

Pavel Kovanic

COMPARISON OF METHODS

KERNEL ESTIMATES OF DISTRIBUTIONS

Principle: Physics, Mathematics (convolution)

Statistical Parzen's kernel is a finite valued function $k(a)$, symmetrical in a , absolutely integrable (integral = 1), satisfying the condition

$$\lim(axk(a))=0 \text{ (for } a \rightarrow \infty).$$

Parzen's kernels are attached to each data item and superposed: their mean is Parzen's estimate of the probability **density**, integral of which is the kernel estimate of probability **distribution**.

There exists *infinite number* of "suitable" Parzen's kernels.

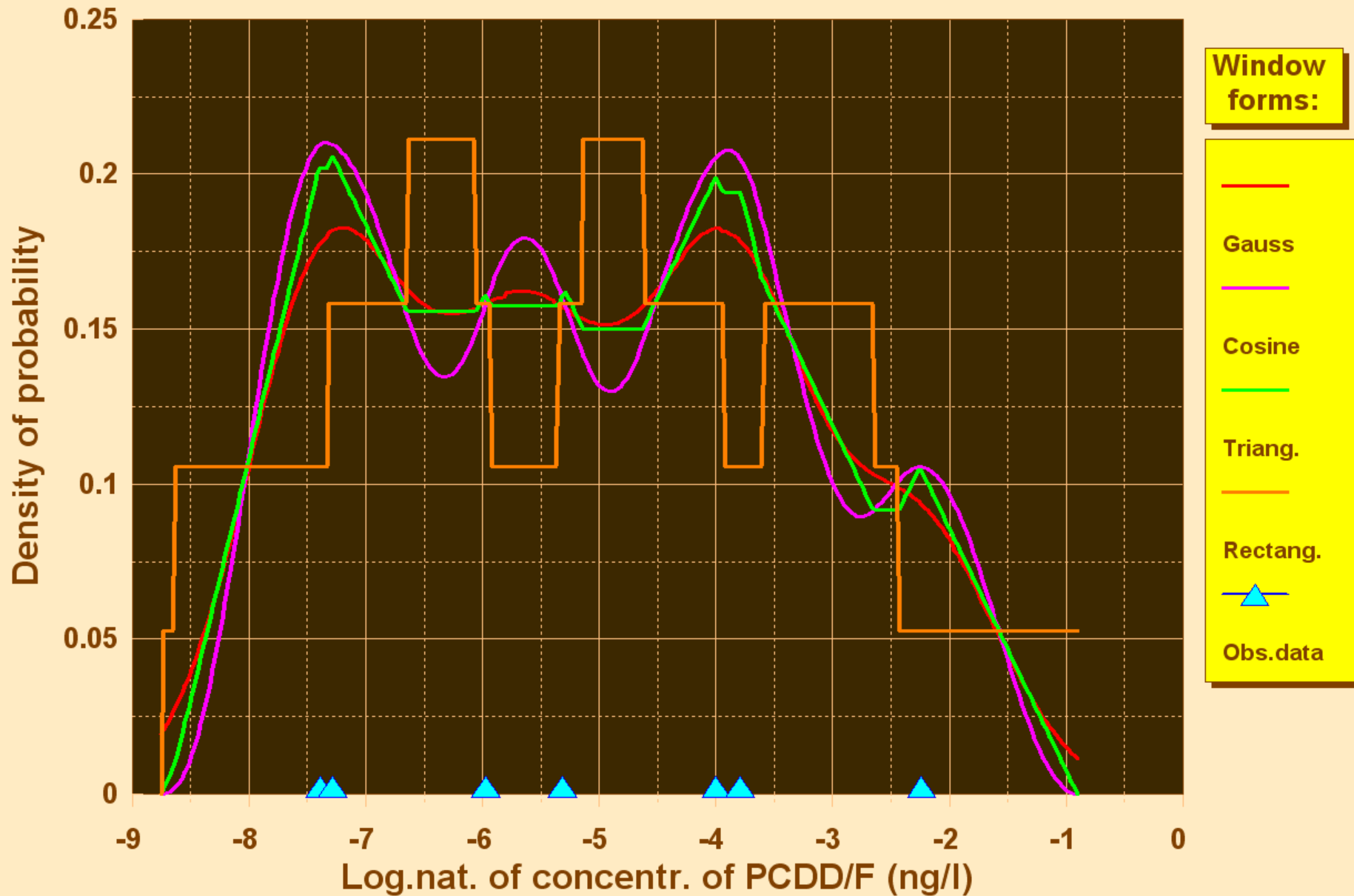
In **gnostics**, the role of Parzen's kernel is played by the weight *we*. For the optimum S is the *we* the **unique** and **optimal kernel**.

KERNEL ESTIMATE IN R-PROJECT

```
dP <- density(x, bw = "nrd0", adjust = 1,  
  kernel = c("gaussian", "epanechnikov",  
  "rectangular", "triangular", "biweight", "cosine",  
  "optcosine"), weights = NULL, window = kernel,  
  width, give.Rkern = FALSE, n = 512, from, to,  
  cut = 3, na.rm = FALSE, ...)
```

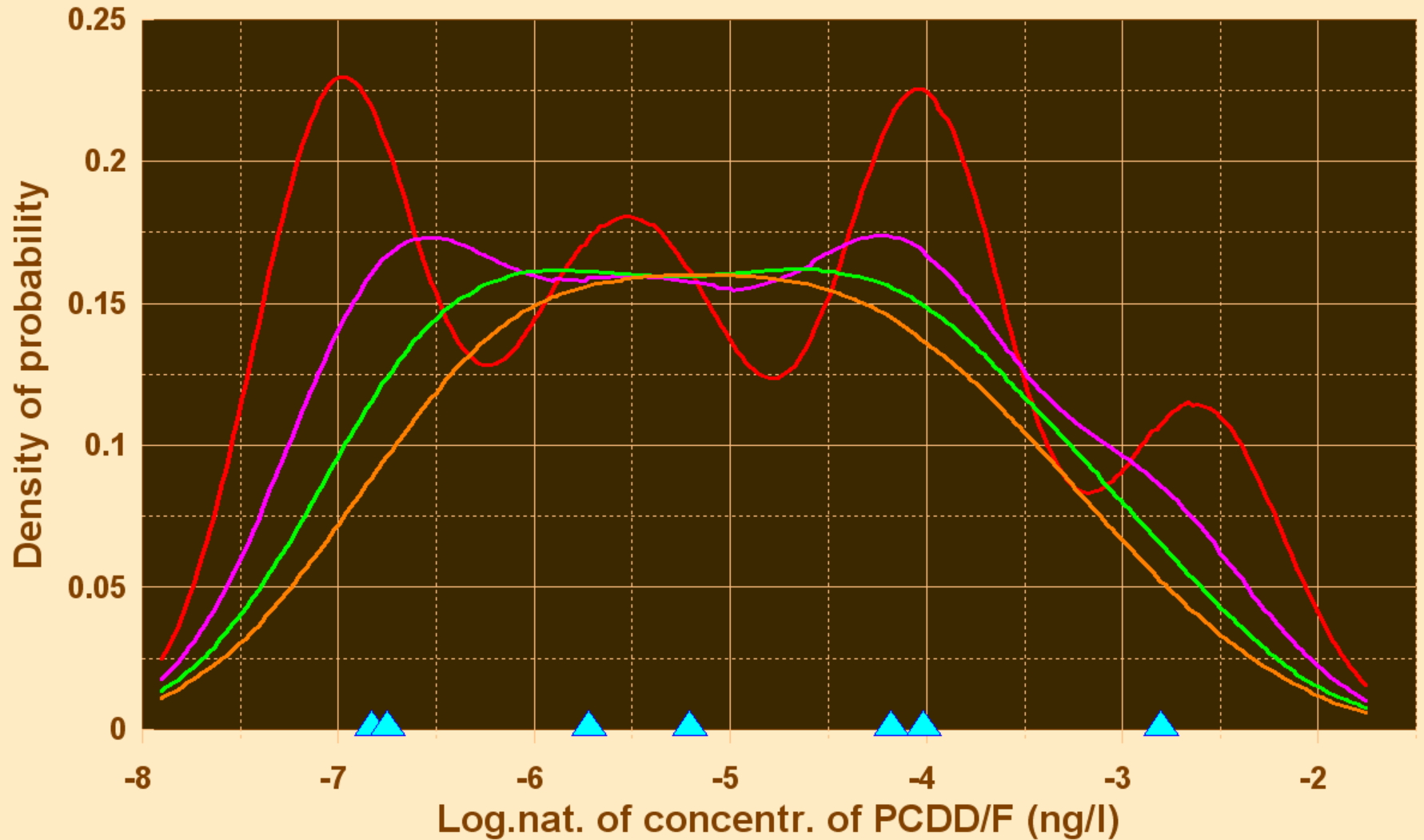
The Sum of PCDD/F in Rivers

Probability density f-s by 4 methods



The Sum of PCDD/F in Rivers

Gauss kernel estimation, 4 widths



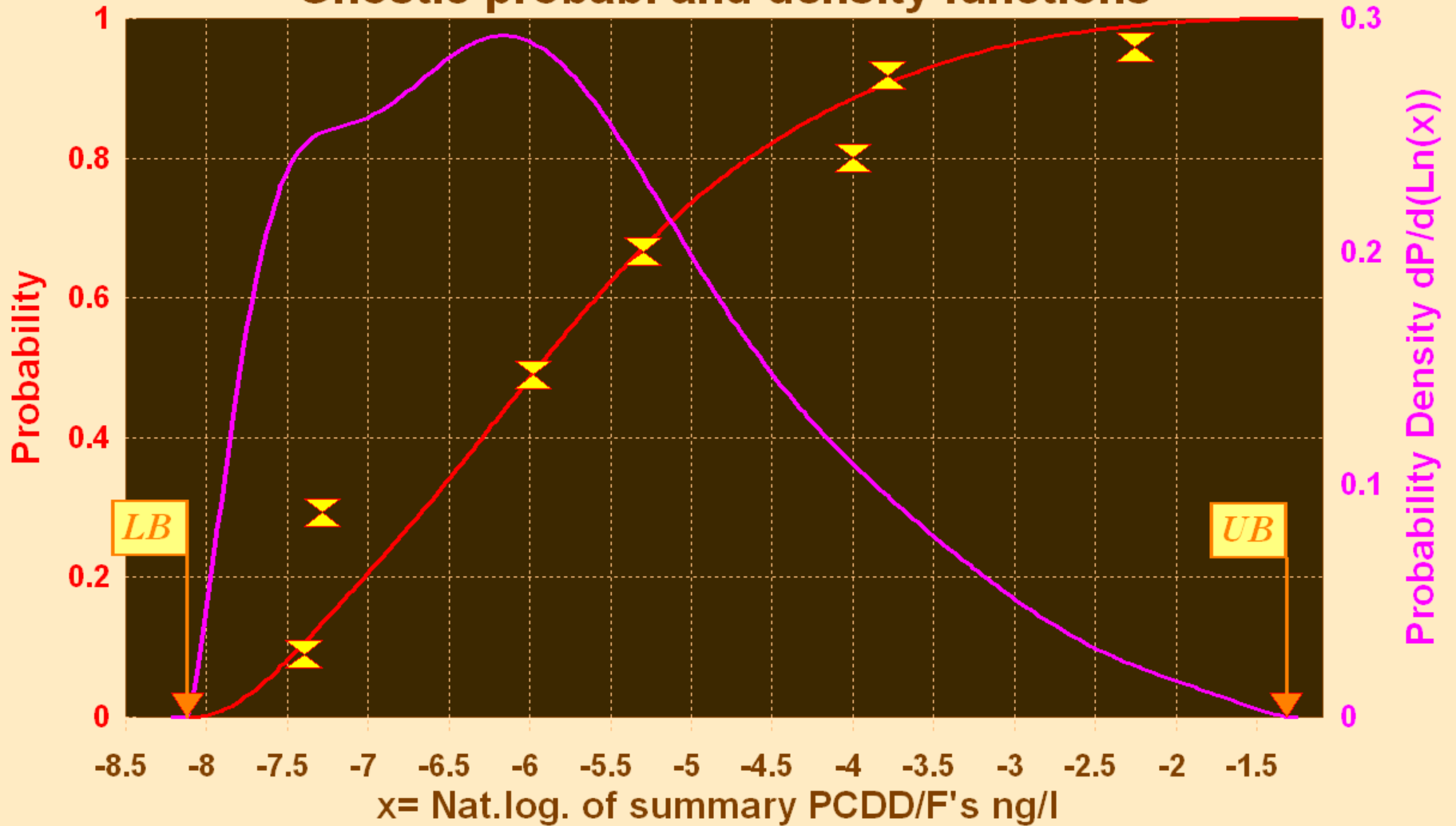
Width=2 Width=3 Width=4 Width=5 Obs.data

SPECIFICS OF GNOSTIC DISTRIBUTIONS

- **Theoretically based and unique kernel**
- **Theoretically based data composition law**
- **No subjective assumption and choices**
- **Optimizing the scale/kernel's width parameters**
- **Optimizing the bounds of data support**
- **Robustness of two kinds**
- **Natural smoothness of distributions**

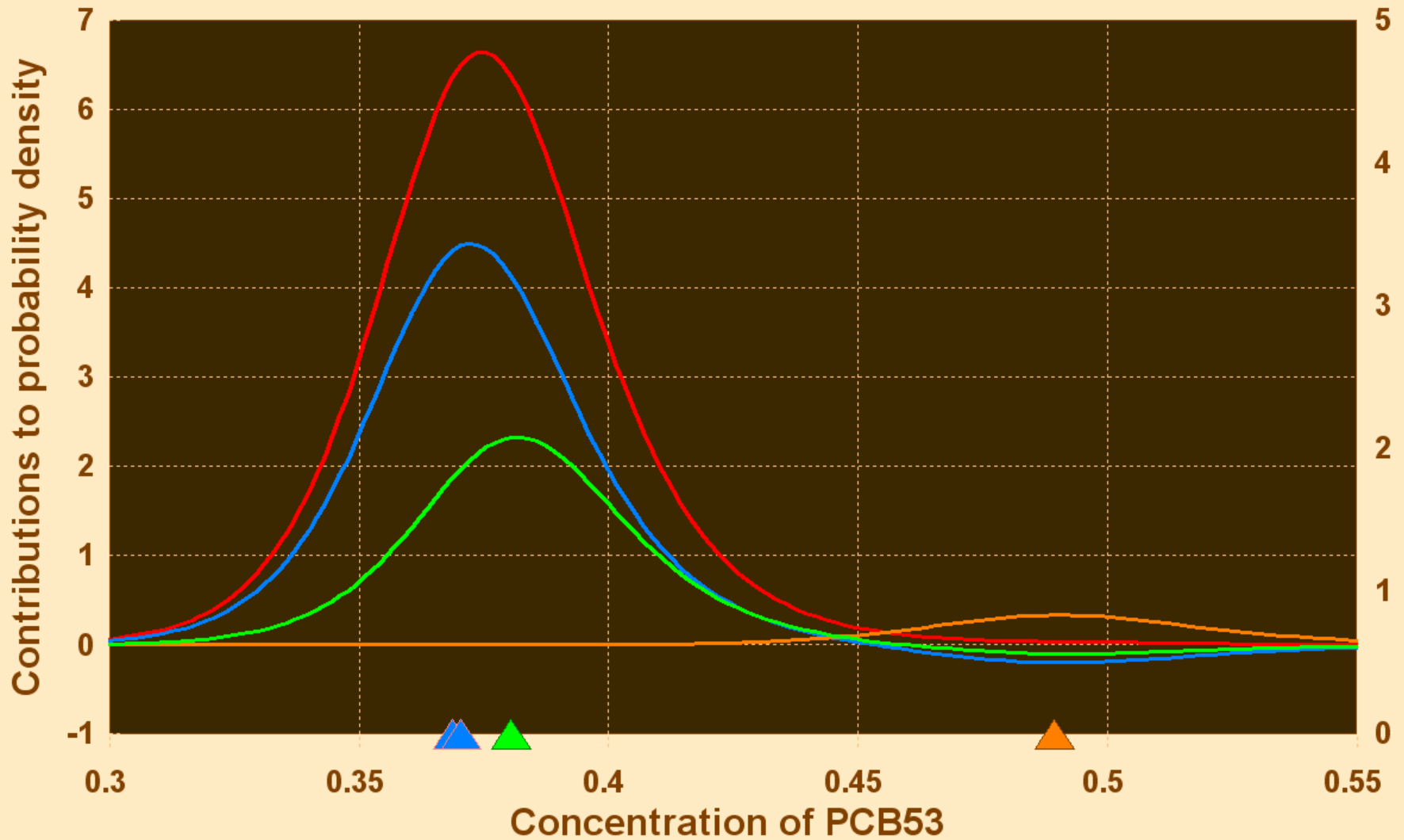
The Sum of PCDD/F in Rivers

Gnostic probab. and density functions



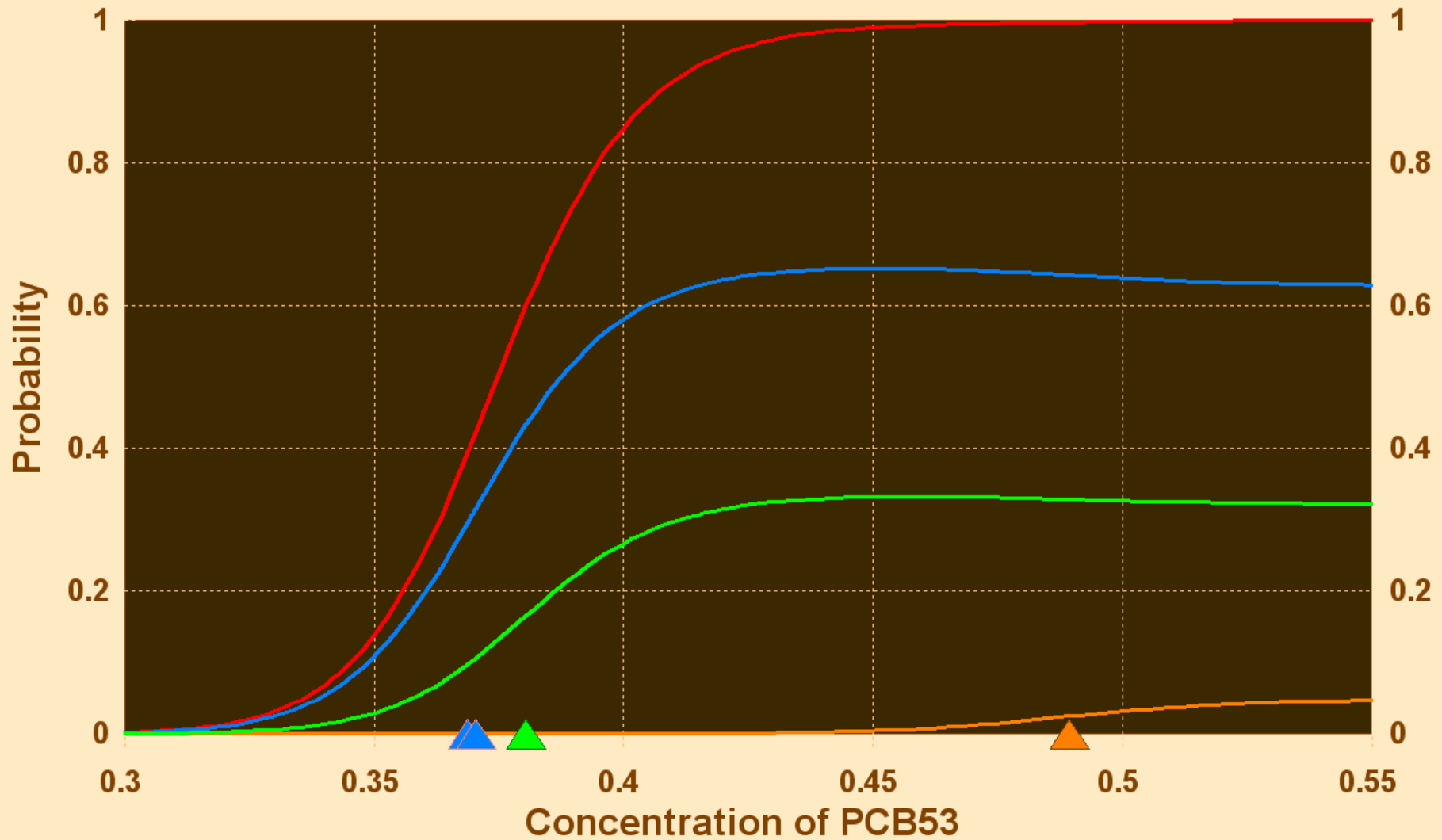
GNOSTIC KERNELS OF PROB. DENSITY

Distribution of PCB53 in 4 sources



GNOSTIC KERNELS OF PROBABILITY

Distribution of PCB53 in 4 sources



— All 4 data

— D1+D2 (0.37+0.37)

— D3 (0.49)

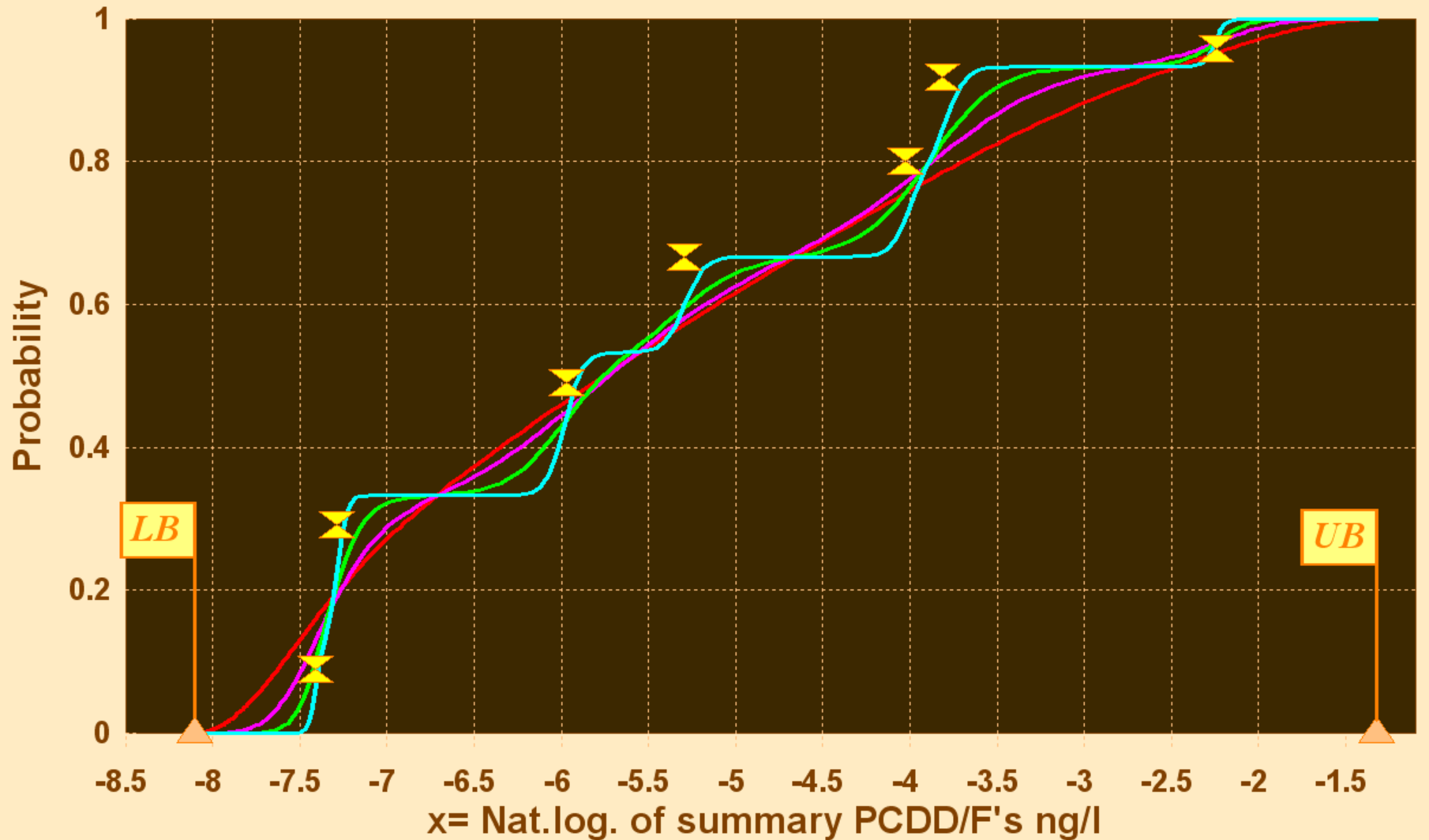
— D4 (0.38)

GNOSTIC LOCAL D.F.

- Local, detailed view on data structure
- Differentiable
- Unlimited flexibility controlled by the scale parameter S
- Applicable to both homogeneous and non-homogeneous data samples
- Applicable to marginal cluster analysis

The Sum of PCDD/F in Rivers

Gnostic local d.f. of probability



— S=2

— S=1

— S=0.5

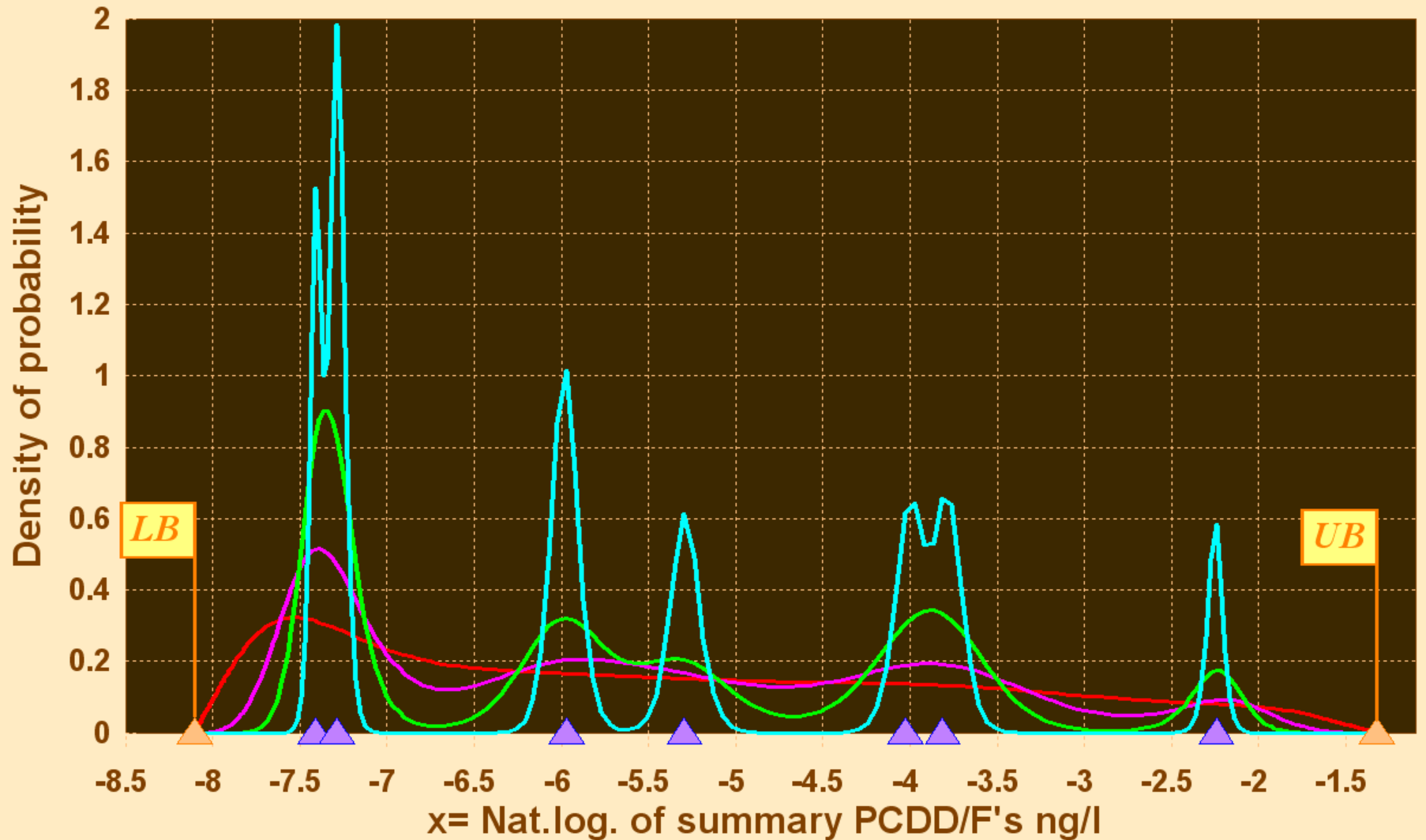
— S=0.15

— X — Empir. d.f.

— ▲ — Data bounds

The Sum of PCDD/F in Rivers

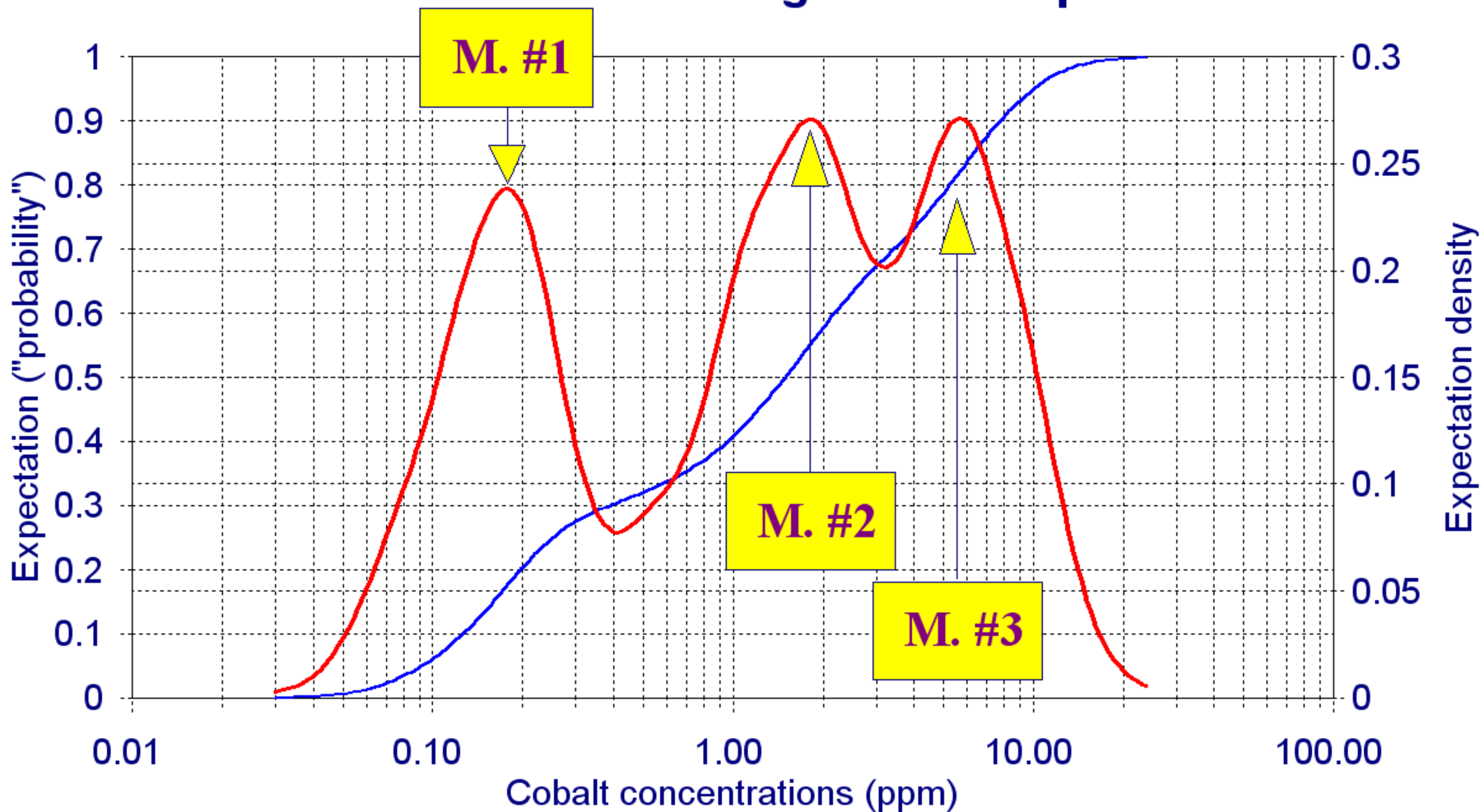
Gnostic local density function



— S=2 — S=1 — S=0.5 — S=0.15 ▲ Obs.data ▲ Data bounds

3 METHODS OF CHEMICAL ANALYSES

Cobalt contents in the granite samples

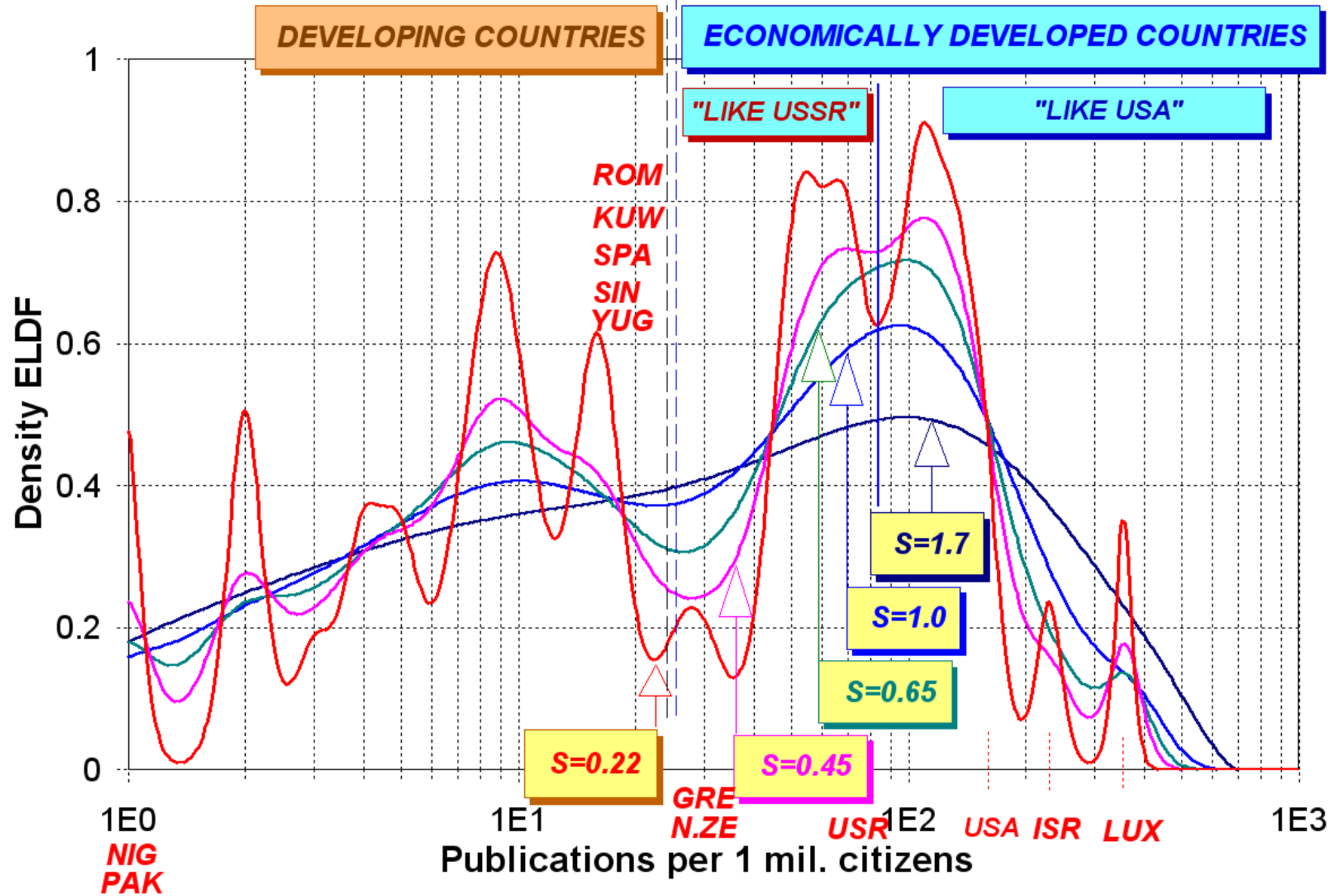


Observed data

— Local distribution

— Local density

PUBLICATIONS IN PHYSICS (1983)



The role of the scale parameter in marginal analysis

TREATMENT OF CENSORED DATA

Methods to compare

(LCD ... Left-censored data, LOD ... Limit of Detection)

A) Heuristic

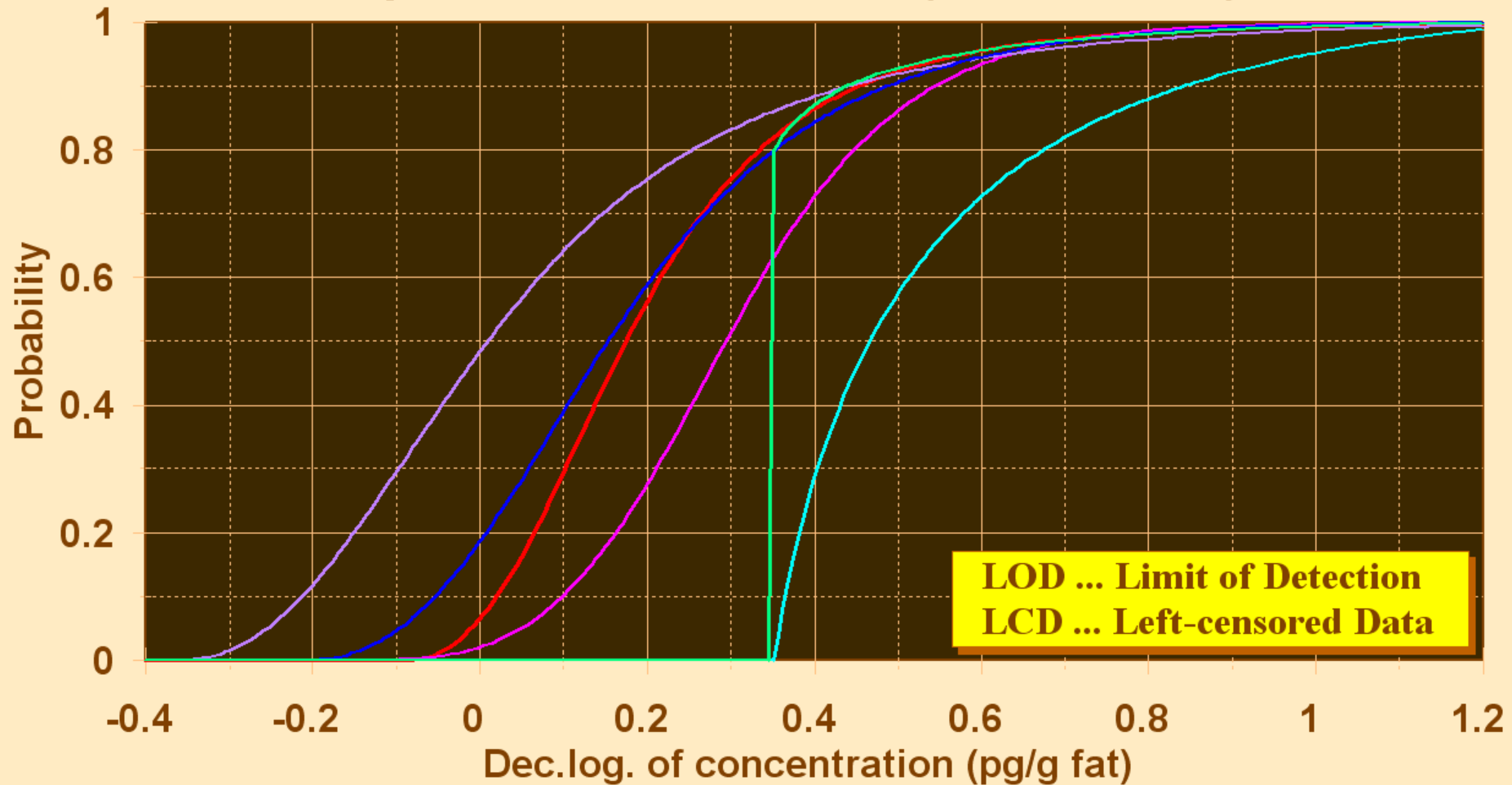
LCD = LOD, LCD = LOD/2, LCD = LOD/sqrt(2),
LCD = 0, Neglected LCD < LOD.

B) Scientific

Kaplan-Meier survival model
Gnostic method

HANDLING OF LEFT-CENSORED DATA

Comparison of 6 methods (case: TCDD)



- 1 Properly censored
- 2 $LCD=LOD$
- 3 $LCD=LOD/2$
- 4 $LCD=LOD/\sqrt{2}$
- 5 ONLY $>LOD$
- 6 $LCD=0$

Fig.7: ANALYSIS OF LEFT-CENSORED DATA

1 - Pr{GammaHCH} in rivers

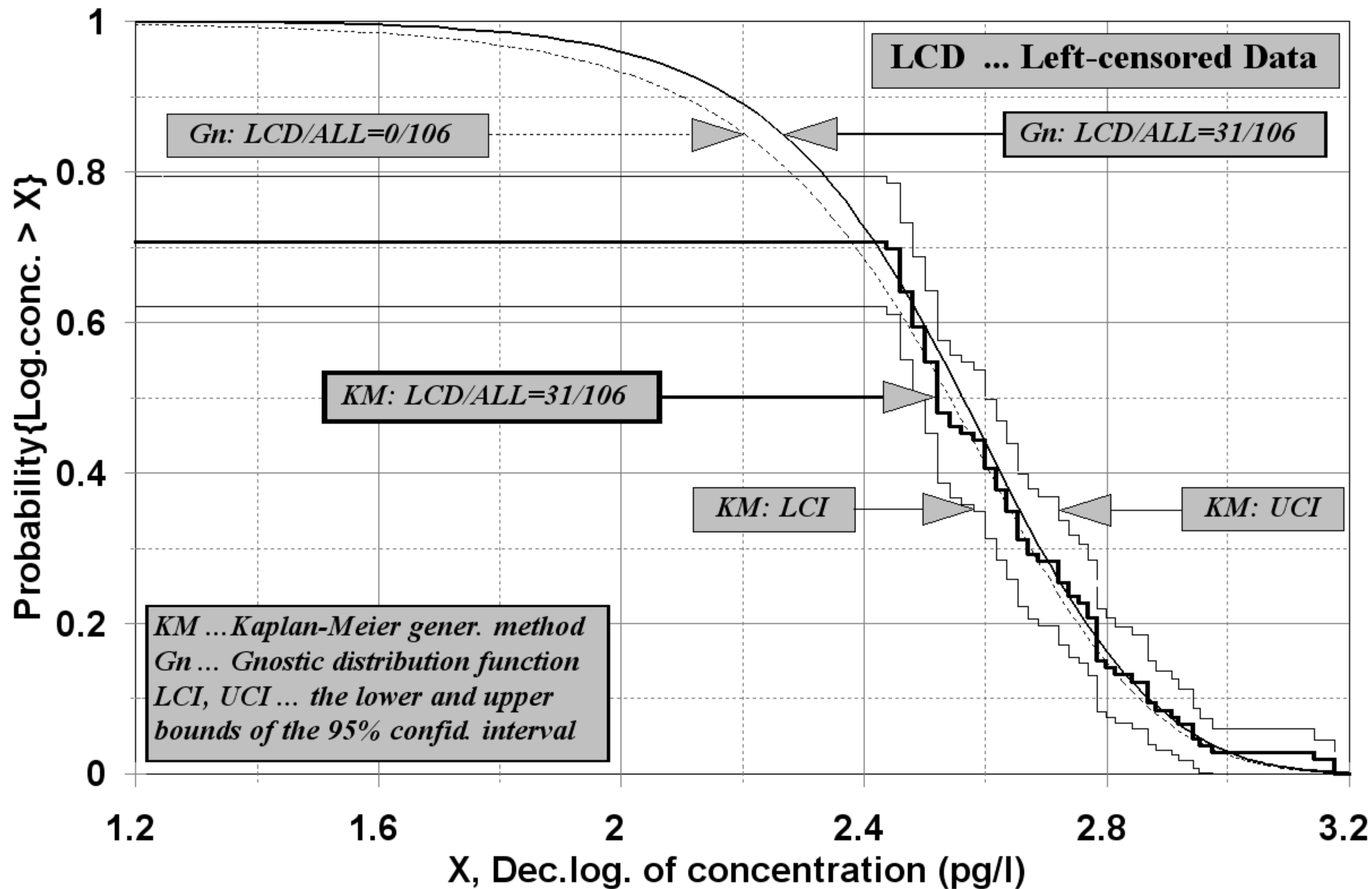
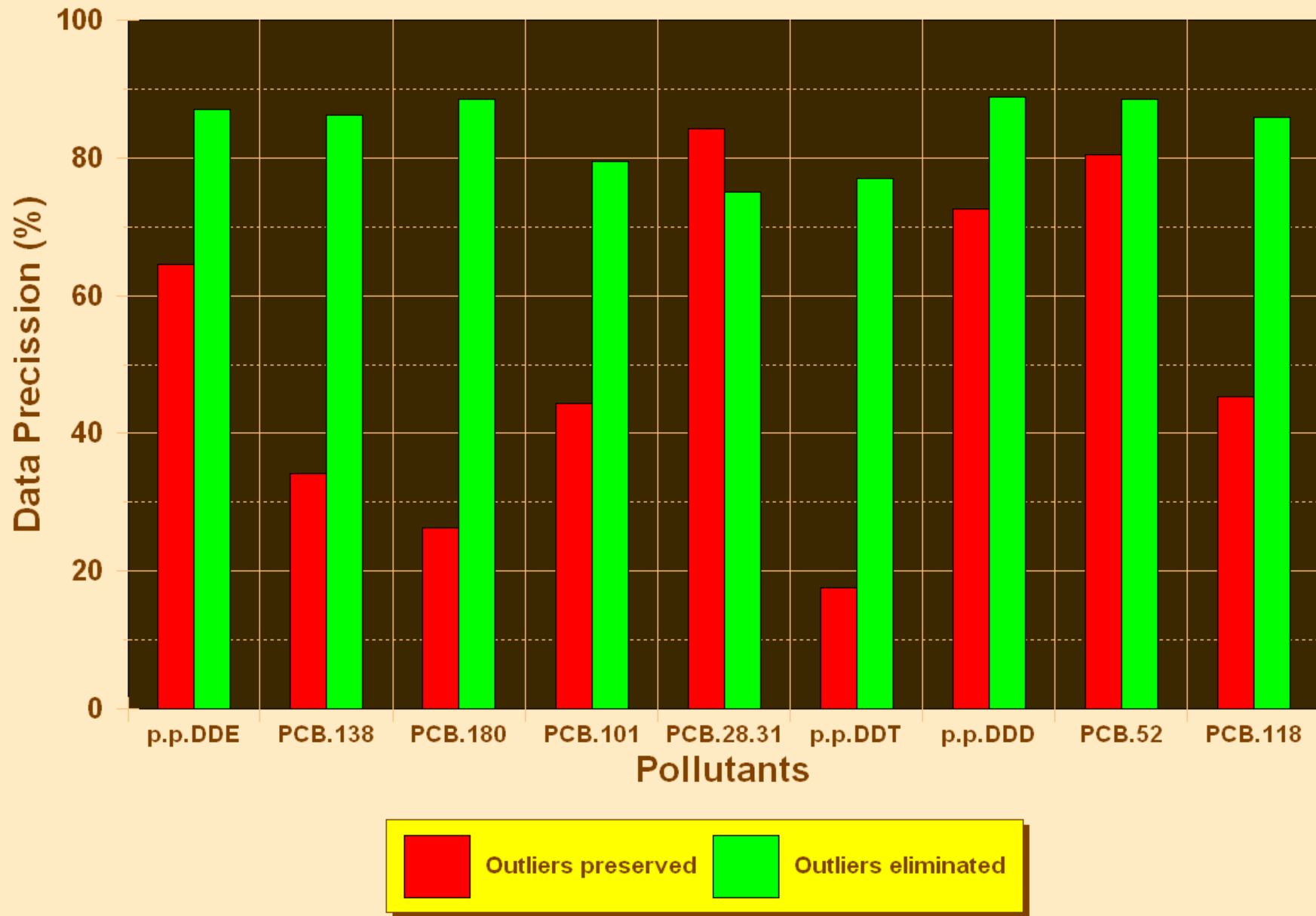


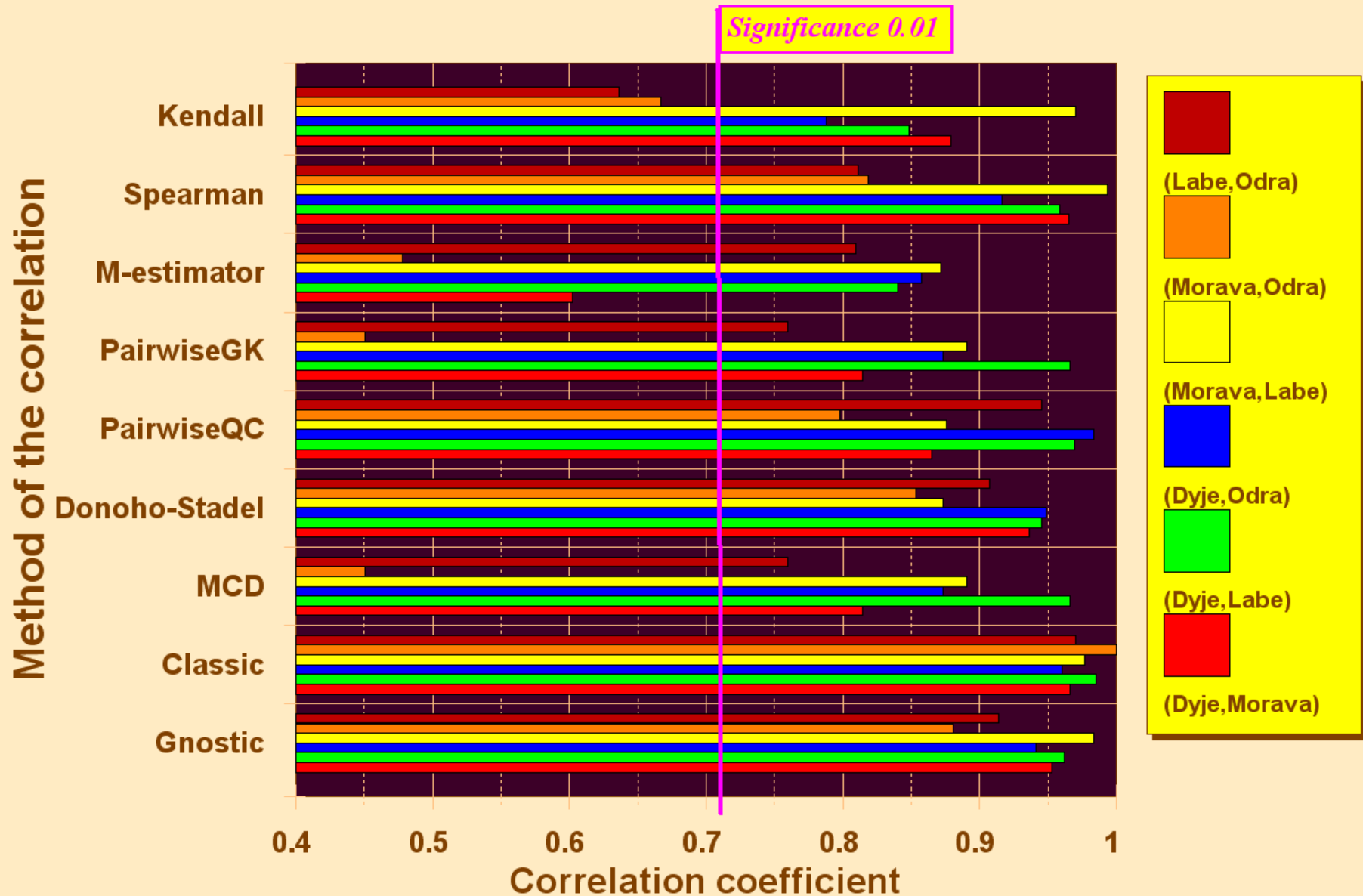
Fig.17: RELATIVE DATA PRECISION

Effects of Outliers



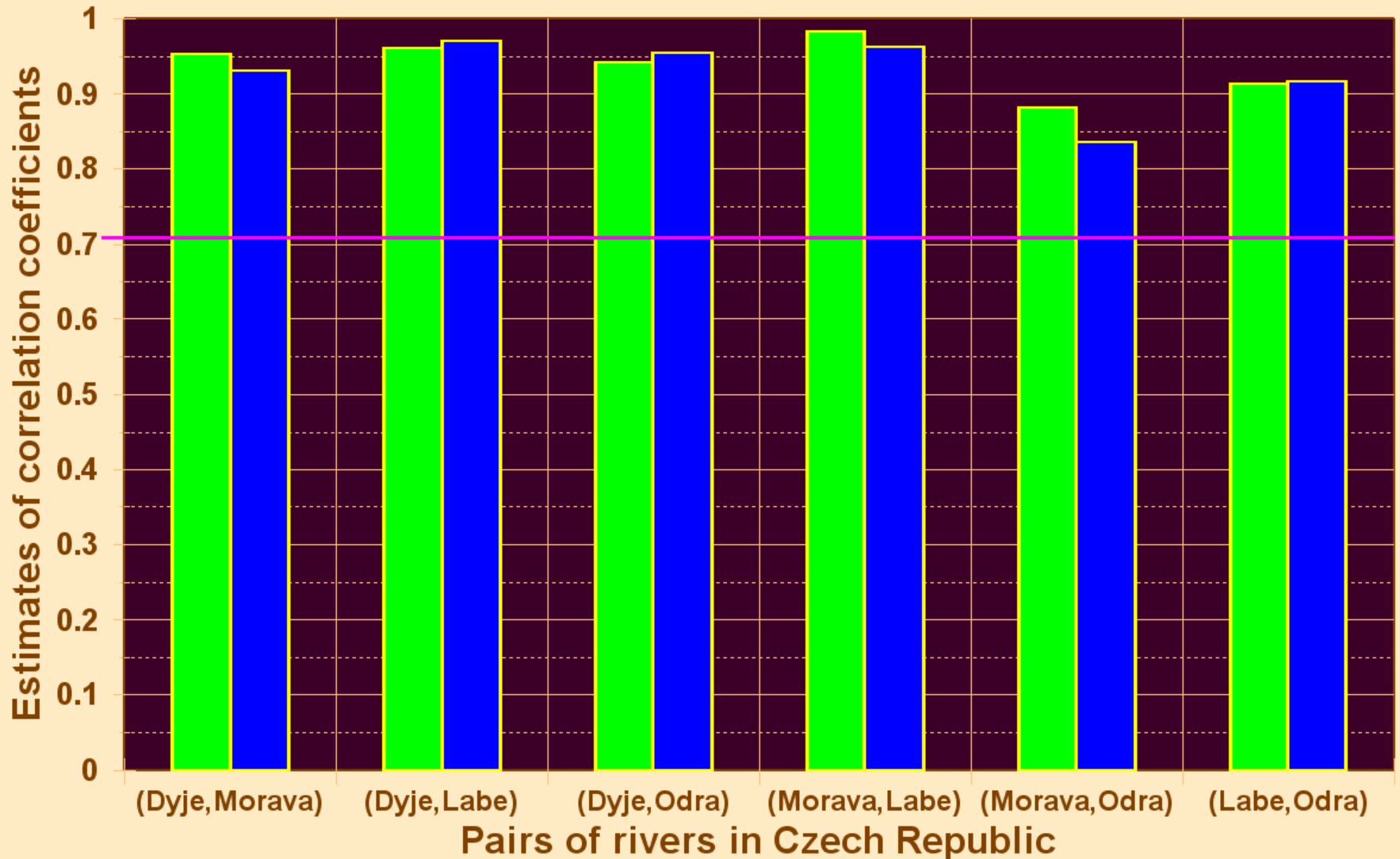
COMPARISON OF ROBUST CORRELATIONS

Contamination similarity in 4 rivers



SIMILARITY OF CONTAMINATION IN RIVERS

Gnostic and 8 statistical correlations



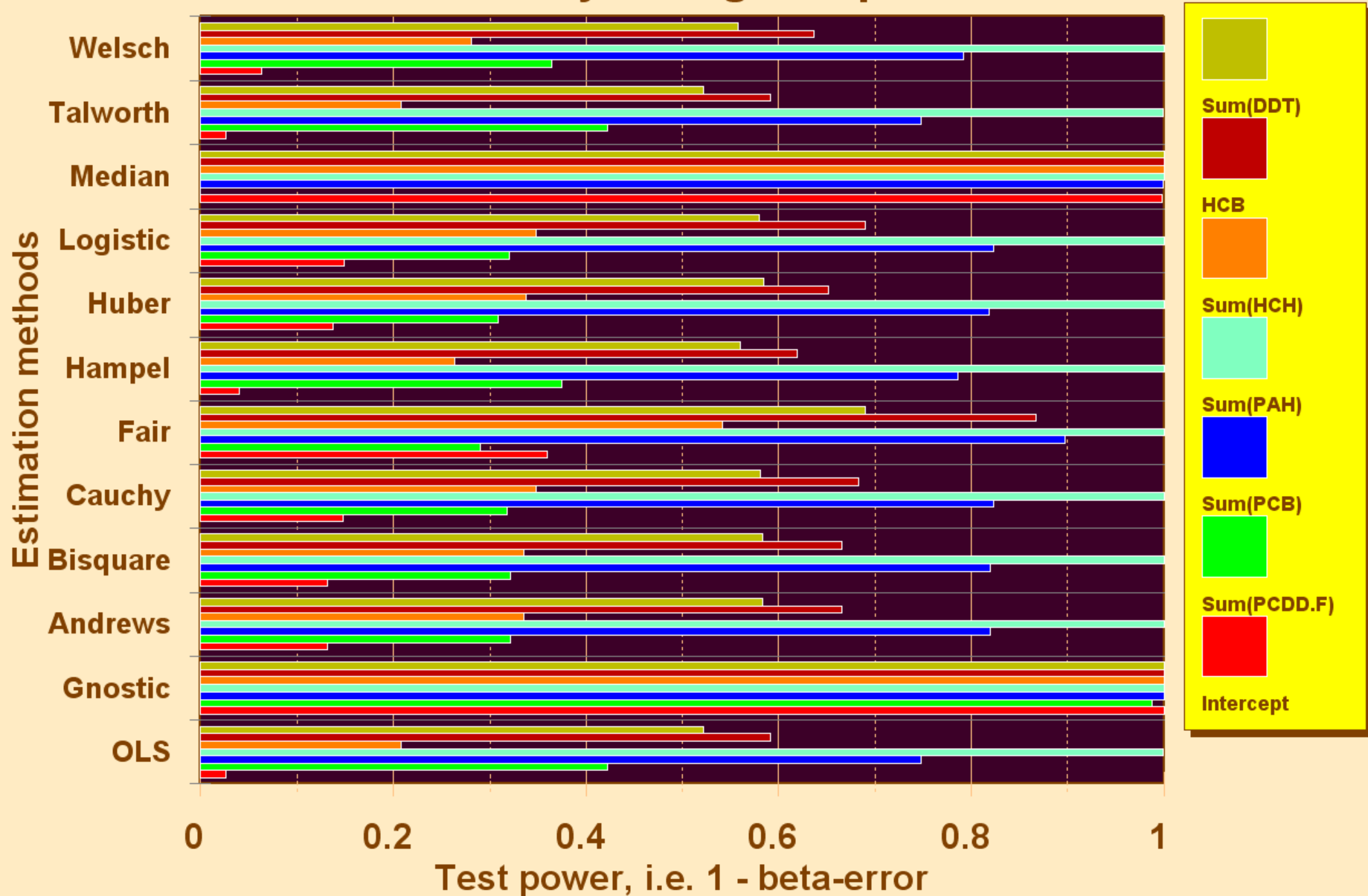
— Significance 0.01

■ Gnostic correlations

■ Geom. mean of 8 statistics

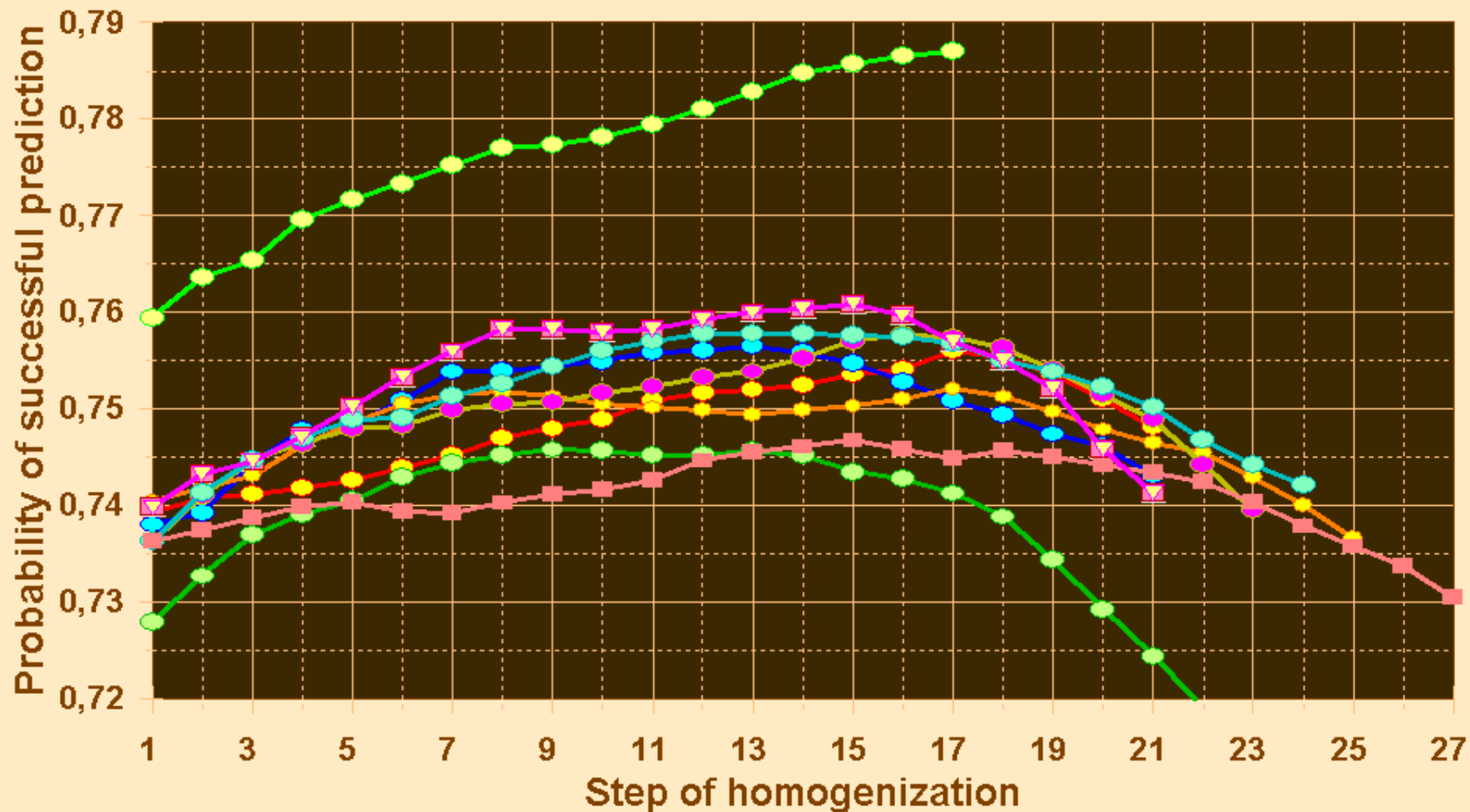
TEST POWER OF M-ESTIMATORS

Task: Toxicity of organic pollutants



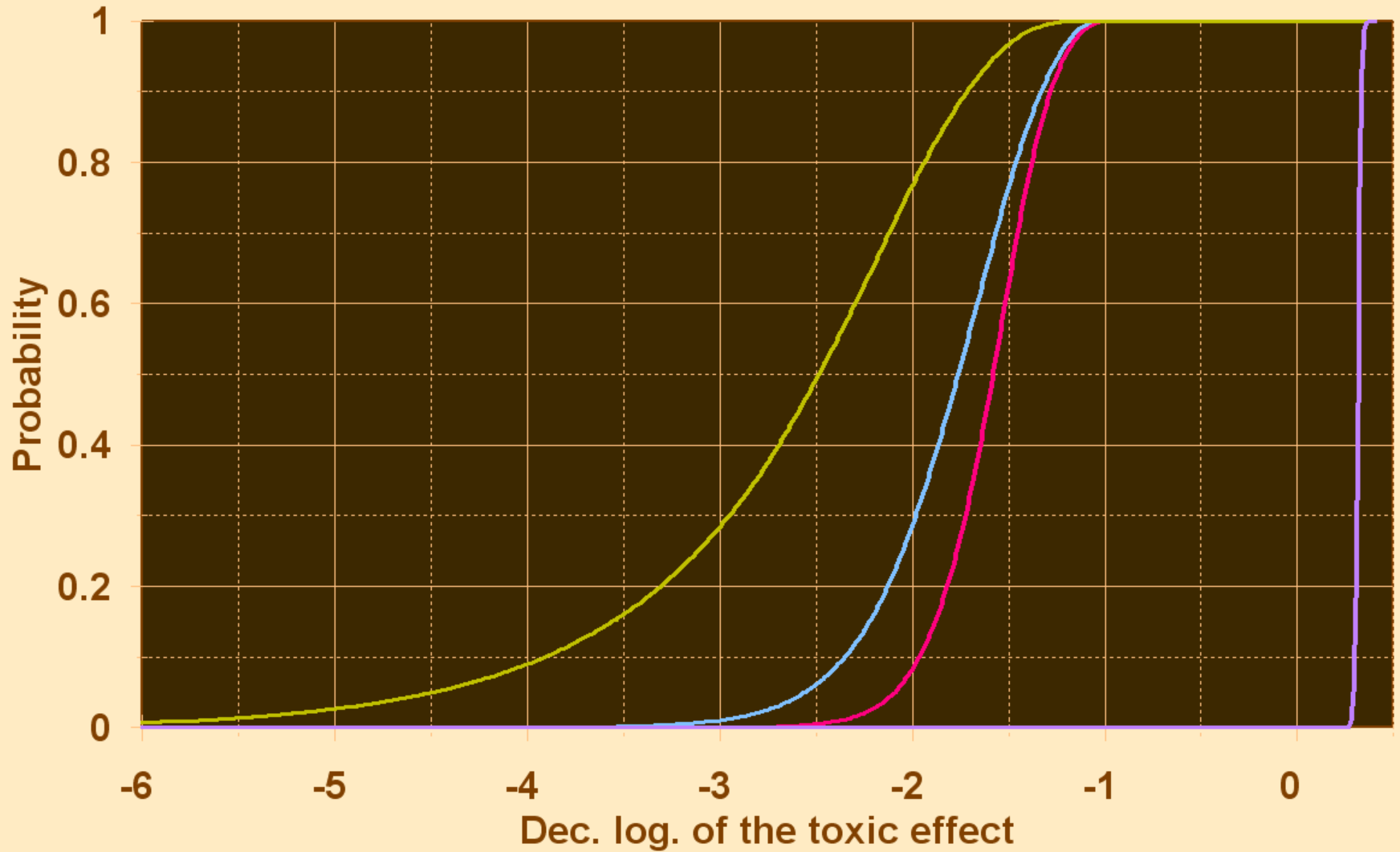
COMPARISON OF ROBUST REGRESSIONS

Improvements by MD-homogenization



TOXIC EFFECTS

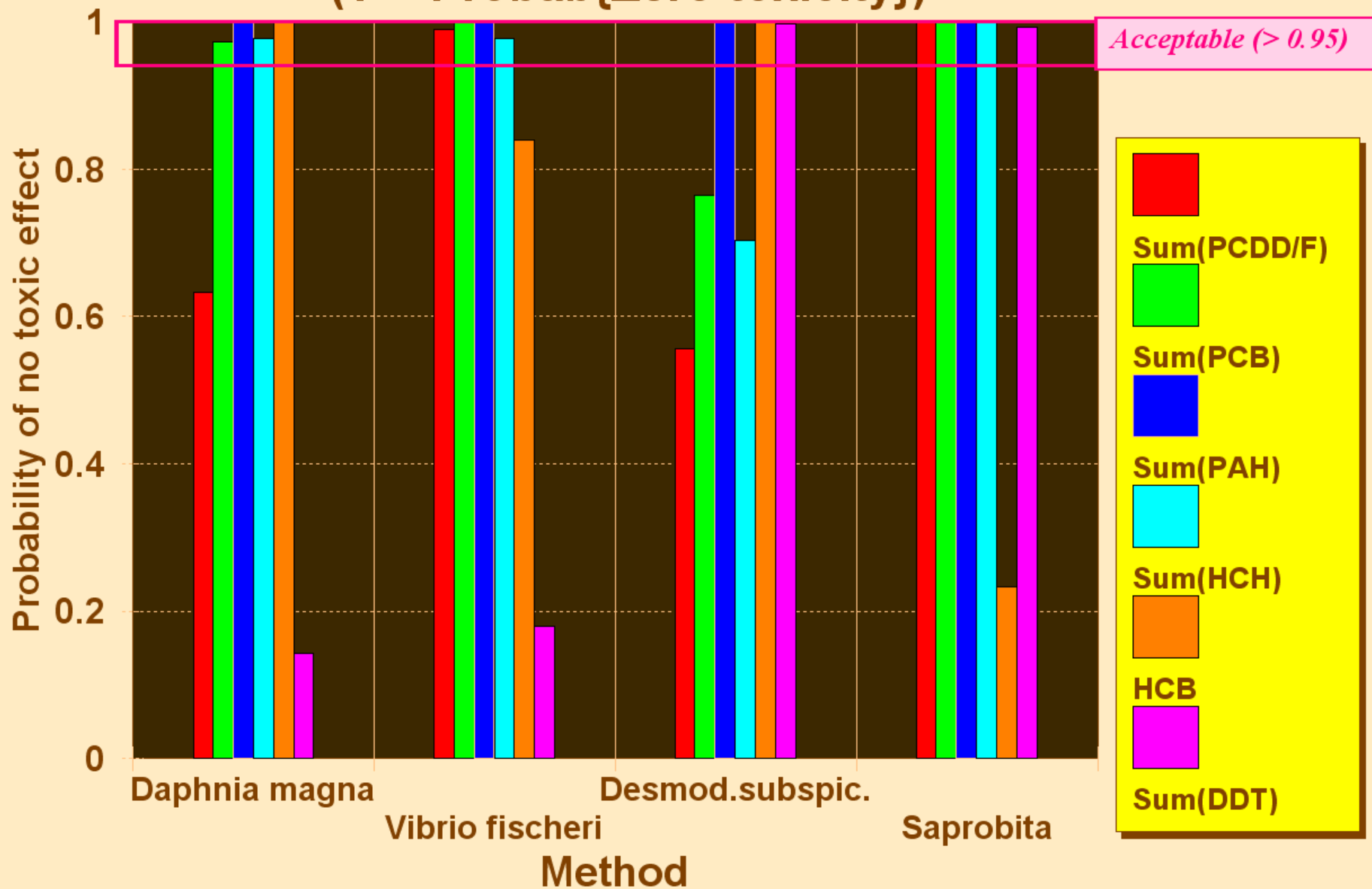
Four measuring methods



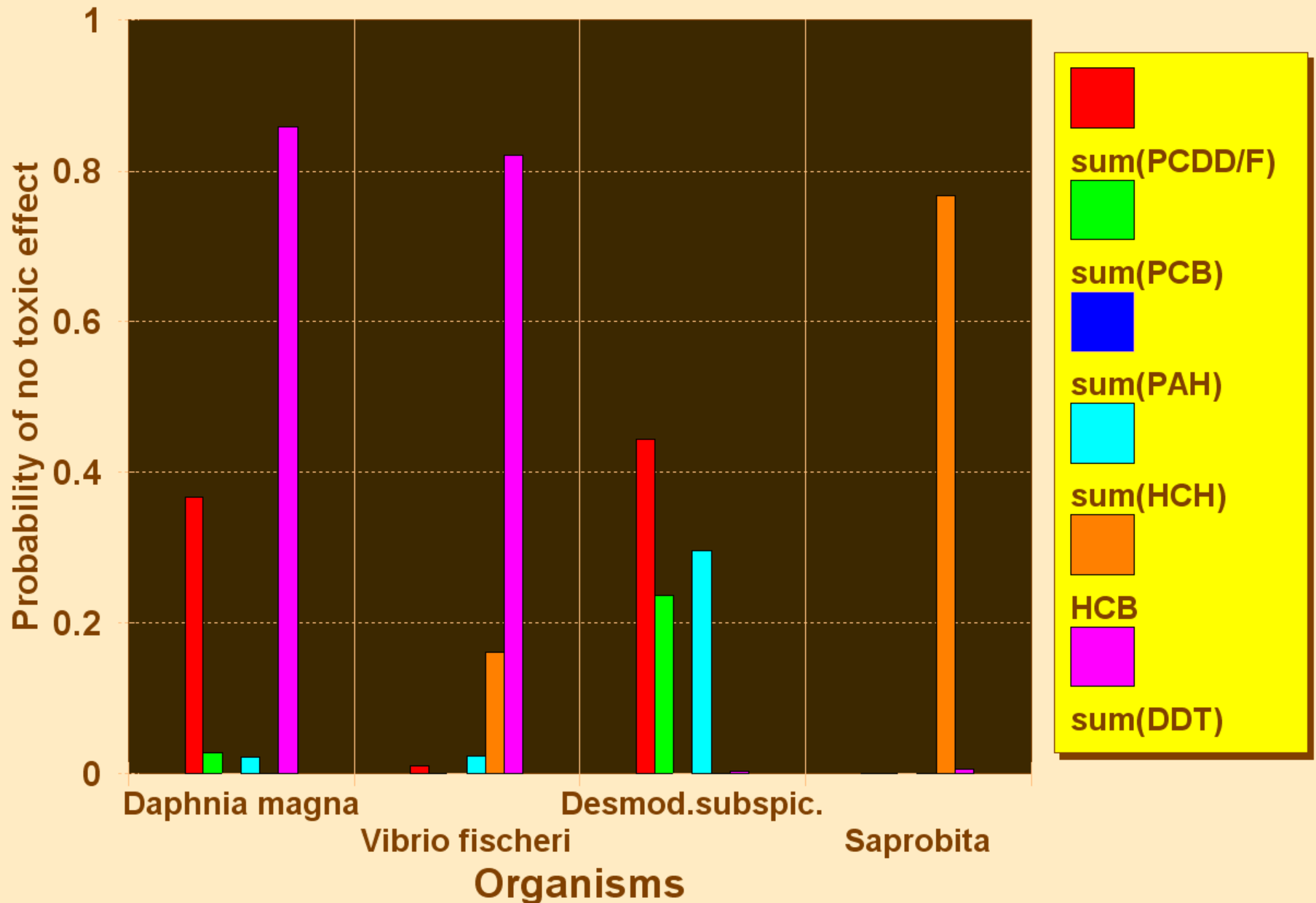
— Daphnia magna (L/d) — Vibrio Fischeri (min.) — Desmod. subsp. — Saprobita

POWER OF THE TOXICITY TEST

(1 -- Probab{Zero toxicity})

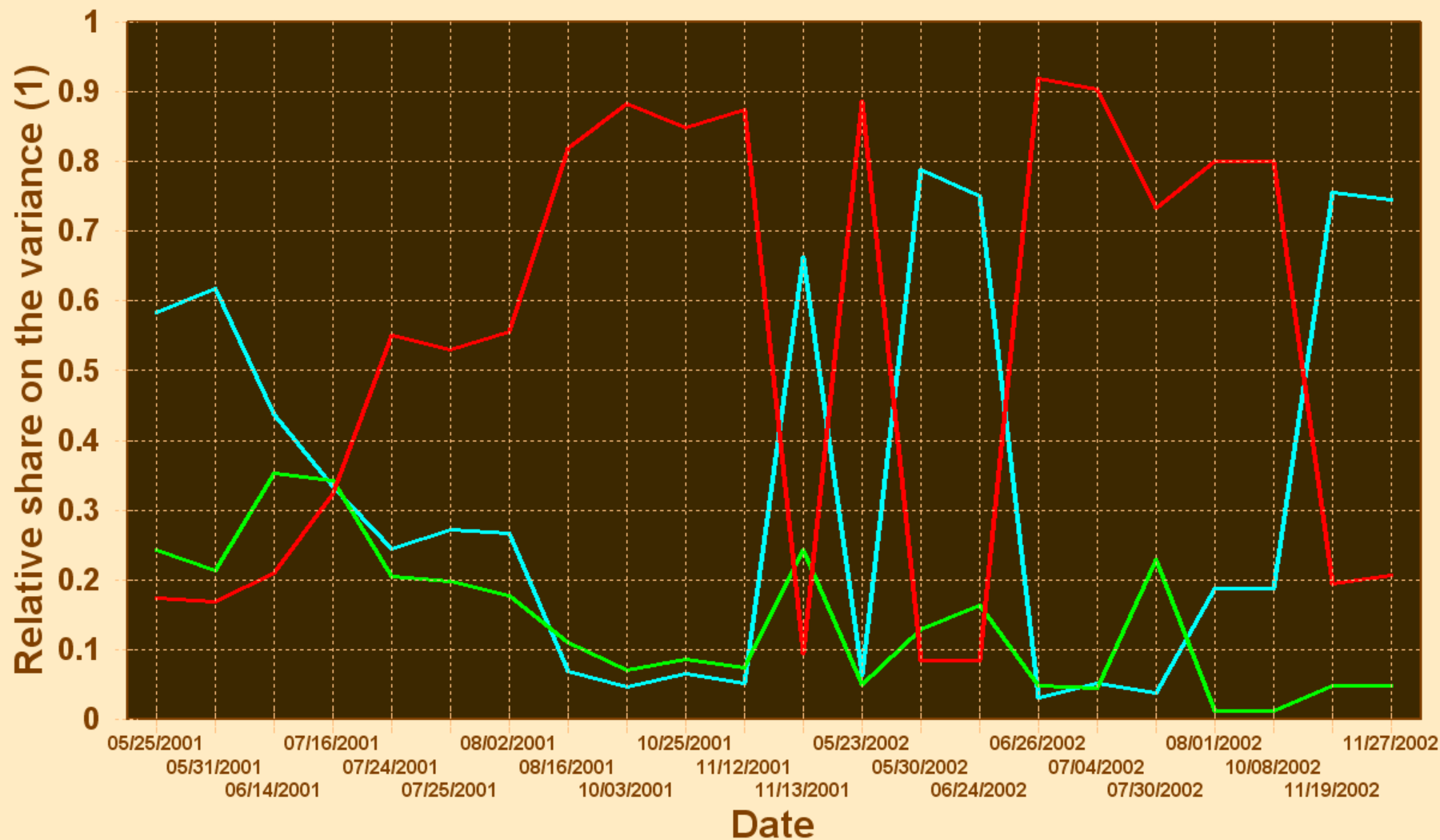


PROBABILITIES OF ZERO TOXICITY



TIME-SERIES OF 3-D MODELS OF EMISSIONS

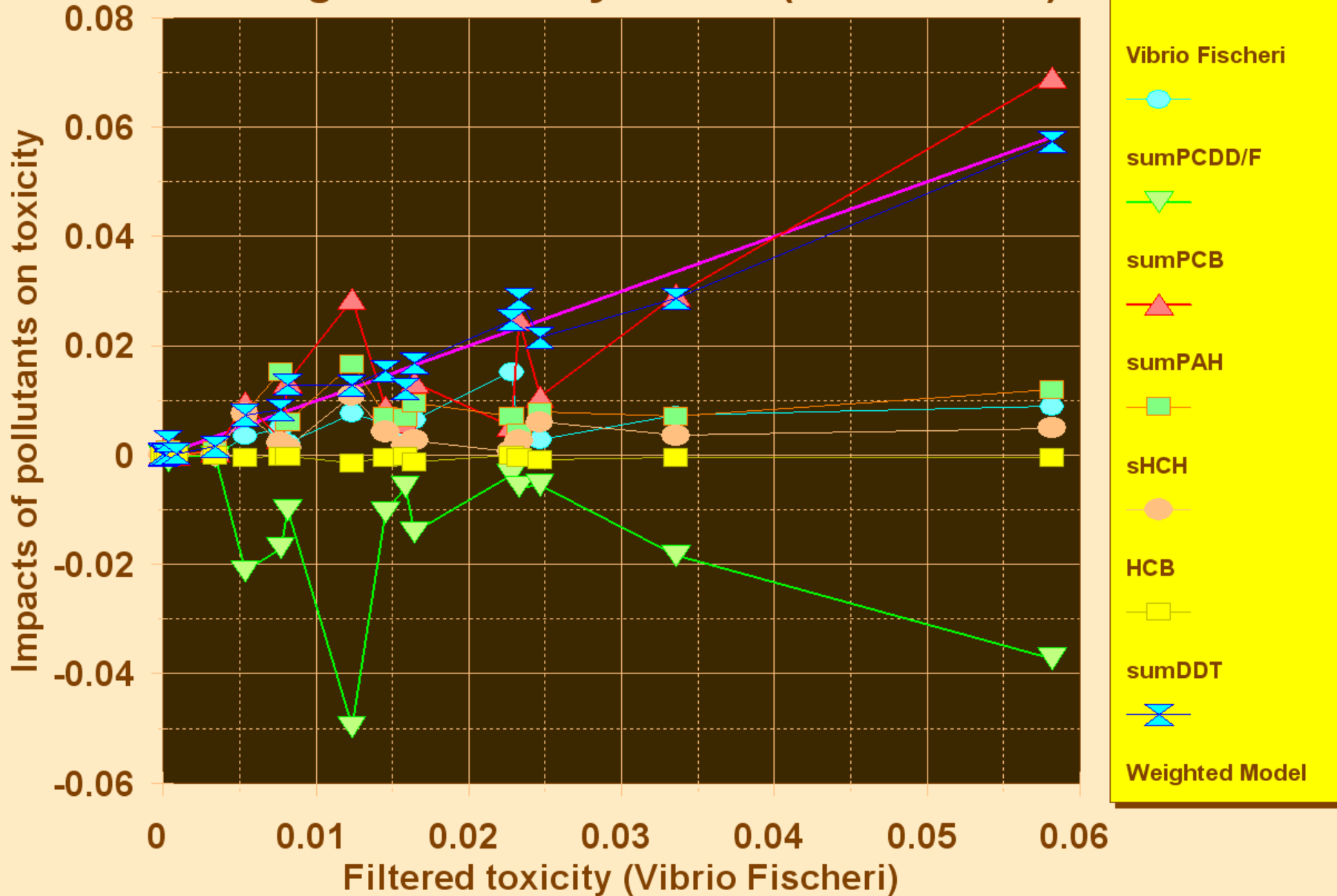
Shares on the variance of model coeffs



— pp.DDD — pp.DDE — pp.DDT

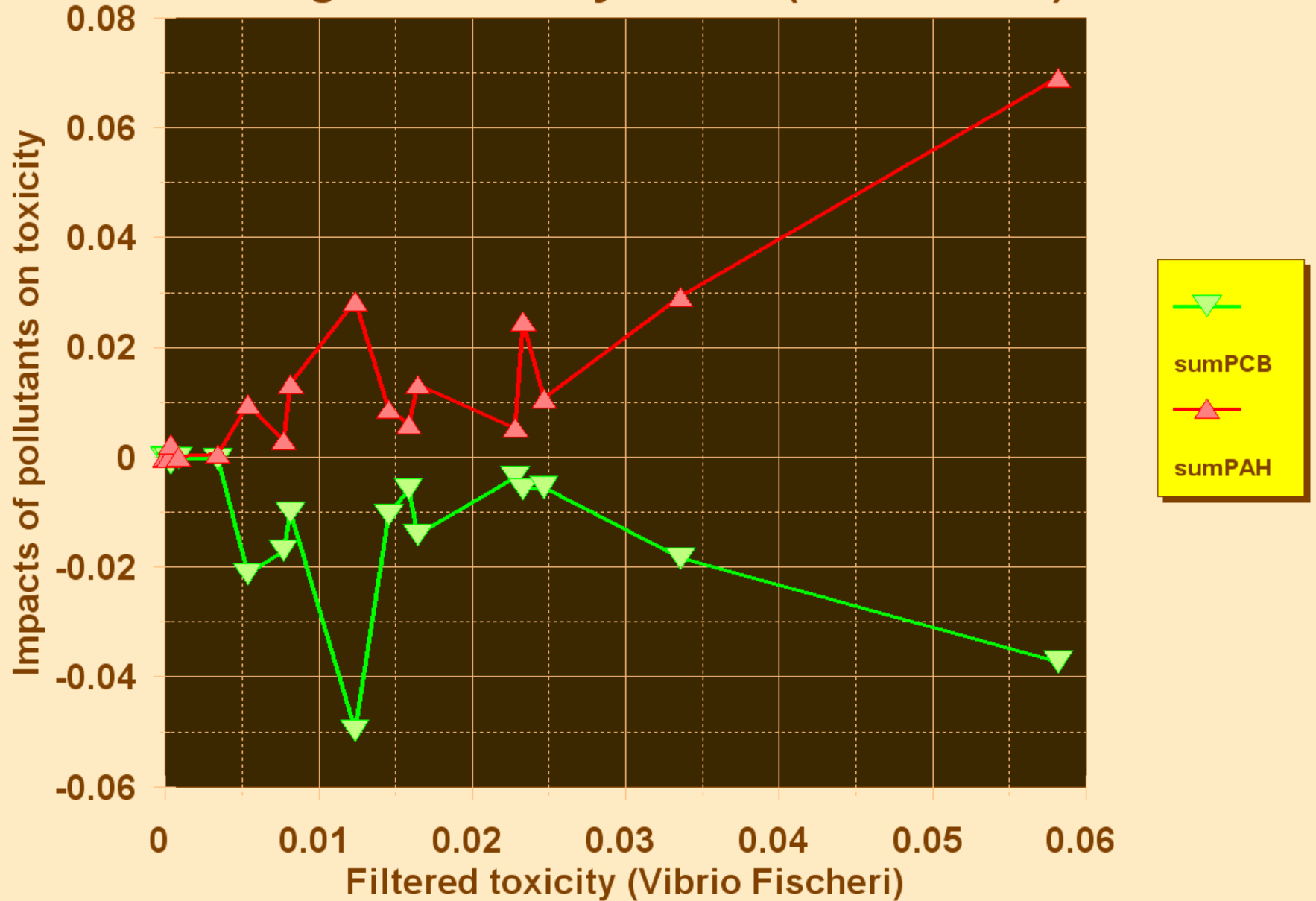
TOXICITY (VIB.F.) ROBUST 6D-MODEL

Weighted contrary effects (PCB to PAH)



TOXICITY (VIB.F.) ROBUST 6D-MODEL

Weighted contrary effects (PCB to PAH)



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

Fig.19: COMPARISON OF METHODS

Mean features in per cents

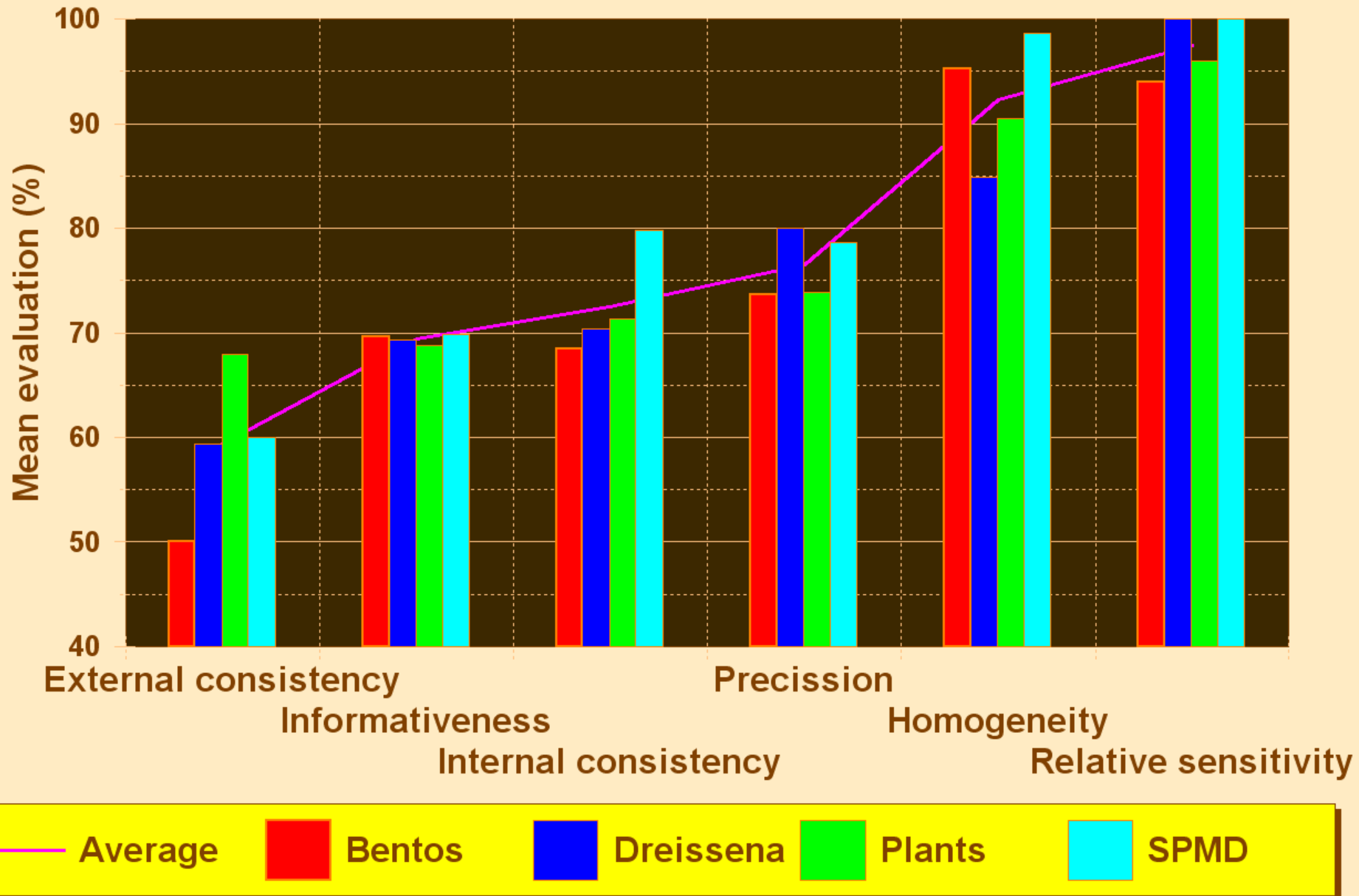
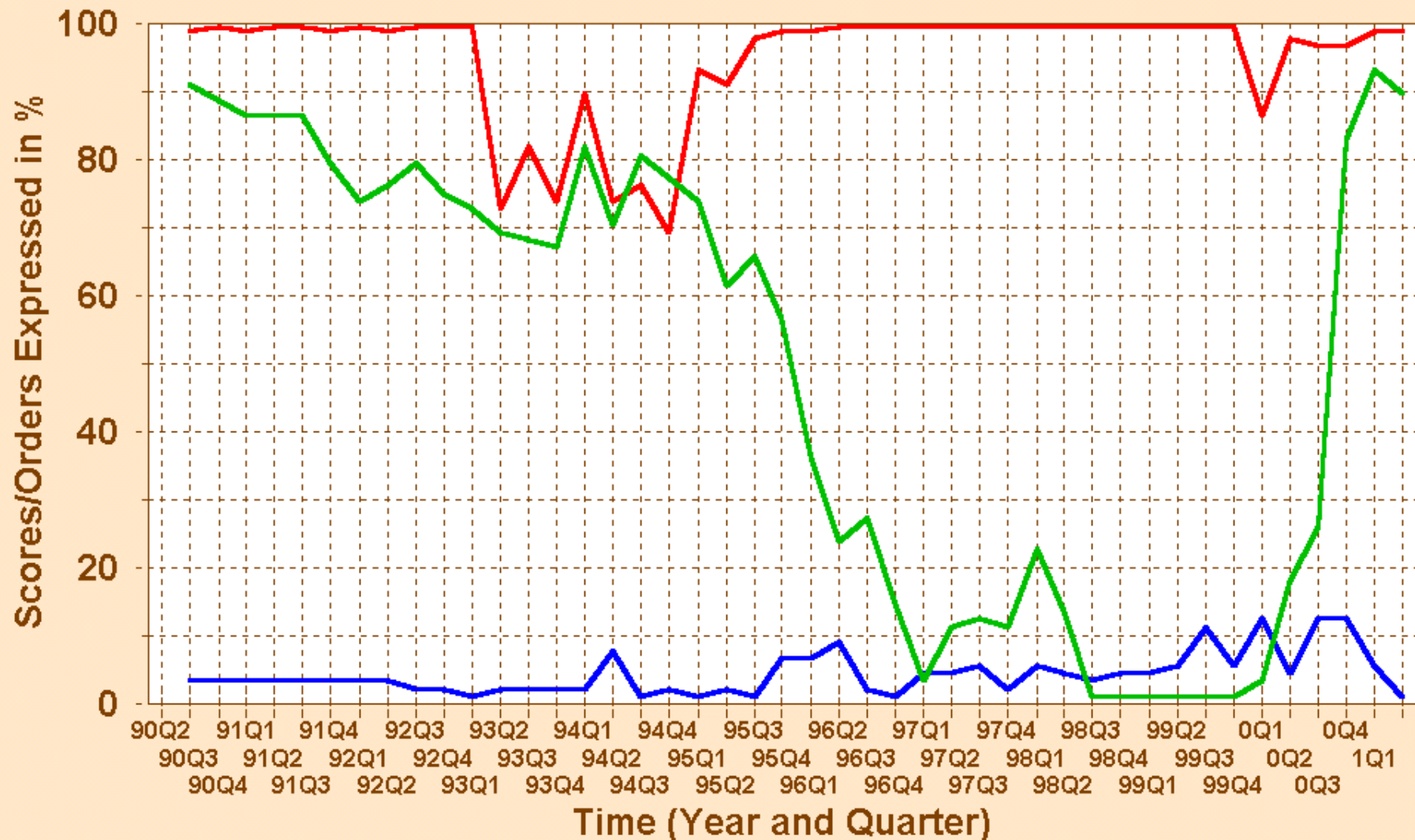


Fig.24.2: M-RATINGS OF THREE STOCKS

Data: 88 Stocks of 2 US Industries



— Keane Inc. — Xerox Corp. — Cirrus Logic

MD-MONITORING OF AN OBJECT

Changes in the MD-model of CCG_BAS

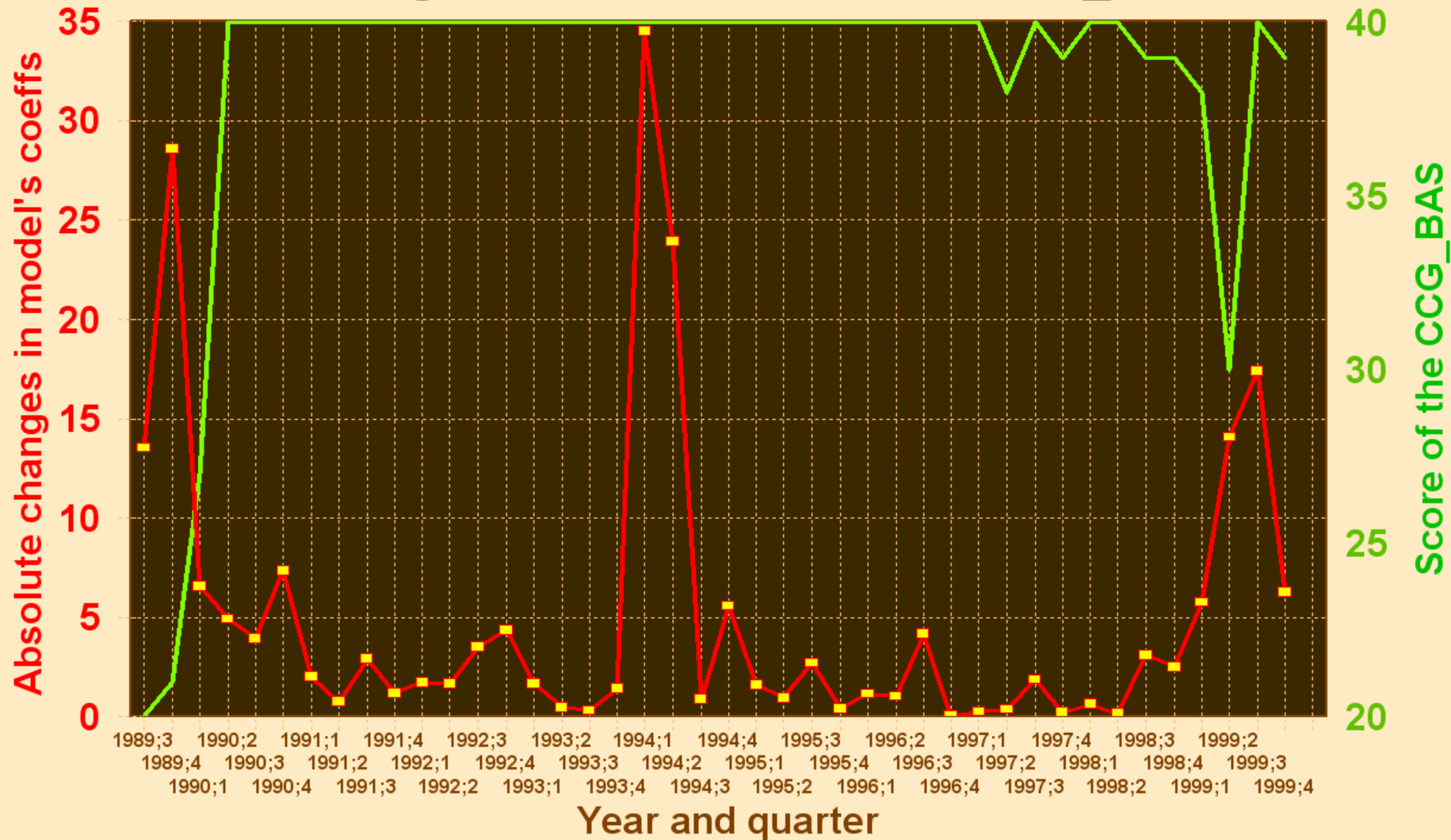
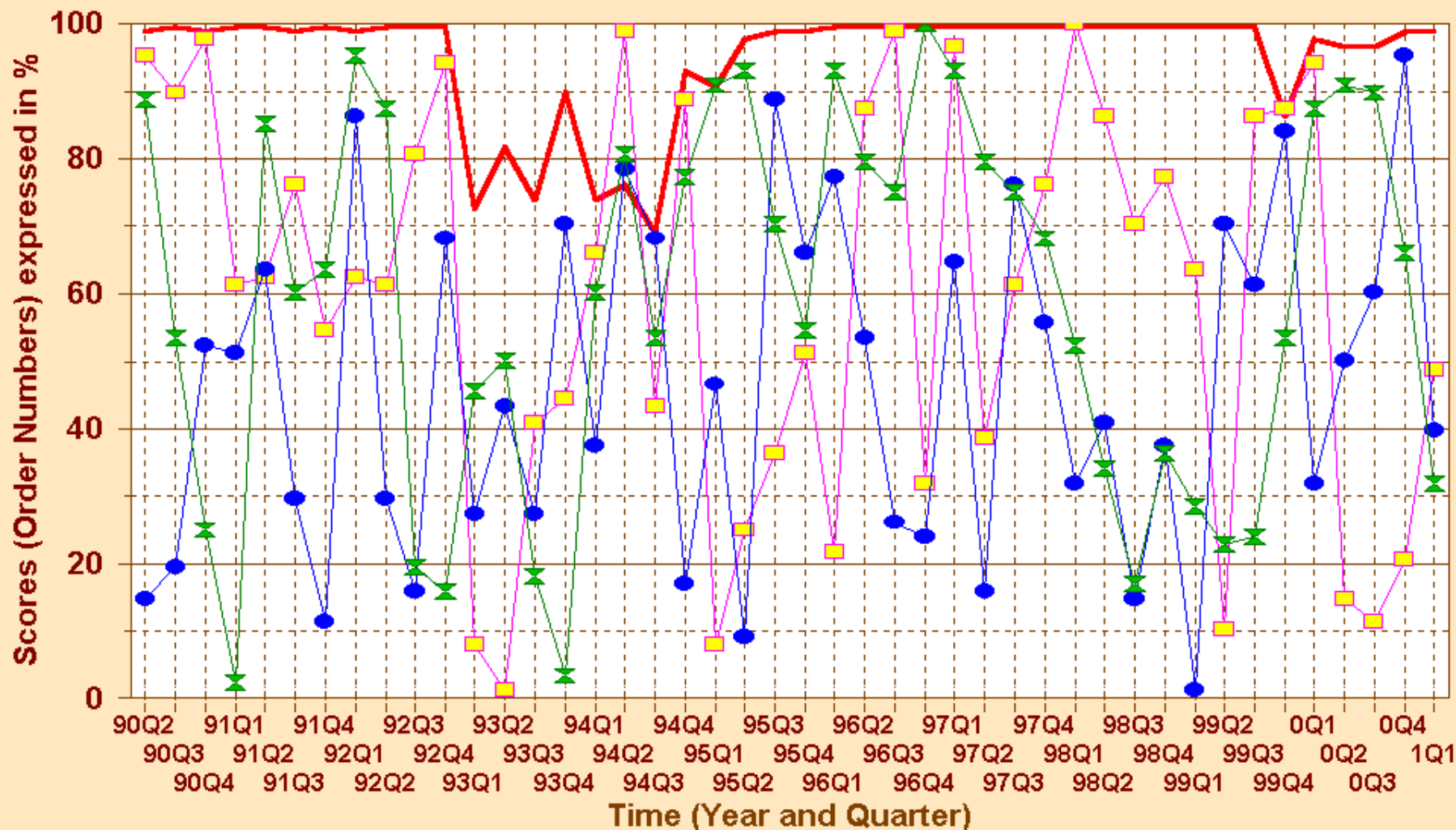


Fig.24.8: MARKET GENERATES CHAOS

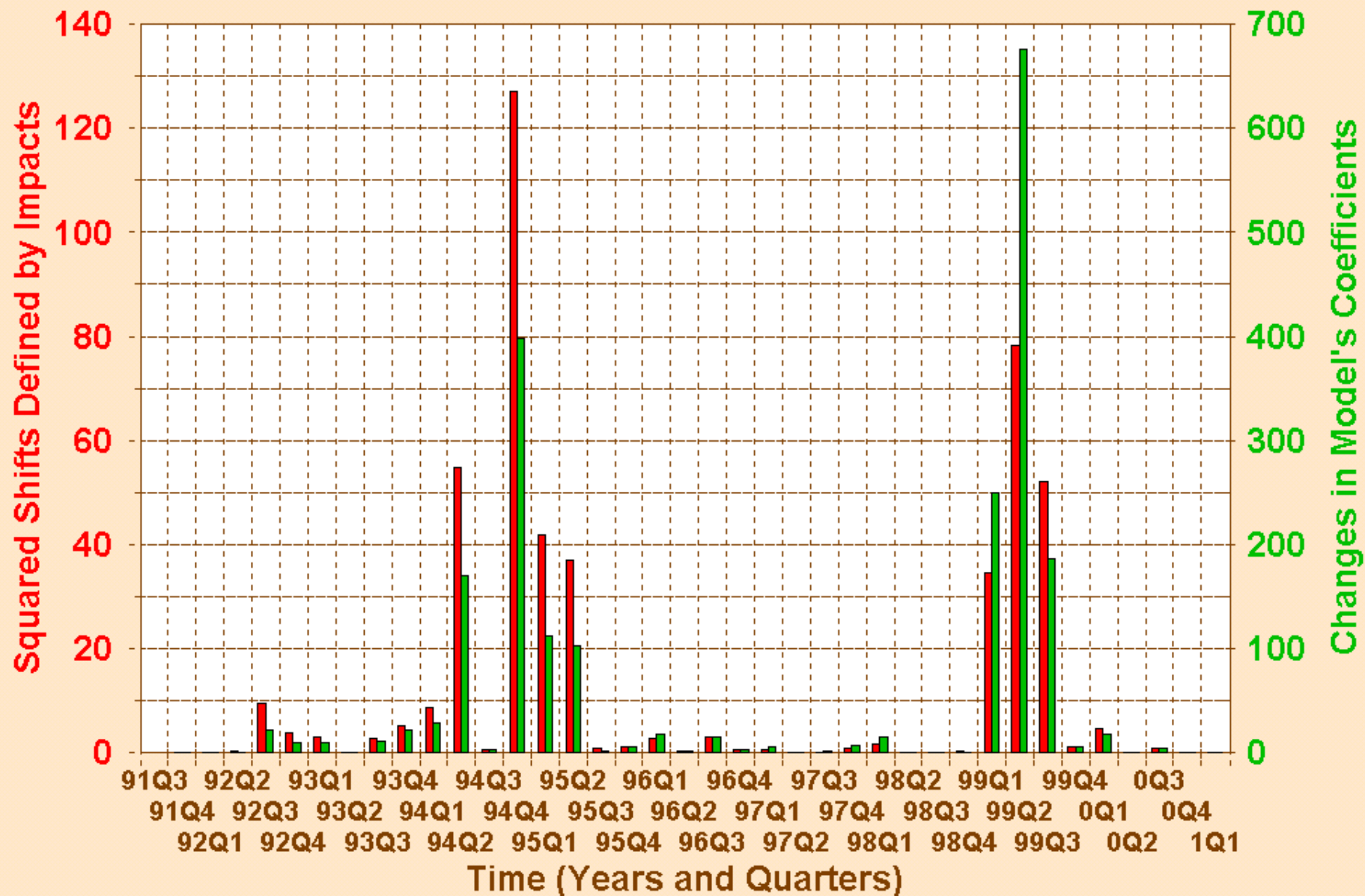
Scores of the Keane, Inc.



Inner M-Rating % Ext./In. M-Rating % Ord(Earn./Price) % Ord(Total Return) %

Fig.24.11: TIME-SERIES MD-MONITORING

Model Changes of Cirrus Logic (CRUS)



CONCLUSIONS

- There exist a lot of methods of robust statistical theory suitable under specific data conditions.
- It is difficult to decide which of these methods is to be applied to a particular case.
- Mathematical gnostics is based on axioms satisfied by all really measured data. Its applicability is therefore more universal. Moreover, it maximizes the resulting information.