IAEA Vienna 8 september 2010

ANALYSIS OF THE CANADIAN BENTHIC DATABASE

Claire Della - Vedova (magelis company)
Jacqueline Garnier – Laplace (IRSN)

EMRAS II Meeting, Vienna, Austria 06-09 September 2010

magelis conseil assistance IRSN

1

RECALL OF THE CONTEXT

- The Fisheries Act of Canada requires the Minister to protect fish populations, their habitat and their resources (Edible fish tissue).

- Benthos is seen as food item for fish and must be protected or not affected by releases of deleterious substances

→ Benthic database from Uranium mines and mill operation :

- Benthos is identified at the family level but most generally at the gender/species level (presence/absence – number of individuals ?)

- This is coupled with measurements of radiological and non radiological contaminants in water and sediment

- The study areas are mostly located in Saskatchewan and Ontario, Canada

WHAT HAD ALREADY BEEN DONE

IAEA Vienna 8 september 2010

Univariate approach (contaminant by contaminant) :

Environmental Monitoring and Assessment (2005) **110:** 71–85
DOI: 10.1007/s10661-005-6291-0

DERIVATION AND USE OF SEDIMENT QUALITY GUIDELINES FOR ECOLOGICAL RISK ASSESSMENT OF METALS AND RADIONUCLIDES RELEASED TO THE ENVIRONMENT FROM URANIUM MINING AND MILLING ACTIVITIES IN CANADA

P. A. THOMPSON*, J. KURIAS and S. MIHOK
Canadian Nuclear Safety Commission, Station B, 280 Slater Street, Ottawa, Canada
*(*author for correspondence, e-mail: thompsonp@cnsc-ccsn.gc.ca)*

(Received 5 August 2004; accepted 12 November 2004)

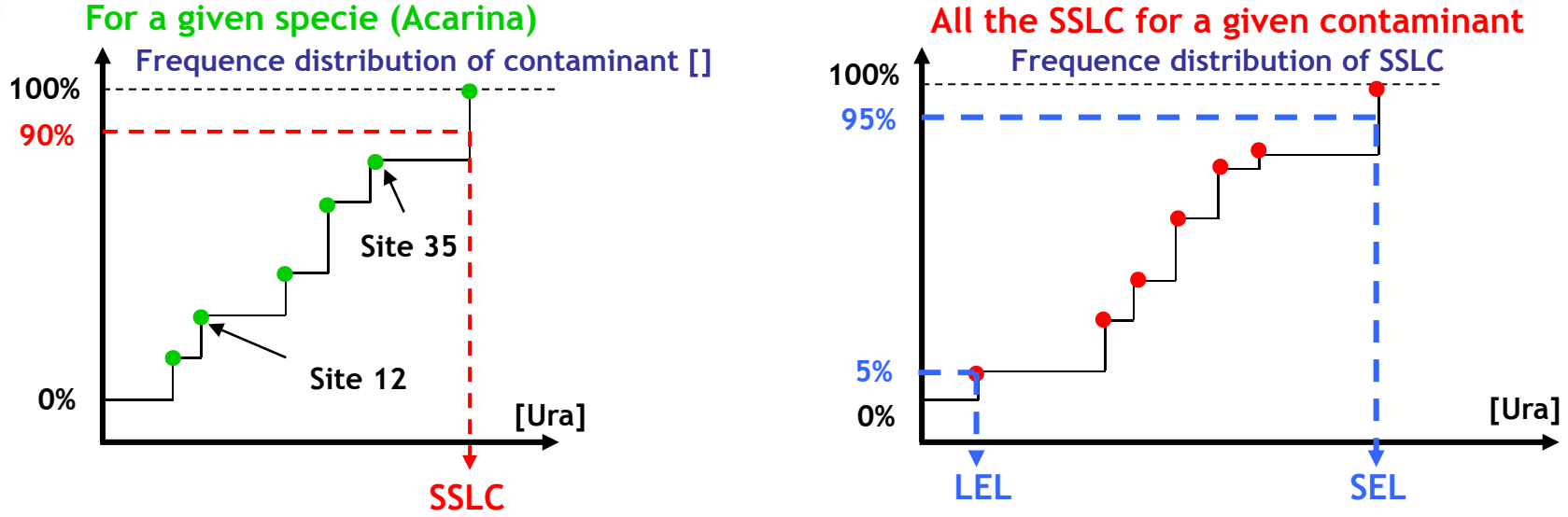
Screening Level Concentration (SLC) method was used to derive Lowest Effect Level (LEL) and Severe Effect Level (SEL) concentration for 9 metals and 3 radionuclides

magelis conseil assistance IRSN

WHAT HAD ALREADY BEEN DONE

IAEA Vienna 8 september 2010

Screening Level Concentration approach (univariate, i.e contaminant by contaminant) :



➔If the sediment contamination is below the Lowest effect level (LEL), environmental risk is likely to be minimal and overruled

➔If the sediment contamination is above the Severe effects level, there is concern about environmental risks

magelis conseil assistance IRSN

WHAT HAVE WE DONE ?

IAEA Vienna 8 september 2010

Multivariate and global approach :

We tried to demonstrate whether changes in species diversity of the benthos community may be explained by changes in contaminants concentrations in sediment.

Plan :

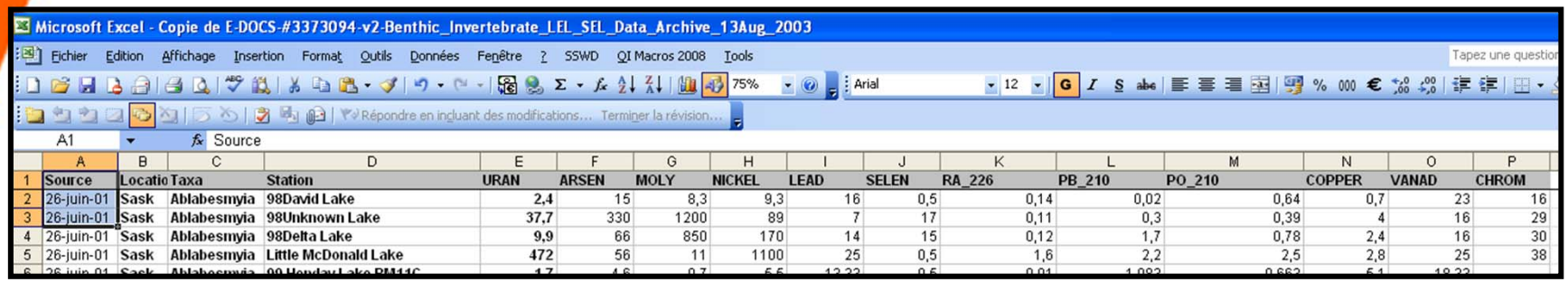
- Methods used: matrix building, statistical analyses (R software, package {vegan})
- Obtained results
- Conclusion
- Future work

magelis conseil assistance IRSN

METHOD - DATABASES

IAEA Vienna 8 september 2010

Original database – sent by the CNSC (Steve Mihok) :



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Source	Locatio	Taxa	Station	URAN	ARSEN	MOLY	NICKEL	LEAD	SELEN	RA_226	PB_210	PO_210	COPPER	VANAD	CHROM
2	26-juin-01	Sask	Ablabesmyia	98David Lake	2,4	15	8,3	9,3	16	0,5	0,14	0,02	0,64	0,7	23	16
3	26-juin-01	Sask	Ablabesmyia	98Unknown Lake	37,7	330	1200	89	7	17	0,11	0,3	0,39	4	16	29
4	26-juin-01	Sask	Ablabesmyia	98Delta Lake	9,9	66	850	170	14	15	0,12	1,7	0,78	2,4	16	30
5	26-juin-01	Sask	Ablabesmyia	Little McDonald Lake	472	56	11	1100	25	0,5	1,6	2,2	2,5	2,8	25	38

2270 lines, 1 line per Taxa and Station

Species matrix (presence / absence)

	Specie1	.....	Specie209
Station 1	0	.....	1
.....	1	.....	0
Sta 196	1	.....	1

Contaminants matrix [ ]$_{sediment}$)

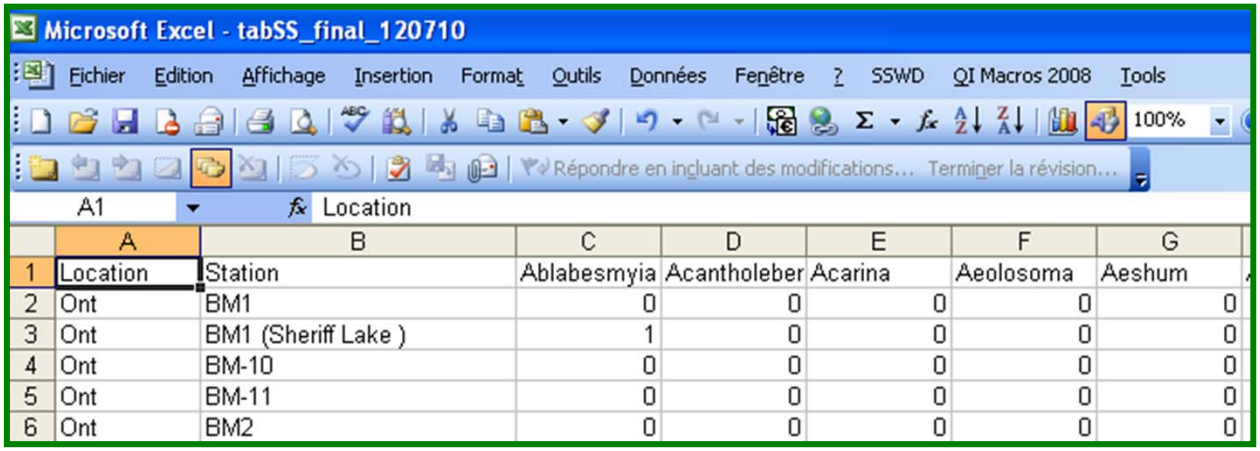
	Cont1	.....	Cont12
Sta 1	[ ]	.....	[ ]
.....	[ ]	.....	[ ]
Sta 196	[ ]	.....	[ ]

magelis conseil assistance IRSN

METHOD –SPECIES MATRIX

IAEA Vienna 8 september 2010

Features of the species matrix (presence / absence)



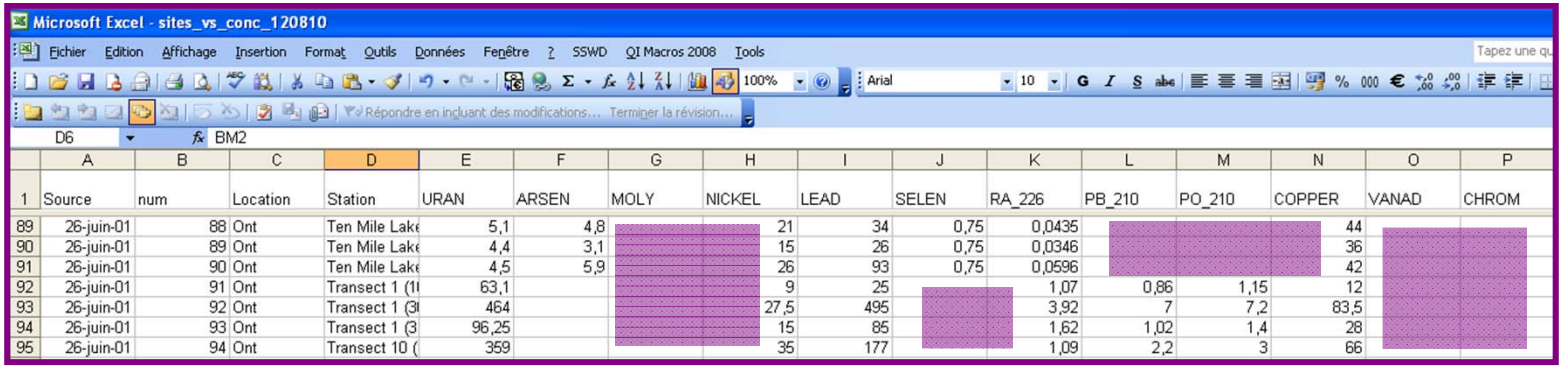
	A	B	C	D	E	F	G
1	Location	Station	Ablabesmyia	Acantholeber	Acarina	Aeolosoma	Aeshum
2	Ont	BM1	0	0	0	0	0
3	Ont	BM1 (Sheriff Lake)	1	0	0	0	0
4	Ont	BM-10	0	0	0	0	0
5	Ont	BM-11	0	0	0	0	0
6	Ont	BM2	0	0	0	0	0

- **196 stations, 209 benthic species (Ablabesmyia – Zapada)**
- **96% of values equal to “0”**

magelis conseil assistance IRSN

METHOD –CONTAMINANT MATRIX

Features of the contaminants matrix ($[]_{sediment}$) - 1



Microsoft Excel - sites_vs_conc_120810

Fichier Edition Affichage Insertion Format Outils Données Fenêtre ? SSWD QI Macros 2008 Tools

D6 BM2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Source	num	Location	Station	URAN	ARSEN	MOLY	NICKEL	LEAD	SELEN	RA_226	PB_210	PO_210	COPPER	VANAD	CHROM
89	26-juin-01	88	Ont	Ten Mile Lake	5,1	4,8		21	34	0,75	0,0435			44		
90	26-juin-01	89	Ont	Ten Mile Lake	4,4	3,1		15	26	0,75	0,0346			36		
91	26-juin-01	90	Ont	Ten Mile Lake	4,5	5,9		26	93	0,75	0,0596			42		
92	26-juin-01	91	Ont	Transect 1 (1	63,1			9	25		1,07	0,86	1,15	12		
93	26-juin-01	92	Ont	Transect 1 (3	464			27,5	495		3,92	7	7,2	83,5		
94	26-juin-01	93	Ont	Transect 1 (3	96,25			15	85		1,62	1,02	1,4	28		
95	26-juin-01	94	Ont	Transect 10 (	359			35	177		1,09	2,2	3	66		

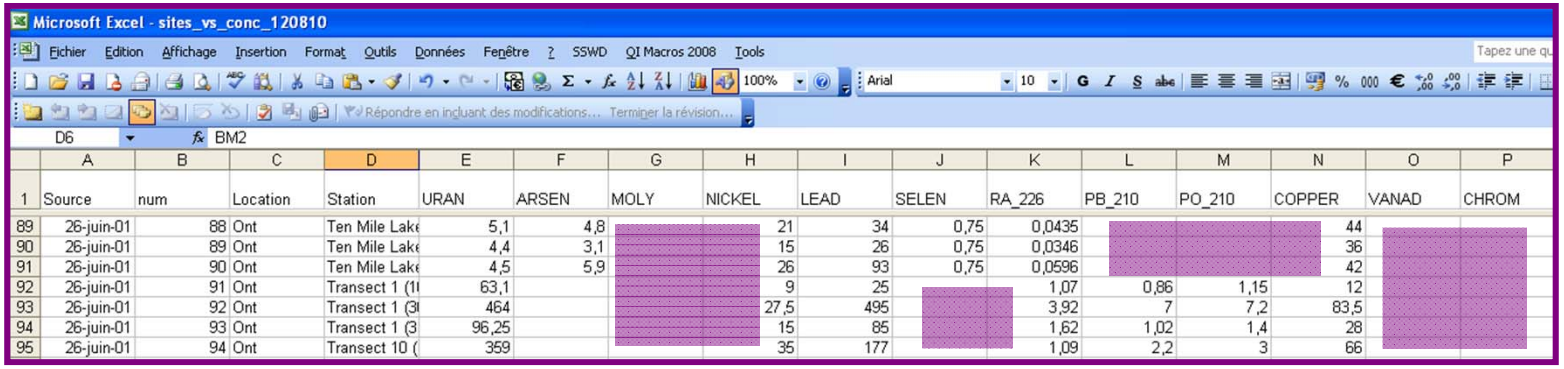
- 196 stations, 12 contaminants (uranium, arsenic, molybdenum, nickel, lead, selenium,radium^{226}, lead^{210}, polonium^{210})
- 27% of NA (empty cells) but with big difference between contaminants :

-uranium : 2.04%, arsenic : 15.82%, molybdenum : 52.55%, nickel :0%, lead : 5.1%, selenium : 35.71%, radium^{226} : 0.51%, lead^{210} : 44.9%, polonium^{210} : 50%, copper : 51%, vanadium : 15.82%, chrome : 62.24%

magelis conseil assistance IRSN

METHOD –SPECIES MATRIX

Features of the contaminants matrix ($[]_{sediment}$) - 2



Microsoft Excel - sites_vs_conc_120810

D6 BM2

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Source	num	Location	Station	URAN	ARSEN	MOLY	NICKEL	LEAD	SELEN	RA_226	PB_210	PO_210	COPPER	VANAD	CHROM
89	26-juin-01	88	Ont	Ten Mile Lake	5,1	4,8		21	34	0,75	0,0435			44		
90	26-juin-01	89	Ont	Ten Mile Lake	4,4	3,1		15	26	0,75	0,0346			36		
91	26-juin-01	90	Ont	Ten Mile Lake	4,5	5,9		26	93	0,75	0,0596			42		
92	26-juin-01	91	Ont	Transect 1 (1	63,1			9	25		1,07	0,86	1,15	12		
93	26-juin-01	92	Ont	Transect 1 (3	464			27,5	495		3,92	7	7,2	83,5		
94	26-juin-01	93	Ont	Transect 1 (3	96,25			15	85		1,62	1,02	1,4	28		
95	26-juin-01	94	Ont	Transect 10 (	359			35	177		1,09	2,2	3	66		

- **Finally, only 31 stations out of 196 (i.e 16%) have measurements for all of the 12 contaminants ➔ We can consider two datasets**

  - a first one built-on with the species and chemical matrix corresponding of the 31 sites without NA values (in chemical matrix) = "complete data"

  -a second one built on with the species and chemical matrix of all the 196 sites = "all data"

METHOD –STATISTICAL ANALYSIS

IAEA Vienna 8 september 2010

What we decided to do :

1. Investigate the contaminants which influence the distribution of the species by means of ordination methods classically used in this situation but which can be applied only to datasets containing no missing data, so to our "compete data" set :
   a) constrained ordination method (Redundancy Analysis - RDA) and
   b) unconstrained ordination method (Principal Components Analysis-PCA) with vectors fitting approach

2. And to develop a method allowing to bring to light the contaminants which influence the distribution of the benthos, even when the dataset contains missing data
   a) use the developed method with "all data" set
   b) Use the developed method with "complete data" set

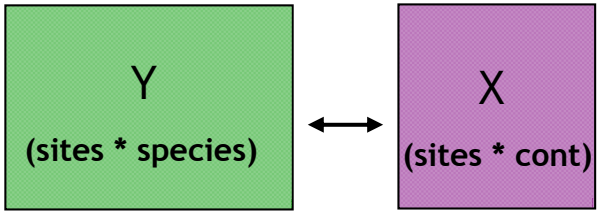
magelis conseil assistance IRSN

RDA – Recall1

IAEA Vienna 8 september 2010

Redundancy analysis (RDA)

1. Redundancy analysis (RDA) is an asymmetric form of canonical analysis where a response matrix (here : species) has to be explained by an explanatory matrix (here : contaminants).



magelis conseil assistance IRSN

RDA – Recall2

IAEA Vienna 8 september 2010

Redundancy analysis (RDA)

2. Redundancy analysis (RDA) is a constrained ordination method

"Ordination is the arrangement of units of some order. (…) For the purpose of analysis, ecologists therefore project the multidimensional scatters diagram onto bivariate graph whose axes are known to be of particular interest. The axes of these graphs are chosen to represent a large fraction of the variability of the multidimensionnal data matrix in a space with reduced dimensionality relative to the original dataset (Legendre et al 1998)".

3. RDA is constrained ordination : it means that only the variation that can be explained by the use of environmental variables, or constraints is displayed.
4. PCA is based on Euclidean distances.

magelis conseil assistance IRSN

RDA – Recall3

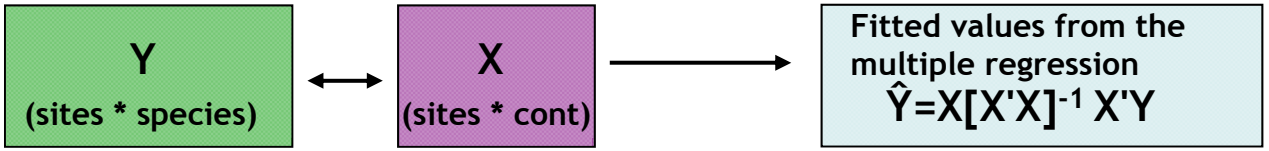
IAEA Vienna 8 september 2010

Redundancy analysis (RDA)

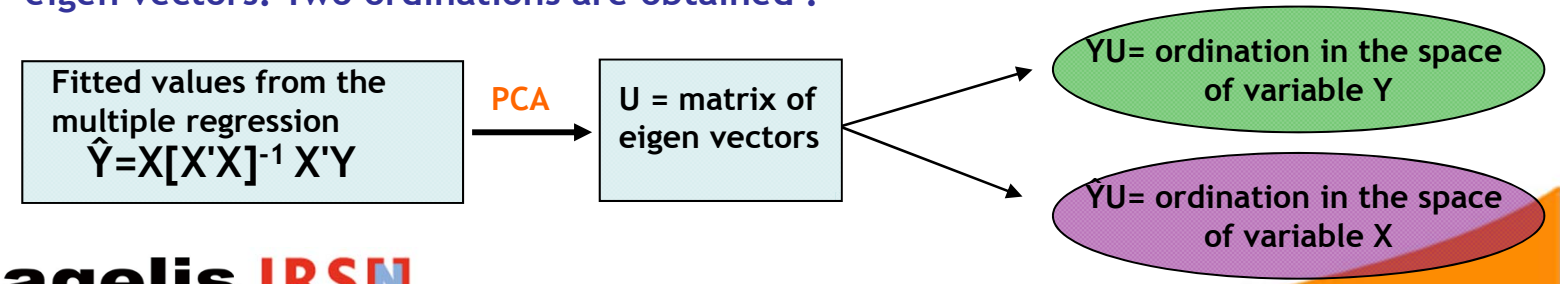
RDA can be seen as an extension of Principal Components analysis (PCA) because the canonical ordination vectors are linear combinations of the response variable Y.

RDA may be understood as a 2-step process :

1. **Regress each variable in Y on all variables in X and compute the fitted values**



2. **Carry out a PCA of the matrix of the fitted values to obtain the eigen values and eigen vectors. Two ordinations are obtained :**



magelis conseil assistance IRSN

RDA with "complete data" set

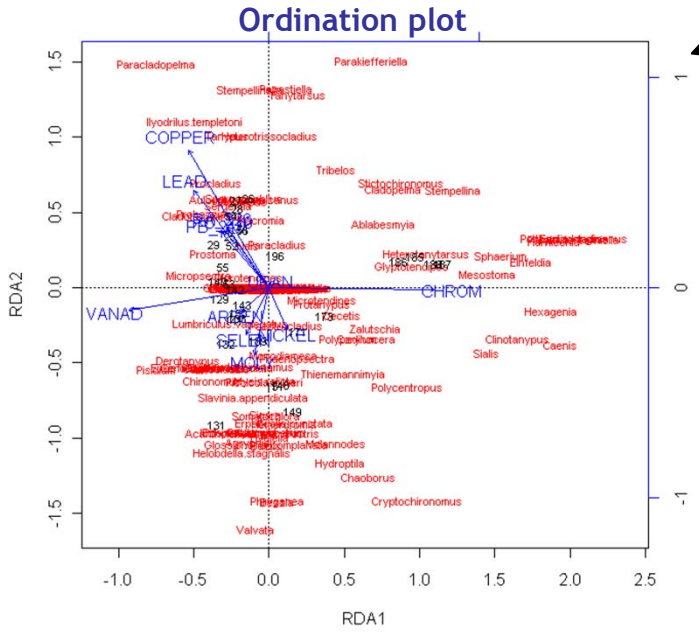
IAEA Vienna 8 september 2010

1st step = use of hellinger transformation for the species matrix

Raw data (sites X species) → **Hellinger transformation*** → **Y=Transformed data** (sites X species) ↔ **X =** (sites X contaminants)

RDA

Representation of elements :
- sites : black number
- species : red names
- contaminants : blue arrows

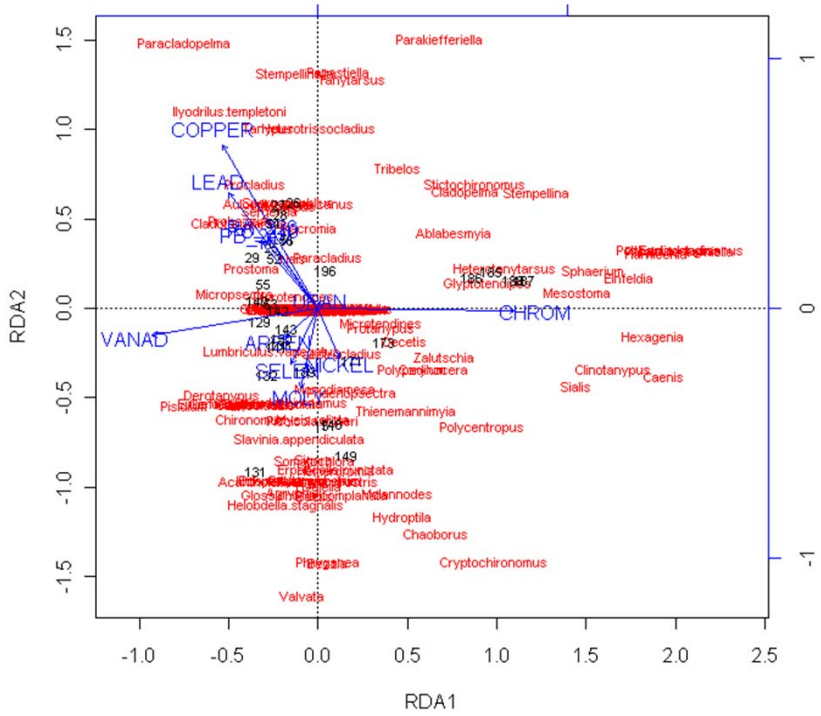


magelis conseil assistance IRSN

RDA with "complete data" set

IAEA Vienna 8 septembre 2010

Rules of interpretation :



1. The orthogonal projection of an object (site) on a quantitative explanatory variable (contaminant) approaches the value of this variable in the object (site).
2. The angle between a response variable (benthic species) and an explanatory variable (contaminant) in the biplot reflects their correlation. Those between 2 explanatory variables has no meaning ; the same between two response variables (species).
3. Distances between objects are approximations of their Euclidean distances in the multidimensional space.

magelis conseil assistance IRSN

RDA with "complete data" set

2nd step : building the optimal model by selection based on AIC

Increasing the number of constraints actually means relaxing constraints : the ordination becomes more similar to the unconstrained ordination. (…) In constrained ordination it is best to reduce the number of constraints to just a few, say 3 or 5 (Oksanen 2010)

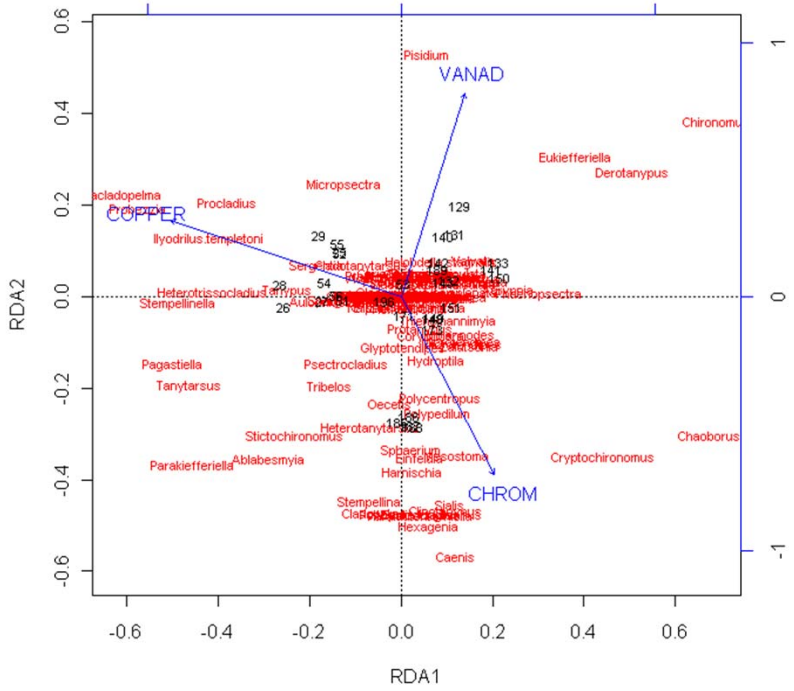
```
Start:  AIC=-11.38
tabSS2.Hell ~ 1
          Df     AIC      F N.Perm Pr(>F)
+ COPPER    1 -12.204 2.7693    199  0.005 **
+ VANAD     1 -11.723 2.2803    199  0.005 **
+ LEAD      1 -11.677 2.2341    199  0.005 **
+ CHROM     1 -11.610 2.1662    199  0.005 **
<none>        -11.376
+ PB_210    1 -11.216 1.7729    199  0.020 *
+ RA_226    1 -11.173 1.7301    199  0.020 *
+ PO_210    1 -11.128 1.6859    199  0.020 *
+ MOLY      1 -11.125 1.6831    199  0.015 *
+ SELEN     1 -11.046 1.6044    199  0.010 **
+ ARSEN     1 -10.886 1.4473    199  0.040 *
+ NICKEL    1 -10.783 1.3461    199  0.100 .
+ URAN      1 -10.683 1.2486     99  0.170
```

```
Step:  AIC=-12.2
tabSS2.Hell ~ COPPER
          Df     AIC      F N.Perm Pr(>F)
+ VANAD     1 -12.875 2.5197    199  0.005 **
+ CHROM     1 -12.252 1.9126    199  0.005 **
<none>        -12.204
+ PB_210    1 -11.941 1.6137    199  0.025 *
+ SELEN     1 -11.925 1.5990    199  0.025 *
+ MOLY      1 -11.872 1.5484    199  0.040 *
+ LEAD      1 -11.861 1.5374    199  0.045 *
+ NICKEL    1 -11.759 1.4408    199  0.060 .
+ ARSEN     1 -11.727 1.4106    199  0.070 .
+ URAN      1 -11.723 1.4070    199  0.080 .
+ RA_226    1 -11.697 1.3818    199  0.070 .
+ PO_210    1 -11.677 1.3629    199  0.110
-COPPER     1 -11.376 2.7693    199  0.005 **
```

RDA with "complete data" set

IAEA Vienna 8 september 2010

Results - 1



1. Three significant contaminant variables: vanadium, copper, chrome

2. Partitioning of variance :

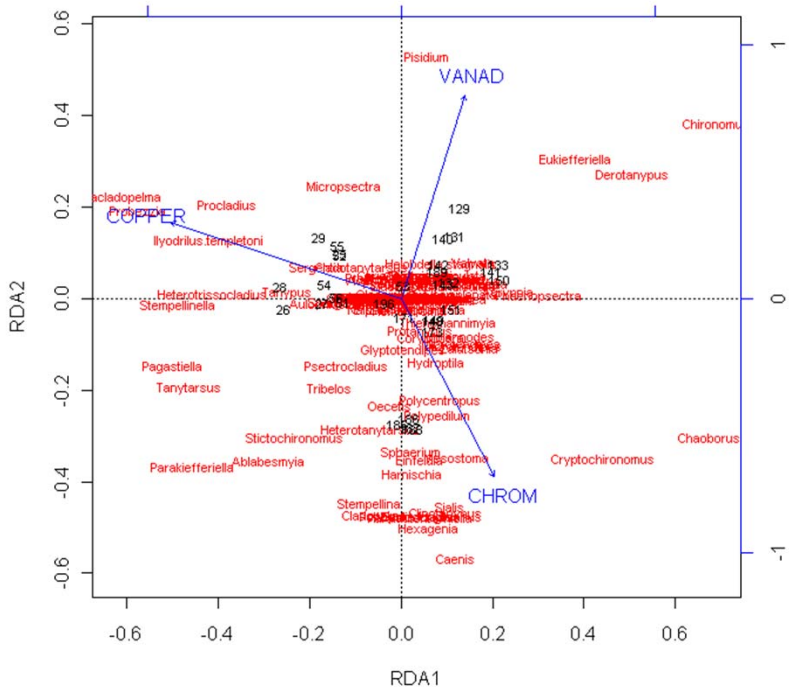
Inertia	Proportion	
Total	0.6712	1.0000
Constrained	0.1460	0.2175
Unconstrained	0.5252	0.7825

3. Contribution to the variance :
   - RDA1 = 0.092 of total variation and 0.42 of constrained variation
   - RDA2 = 0.42 of total variation and 0.40 of constrained variation

RDA with "complete data" set

IAEA Vienna 8 september 2010

Results - 2



4. Observed 'contaminant-species' associations

Vanadium:

Corr + : **Pisidium, Microspectra, Eukiefferellia, …**

Corr - : **Parakiefferiella,Ablamesmyia, Stichtochironomus, …**

Chrome:

Corr + : **Caenis, Hexagenia, clinotanypus, …**

Corr - :**microspectra, procladius, …**

Copper:

Corr + : **Paracladopelma, Probezzia, Ilyodrilus,templetoni, …**

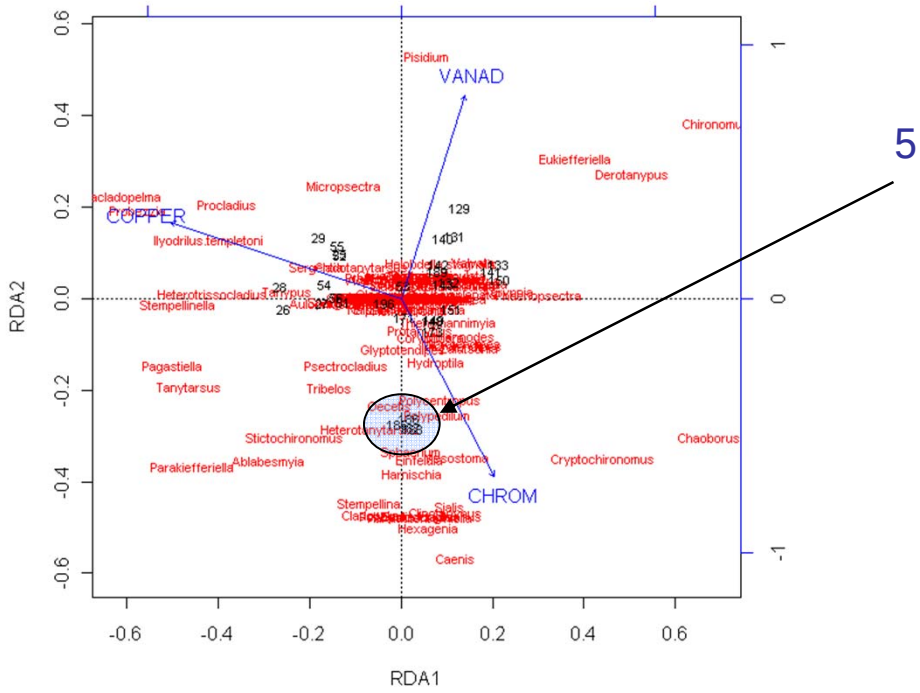
Corr - :**Chaoborus, Cryptochironomus,**

magelis conseil assistance IRSN

RDA with "complete data" set

IAEA Vienna 8 september 2010

Results-3



5. Some groups of sites are geographically closed ➔ possible confusing variables ?

   - **185 : Wheeler River System Zone 1**
   - **186 : Wheeler River System Zone 2**
   - **187 : Wheeler River System Zone 3**
   - **188 : Wheeler River System Zone 4**

magelis conseil assistance IRSN

RDA with "complete data" set

IAEA Vienna 8 september 2010

Conclusion

1. RDA seems to be an appropriate method because it permits to highlight the relationship between some species (presence and absence) and some contaminants

2. But what is the robustness of these results since the analysis is only done with 31 sites ?

   a) **Are they representative of all the data knowing that the Ontario/Saskatchewan percentage is inverted with regard to the "all data" set ? (35%/65% for "complete data" vs 65%/35% for "all data")**

   b) **Is there a possible confusing variable since some groups of sites are just geographically closed sites?**

3. What would be the results using the other method : non constrained ordination (PCA) with environmental vectors fitting?

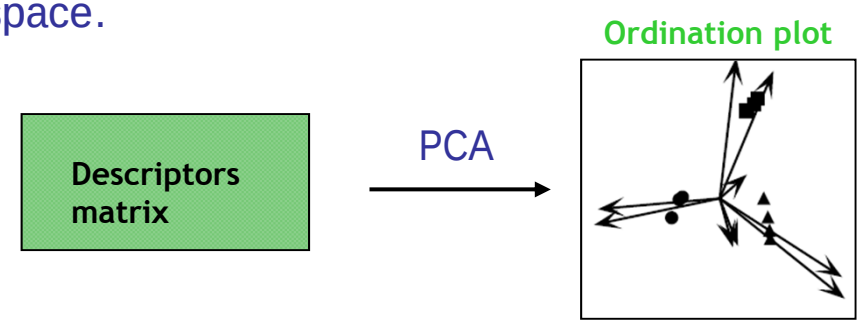
magelis conseil assistance IRSN

PCA and vectors fitting

IAEA Vienna 8 september 2010

Principal Components analysis (PCA)

1. Principal Components Analysis is a powerful technique for ordination in reduce space.



2. Ordination vectors are linear combinations of the response variable Y.

3. PCA is unconstrained ordination : it means that all the compositional variation is display (and not only the variation that can be explained by the used of environmental variables as in RDA)

4. PCA is based on Euclidean distances.

magelis conseil assistance IRSN

PCA and vectors fitting

IAEA Vienna 8 september 2010

1st step = use of Hellinger transformation for the species matrix

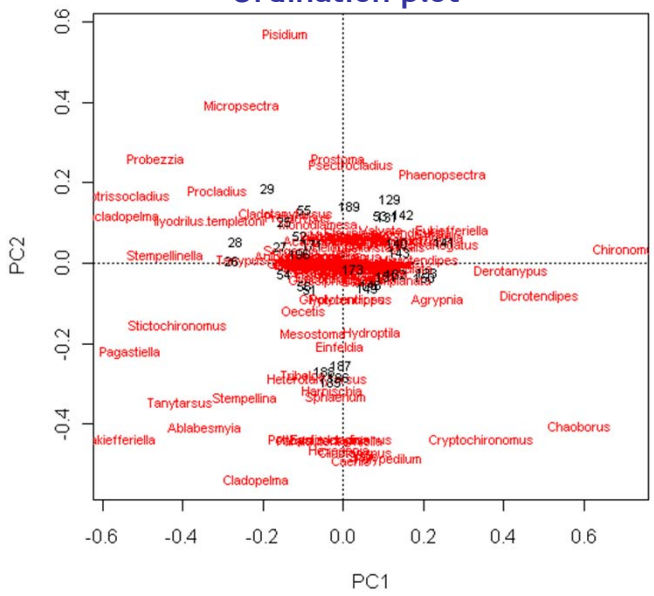
Raw data
(sites X species)

→ **Hellinger transformation***

Y=Transformed data
(sites X species)

→ **PCA**

Ordination plot



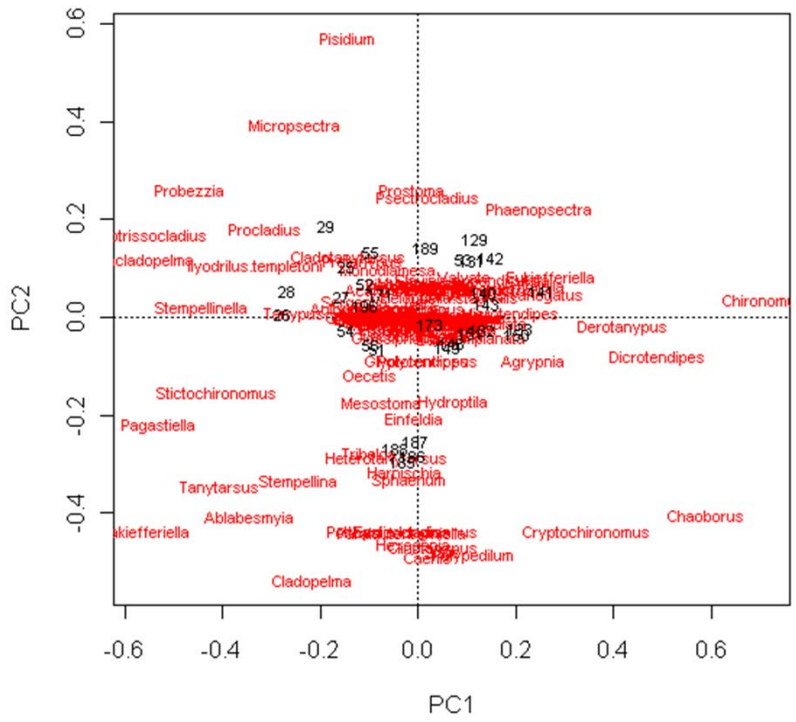
Representation of elements :
- sites : black number
- species : red names

magelis conseil assistance IRSN

PCA and vectors fitting

IAEA Vienna 8 septembre 2010

PCA : Rules of interpretation



1. The distances between objects (sites) in the biplot are estimates of their Euclidian distances in the multidimensional space.

   ➔ **arrangement of the sites according to their species distribution resemblance**

2. Angles between vectors representing descriptors (species) have no sense.

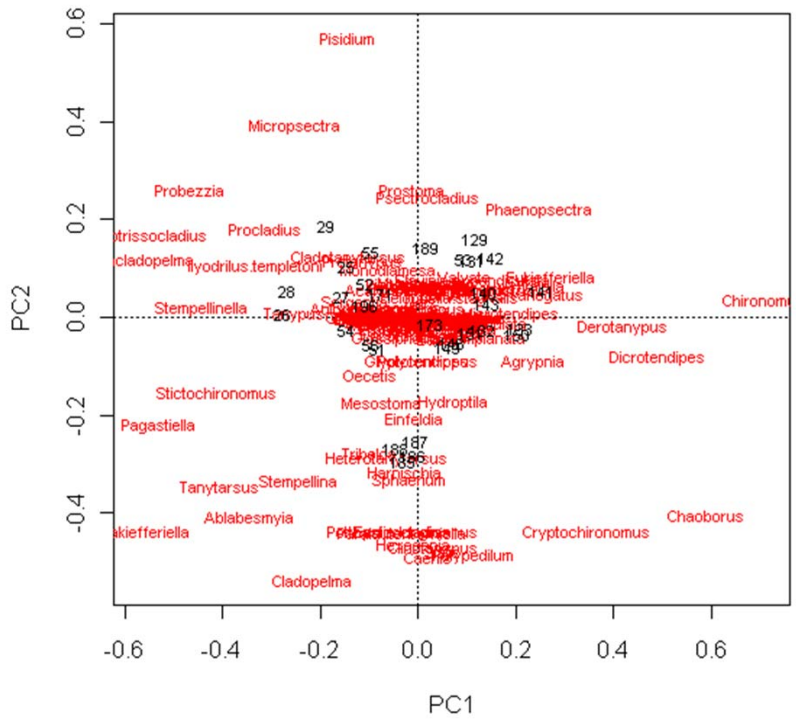
3. The orthogonal projection of a point object (site) on a vector descriptor (species) approaches the value of the variable in the considered object.

magelis conseil assistance IRSN

PCA and vectors fitting

IAEA Vienna 8 septembre 2010

PCA : Results



1. Contribution to the variance:
   - PC1 : 13.15%
   - PC2 : 11.05%

   } 24.2% (not so bad)

   ➔ **arrangement of the sites according their species distribution resemblance**

2. Meaning of the principal components ?

   ➔ need of benthos expert

magelis conseil assistance IRSN

PCA and vectors fitting

IAEA Vienna 8 september 2010

2nd step : fit of the environmental vectors onto the biplot ordination (PCA)

a) *Prediction of contaminant variable separately from the ordination scores (two first axes of the PCA) by least square regression :*

$[Uran]_i = a\ scorePC1_i + b\ scorePC2_i$ (i = site index)

	PC1	PC2	r2	Pr(>r)	
URAN	-0.8641110	0.5033013	0.0326	0.651	
ARSEN	0.9999984	0.0017619	0.0908	0.251	
MOLY	0.9720080	-0.2349476	0.1378	0.100	.
NICKEL	-0.4468266	0.8946206	0.0036	0.959	
LEAD	-0.8879775	0.4598870	0.3794	0.006	**
SELEN	0.9792747	-0.2025364	0.1100	0.196	
RA_226	-0.8806784	0.4737146	0.1274	0.147	
.....					

1. PC1 and PC2 = directions of the vectors regarding the principal components
2. r2 = squared correlation coefficient between ordination and environmental variables
3. Pr(>r) assesses the significance of r^2 (obtained by permutation)

magelis conseil assistance IRSN

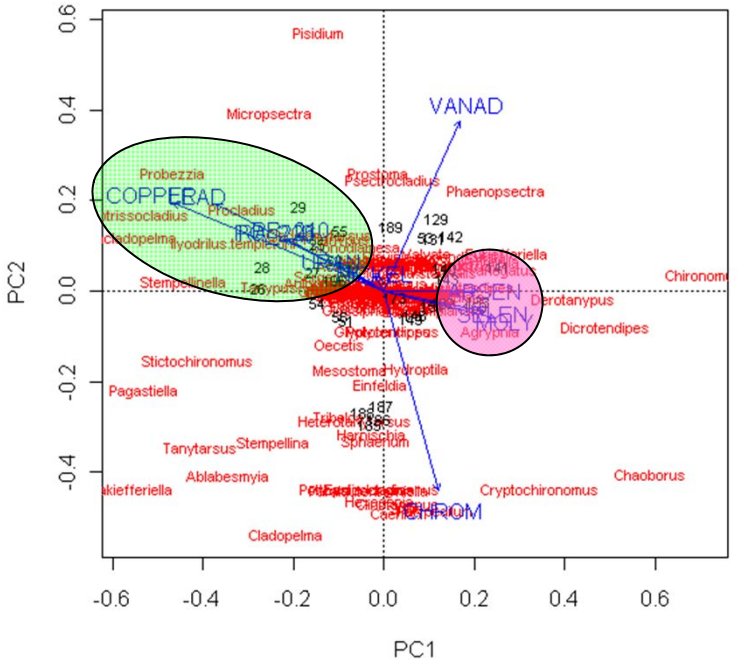
PCA and vectors fitting

IAEA Vienna 8 septembre 2010

2nd step : Fit the environmental vectors onto the biplot ordination (PCA) – all vectors

b) Plot the environmental (contaminant) vectors onto the ordination plot

	PC1	PC2	r2	Pr(>r)	
URAN	-0.8641110	0.5033013	0.0326	0.651	
ARSEN	0.9999984	0.0017619	0.0908	0.251	
MOLY	0.9720080	-0.2349476	0.1378	0.100	.
NICKEL	-0.4468266	0.8946206	0.0036	0.959	
LEAD	-0.8879775	0.4598870	0.3794	0.006	**
SELEN	0.9792747	-0.2025364	0.1100	0.196	
RA_226	-0.8806784	0.4737146	0.1274	0.147	
PB_210	-0.8380518	0.5455907	0.1073	0.200	
PO_210	-0.8850019	0.4655873	0.1309	0.141	
COPPER	-0.9223579	0.3863364	0.5535	0.001	***
VANAD	0.4051651	0.9142435	0.3652	0.003	**
CHROM	0.2613504	-0.9652440	0.4510	0.001	***

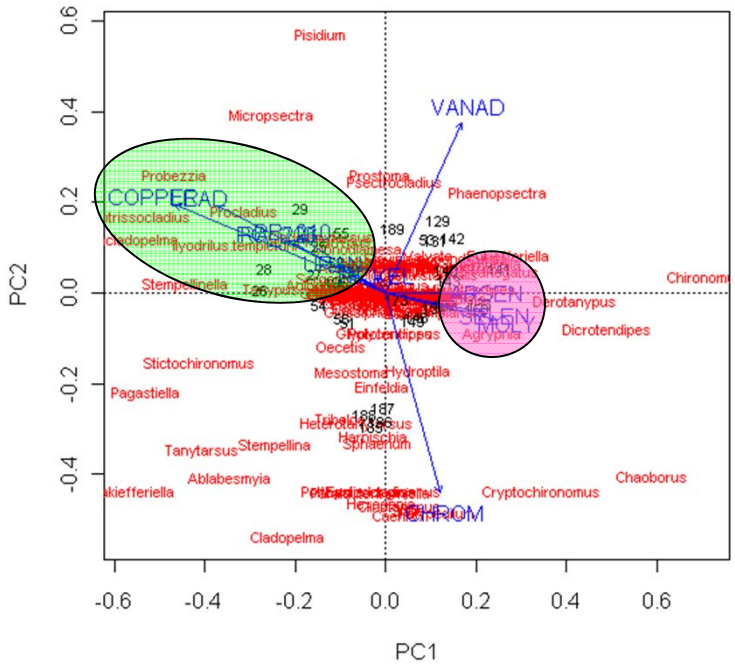


PCA and vectors fitting

IAEA Vienna 8 september 2010

2nd step : Fit the environmental vectors onto the biplot ordination (PCA) – all vectors

b) Plot the environmental (contaminant) vectors onto the ordination plot



Rules of interpretation

1. The length of the arrow is proportional to the correlation between ordination and environmental variable

2. Interpretation of the species presence/absence in connection to the environmental vectors seems to be done by angles

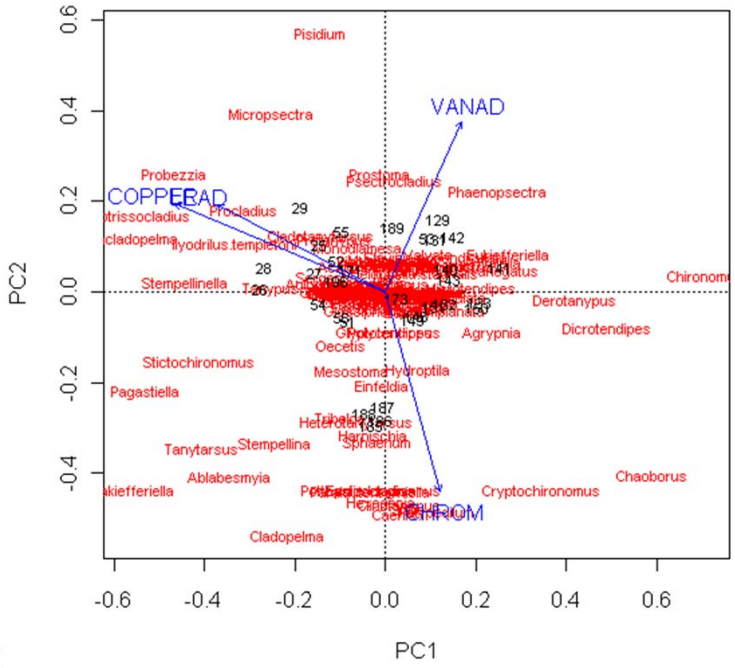
magelis conseil assistance IRSN

PCA and vectors fitting

IAEA Vienna 8 september 2010

2nd step : Fit the environmental vectors onto the biplot ordination (PCA) – only the significant vectors

c) Plot the only significant environmental (contaminant) vectors

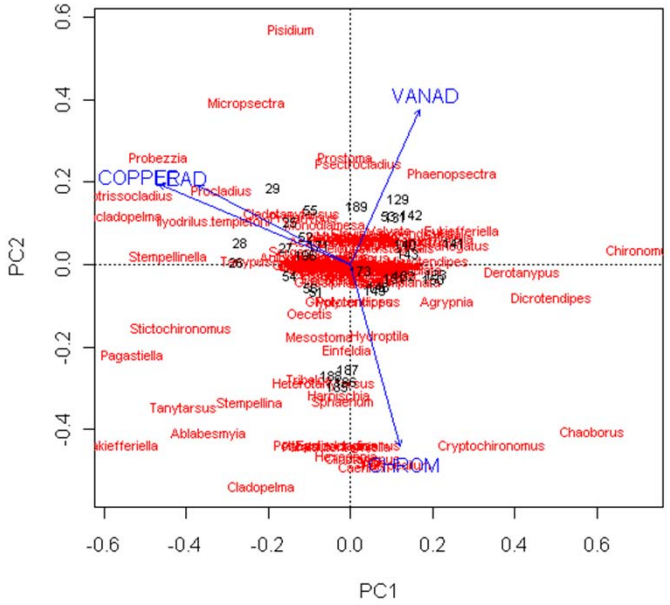


PCA vs RDA

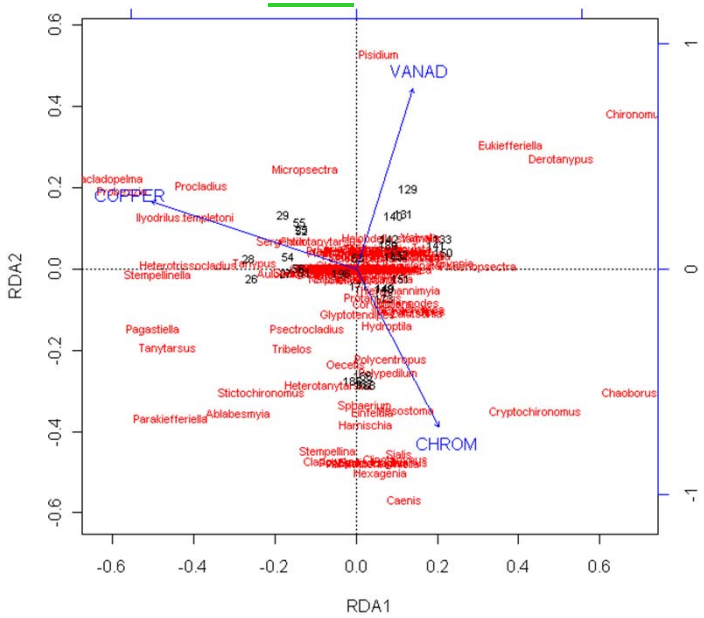
IAEA Vienna 8 september 2010

Results

PCA and vectors fitting



RDA

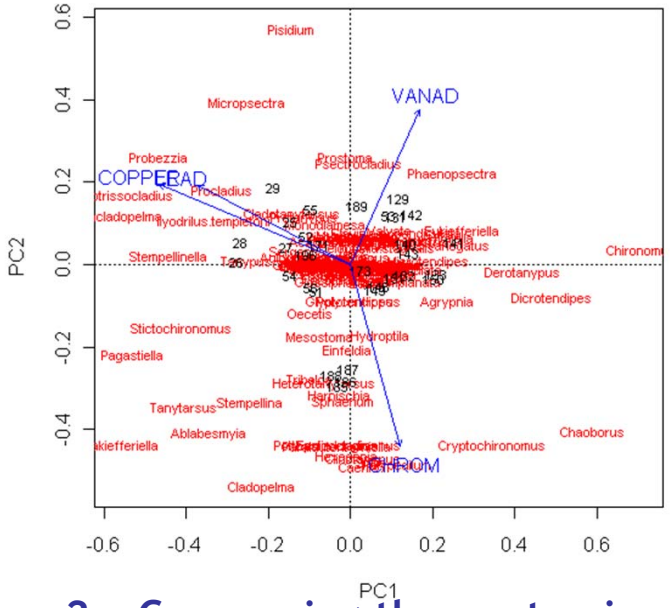


1. **Concerning the sites, these two methods give nearly the same results, they are globally situated at the same place**

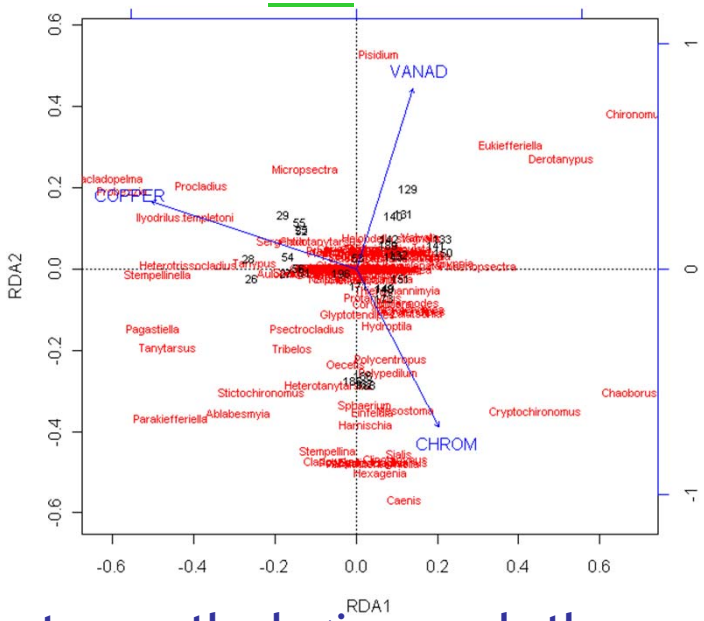
PCA vs RDA

Results

PCA and vectors fitting



RDA



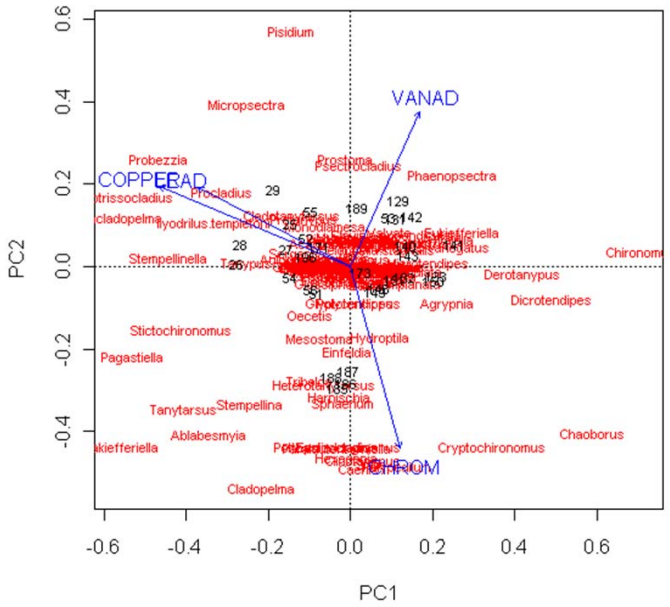
2. **Concerning the contaminants : these two methods give nearly the same results with an added (lead) significant vector for the "PCA and vector fit" method**

magelis conseil assistance IRSN

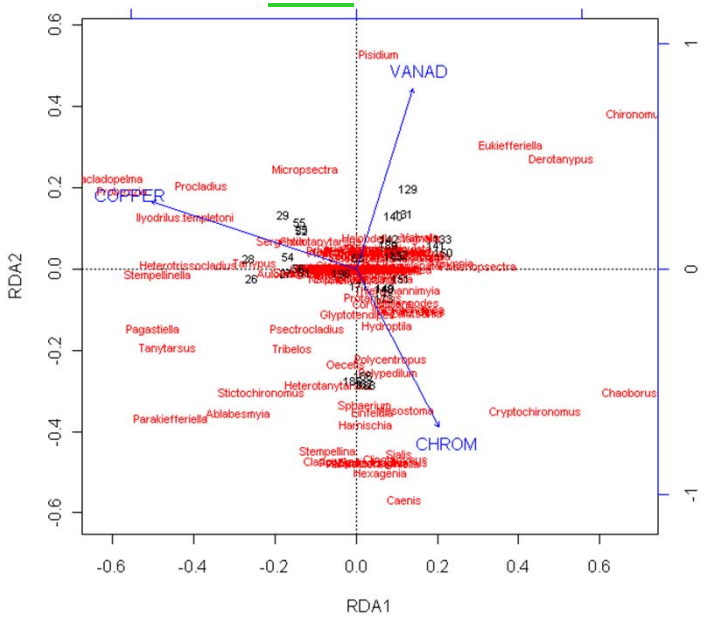
PCA vs RDA

Results

PCA and vectors fitting



RDA



3. **The species are globally situated in the same place, with some difference**

PCA vs RDA

IAEA Vienna 8 september 2010

Conclusion

1. Globally, RDA and PCA +vectors fitting give similar results.
2. Using RDA it seems that we establish a causality link between the contaminants and the distribution of the benthic species.
   Using PCA with environmental vectors fitting it's not causality link, but only correlation. Nevertheless we take in account all the variation.
3. These two methods permit to order the contaminant in regard of their power.
4. The results obtained by these two methods have the same drawbacks because of the 31 sites used:
   a) Representativity of the 31 sites in regard of all the data since (Ontario/Saskatchewan) proportion is inverted
   b) Presence of confusing variables since into some groups sites are correlated

magelis conseil assistance IRSN

THE METHOD WE DEVELOPED with "all data" set

IAEA Vienna 8 september 2010

1st step = use of Hellinger transformation for the species matrix and PCA

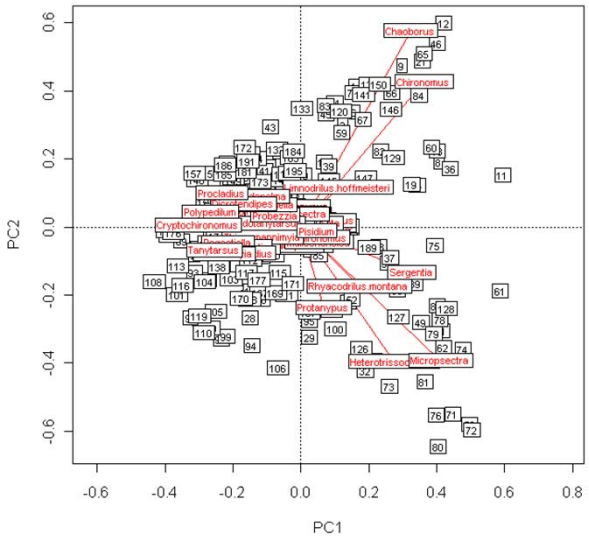
Raw data
(sites X species)

Hellinger transformation*

Y=Transformed data
(sites X species)

PCA

Ordination plot



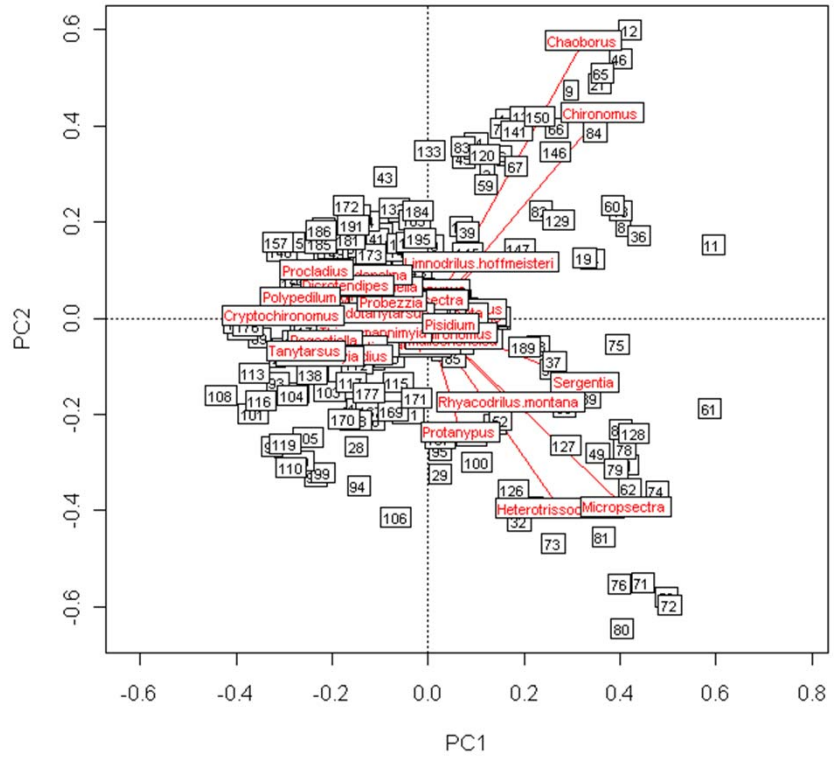
Representation of elements :
- sites : black number
- species : red arrow

magelis conseil assistance IRSN

THE METHOD WE DEVELOPED with "all data" set

IAEA Vienna 8 september 2010

1st step : ordination (PCA) from the species matrix (with Hellinger transformation)



Same rules of interpretation :

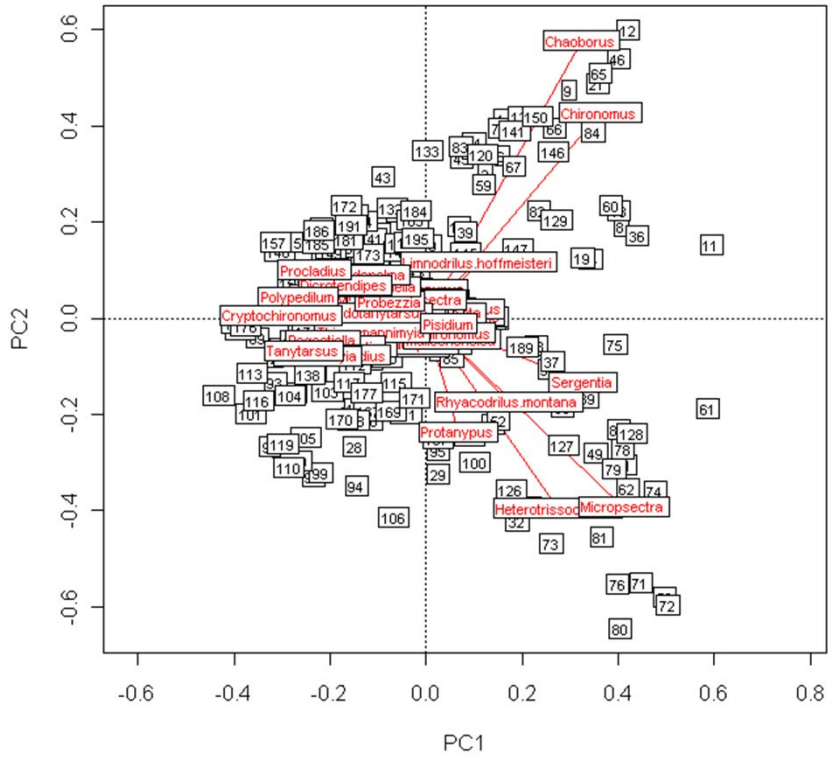
1. The distances between objects (sites) in the biplot are estimates of their Euclidian distances in the multidimensional space.
   ➔ **arrangement of the sites according to their species distribution resemblance**
2. Angles between vectors representing descriptors (species) have no sense.
3. The orthogonal projection of a point object (site) on a vector descriptor (species) approaches the value of the variable in the considered object.

magelis conseil assistance IRSN

THE METHOD WE DEVELOPED with "all data" set

IAEA Vienna 8 september 2010

1st step : ordination (PCA) from the species matrix (with Hellinger transformation)



Results :

1. Contribution to the variance:
   - PC1 : 10.74%
   - PC2 : 8.63%

   } 19.37%

2. Meaning of the principal components ?
   ➔ need of benthos expert

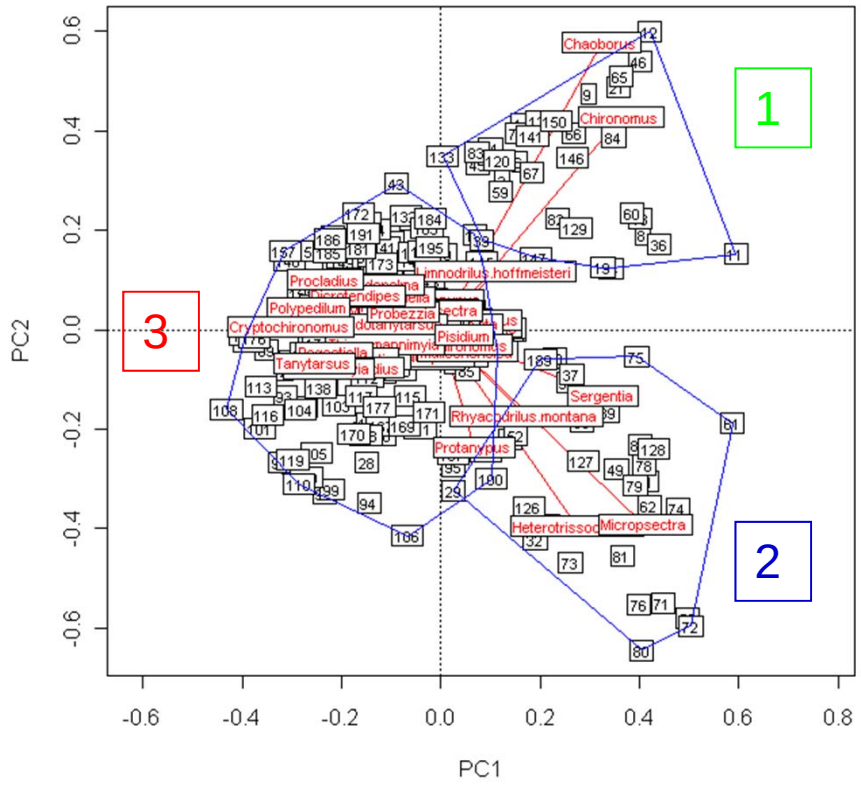
3. Sites seems to be distributed according 3 groups
   ➔ 2nd step : partioning of sites into 3 groups by *kmeans* method

magelis conseil assistance IRSN

THE METHOD WE DEVELOPED with "all data" set

IAEA Vienna 8 september 2010

2nd step : partioning sites into 3 group by kmeans



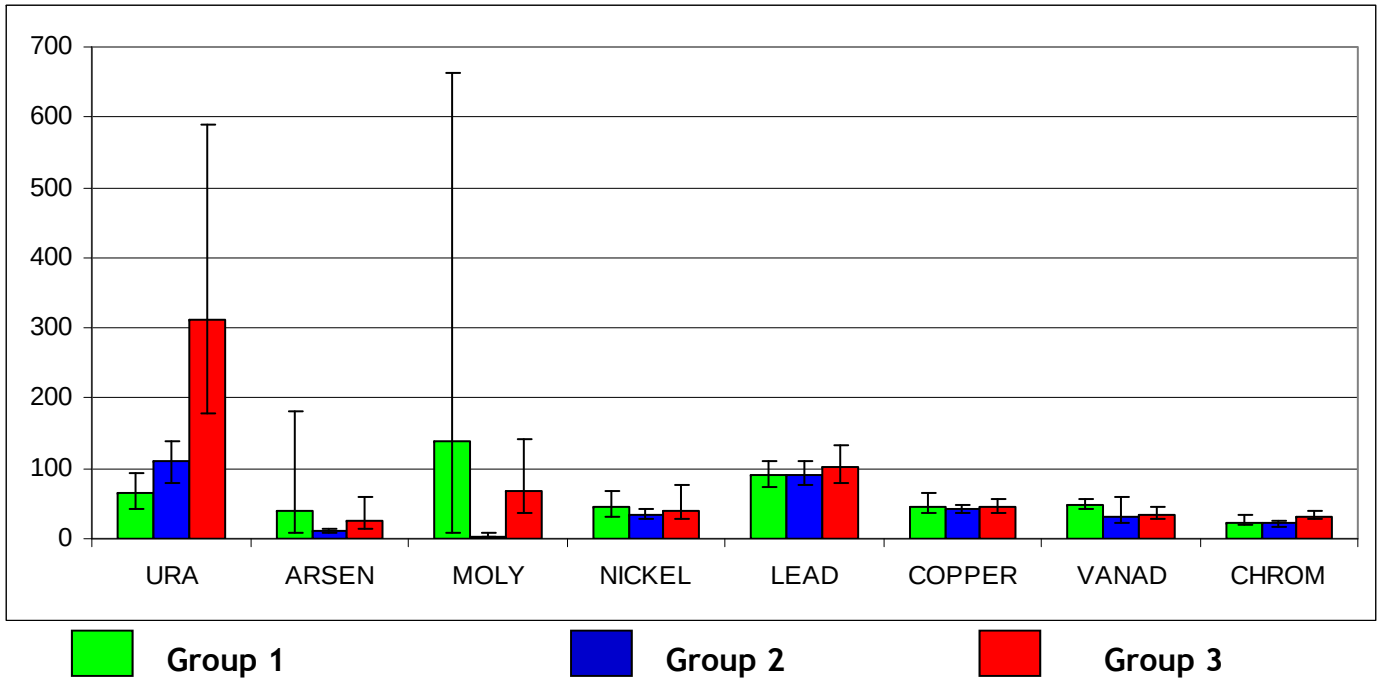
3rd step : characterization of these 3 groups in terms of contaminants concentration

4th step : characterization of these 3 groups in terms of species

magelis conseil assistance IRSN

THE METHOD WE DEVELOPED with "all data" set

3rd step : charaterization of the groups by [] contaminants (means + IC)

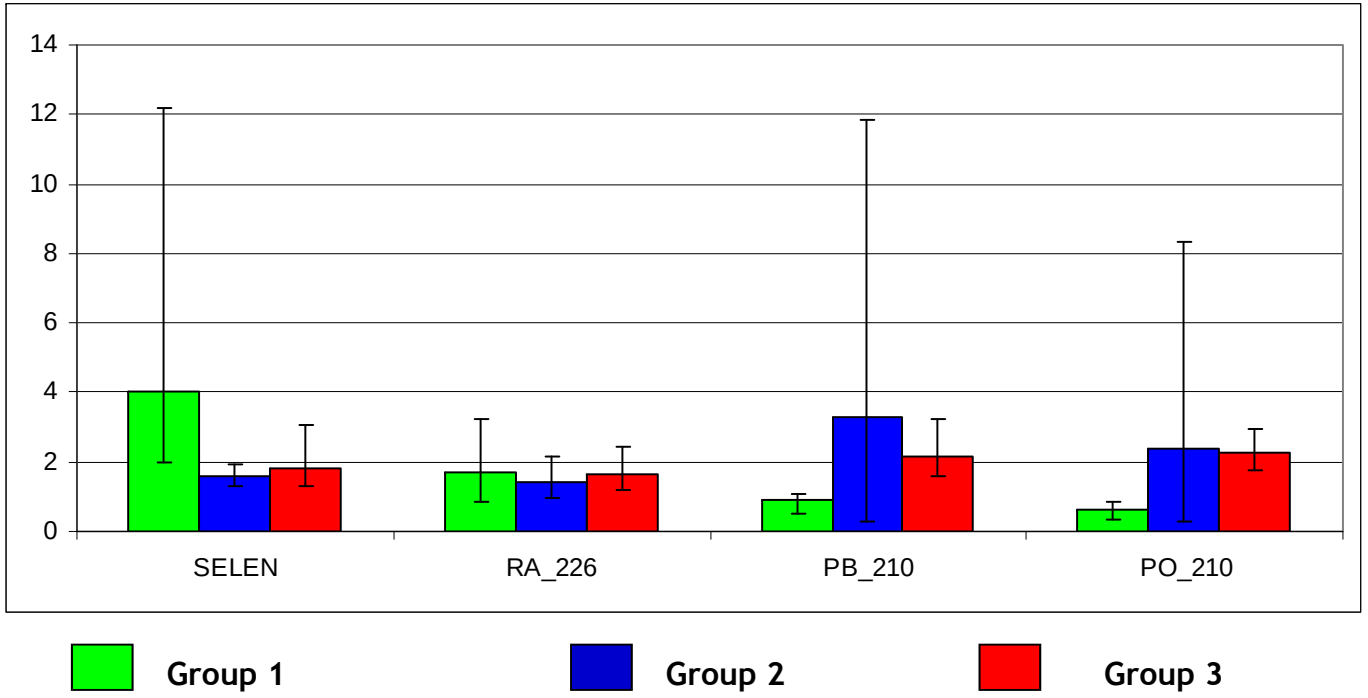


magelis conseil assistance IRSN

IAEA Vienna 8 september 2010

THE METHOD WE DEVELOPED with "all data" set

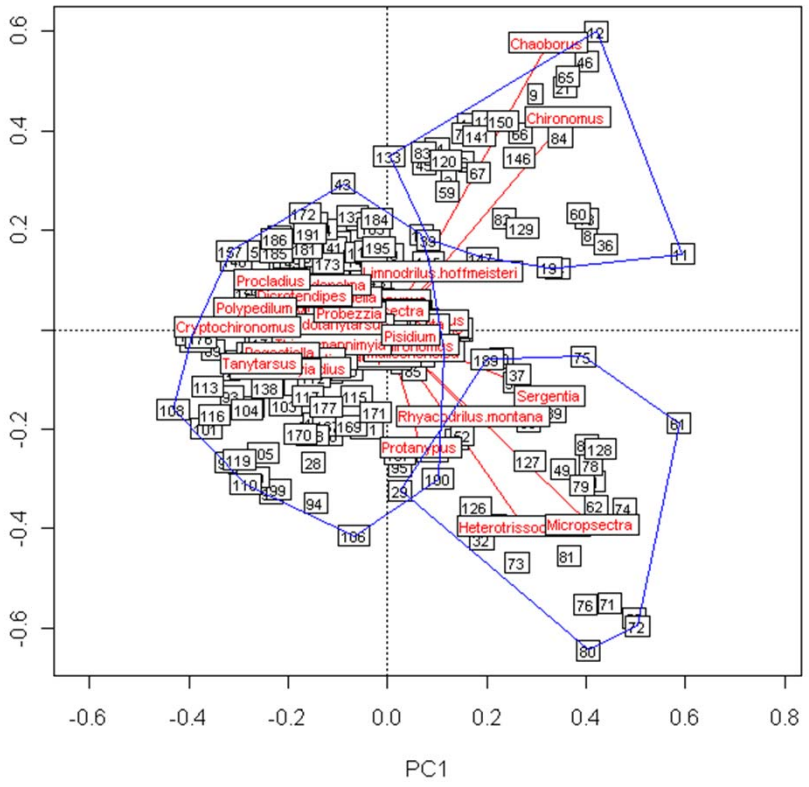
3rd step : charaterization of the groups by [] contaminants (means + IC)



magelis conseil assistance IRSN

THE METHOD WE DEVELOPED with "all data" set

3rd step : charaterization of the groups by [] contaminants (means + IC)



[u] - [PB210]- [PO210]- [Se]+

[u]+ [PB210]+ [PO210]+

[u] - [Se]-

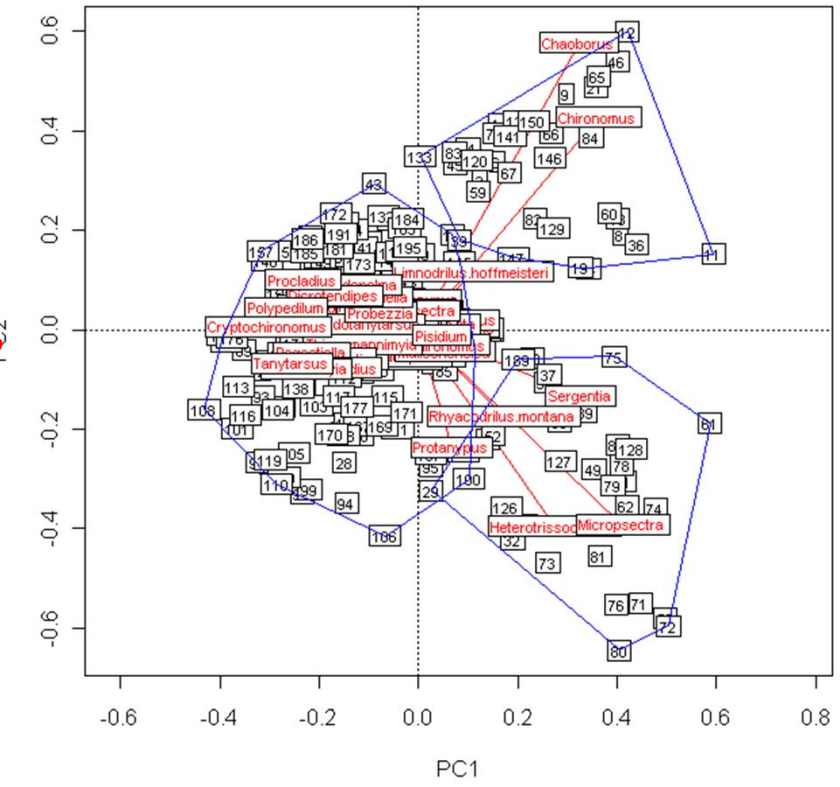
IAEA Vienna 8 september 2010

THE METHOD WE DEVELOPED with "all data" set

IAEA Vienna 8 septembre 2010

4th step : charaterization of the groups by the presence of species

Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%



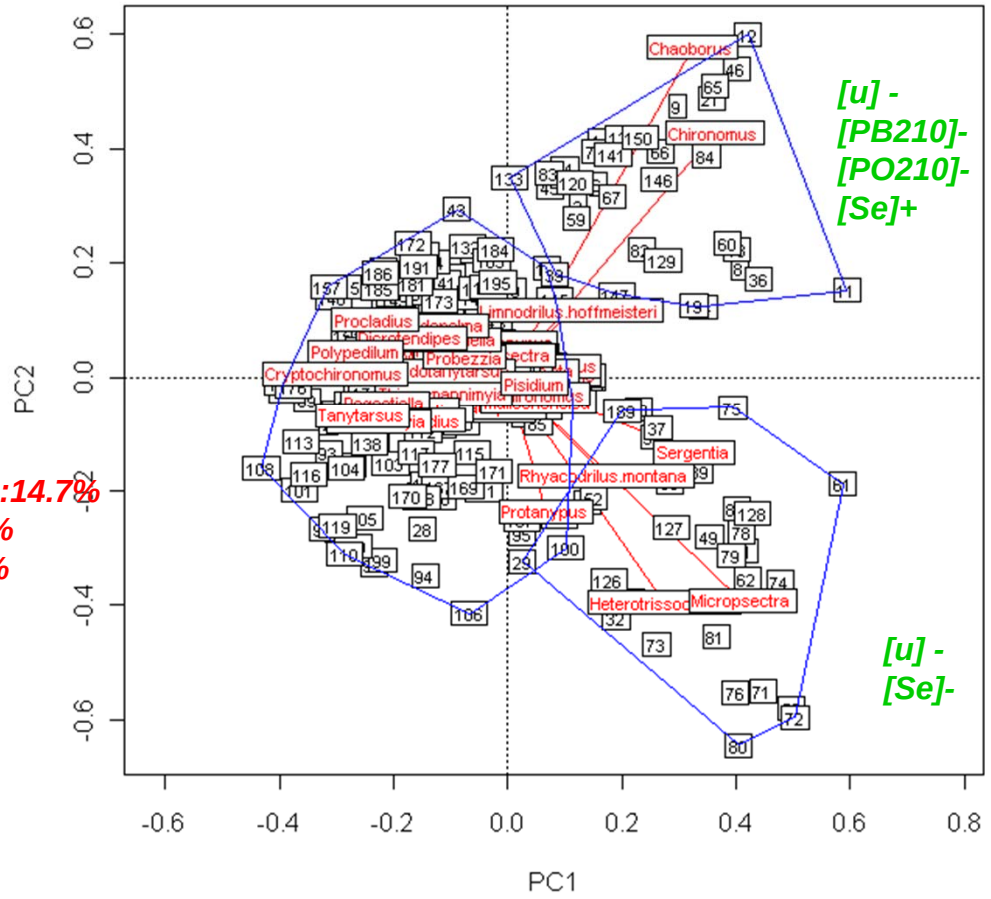
Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

magelis conseil assistance IRSN

THE METHOD WE DEVELOPED with "all data" set

All the results



[u] + [PB210]+ [PO210]+

Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%

[u] - [PB210]- [PO210]- [Se]+

Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

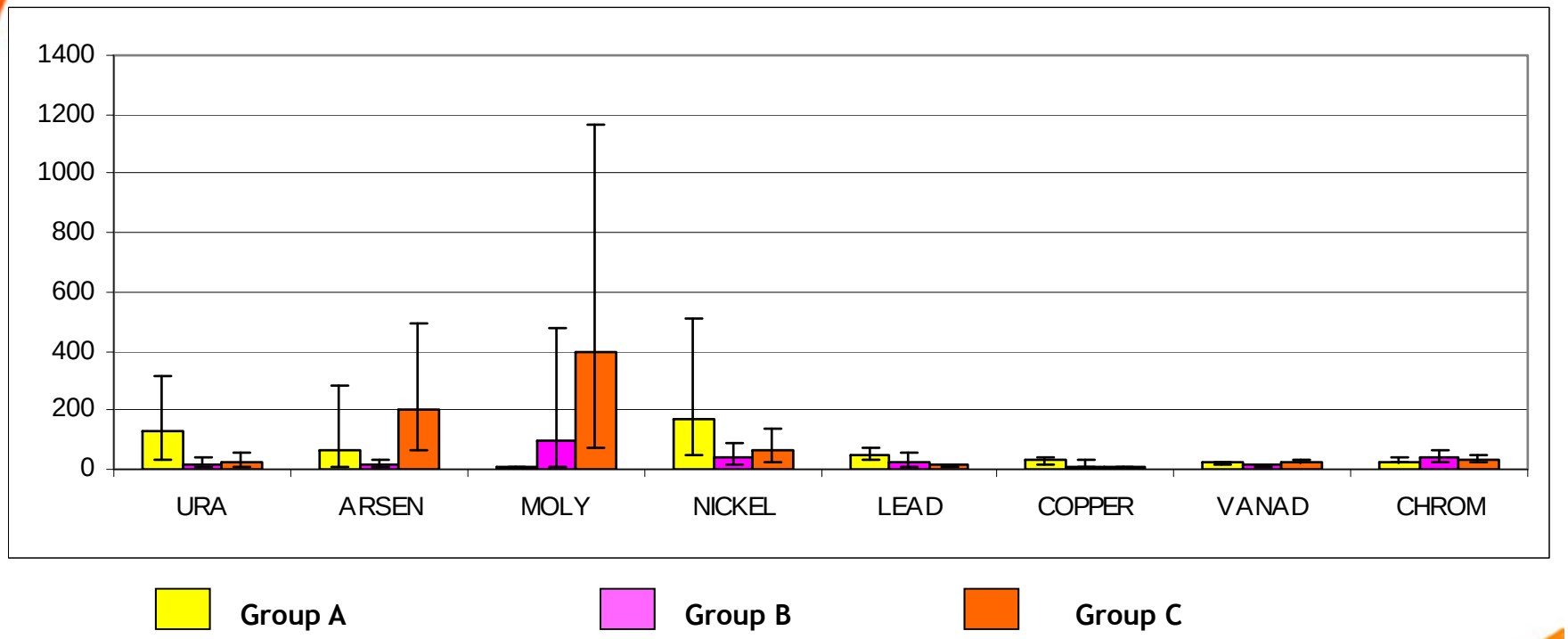
[u] - [Se]-

Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

magelis conseil assistance IRSN

THE METHOD WE DEVELOPED with "complete data" set

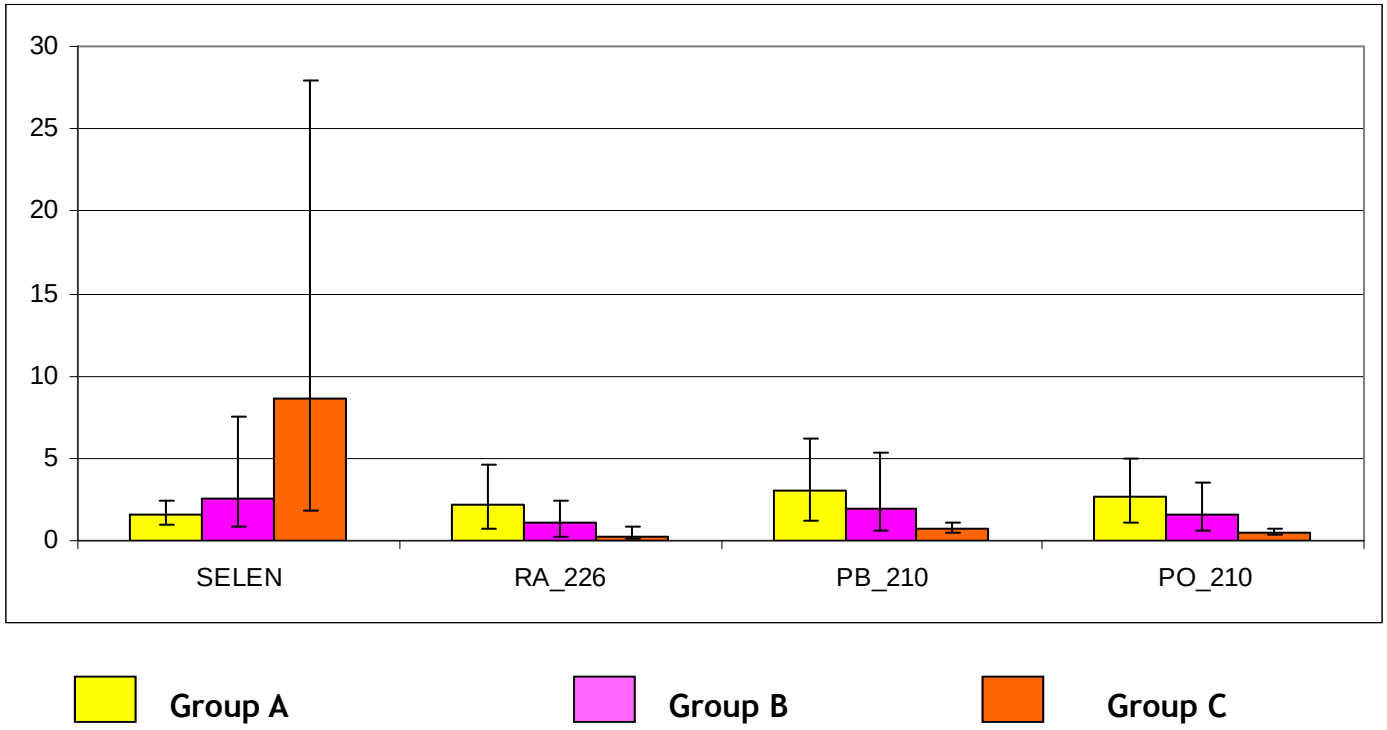
3rd step : charaterization of the groups by [] contaminants (means + IC)



THE METHOD WE DEVELOPED with "complete data" set

IAEA Vienna 8 september 2010

3rd step : charaterization of the groups by [] contaminants (means + IC)



magelis conseil assistance IRSN

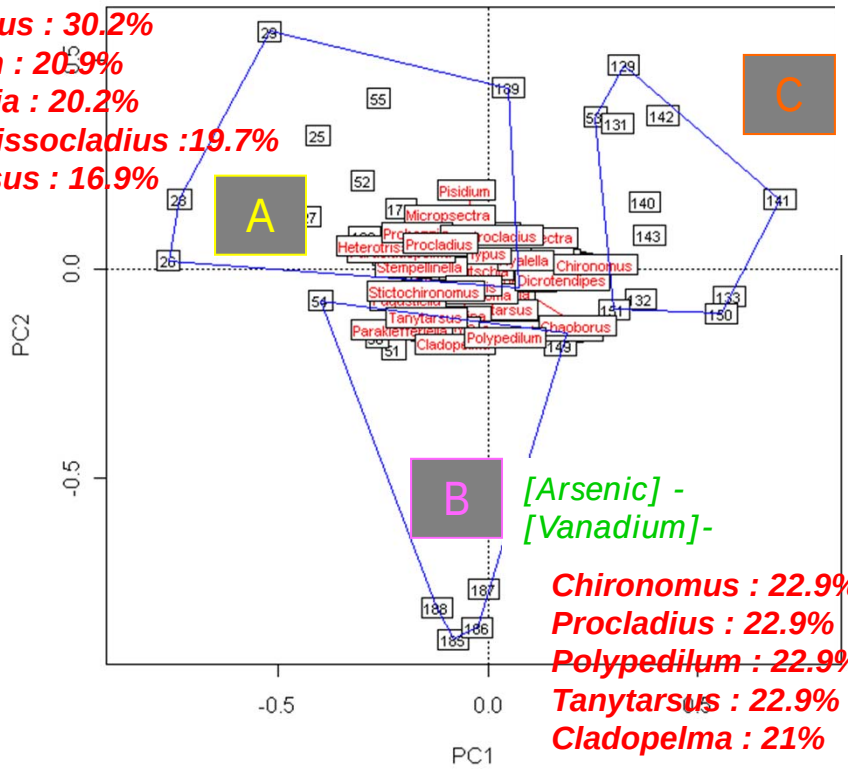
THE METHOD WE DEVELOPED with "complete data" set

IAEA Vienna 8 september 2010

All the results

[Lead]+
[Copper]+
[PB^{210}]+
[P0^{210}]+
[Moly] -

Procladius : 30.2%
Pisidium : 20.9%
Probezzia : 20.2%
Heterotrissocladius :19.7%
Tanytarsus : 16.9%



[Arsenic] +
[Moly]+
[vanadium]+
[PB210]-
[P0210]-
[Lead]-
[Copper]-

Chironomus : 34.3%
Dicrotendipes : 23.4%
Procladius : 20.12%
Chaoborus : 18.2%
Pisidium : 15.7%

[Arsenic] -
[Vanadium]-

Chironomus : 22.9%
Procladius : 22.9%
Polypedilum : 22.9%
Tanytarsus : 22.9%
Cladopelma : 21%

magelis conseil assistance IRSN

RESULTS - 1

IAEA Vienna 8 september 2010

Comparison of the patterns observed

"all data"

[u] +
[PB210]+
[PO210]+
3

Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%

[u] -
[Se]-
2

Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

[u] -
[PB210]-
[PO210]-
[Se]+
1

Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

"complete data"

[Lead]+
[Copper]+
[PB210]+
[P0210]+
[Moly] -
A

Procladius : 30.2%
Pisidium : 20.9%
Probezzia : 20.2%
Heterotrissocladius :19.7%
Tanytarsus : 16.9%

[Arsenic] -
[Vanadium]-
B

Chironomus : 22.9%
Procladius : 22.9%
Polypedilum : 22.9%
Tanytarsus : 22.9%
Cladopelma : 21%

[Arsenic] +
[Moly]+
[vanadium]+
[PB210]-
[P0210]-
[Lead]-
[Copper]-
C

Chironomus : 34.3%
Dicrotendipes : 23.4%
Procladius : 20.12%
Chaoborus : 18.2%
Pisidium : 15.7%

magelis conseil assistance IRSN

RESULTS - 1

IAEA Vienna 8 september 2010

Comparison of the patterns observed

"all data"

3

[u] +
[PB210]+
[PO210]+
Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%

2

[u] -
[Se]-
Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

1

[u] -
[PB210]-
[PO210]-
[Se]+
Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

The only comparisons we can do is between groups having relative high PB^{210} and PO^{210} concentrations and

"complete data"

A

[Lead]+
[Copper]+
[PB210]+
[P0210]+
[Moly] -
Procladius : 30.2%
Pisidium : 20.9%
Probezzia : 20.2%
Heterotrissocladius :19.7%
Tanytarsus : 16.9%

B

[Arsenic] -
[Vanadium]-
Chironomus : 22.9%
Procladius : 22.9%
Polypedilum : 22.9%
Tanytarsus : 22.9%
Cladopelma : 21%

C

[Arsenic] +
[Moly]+
[vanadium]+
[PB210]-
[P0210]-
[Lead]-
[Copper]-
Chironomus : 34.3%
Dicrotendipes : 23.4%
Procladius : 20.12%
Chaoborus : 18.2%
Pisidium : 15.7%

magelis conseil assistance IRSN

RESULTS - 1

Comparison of the patterns observed

"all data"

3

[u] +
[PB210]+
[PO210]+
Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%

2

[u] -
[Se]-
Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

1

[u] -
[PB210]-
[PO210]-
[Se]+
Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

The only comparisons we can do is between groups having relative high PB^{210} and PO^{210} concentrations and low PB^{210} and PO210 concentrations.

"complete data"

A

[Lead]+
[Copper]+
[PB210]+
[PO210]+
[Moly] -
Procladius : 30.2%
Pisidium : 20.9%
Probezzia : 20.2%
Heterotrissocladius :19.7%
Tanytarsus : 16.9%

B

[Arsenic] -
[Vanadium]-
Chironomus : 22.9%
Procladius : 22.9%
Polypedilum : 22.9%
Tanytarsus : 22.9%
Cladopelma : 21%

C

[Arsenic] +
[Moly]+
[vanadium]+
[PB210]-
[P0210]-
[Lead]-
[Copper]-
Chironomus : 34.3%
Dicrotendipes : 23.4%
Procladius : 20.12%
Chaoborus : 18.2%
Pisidium : 15.7%

magelis conseil assistance IRSN

RESULTS - 1

IAEA Vienna 8 september 2010

Comparison of the patterns observed

"all data"

3

[u] +
[PB210]+
[PO210]+
Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%

2

[u] -
[Se]-
Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

1

[u] -
[PB210]-
[PO210]-
[Se]+
Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

The only comparisons we can do is between groups having relative high PB^{210} and PO^{210} concentrations and low PB^{210} and PO^{210} concentrations. Moreover these groups are characterized by the presence of other different contaminants.

"complete data"

A

[Lead]+
[Copper]+
[PB210]+
[PO210]+
[Moly] -
Procladius : 30.2%
Pisidium : 20.9%
Probezzia : 20.2%
Heterotrissocladius :19.7%
Tanytarsus : 16.9%

B

[Arsenic] -
[Vanadium]-
Chironomus : 22.9%
Procladius : 22.9%
Polypedilum : 22.9%
Tanytarsus : 22.9%
Cladopelma : 21%

C

[Arsenic] +
[Moly]+
[vanadium]+
[PB210]-
[P0210]-
[Lead]-
[Copper]-
Chironomus : 34.3%
Dicrotendipes : 23.4%
Procladius : 20.12%
Chaoborus : 18.2%
Pisidium : 15.7%

magelis conseil assistance IRSN

RESULTS - 1

Comparison of the patterns observed

"all data"

3

[u] +
[PB^{210}]+
[PO^{210}]+
Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%

2

[u] -
[Se]-
Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

1

[u] -
[PB^{210}]-
[PO^{210}]-
[Se]+
Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

The only comparisons we can do is between groups having relative high PB^{210} and PO^{210} concentrations and low PB^{210} and PO^{210} concentrations. Moreover these groups are characterized by the presence of other different contaminants. Nevertheless some same combinations can be observed:

"complete data"

A

[Lead]+
[Copper]+
[PB^{210}]+
[PO^{210}]+
[Moly] -
Procladius : 30.2%
Pisidium : 20.9%
Probezzia : 20.2%
Heterotrissocladius :19.7%
Tanytarsus : 16.9%

B

[Arsenic] -
[Vanadium]-
Chironomus : 22.9%
Procladius : 22.9%
Polypedilum : 22.9%
Tanytarsus : 22.9%
Cladopelma : 21%

C

[Arsenic] +
[Moly]+
[vanadium]+
[PB210]-
[P0210]-
[Lead]-
[Copper]-
Chironomus : 34.3%
Dicrotendipes : 23.4%
Procladius : 20.12%
Chaoborus : 18.2%
Pisidium : 15.7%

magelis conseil assistance IRSN

RESULTS - 1

Comparison of the patterns observed

"all data"

3

[u] +
[PB210]+
[PO210]+
Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%

2

[u] -
[Se]-
Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

1

[u] -
[PB210]-
[PO210]-
[Se]+
Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

PO210+, PB210+ : procladius (25.4% and 30.2%)

"complete data"

A

[Lead]+
[Copper]+
[PB210]+
[P0210]+
[Moly] -
Procladius : 30.2%
Pisidium : 20.9%
Probezzia : 20.2%
Heterotrissocladius :19.7%
Tanytarsus : 16.9%

B

[Arsenic] -
[Vanadium]-
Chironomus : 22.9%
Procladius : 22.9%
Polypedilum : 22.9%
Tanytarsus : 22.9%
Cladopelma : 21%

C

[Arsenic] +
[Moly]+
[vanadium]+
[PB210]-
[P0210]-
[Lead]-
[Copper]-
Chironomus : 34.3%
Dicrotendipes : 23.4%
Procladius : 20.12%
Chaoborus : 18.2%
Pisidium : 15.7%

magelis conseil assistance IRSN

RESULTS - 1

Comparison of the patterns observed

"all data"

[u] +
[PB210]+
[PO210]+
3

Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%

[u] -
[Se]-
2

Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

[u] -
[PB210]-
[PO210]-
[Se]+
1

Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

PO210+, PB210+ : procladius (25.4% and 30.2%)

PO210-, PB210- : chironomus (39.1% and 34.3%), chaoborus (45.6% and 18.2%),pisidium (10 and 15.7%)

"complete data"

[Lead]+
[Copper]+
[PB210]+
[P0210]+
[Moly] -
A

Procladius : 30.2%
Pisidium : 20.9%
Probezzia : 20.2%
Heterotrissocladius :19.7%
Tanytarsus : 16.9%

[Arsenic] -
[Vanadium]-
B

Chironomus : 22.9%
Procladius : 22.9%
Polypedilum : 22.9%
Tanytarsus : 22.9%
Cladopelma : 21%

[Arsenic] +
[Moly]+
[vanadium]+
[PB210]-
[P0210]-
[Lead]-
[Copper]-
C

Chironomus : 34.3%
Dicrotendipes : 23.4%
Procladius : 20.12%
Chaoborus : 18.2%
Pisidium : 15.7%

magelis conseil assistance IRSN

RESULTS - 1

IAEA Vienna 8 september 2010

Comparison of the patterns observed

"all data"

3

[u] +
[PB210]+
[PO210]+
Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%

2

[u] -
[Se]-
Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

1

[u] -
[PB210]-
[PO210]-
[Se]+
Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

PO210+, PB210+ : procladius (25.4% and 30.2%)

PO210-, PB210- : Chironomus (39.1% and 34.3%), Chaoborus (45.6% and 18.2%), Pisidium (10 and 15.7%) but Procaldius (21.5% and 20.12%)

"complete data"

A

[Lead]+
[Copper]+
[PB210]+
[P0210]+
[Moly] -
Procladius : 30.2%
Pisidium : 20.9%
Probezzia : 20.2%
Heterotrissocladius :19.7%
Tanytarsus : 16.9%

B

[Arsenic] -
[Vanadium]-
Chironomus : 22.9%
Procladius : 22.9%
Polypedilum : 22.9%
Tanytarsus : 22.9%
Cladopelma : 21%

C

[Arsenic] +
[Moly]+
[vanadium]+
[PB210]-
[P0210]-
[Lead]-
[Copper]-
Chironomus : 34.3%
Dicrotendipes : 23.4%
Procladius : 20.12%
Chaoborus : 18.2%
Pisidium : 15.7%

magelis conseil assistance IRSN

RESULTS - 1

IAEA Vienna 8 september 2010

Comparison of the patterns observed

"all data"

3

[u] +
[PB210]+
[PO210]+
Procladius : 25.4%
Tanytarsus : 21%
Cryptochironomus :14.7%
Polypedilum : 13.1%
Chironomus : 11.3%

2

[u] -
[Se]-
Microspectra : 39.9%
Heterotrissocladius :33%
Sergentia :21.4%
Chironomus : 19.9%
Ryacodrilus montana :15.2%

1

[u] -
[PB210]-
[PO210]-
[Se]+
Chaoborus : 45.6%
Chironomus : 39.1%
Procladius : 21.5%
Limnodrilus : 12.6%
Pisidium :10.3%

PO210+, PB210+ : procladius (25.4% and 30.2%) but chironomus (11.3 (all data))

PO210-, PB210- : Chironomus (39.1% and 34.3%), Chaoborus (45.6% and 18.2%), Pisidium (10 and 15.7%) but Procaldius (21.5% and 20.12%)

"complete data"

A

[Lead]+
[Copper]+
[PB210]+
[PO210]+
[Moly] -
Procladius : 30.2%
Pisidium : 20.9%
Probezzia : 20.2%
Heterotrissocladius :19.7%
Tanytarsus : 16.9%

B

[Arsenic] --
[Vanadium]--
Chironomus : 22.9%
Procladius : 22.9%
Polypedilum : 22.9%
Tanytarsus : 22.9%
Cladopelma : 21%

C

[Arsenic] +
[Moly]+
[vanadium]+
[PB210]-
[PO210]-
[Lead]-
[Copper]-
Chironomus : 34.3%
Dicrotendipes : 23.4%
Procladius : 20.12%
Chaoborus : 18.2%
Pisidium : 15.7%

magelis conseil assistance IRSN

SUMMARY OF RESULTS - 1

IAEA Vienna 8 september 2010

Comparison of the results obtained using our developed method

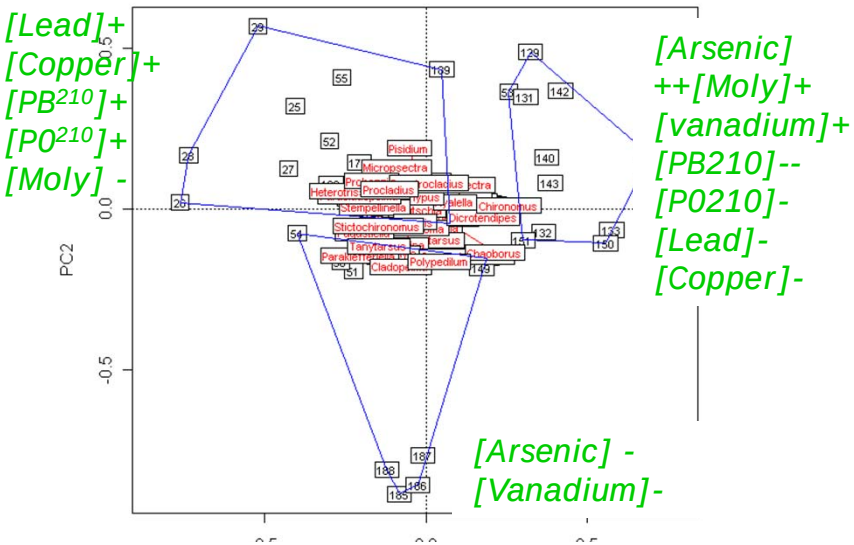
1. The two first principal components represent almost the same percentage of variation for both datasets (19.37 for 'all data' and 24.3 for 'complete data')
2. The comparisons of the 'contaminant-species' associations are weak because
   a) there are only 2 comparisons
   b) they concern the same contaminants (Po^{210}-Pb^{210}) which are at high concentrations in one case and low concentration in the other case
3. Nevertheless some same 'contaminants-species' combinations appear :
   - ➔ **our method is able to detect some 'contaminants-species' combinations**
   - ➔ **our method is able to assess the influence of contaminants on the species diversity distribution**
4. Some species overlap onto two groups :
   - ➔ **lack of accuracy for our method ?**
   - ➔ **due to the mixture of chemicals ?**

magelis conseil assistance IRSN

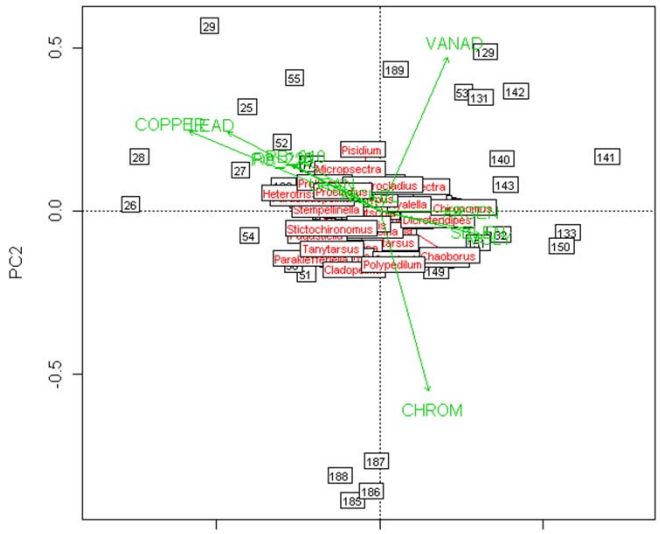
RESULTS - 2

Comparison of the methods (with "complete data" set)

These two methods give two different perceptions. In "PCA+vectors fitting", sites, species and vectors are considered individually, positioned one in regard of the others. In the method we developed, what is considered is groups of sites, which are characterized in terms of contaminants and species.



developed method



"PCA + vectors fitting"

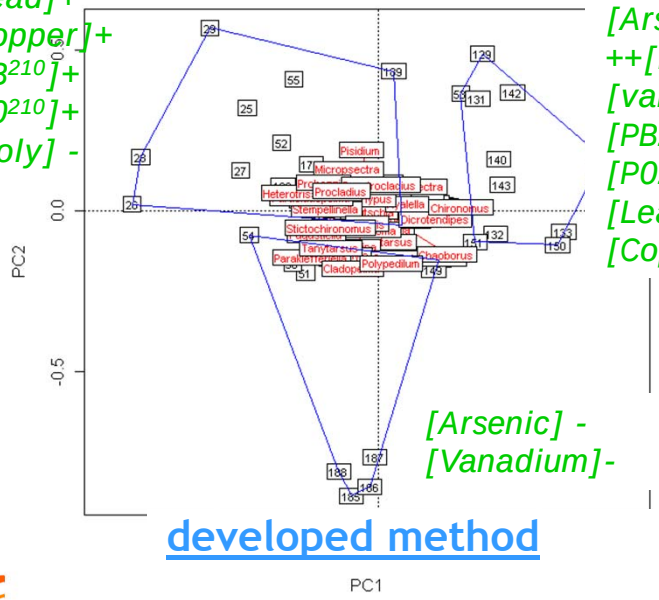
magelis conseil assistance IRSN

RESULTS - 2

Comparison of the methods (with "complete data" set)

Nevertheless, when a contaminant turns out to have a big influence according to the PCA method and when its coordinates on the biplot place it in the middle of a group of sites of the developed method, then this last one is also able of bringing to light it influence (i.e. lead, copper). But when the power is weak it happens that the developed method don't detect it (i.e. radium, selenium, uranium) and when the contaminant vector is between two groups it don't detect it too(i.e chrome)

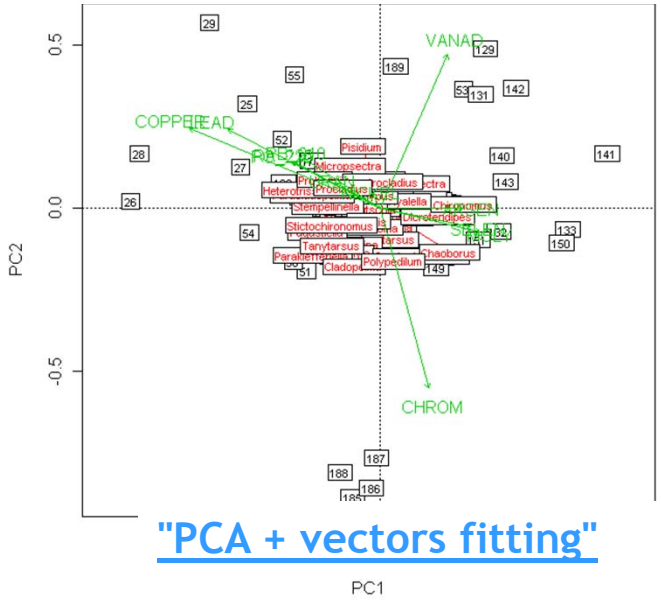
[Lead]+
[Copper]+
[PB^{210}]+
[P0^{210}]+
[Moly] -



[Arsenic]
++[Moly]+
[vanadium]+
[PB210]--
[P0210]-
[Lead]-
[Copper]-

[Arsenic] -
[Vanadium]-

developed method



"PCA + vectors fitting"

CONCLUSION

	Developed method	"PCA + vector fitting" or RDA
Strength	-Take all the available information (less influence of possible confusing variables)	- consider sites, species and contaminants individually -bring to light association of several species with a specific contaminant - consider the presence and the absence of species
Weakness	-doesn't consider sites, species and contaminants individually but into groups -doesn't bring to light association of several species with a specific contaminant but with a mixture of contaminants -Doesn't order the contaminants according to their “power” on the species distribution -For the moment only permits to consider the presence of a species (not the absence)	- can only be used with no NA values, so loss of information and perhaps influence of possible confusing variables on the results

PERSPECTIVES

IAEA Vienna 8 september 2010

About the screening level concentration method

1. Steve Mihok would like to estimate confidence interval of SEL and LEL of contaminants by bootstrap approach

2. In the publication, change in abundance (mean number of species) and species richness (mean number of taxa at genus and species level) are used to categorize 21 studied sites into "severely impacted" , "middly" and "not adversely affected" groups. If we could have the list of these groups, we could use 'linear discriminant analysis" to assess how the contaminant variables contribute to this classification.

magelis conseil assistance IRSN

PERSPECTIVES

IAEA Vienna 8 september 2010

Your views are needed about the possibility of pursuing in the analysis of the whole available information, i.e. in the improvement of the developed method. If yes :

1. We could think about:
   a) how to represent absence of species into a group in regard of another one group of sites
   b) how to choose the number of group determined by *kmeans* method according to an objective criterion

2. We also could try to deal with the missing values, perhaps using imputation techniques.

3. If we could have other information for each site such as abundance (number of observed individuals for each specie) we could try to improve our prediction of "contaminants-species" association too.

magelis conseil assistance IRSN

PERSPECTIVES

IAEA Vienna 8 september 2010

Your views are needed about the possibility to produce a paper on the influence of the contaminants on the benthos distribution

1. We could write a paper explaining the analysis of the 31 complete sites by means of classical ordination methods (RDA, PCA and vectors fitting). If we could have the information concerning the 21 sites previously studied by CNSC (or even these 31 sites) we could add a linear discriminant analysis to assess how the contaminant variables contribute to the "severely impacted" , "middly" and "not adversely affected" classification.

2. The writing of a paper explaining the analysis of the "all data" set by means of the method we developed seems, at present, premature

magelis conseil assistance IRSN

REFERENCES

Bouchon-Navaro, Y., Bouchon, C., Louis, M., Legendre, P., 2005. Biogeographic patterns of coastal fis assemblages in the West Indies. J. exp; Mar. Biol. Ecol. 315, 31-47.

Legendre, P.,Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129, 271-280

Legendre, P.,Legendre, L., 1998. Numerical Ecology. 2nd English ed. Elsevier, Amsterdam

Oksanen, J., 2010. Multivariate Analysis of Ecological Communities in R : vegan tutorial.

Crawley, M.J., 2007. The R Book. Ed Wiley, England

Efron, B., Tibshirani, R. B. 1993. An introduction to the bootstrap. Ed Chapman & Hall/CRS, New York.

More about…

IAEA Vienna 8 september 2010

The Hellinger transformation :

"consists in expressing each presence as a fraction of the total number of species observed at the site and taking the square root of the fraction. Legendre and Gallagher (2001) have shown this transformation makes species presence-absence or abundance data amenable to linear ordination methods such as principal component analysis (PCA) or canonical redundancy analysis (RDA)" (Bouchon_Navaro et al 2005).

Hellinger transformation of the species data allow ecologists to use ordination methods such as PCA and RDA, which are Euclidean-based, with community composition data containing many zeros (long gradient).

magelis conseil assistance IRSN

More about…

IAEA Vienna 8 september 2010

Partioning by kmeans

"Partitioning consists in finding a single partition of a set of objects. (…) given n objects in a p-dimensional space, determine a partition of the objects into K groups, or clusters, such that objects within each cluster are more similar to one another than to objects in the clusters. The number of groups, K, is determined by the user" ((Legendre et al 1998).

"The kmeans function [of R software] operates on a dataframe in which columns are variables and the row are the individuals. Group membership is determined by calculating the centroid for each group. This is the multidimensionnal equivalent of the mean. Each individual is assigned to the group with the nearest centroid. The kmeans function fits a user-specified numbers of cluster centres, such that the within-cluster sum of squares from these centre is minimized, based on Euclidean distance.

magelis conseil assistance IRSN

More about…

IAEA Vienna 8 september 2010

Bootstrap

The Bootstrap is a statistical method of re-sample classically used to estimate standard error or confidence interval of parameters (e.g. mean). Some of bootstrap approaches are parametric (e.g. based on specific distributions), and others are non parametric (e.g. based on empirical distributions).

magelis conseil assistance IRSN