COMPARISON OF SPMDs AND BIOTIC SAMPLERS USING GNOSTIC ANALYSIS

Institute of Public Health, Ostrava, Czech Republic
National reference laboratory for POPs

Tomas Ocelka, Pavel Kovanic

Tomas.Ocelka@zuova.cz

TOPICS

- Sampling methods to be compared
- Objects of measuring
- Problems of analysis
- Gnostic analysis
- Methods’ features to be compared
- Results of comparison

2

Geographic location

3

Centre laboratories, accreditation

- Personnel: over 140, 5+2 workplaces
- According to ČSN EN ISO/IEC 17 025
  - Over 200 parameters, PCDD/Fs, PCBs, OCPs, PBDE, ….
  - Recognized by ILAC, EA, IAF
- Sampling and Testing
  - Integral – water
    - SPMDs
    - DGTs
    - POCIS
  - Biotic organisms
- Intercalibration
  - Czech + International
- Data analysis (univariate/multivariate)
  - Statistical
  - Gnostic

4

Instrumentation
(worth over 6 mil. USD)

GC-MS/MS (ion-trap)

GCQ, Polaris

Since 1996 (starting to POPs issue)

GC-HRMS (POPs)

- MAT 95XP
- since 2003

LC-MS/MS (pharmacy, pesticides)

- ThermoFinigan
- since 2006

5

Data source for comparison of methods

- All rivers within Czech Republic scale (15)
- 21 sampling profiles
- Complementary to biotic sampling system (since 1999) with abiotic (SPMDs, DGTs, POCIS) – since 2003
- Aims
  - Pilot application 2 years before routine application
  - Parallel exposure of Dreissena Polymorpha, Benthos, Plants
  - POPs (basic: OCPs, PCBs)
  - POPs (other: PCBs – cong., PCDD/Fs, PAHs, PBDEs)

SAMPLING METHODS
TO BE COMPARED

Three biotic methods:

- Bentos
- Dreissena
- Plants

One abiotic method: SPMD (Semipermeable Membrane Measuring Device)

7

The selection

Concentrations of selected permanent organic pollutants (POPs) in several locations of Elbe river in Czech Republic:

p.p.DDE, PCB138, PCB180, PCB101, PCB28.31, p.p.DDT, p.p.DDD, PCB52, PCB118

8

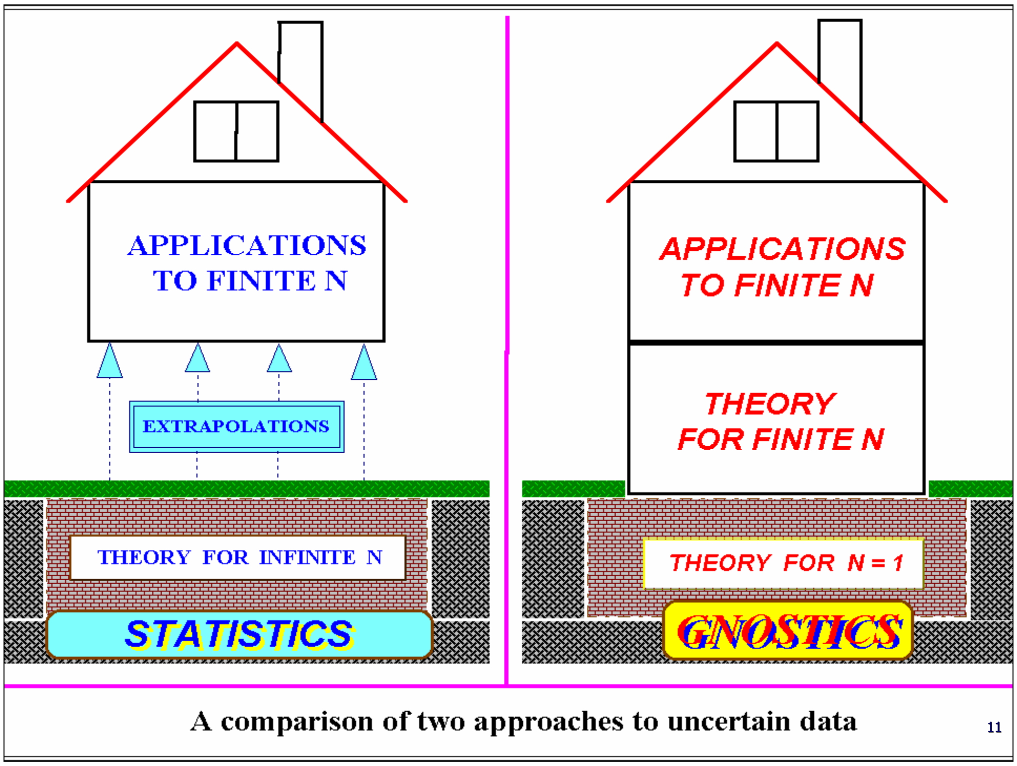
PROBLEMS OF ANALYSIS

- Small data samples
- Different mean concentrations
- Strong variability
- Different length of data vectors
- Data censoring (eg data below the LOD)
- Non-homogeneous and outlying data

SPECIFICS
of MATHEMATICAL GNOSTICS

- Theory of individual data
  and small data samples
- Realistic assumptions
- Uncertainty: a lack of knowledge
- "Let data speak for themselves"
- Results maximizing information
- Natural robustness

10


APPLICATIONS
TO FINITE N
EXTRAPOLATIONS
THEORY FOR INFINITE N
STATISTICS
APPLICATIONS
TO FINITE N
THEORY
FOR FINITE N
THEORY FOR N = 1
GNOSTICS
A comparison of two approaches to uncertain data
11

GNOSTIC DISTRIBUTION FUNCTIONS

- No a priori model (everything from data)
- Maximum information
- Robustness in estimation of probability, quantiles, scale and location parameters, bounds of data support, and membership interval
- Robust correlations

12

GNOSTIC DISTRIBUTION FUNCTIONS II

- Data homogeneity tests
- Marginal cluster analysis
- Cross-section filtering
- Applicability to censored data
- Applicability to heteroscedastic data

13

QUALITY OF METHODS TO BE COMPARED

- Relative sensitivity (treshold, range)
- Homogeneity of results
- Consistency of results
  - Internal (of method's own results)
  - External (mutual consistency of methods)
- Informativeness of results
- Precission

14

RELATIVE SENSITIVITY

Method's relative sensitivity depends:

- On the pollutant's concentration
- On the method's measuring domain

RS = (1 – NC/N) x 100 (%)

NC … number of data in the interval [sensitivity threshold, max(range)]

N … all data of the sample

15

HOMOGENIZATION

TO BE OR NOT TO BE?

Homogeneous data:

- the same origin of true values
- the same nature of the uncertainty

To homogenize?

- **Pros:**
  - More certain main cluster
- **Cons:**
  - Possible loss of information

Rule: homogenize and verify

16

MEASURABILITY

Homogenization … elimination of outliers

Meas = (1 – (NL+NU)/N) x 100 (%)

NL … number of lower outliers

NU … number of upper outliers

N … number of the sample’s data

N – NL – NU … data of the main cluster

17

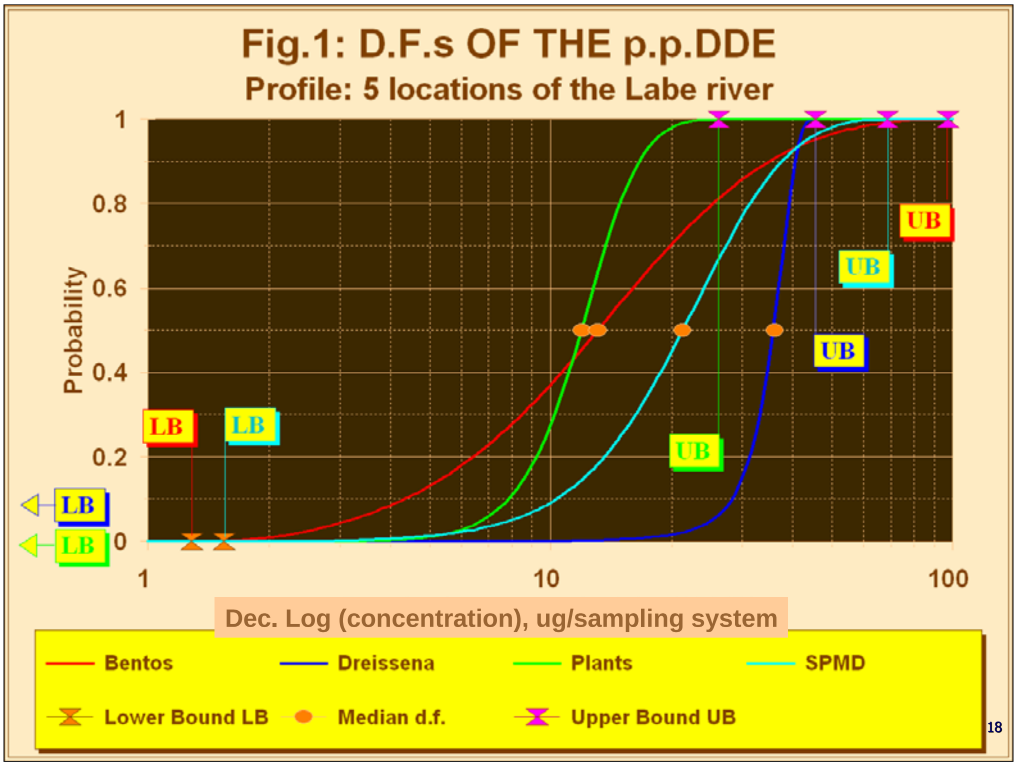


Fig.1: D.F.s OF THE p.p.DDE
Profile: 5 locations of the Labe river
Probability
Dec. Log (concentration), ug/sampling system
LB
UB
Bentos
Dreissena
Plants
SPMD
Lower Bound LB
Median d.f.
Upper Bound UB
18

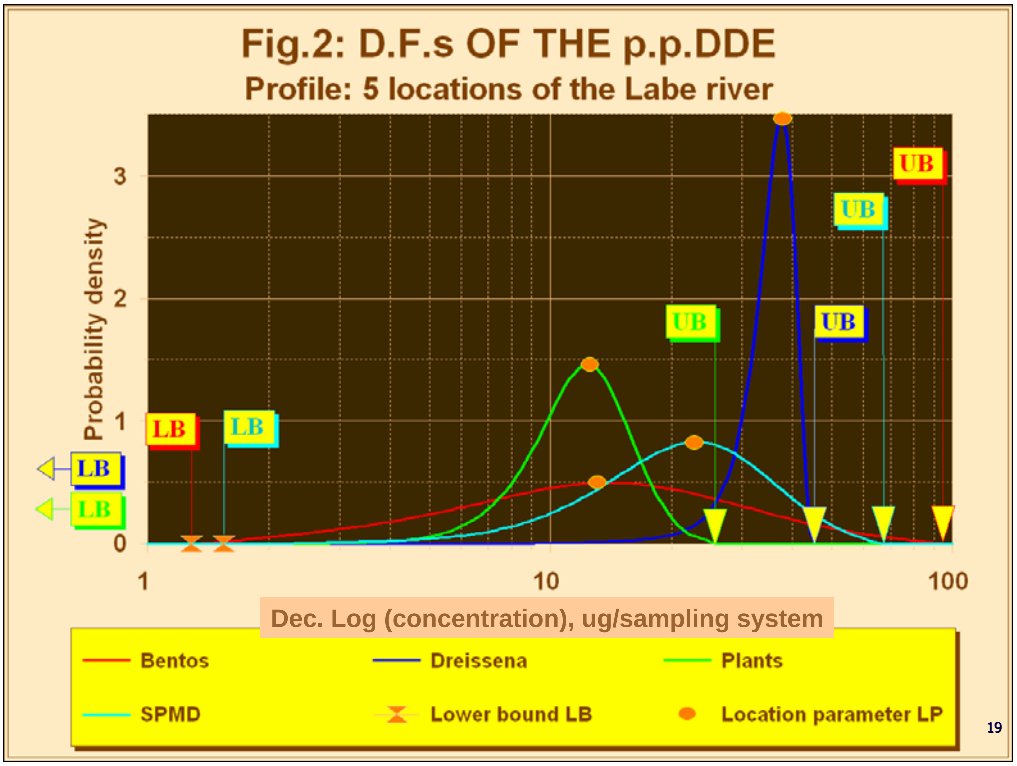


Fig.2: D.F.s OF THE p.p.DDE
Profile: 5 locations of the Labe river
Probability density
0
1
2
3
1
10
100
LB
LB
LB
LB
UB
UB
UB
UB
UB
Dec. Log (concentration), ug/sampling system
Bentos
Dreissena
Plants
SPMD
Lower bound LB
Location parameter LP
19

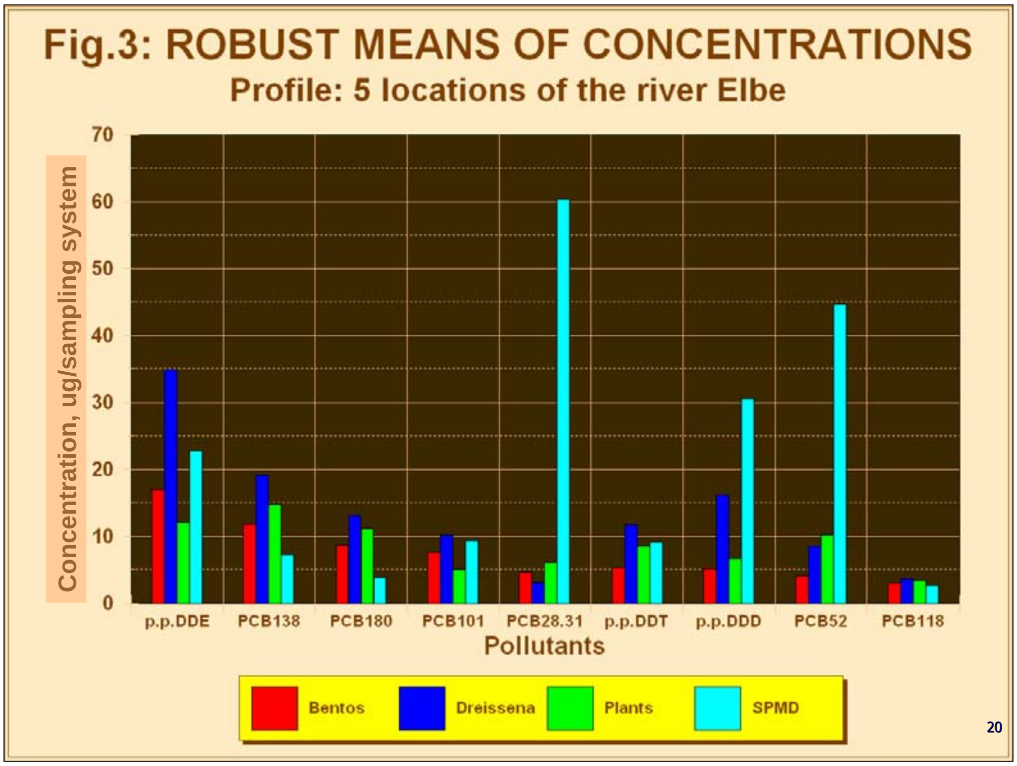


Fig.3: ROBUST MEANS OF CONCENTRATIONS
Profile: 5 locations of the river Elbe
Concentration, ug/sampling system
70
60
50
40
30
20
10
0
p.p.DDE
PCB138
PCB180
PCB101
PCB28.31
p.p.DDT
p.p.DDD
PCB52
PCB118
Pollutants
Bentos
Dreissena
Plants
SPMD
20

DIFFERENCES IN METHODS

- Different accumulation of pollutants:
  - different mean concentrations
  - different variabilities
- Different relations between means
- Rare exception: agreement in PCB118
- Impact of outliers to SPMD? ***NO!***

21

METHOD’S CONSISTENCY

Methods are *consistent* when they give

similar results

Measuring of similarity:

Correlations, or (more generally)

mean angles between vectors of results

SIMcc = 100 x correl.coefficient (%)

SIMqa = 100 x (1 – |Ang|/180) (%)

22

GNOSTIC CORRELATIONS

Data error in gnostic: *irrelevance*

$$ir = (2p - 1)/2$$

p … probability of the data item.

Correlation coefficient of two samples:

$$Gcc(M,N) = cc\{ir(m),ir(n)\}$$

(m in M, n in N), cc{ ..} statist. cor.coef.

Robustness:

$$-1 <= ir <= +1$$

23

SIGNIFICANCE OF CORRELATIONS

- Problems: false statistical model (normality?!, finite data support), small data samples, unrobustness
- Gnostic estimating of significance:
  - fast, auxiliary: using Spearman's robust estimate of significance
  - carefully: distribution function of correlation coefficients

24

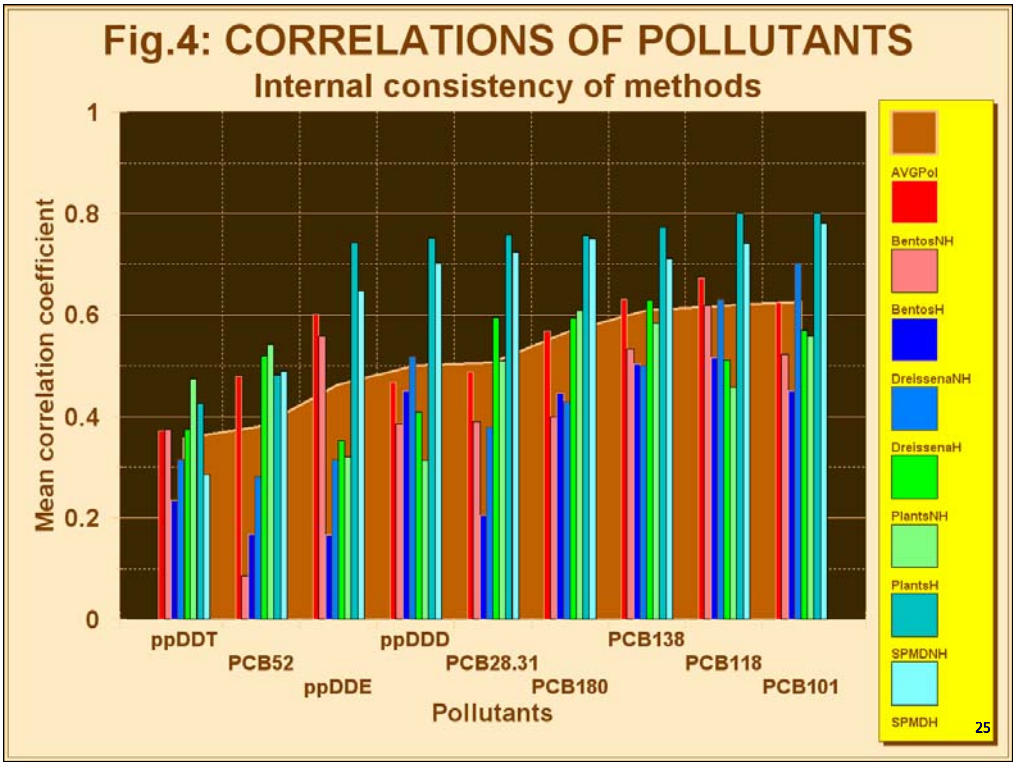


Fig.4: CORRELATIONS OF POLLUTANTS
Internal consistency of methods
Mean correlation coefficient
1
0.8
0.6
0.4
0.2
0
ppDDT
PCB52
ppDDE
ppDDD
PCB28.31
PCB180
PCB138
PCB118
PCB101
Pollutants
AVGPol
BentosNH
BentosH
DreissenaNH
DreissenaH
PlantsNH
PlantsH
SPMDNH
SPMDH
25

QUANTILE VECTORS

- Make sample's distribution function
- Set a series of probabilities *p1,…,pN*
- Find quantiles *q1,…,qN* so that *P{qk}=pk*
- Take *q1,…,qN as a quantile vector*

Advantages:

Robustness, making use of censored data, independence of data amount and of mean data value, filtering effect.

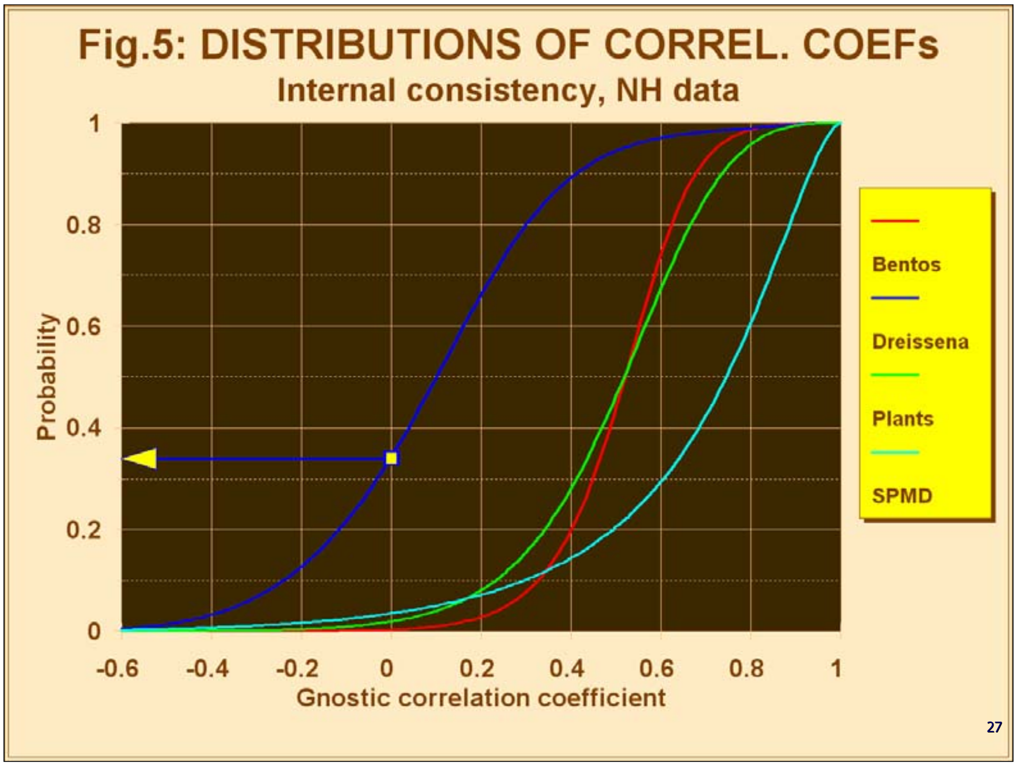


Fig.5: DISTRIBUTIONS OF CORREL. COEFs
Internal consistency, NH data
Probability
1
0.8
0.6
0.4
0.2
0
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
Gnostic correlation coefficient
Bentos
Dreissena
Plants
SPMD
27

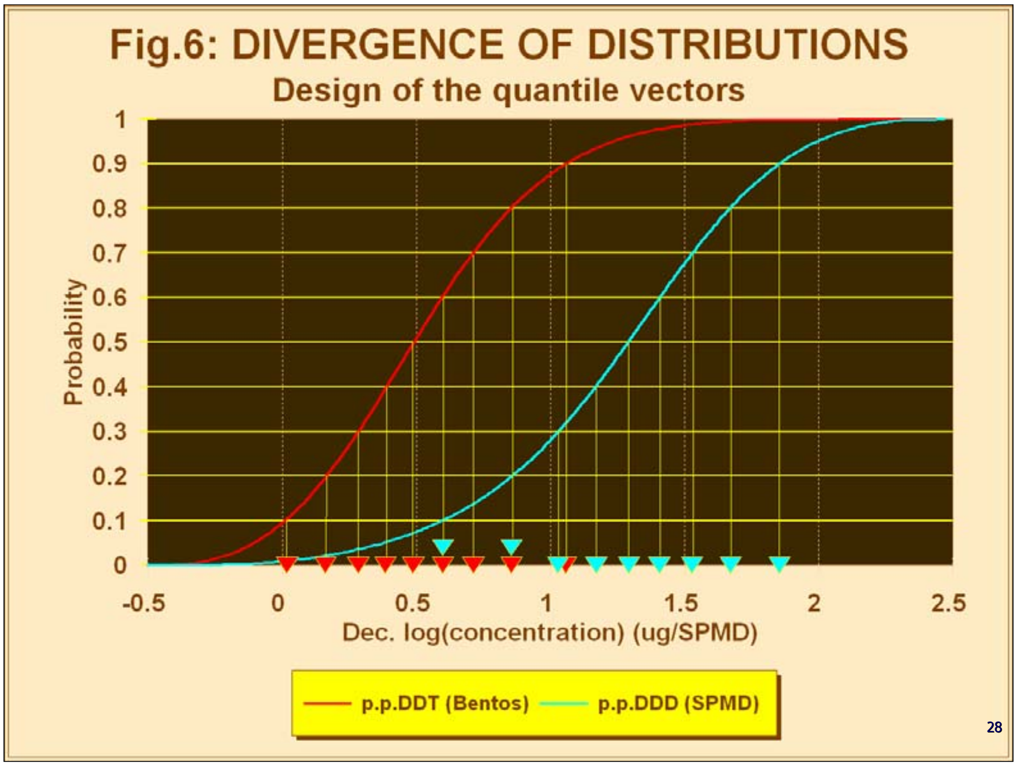


Fig.6: DIVERGENCE OF DISTRIBUTIONS
Design of the quantile vectors
Probability
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.5
0
0.5
1
1.5
2
2.5
Dec. log(concentration) (ug/SPMD)
p.p.DDT (Bentos)
p.p.DDD (SPMD)
28

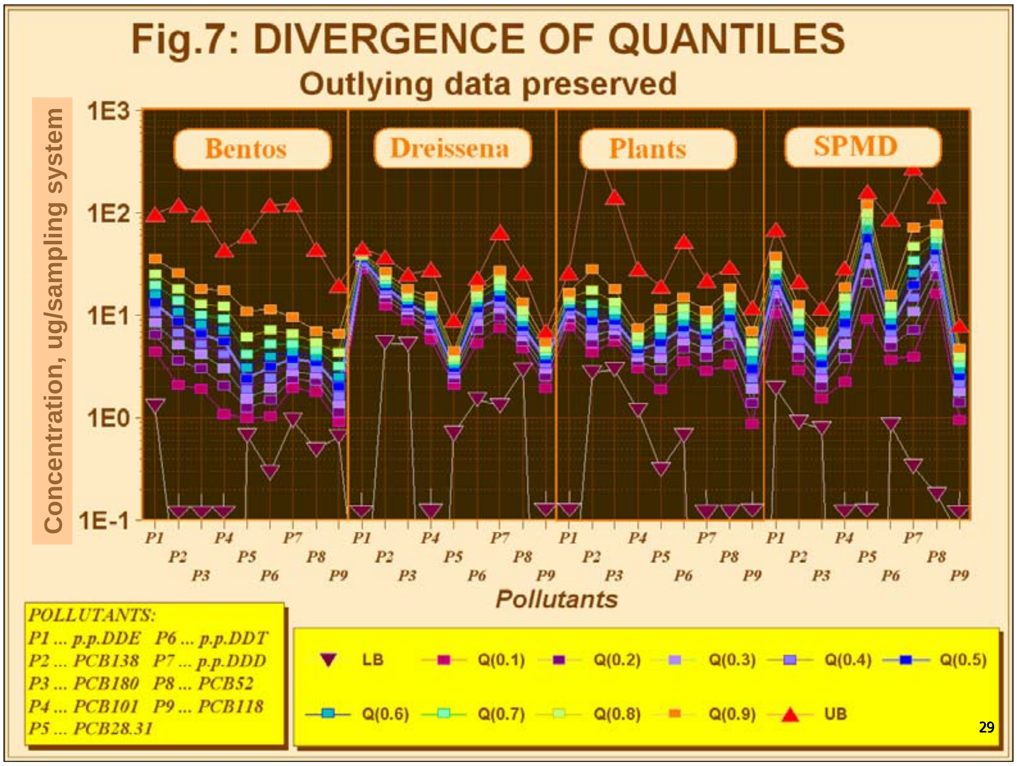


Fig.7: DIVERGENCE OF QUANTILES
Outlying data preserved
Concentration, ug/sampling system
1E3
1E2
1E1
1E0
1E-1
Bentos
Dreissena
Plants
SPMD
P1 P2 P3 P4 P5 P6 P7 P8 P9
Pollutants
POLLUTANTS:
P1 ... p.p.DDE P6 ... p.p.DDT
P2 ... PCB138 P7 ... p.p.DDD
P3 ... PCB180 P8 ... PCB52
P4 ... PCB101 P9 ... PCB118
P5 ... PCB28.31
LB
Q(0.1)
Q(0.2)
Q(0.3)
Q(0.4)
Q(0.5)
Q(0.6)
Q(0.7)
Q(0.8)
Q(0.9)
UB
29

EXTERNAL CONSISTENCY

Approaches:

- Correlations
- Angles between MD-vectors of means
- Angles between quantile vectors
- Conjunction of typical data intervals
- Conjunction of data supports

INTERVAL ANALYSIS

1) Distribution functions
2) Interval analysis:
    a) Data support *(LB, UB)*
    b) Membership interval *(LSB, USB)*
    c) Interval of typical data *(ZL, UL)*
    d) Tolerance interval *(Z0L, Z0U)*
3) Overlapping:

100xconjunction(I1, I2)/union(I1,I2) (%)

31

INFORMATIVENESS

1) Data sample
2) Distribution function
3) Probability *p* of an individual data item
4) Information of the data item:

   Info=(p log(p) + (1-p)log(1-p))/log(1/2)

5) Informativeness of a data sample:

   100 x Mean(Info) (%)

32

EVALUATION OF PRECISION

- Weak variability:

  $$Prec = 100 \times (1 - STD/AVG)\ (\%)$$

  (*STD* … standard deviation, *AVG* … mean)

- Strong uncertainty:

  $$Prec = 100 \times (1 - Mean(GW))\ (\%)$$

  (*GW* … gnostic weight of data; entropy change caused by the uncertainty)

  $$0 <= GW <= 1$$

33

SUMMARY COMPARISON

	Averige of 14 evaluations	
Method	Non-hom.data	Homog. data
Bentos	60.9 %	62.7 %
Dreissena	64.5 %	67.5 %
Plants	64.2 %	68.9 %
SPMD	67.5 %	69.5 %

34

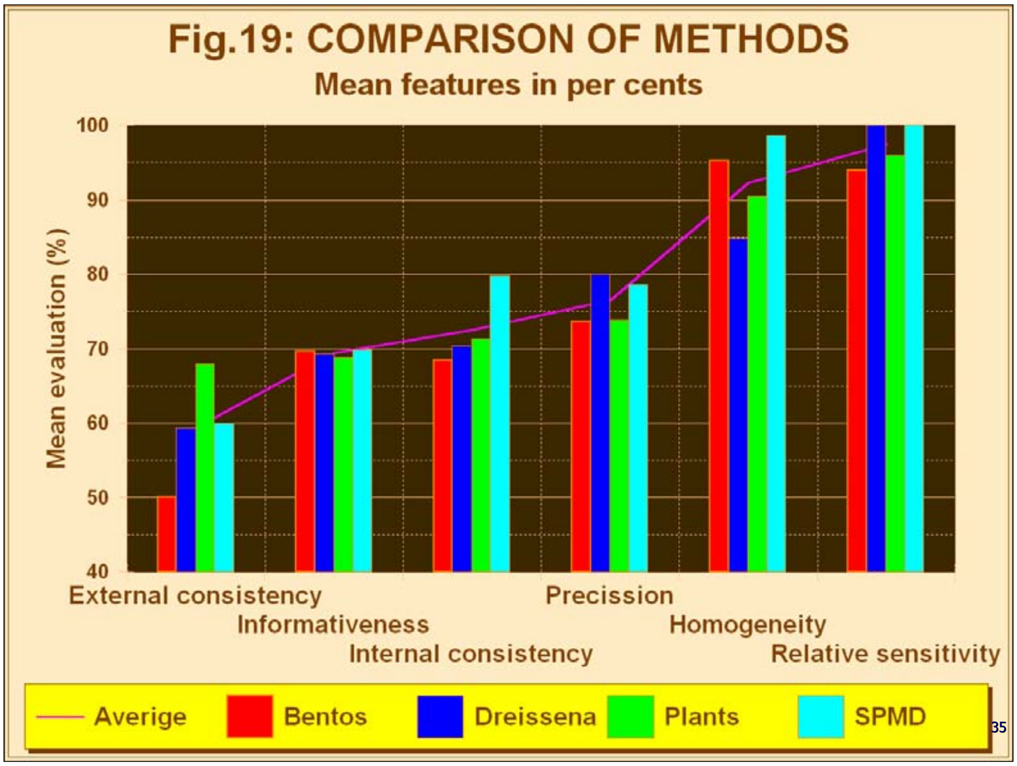


Fig.19: COMPARISON OF METHODS
Mean features in per cents
Mean evaluation (%)
100
90
80
70
60
50
40
External consistency
Informativeness
Internal consistency
Precission
Homogeneity
Relative sensitivity
Averige
Bentos
Dreissena
Plants
SPMD
35

RATING OF METHODS

Feature	Bentos	Dreiss.	Plants	SPMD
Ext.consistency	4	3	1	2
Int.consistency	4	3	2	1
Informativeness	1	3	4	2
Precission	3	1	4	2
Homogeneity	2	4	3	1
Rel.sensitivity	3	1	2	1
Mean rating	**2.8**	**2.5**	**2.7**	**1.5**

36

Conclusions

- Passive sampling, like SPMDs shown the best results; if there are no legal requirements for biota, biotic organisms can be replaced
- Do not forget to analyze data precisely, independently, before your interpretation
  - Do not rely ONLY on functionality of any processing package
  - Statistical approach has some limitations on small data sets (majority of monitoring studies)
- Any headache from analytical tools can be eliminated by experience
  - Try it!

Further intentions

- Finalization of Gnostic analytical tool, with GUI (S-Plus)
- Extension to other platforms by interface
- Linking to databases (LIMS, GIS, …)
- Training and dissemination
- Projects solutions and participations
  - Join us: 2-FUN project, www.2-fun.org

38

... thank you for your attention!

39