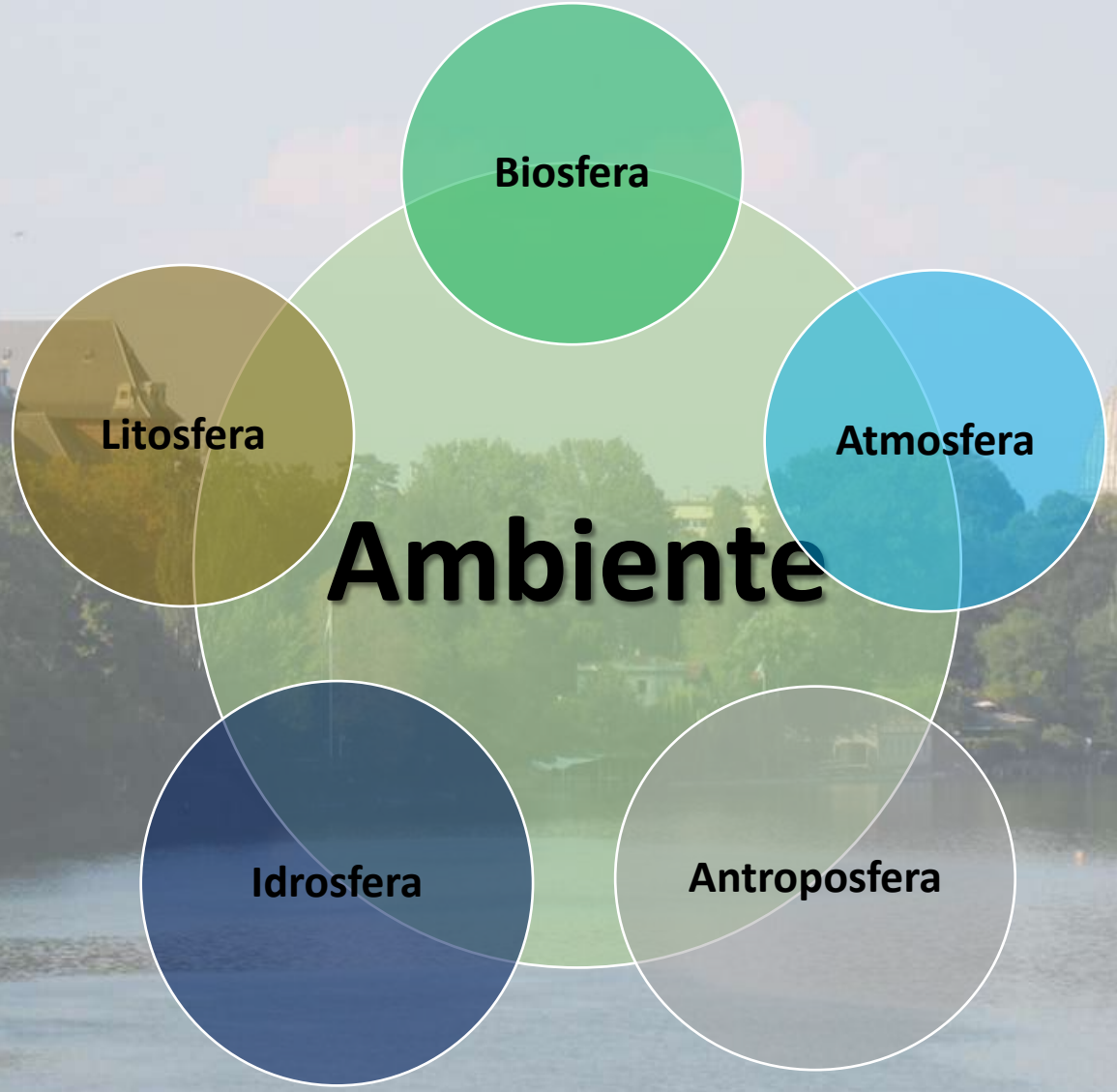


2D → 3D

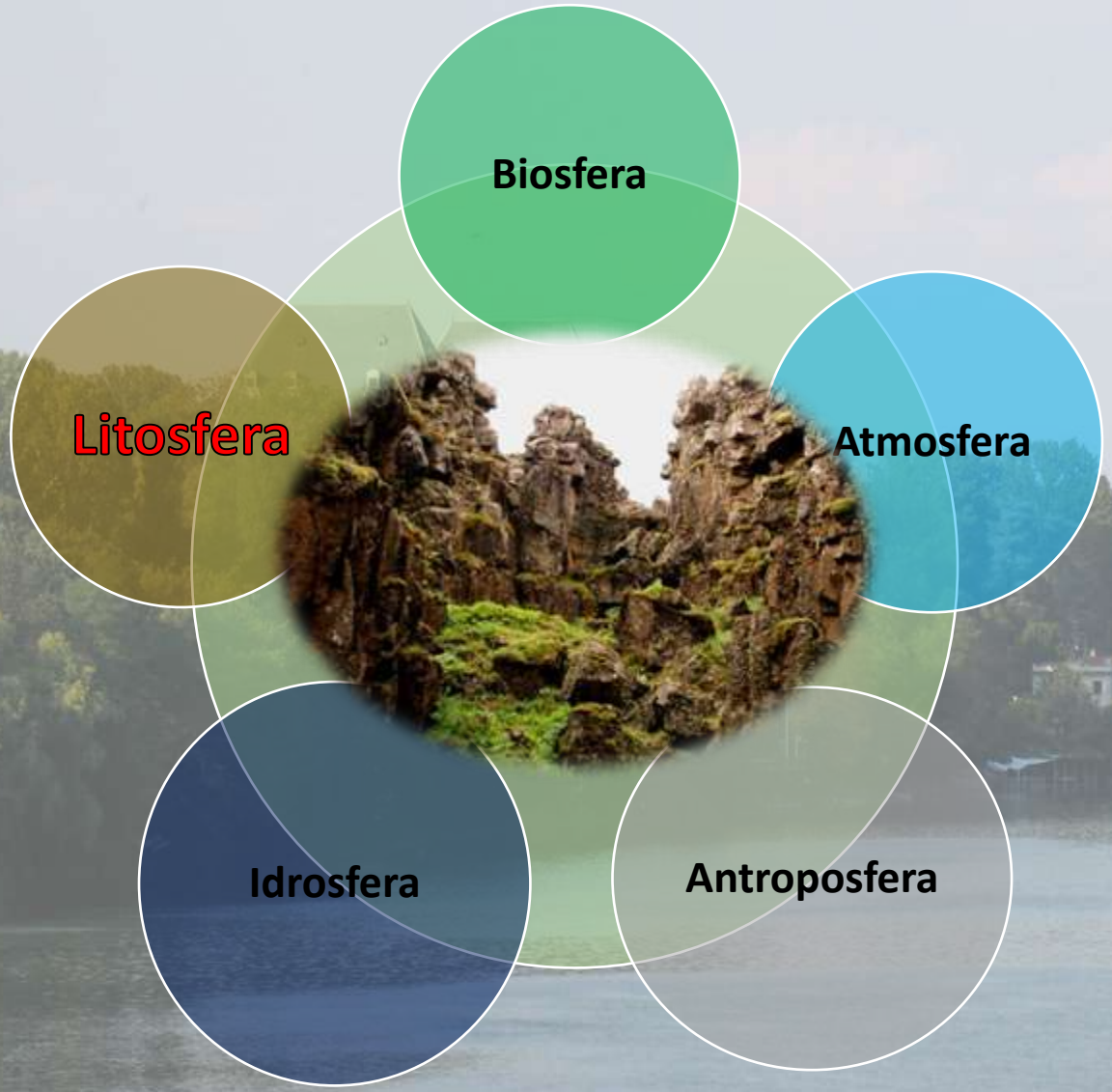


3D → 4D

MA QUALI DATI?



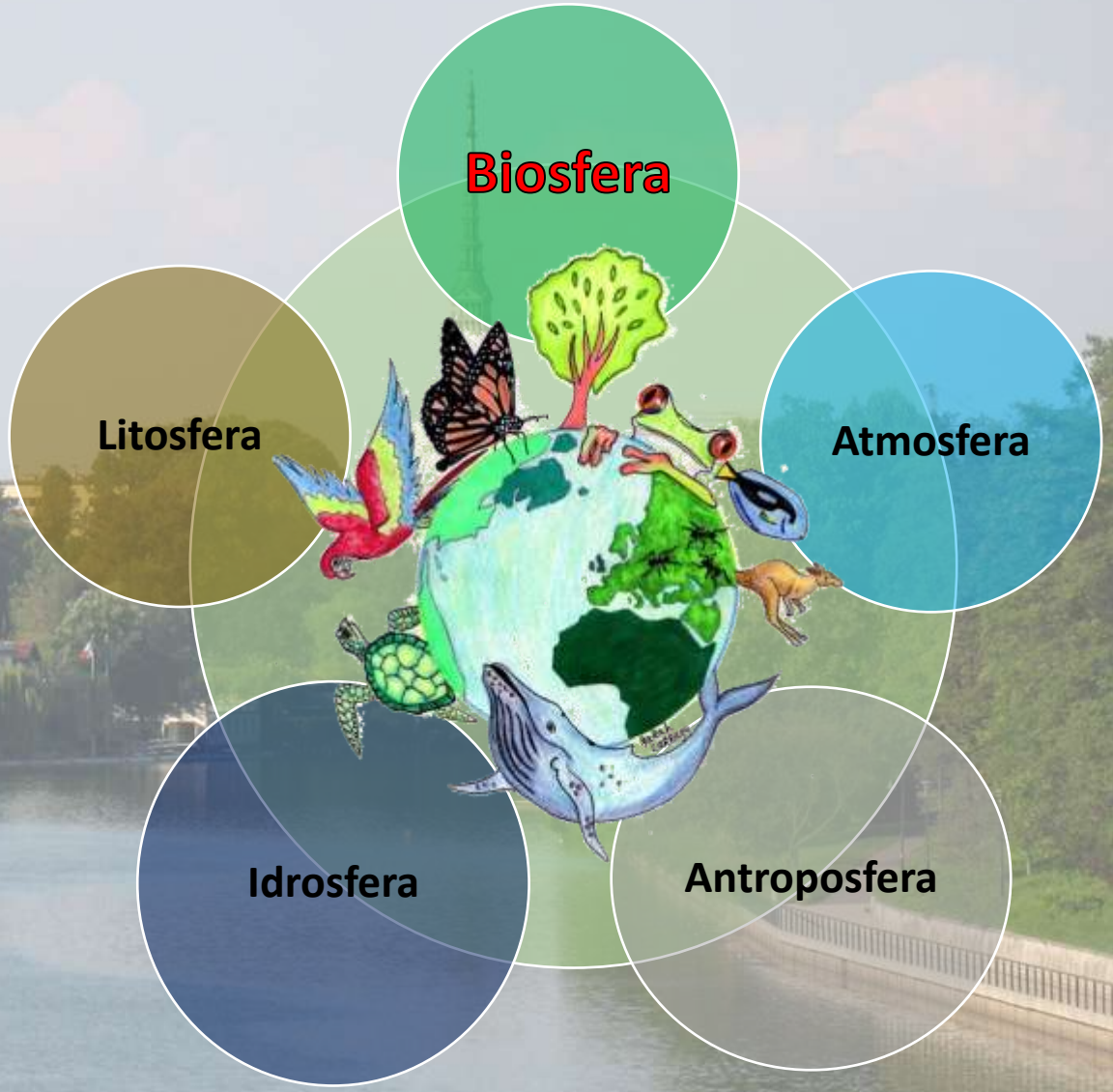
LITOSFERA



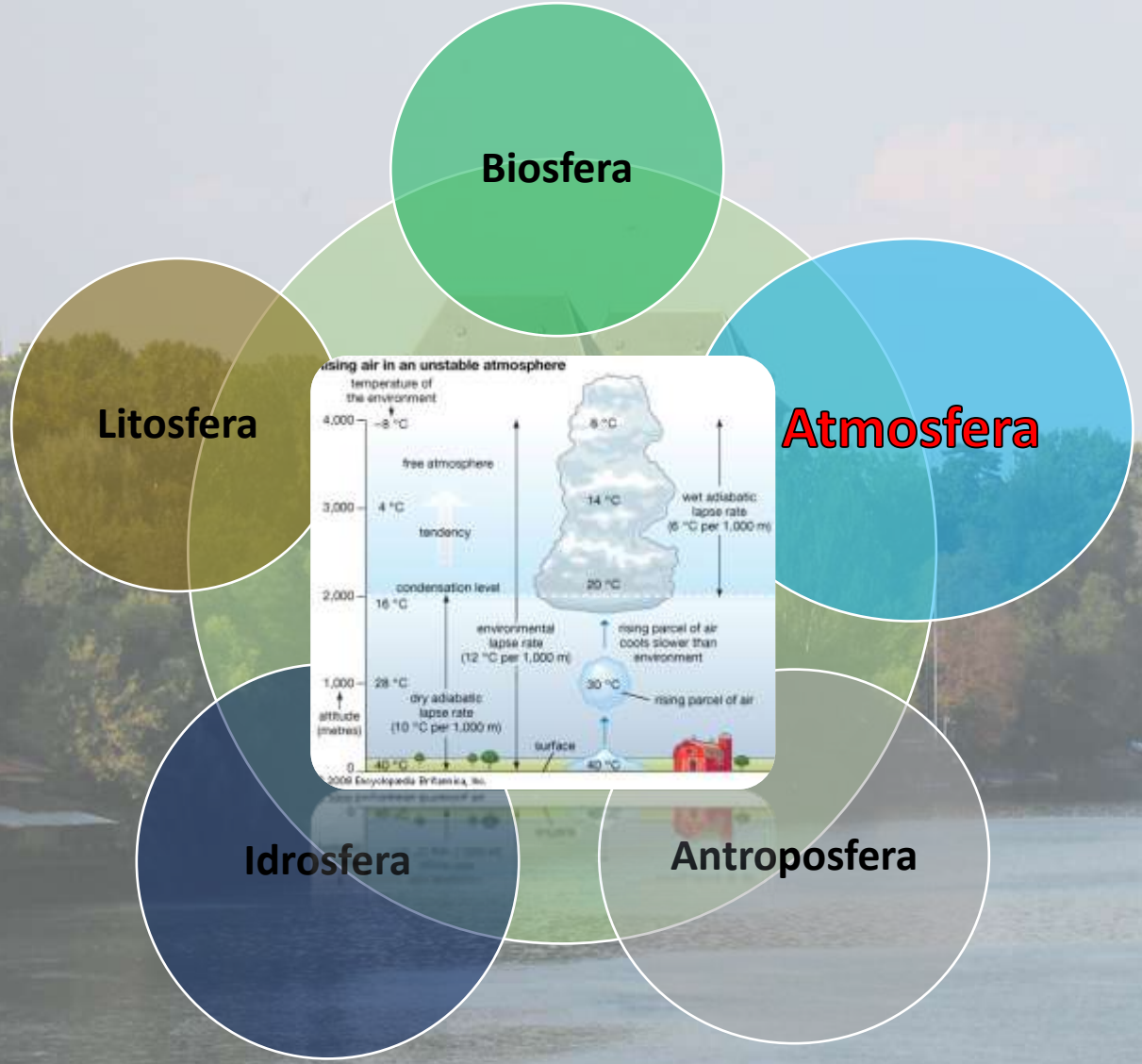
- Parametri del suolo
- Geologia
- Rocce dei letti dei fiumi
- Composizione
- Chimica del suolo

BIOSFERA

- Abbondanza delle specie
- Condizioni fisiologiche
- Patologie
- Contenuto elementare
- Dati microbici



ATMOSFERA

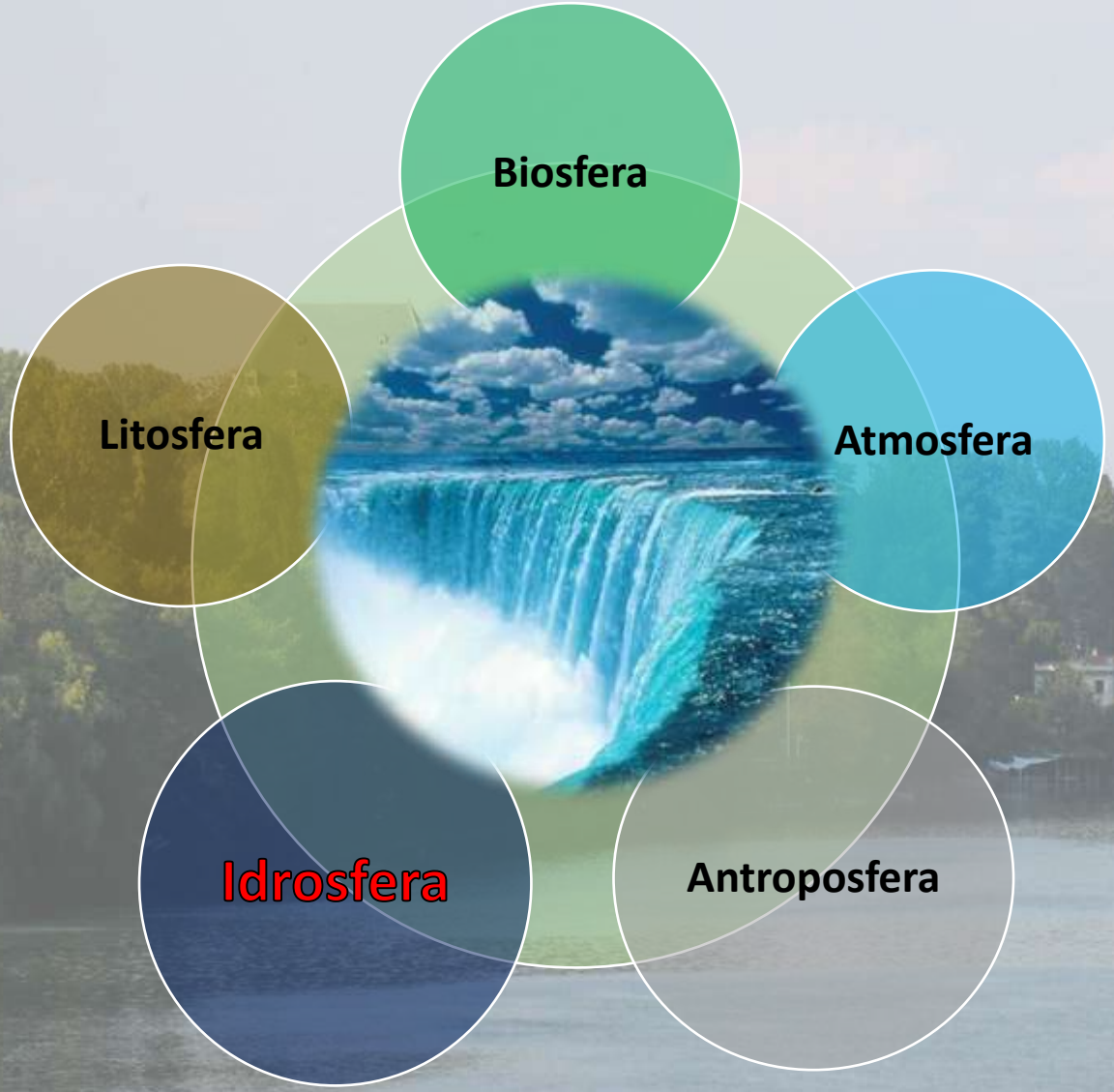


- Concentrazione inquinanti
- Velocità e direzione dei venti
- Emissioni
- Frazioni isotopiche

ANTROPOSFERA

- Dati relativi ai trasporti
- Sfruttamento dei terreni
- Informazioni economiche
- Dati relativi alle discariche

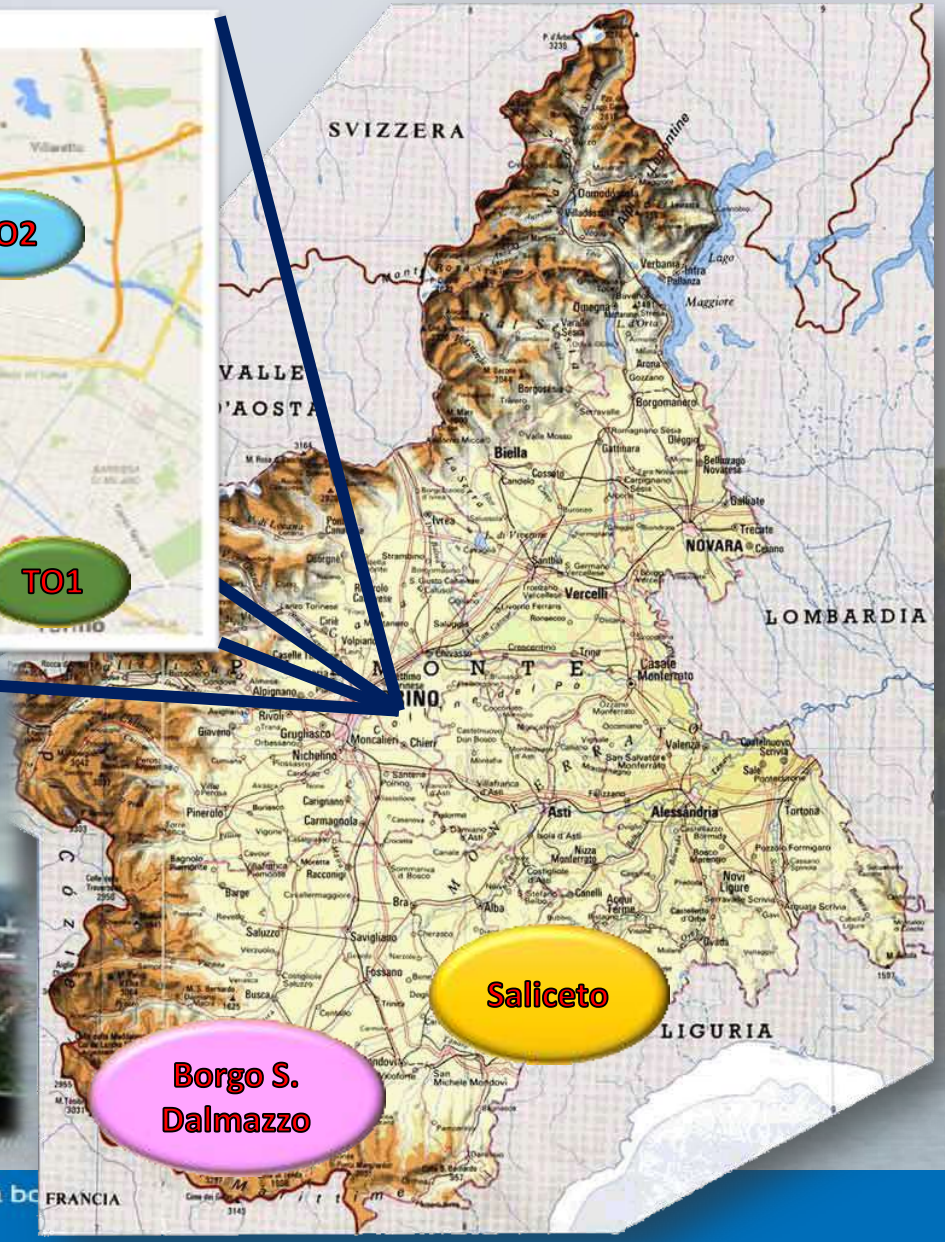
IDROSFERA



- Contaminazione acque di falda
- Flusso in superficie
- Laghi
- Dati relativi alla formazione di sedimenti

1. Atmosfera: Monitoraggio dei metalli in traccia nel particolato (PM10) in aree urbane e rurali in Piemonte

ATMOSFERA – METALLI in PIEMONTE



ATMOSFERA – METALLI in PIEMONTE

Raccolta campioni:

Ogni giorno (2011) da ARPA Piemonte;

- 5 stazioni selezionate;
- 28 campioni per ogni stazione;
- 7 per ogni stagione (febbraio, maggio, agosto e novembre)

ATMOSFERA – METALLI in PIEMONTE

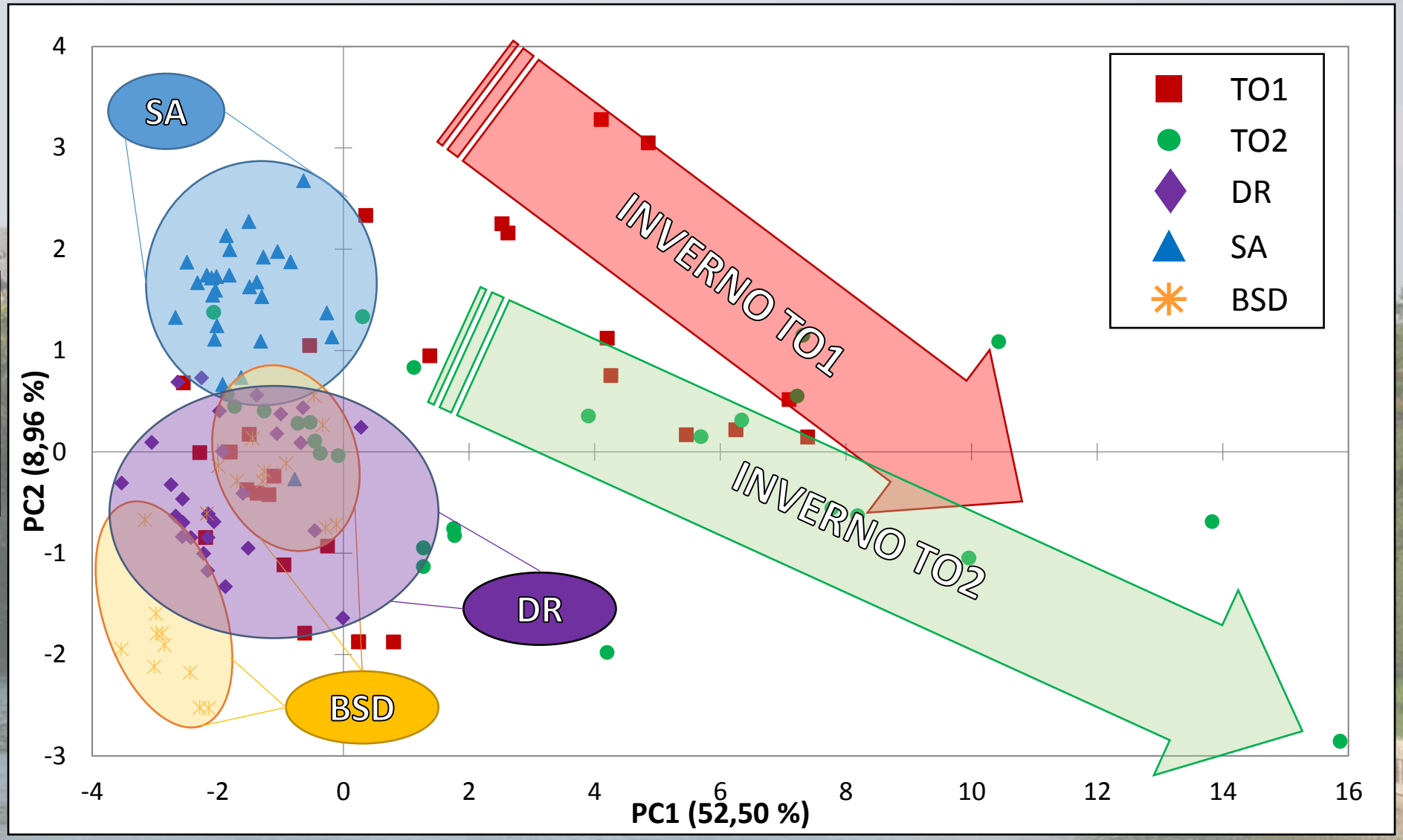
Medie ed intervalli per ogni metallo nei campioni di PM10 per ogni stazione.
Tutti i valori sono in ng/m³, ad eccezione del PM10 espresso in µg/m³.

	TO1		TO2		DR		BSD		SA	
	Mean	Range	Mean	Range	Mean	Range	Mean	Range	Mean	Range
PM_{10}	58	23-110	65	16-125	32	7-66	26	15-48	32	15-72
Mg	254	19-738	522	53-1592	148	6.77-314	400	48.9-1891	173	53.7-367
Al	203	17,8-428	383	11.9-1127	137	4.08-531	211	7.44-824	123	10.4-344
K	486	79.4-1285	395	130-909	232	56.8-472	325	60.8-851	459	128-1143
Ca	749	18.7-1517	1041	271-2721	482	146-1460	1925	328-7332	735	279-1416
Fe	1316	272-3164	2487	431-6810	356	57.4-658	382	105-848	259	93-391
Ti	23.5	2.29-54	26.7	5.7-65.4	18.5	4.99-33.7	23.9	6.33-63.2	19	8.2-49.5
Mn	17.3	3.63-37.5	28.4	6.2-95.8	8.23	2.04-31.3	9.61	3.51-45.8	7.14	3.9-12.5
Cu	22.3	5.04-52.4	67.3	13.4-176.4	3.9	0-15.1	11.4	3.37-33.2	1.73	0-8.46
Zn	64.3	19.6-164	83	14.6-228.8	37.9	4.81-79.3	30.9	0-79.4	26.1	6.4-64
Sb	11	1.9-48.3	10.1	2.6-27.5	2.03	0-5.92	4.23	0.6-12.8	0.16	0-1.2
Ba	32.1	12.4-76.9	40.2	7.4-101.5	16.7	7.39-54.5	13.8	7.74-40.3	6.01	0-15.5
Pb	16.3	3.8-44.6	13.7	0-39.9	6.06	1.7-13.2	10.9	1.13-171	6.18	1.91-18.8
Cr	8.24	1.9-14.3	12,1	2.6-34.3	5.22	0.51-12.5	3.58	0.83-10.1	3.21	1.84-7.82
V	3.15	1.06-5.73	2.77	0.73-5.71	1.44	0.67-2.5	1.97	0.39-7.83	3.21	0.6-8.93
Co	0.7	0.06-1.85	0.65	0.09-1.54	0.18	0-0.87	0.3	0.05-2.43	0.11	0.02-0.52
Ni	5.74	0-16.2	5.97	2.2-15.1	6.45	0-41.9	8.88	0.68-94.2	3.3	1.97-631
As	0.71	0.18-1.48	1.03	0.19-4.66	0.38	0.14-0.7	0.43	0.12-1.04	0.39	0.1-4.16
Sr	1.88	0.04-4.37	239	0.44-11.15	0.72	0-1.39	0.33	0-1.98	0.78	0.34-1.22
Zr	1.66	0.29-4.01	3.21	0.55-7.96	0.38	0-3.43	0.43	0-1.9	0.36	0.14-0.62
Mo	4.86	083-16.02	5.35	1.3-14.21	1.09	0.24-3.03	0.48	0.06-1.51	0.92	0.56-1.73
Cd	0.62	0,19-1,.8	1.1	0.32-5.51	0.31	0.05-1.01	0.31	0.01-1.36	0.33	0.15-0.5
Ce	0.22	0-0.88	2.57	2.12-5.95	0.15	0-1.53	1.52	0.43-3.71	2.77	2.55-3.23
La	<0.007	<0.007	2.39	2.09-4	0.02	0-0.47	1.16	0.05-2.76	3.08	2.83-3.41
Hg	0.44	0-2.77	4.49	0.45-6.9	0.91	0-3.42	0.75	0-9.4	4.13	232-16.44

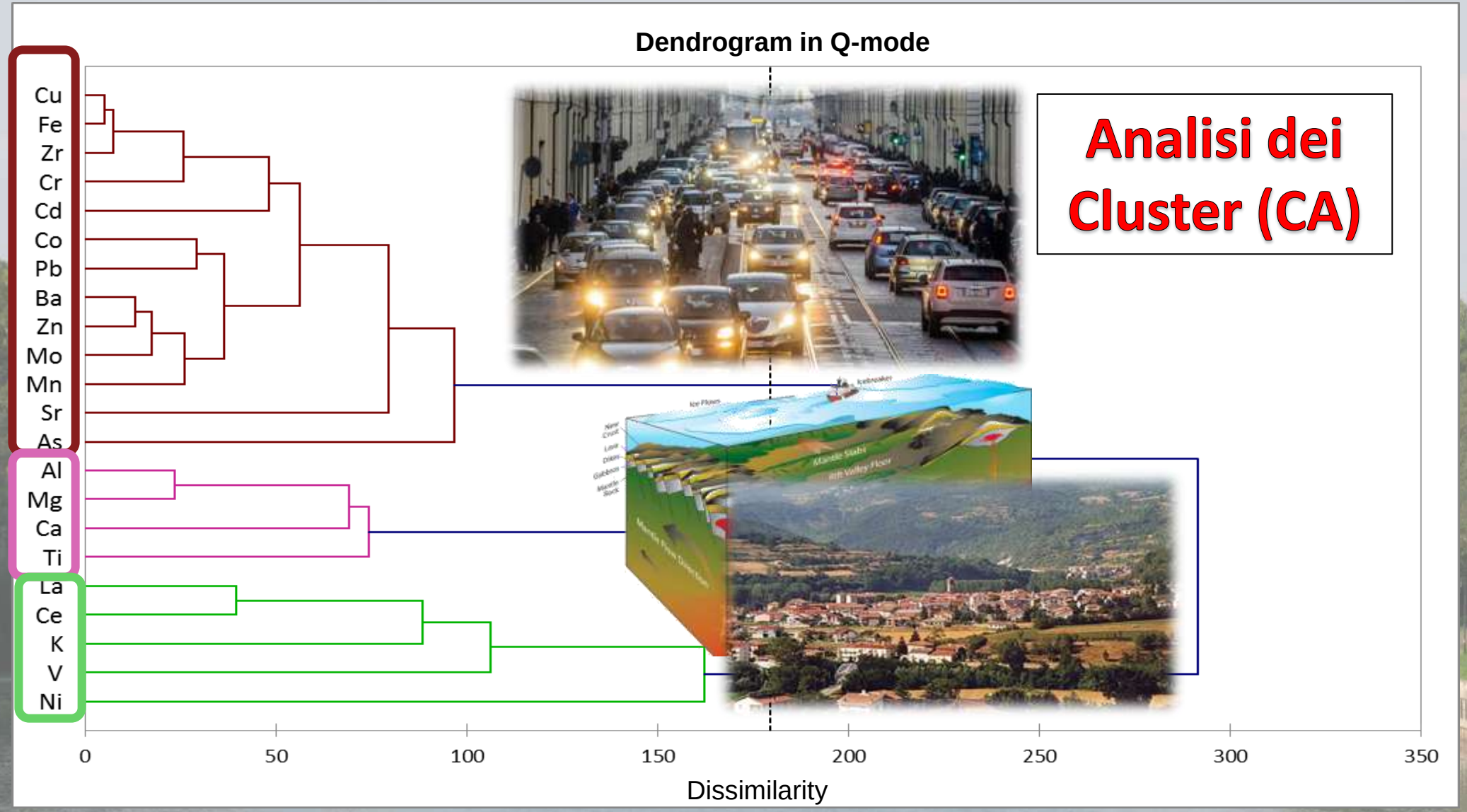
Analisi Multivariata

ATMOSFERA – METALLI in PIEMONTE

Analisi delle Componenti Principali (PCA)

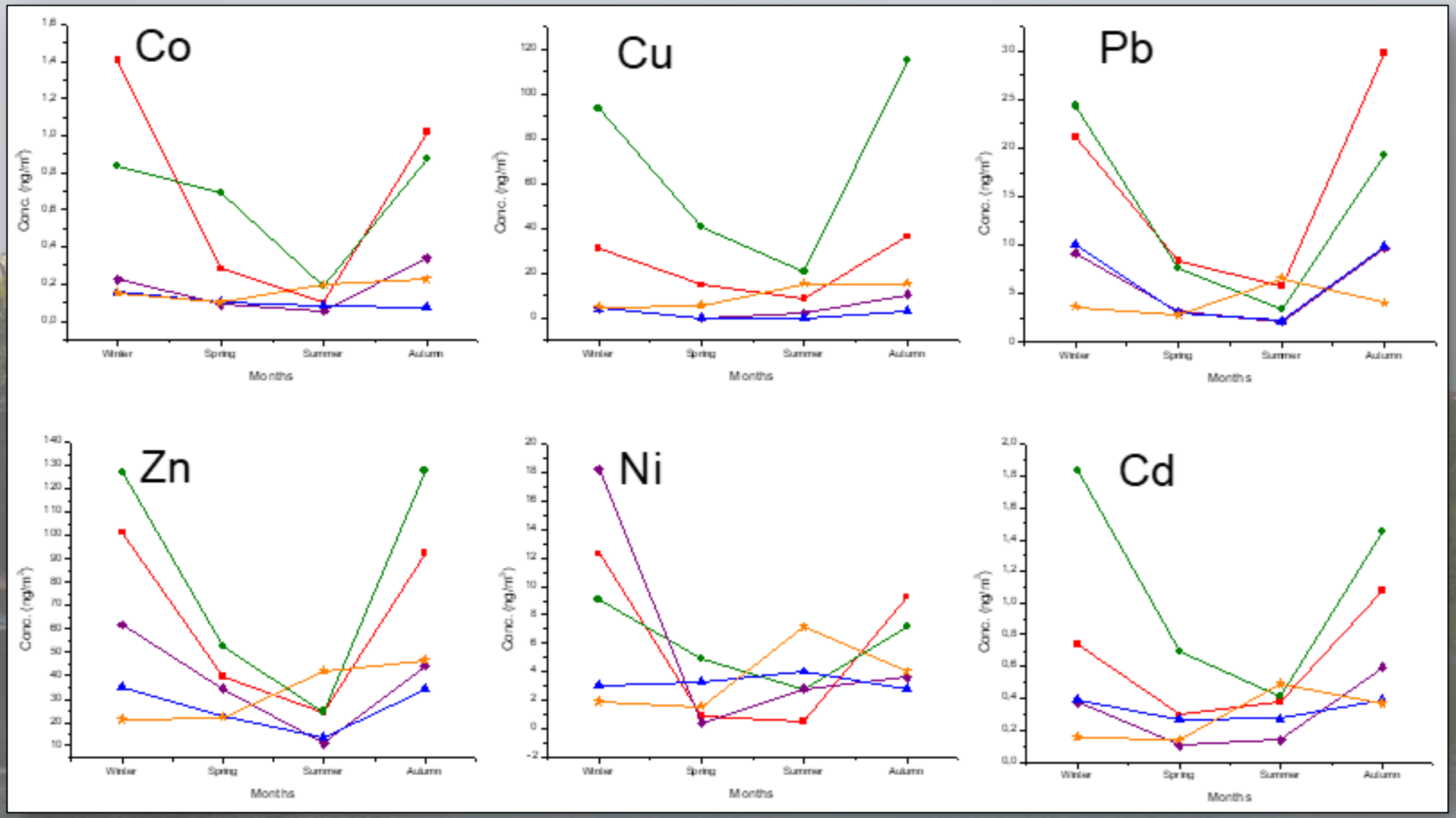


ATMOSFERA – METALLI in PIEMONTE

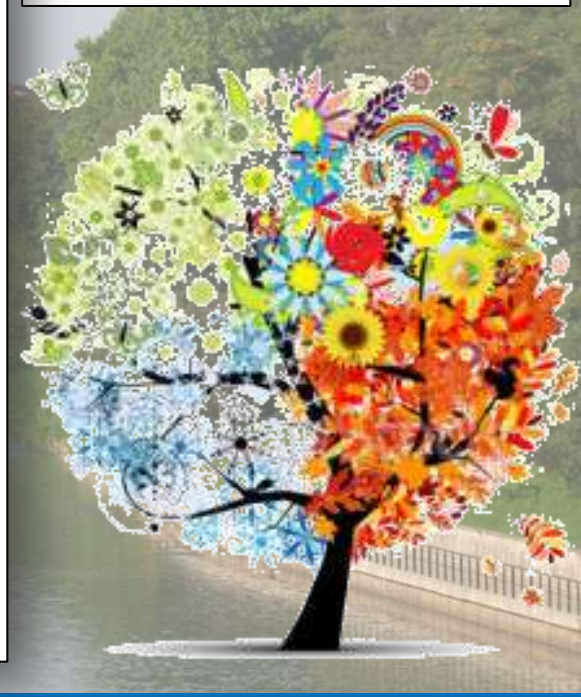


Analisi dei Cluster (CA)

ATMOSFERA – METALLI in PIEMONTE



Stagionalità: tendenze nel tempo



2. Idrosfera: Monitoraggio di parametri chimico-fisici in fiumi piemontesi e svizzeri

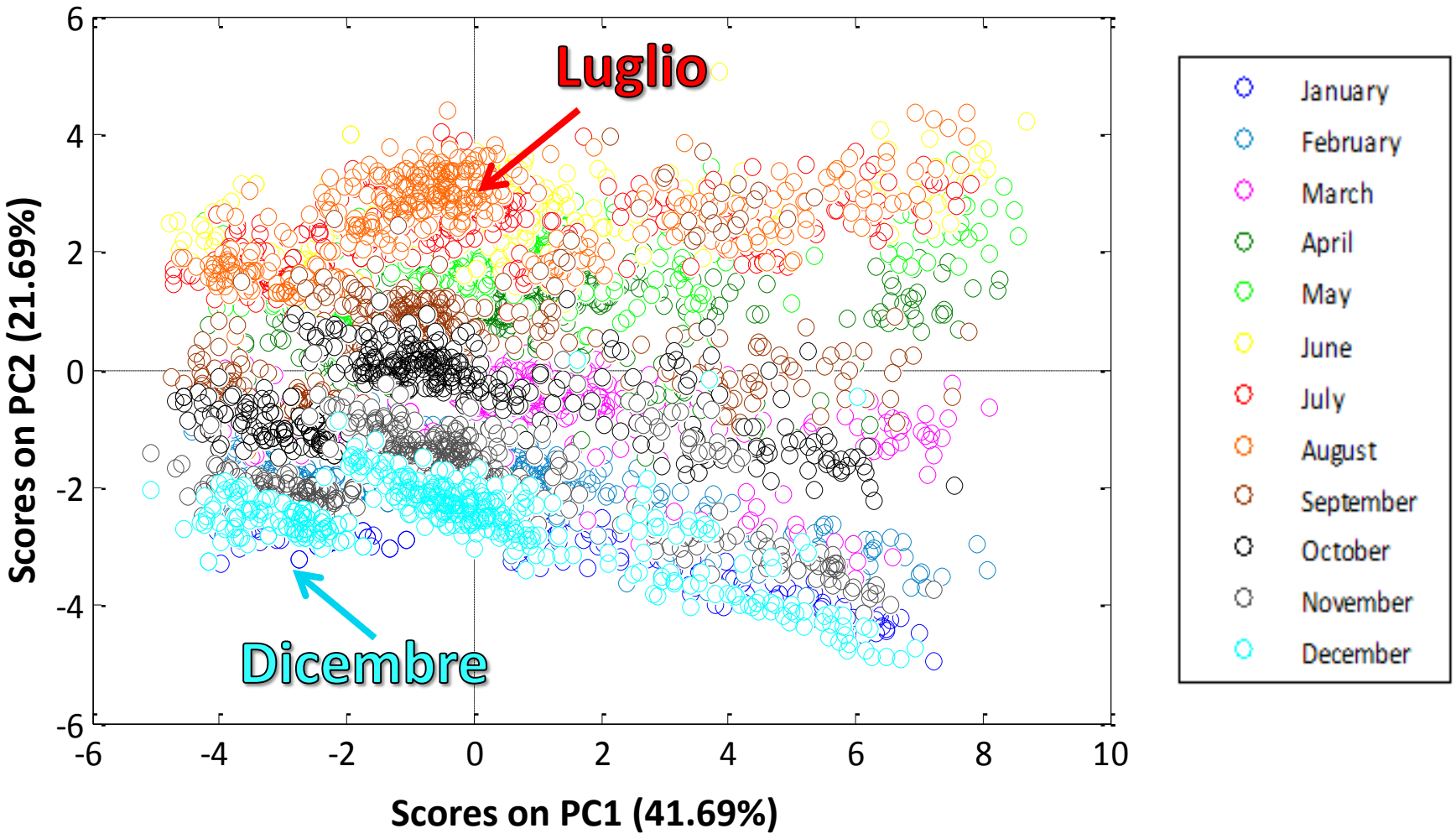
ESEMPI di ANALISI di DATI AMBIENTALI

- 15 fiumi;
- 19 variabili;
- 39 anni (1977-2016)

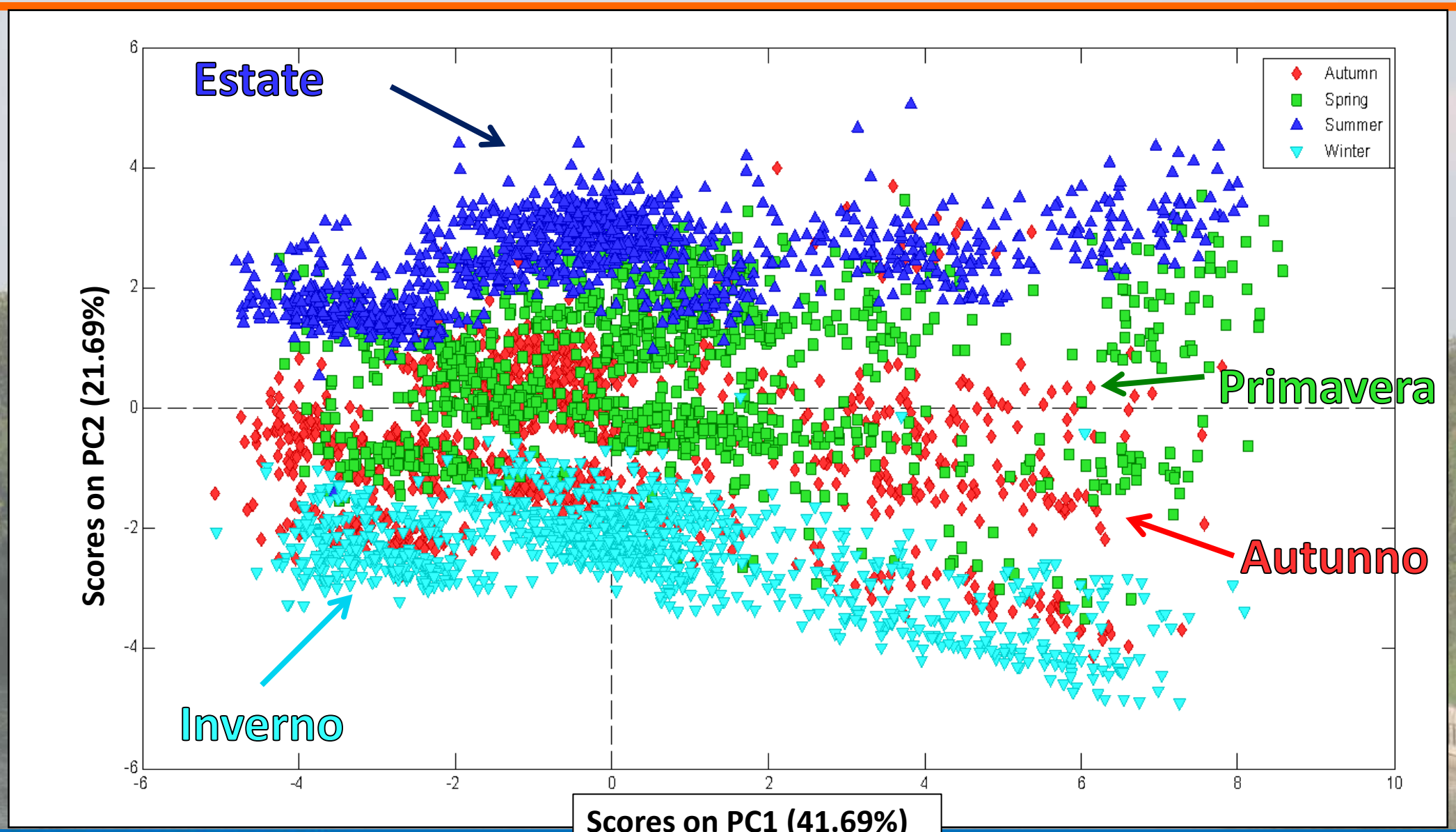
Ricerca di tendenze ed andamenti nei dati raccolti

NO_3^-, M
NPOC, mg C/L
HCO_3^-, M
CO_3^{2-}, M
pH
Alkalinity, eq L^{-1}
Conductivity, 25°C (LHG), µS cm^{-1}
Ca^{+2}, mg L^{-1}
Mg^{+2}, mg L^{-1}
Cl^-, mg L^{-1}
H_4SiO_4, mg L^{-1}
SO_4^{2-}, mg L^{-1}
T, °C
Flow, m^3/s
$^{\cdot}OH$, M
$CO_3^{\cdot-}$, M
$^1O_2$, M
$^3CDOM^*$, M

ESEMPI di ANALISI di DATI AMBIENTALI



ESEMPI di ANALISI di DATI AMBIENTALI



CONCLUSIONI

L'Analisi Multivariata dei Dati permette:

- **Trattare i Big Data;**
- Valutare contemporaneamente variabili di natura differente (Chimica, Fisica, Economia, Sociologia, …);
- Estrarre le importanti informazioni da dati complessi;
- Trovare andamenti, gruppi, differenze, …

TAKE-HOME MESSAGE

"Avere moltissimi dati raccolti nel migliore dei modi, senza interpretarli correttamente, è come avere un meraviglioso pianoforte e suonarlo con un dito solo"

(prof. Riccardo Leardi)

Climathon Torino

chimica

Grazie per l'attenzione!

ealladio@unito.it