17. International Diffuse Reflectance Conference, Aug 2014

Process sampling/variographics: Quality control of complete on-line measurement systems

Kim H. Esbensen
Geological Survey of Denmark and Greenland (GEUS)
Copenhagen, Denmark

Jaco Minnaar
Timmerman Analytical, Belville, South Africa

Timmerman Analytical

Who / What / When ….

Kim H. Esbensen, research professor at Denmark & Greenland Geological Surveys (GEUS) (*chemometrics and sampling*), 2010

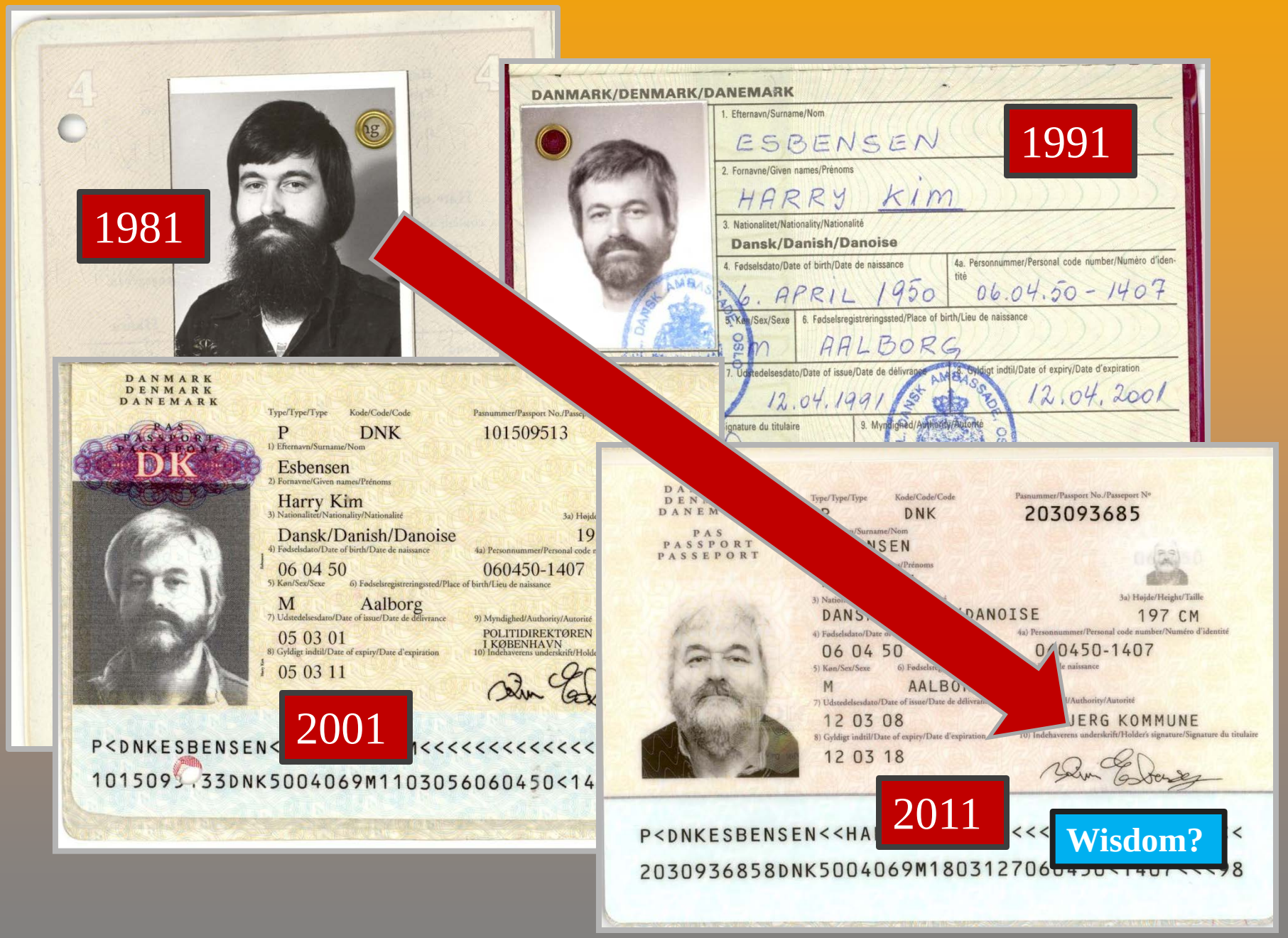
Aalborg University, prof. (*chemometrics & sampling*), 2001

Telemark University of Process Technology (HIT), prof. 1991

Norwegian Computing Center (NCC) & SINTEF, 1985

Terra Swede (exploration), 1982

Technical University of Denmark (DTH), Ph.D. 1981
Århus University, Denmark: M.Sc. (geology), 1979

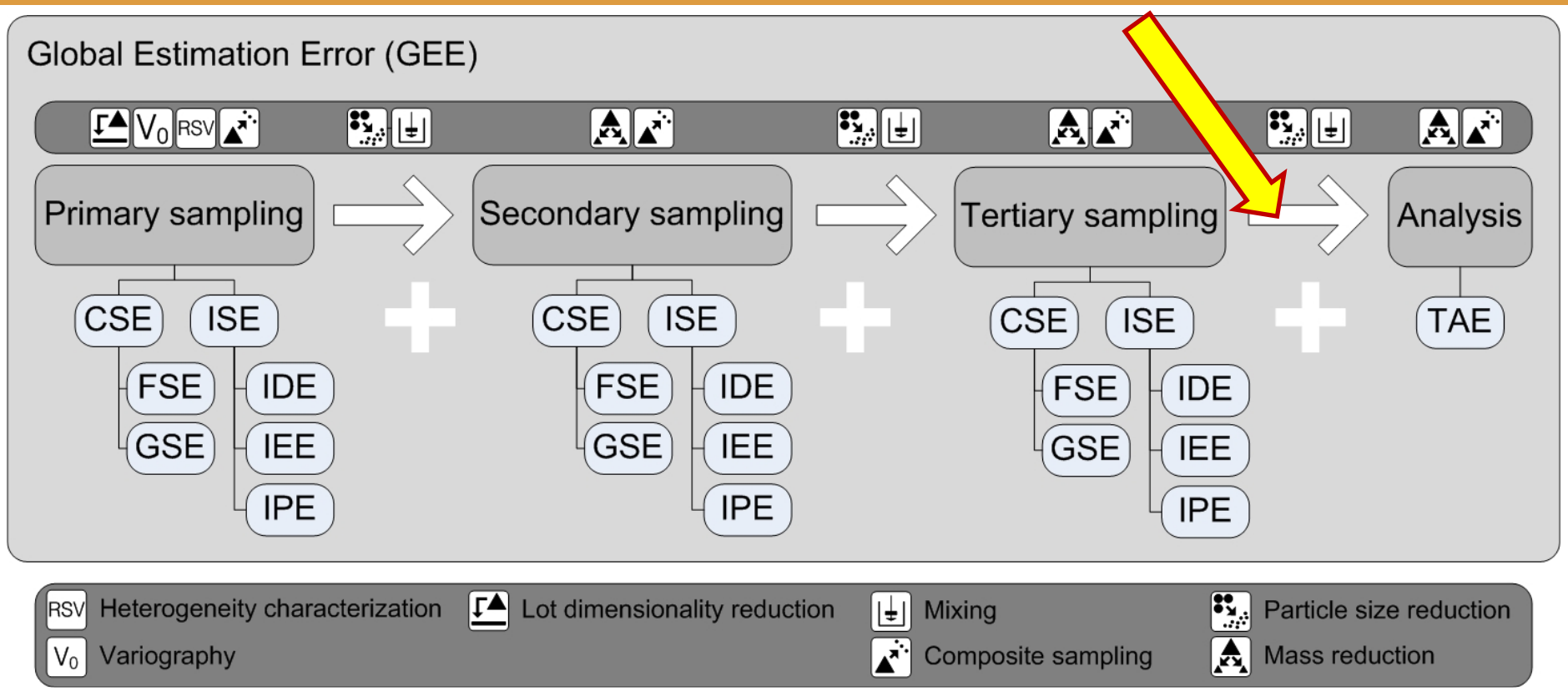

1981
DANMARK/DENMARK/DANEMARK
1. Efternavn/Surname/Nom
ESBENSEN
1991
2. Fornavne/Given names/Prénoms
HARRY kim
3. Nationalitet/Nationality/Nationalité
Dansk/Danish/Danoise
4. Fødselsdato/Date of birth/Date de naissance
6. APRIL 1950
4a. Personnummer/Personal code number/Numéro d'identité
06.04.50-1407
5. Køn/Sex/Sexe
M
6. Fødselsregistreringssted/Place of birth/Lieu de naissance
AALBORG
7. Udstedelsesdato/Date of issue/Date de délivrance
12.04.1991
8. Gyldigt indtil/Date of expiry/Date d'expiration
12.04.2001
DANMARK
DENMARK
DANEMARK
PAS
PASSPORT
PASSEPORT
DK
Type/Type/Type
P
Kode/Code/Code
DNK
Pasnummer/Passport No./Passep
101509513
1) Efternavn/Surname/Nom
Esbensen
2) Fornavne/Given names/Prénoms
Harry Kim
3) Nationalitet/Nationality/Nationalité
Dansk/Danish/Danoise
3a) Højde
19
4) Fødselsdato/Date of birth/Date de naissance
06 04 50
4a) Personnummer/Personal code
060450-1407
5) Køn/Sex/Sexe
6) Fødselsregistreringssted/Place of birth/Lieu de naissance
M
Aalborg
7) Udstedelsesdato/Date of issue/Date de délivrance
05 03 01
9) Myndighed/Authority/Autorité
POLITIDIREKTØREN
I KØBENHAVN
8) Gyldigt indtil/Date of expiry/Date d'expiration
05 03 11
10) Indehaverens underskrift/Holde
P<DNKESBENSEN<
<<<<<<<<<<<<<<
1015095
33DNK5004069M1103056060450<14
2001
Type/Type/Type
Kode/Code/Code
DNK
Pasnummer/Passport No./Passeport N°
203093685
PAS
PASSPORT
PASSEPORT
DANOISE
3a) Højde/Height/Taille
197 CM
4) Fødselsdato/Date
06 04 50
4a) Personnummer/Personal code number/Numéro d'identité
0
0450-1407
5) Køn/Sex/Sexe
M
AALBO
7) Udstedelsesdato/Date of issue/Date de délivra
12 03 08
JERG KOMMUNE
8) Gyldigt indtil/Date of expiry/Date d'expiration
12 03 18
10) Indehaverens underskrift/Holder's signature/Signature du titulaire
P<DNKESBENSEN<<HA
<<<
2030936858DNK5004069M1803127
2011
Wisdom?

Representative Sampling: Theory of Sampling (TOS)

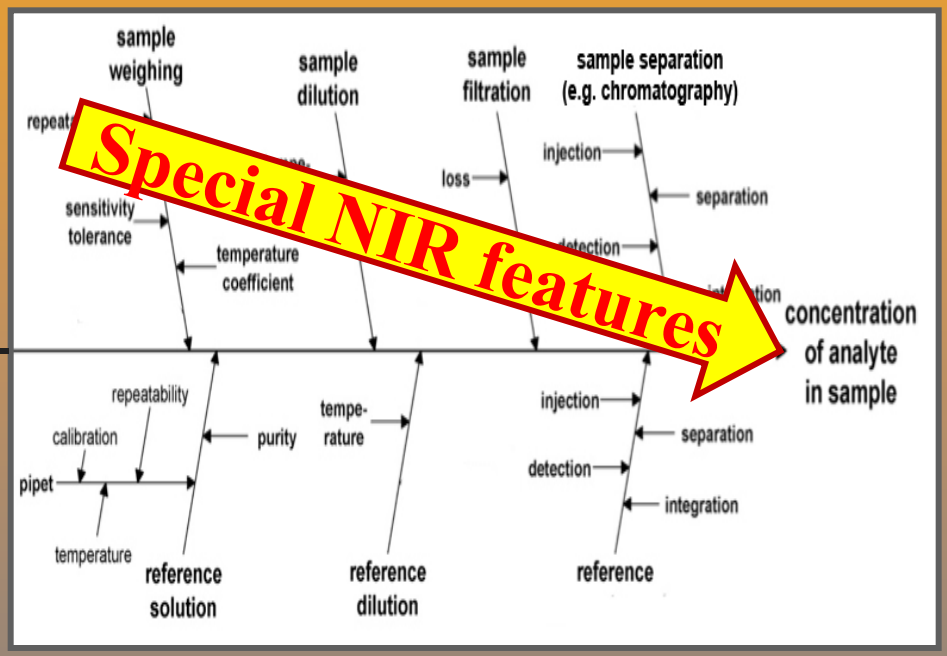
Complete, axiomatic exposé of TOS

1. FSP: Fundamental Sampling Principle
2. SSI: Sampling Scale Invariance
3. PSC: Sampling Correctness (bias-free sampling)
4. PSS: Sampling Simplicity (primary sampling + mass-reduction)
5. LDT: Lot Dimensionality Transformation
6. Process sampling quality assurance: 1-D lots
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

“HORIZONTAL -
a matrix-independent standard for representative sampling”

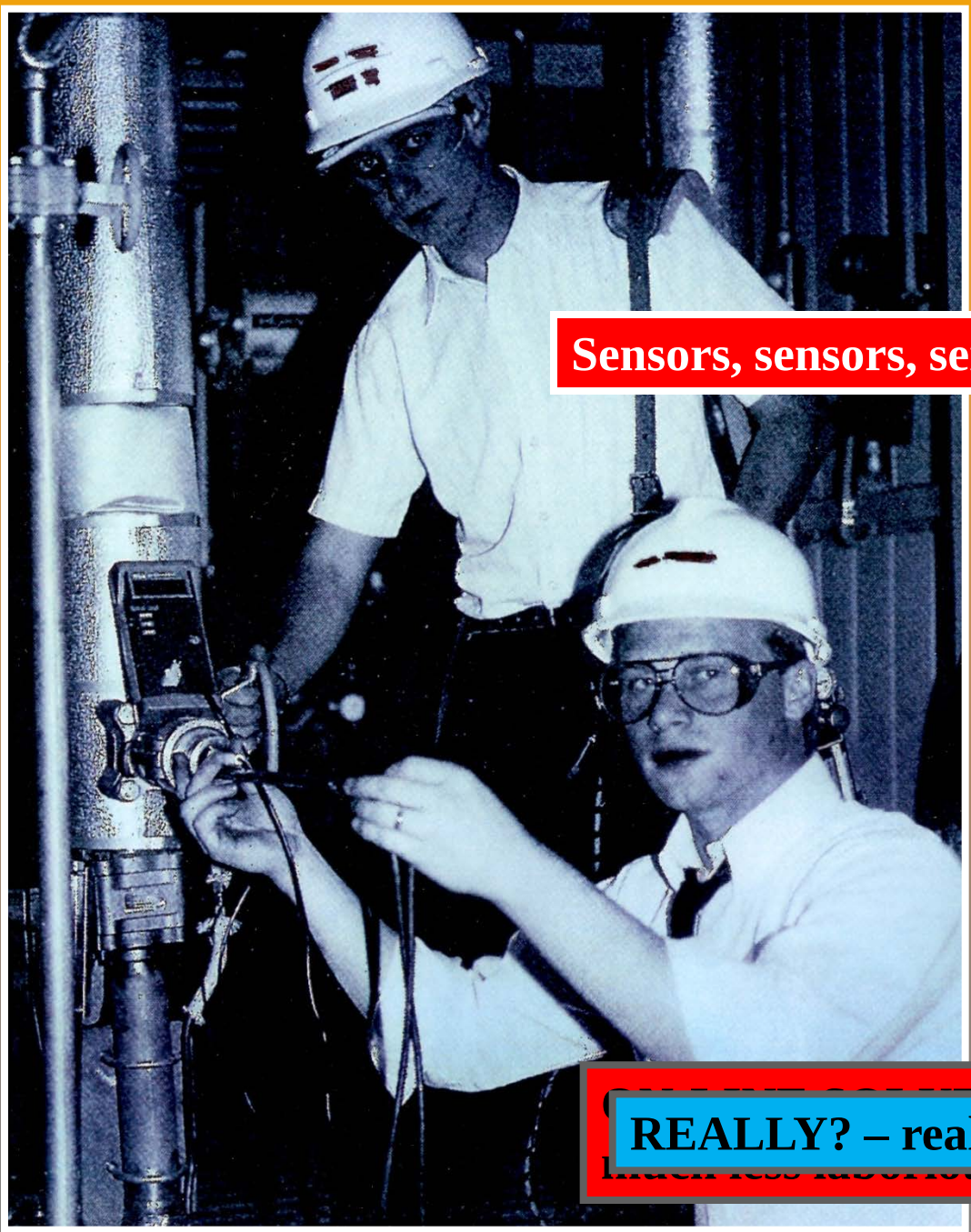


Analysis (incl. NIR analysis)



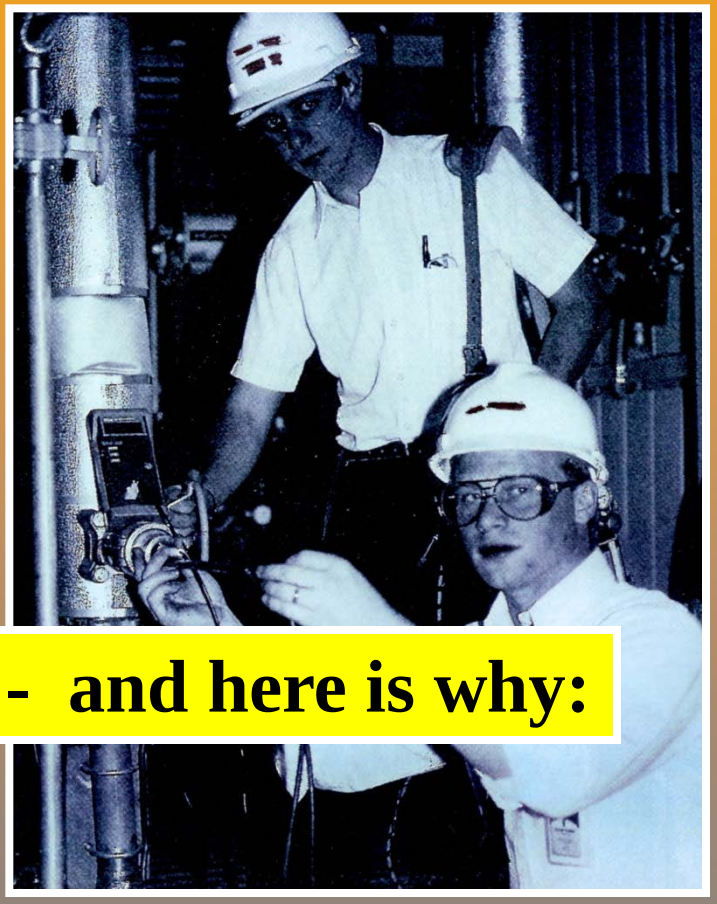
Special NIR features

$$Mu_{total} = MU_{sampling} + MU_{analysis}$$



Sensors, sensors, sensors & insert probes !

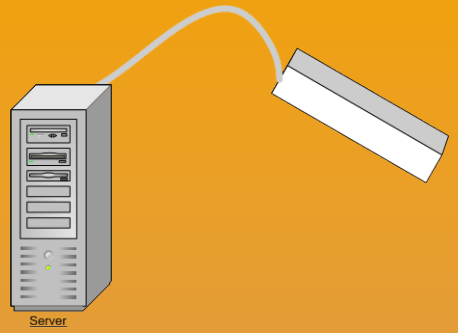
REALLY? – really? (”no sampling”)??

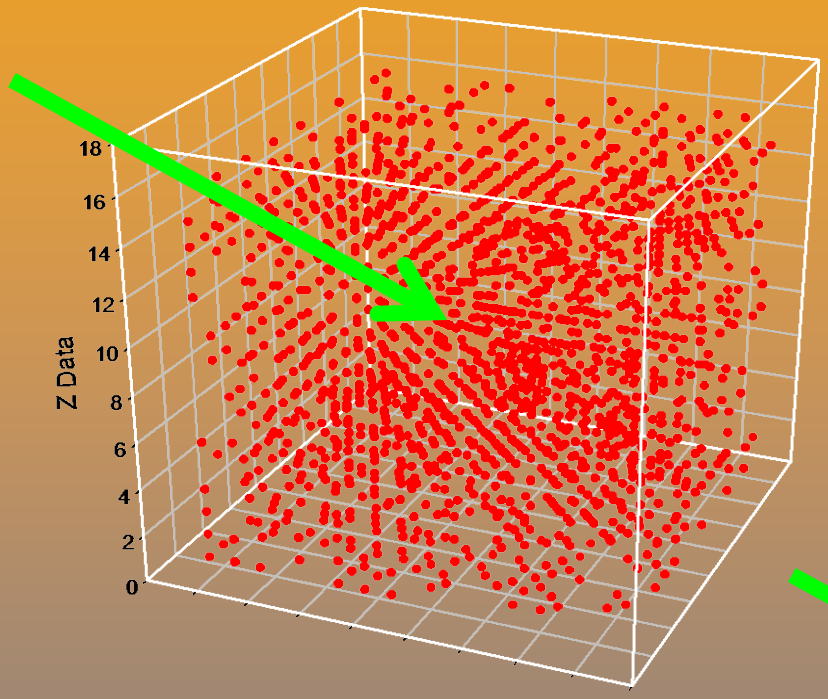


Well - not so fast - and here is why:

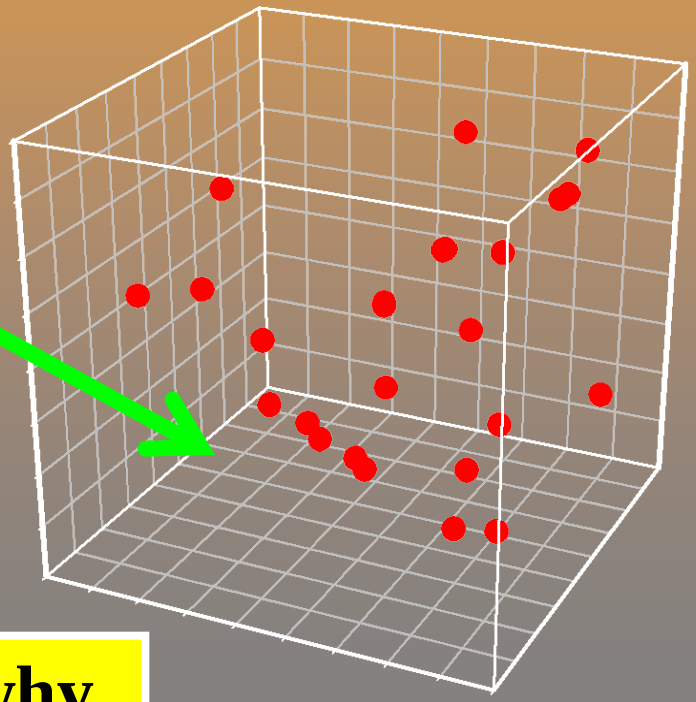
The so-called ”PAT revolution” a.o. hailed as such due to there be no need for sampling: Process SENSORS!!

Tacit assumption: ” most likely a reasonably homogenous material”





The harsh reality: … WHAT IF? - signifiantly heterogenous materials

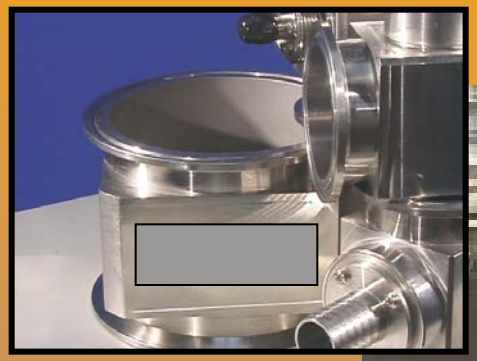


Well - not so fast - and here is why

“Elimination of sampling” ….

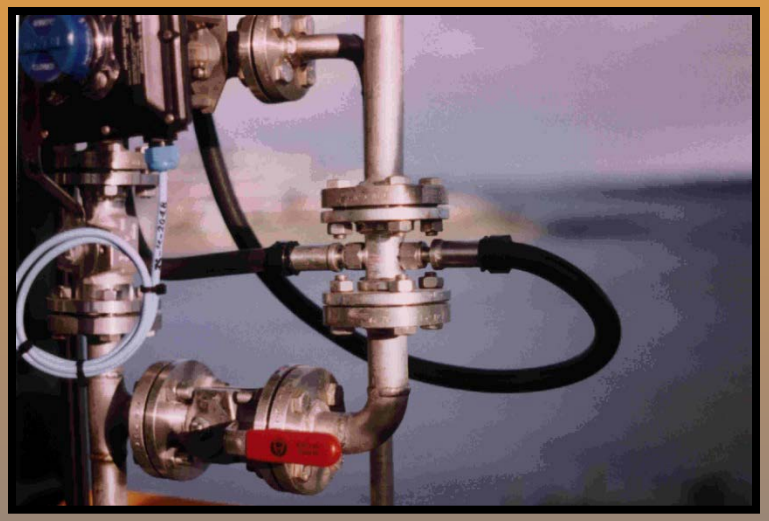
Tacit assumption: ”most likely a reasonably homogenous material”

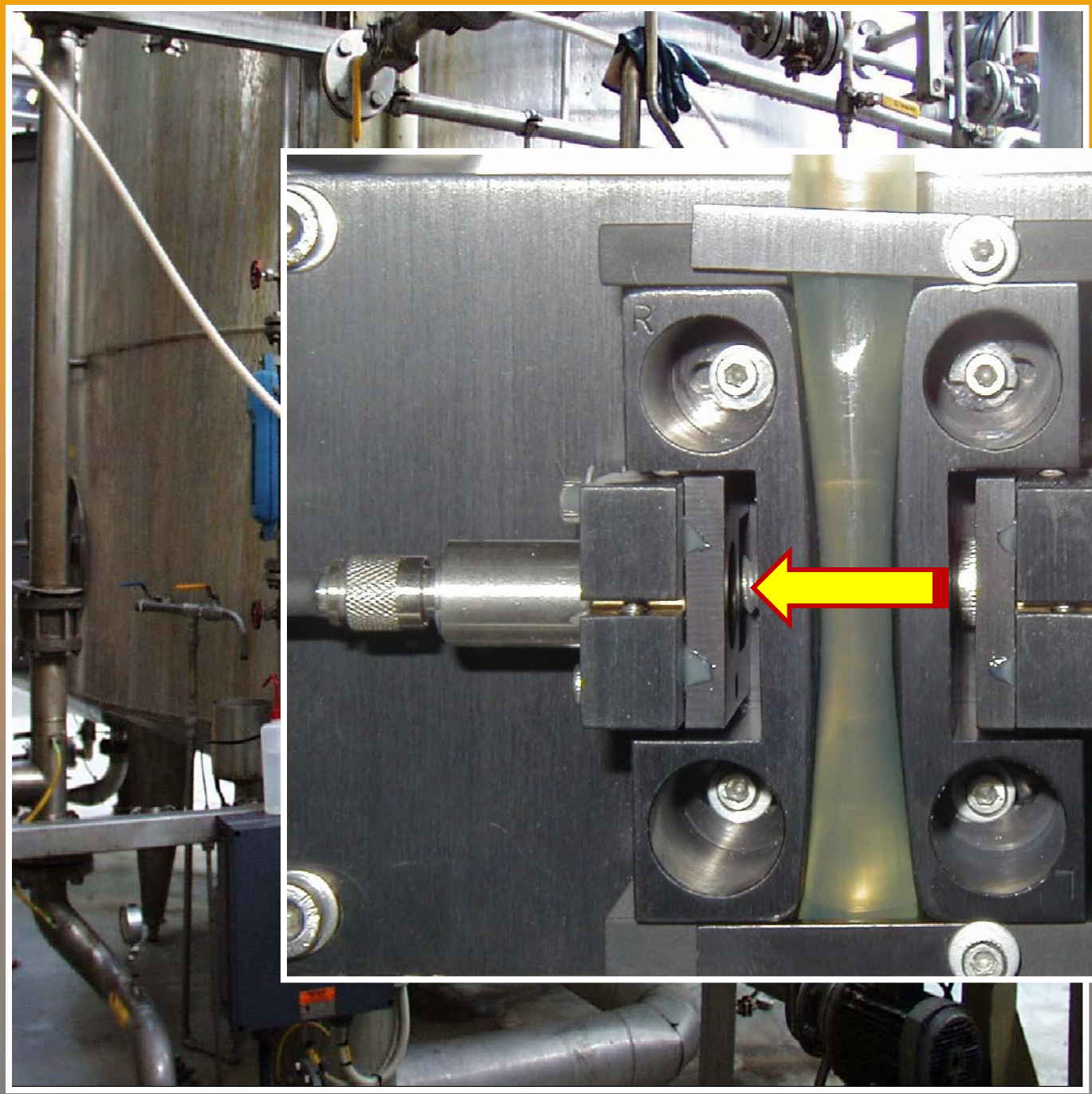
On-line/in-line flow-cells and sensors

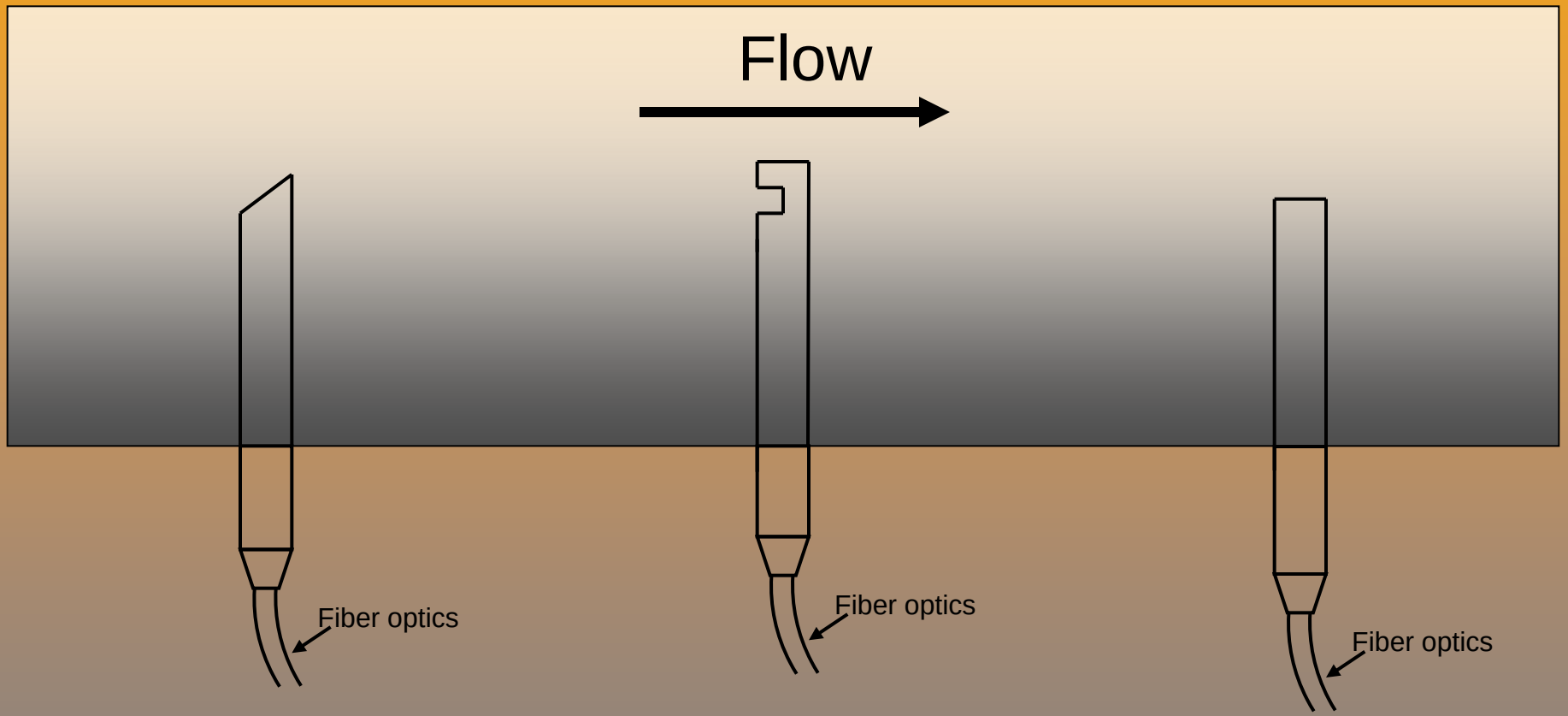


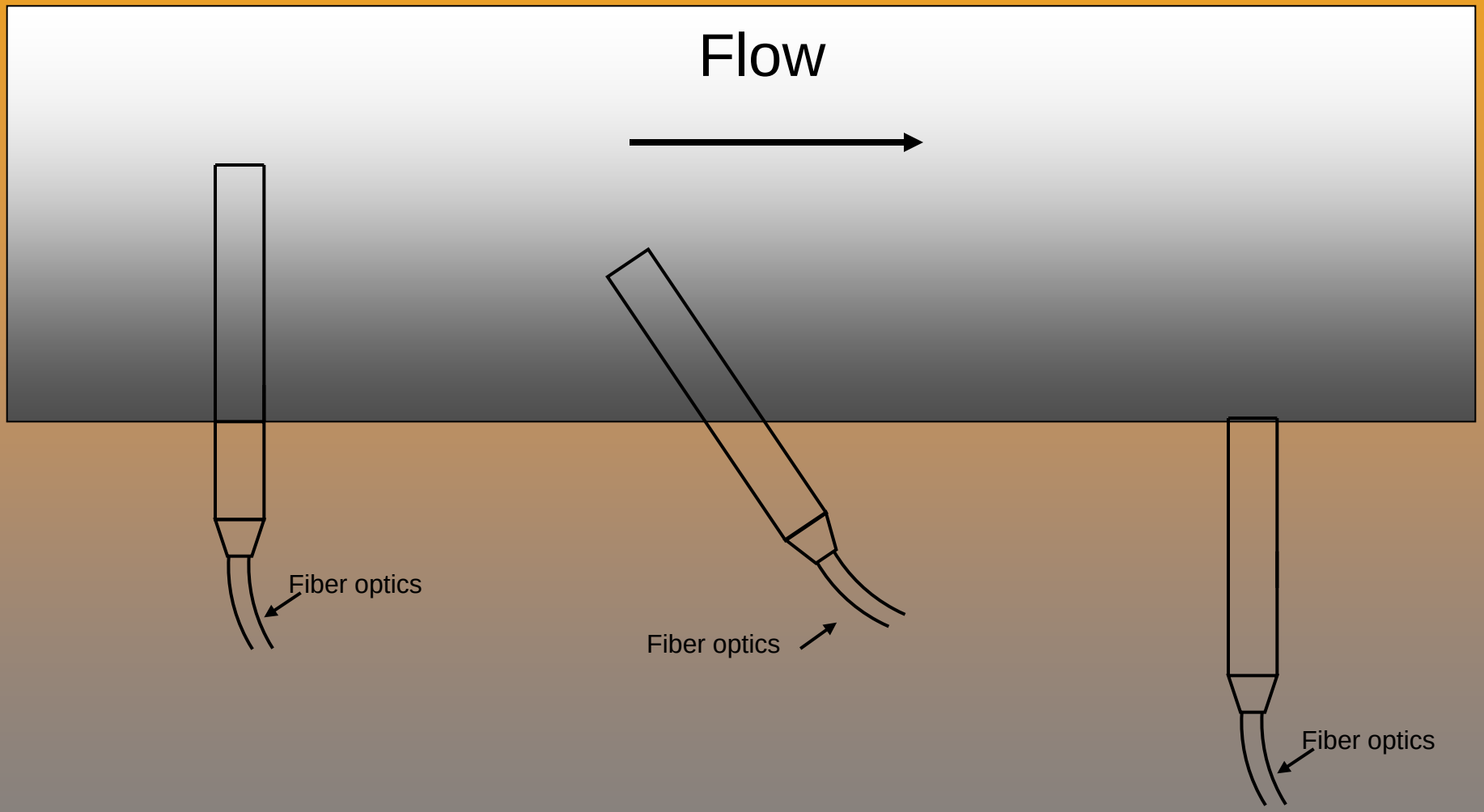
On-line / in-line PAT revolution

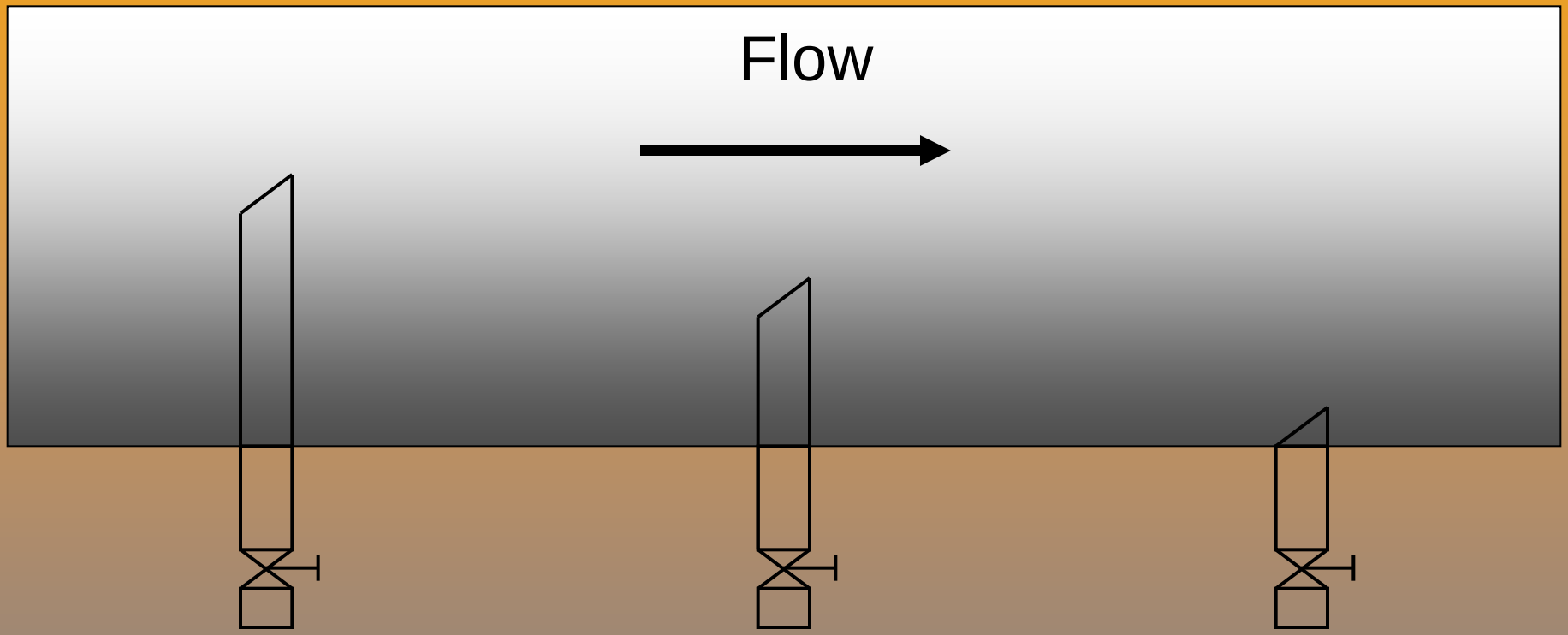
ON-line NIR etc - good stuff



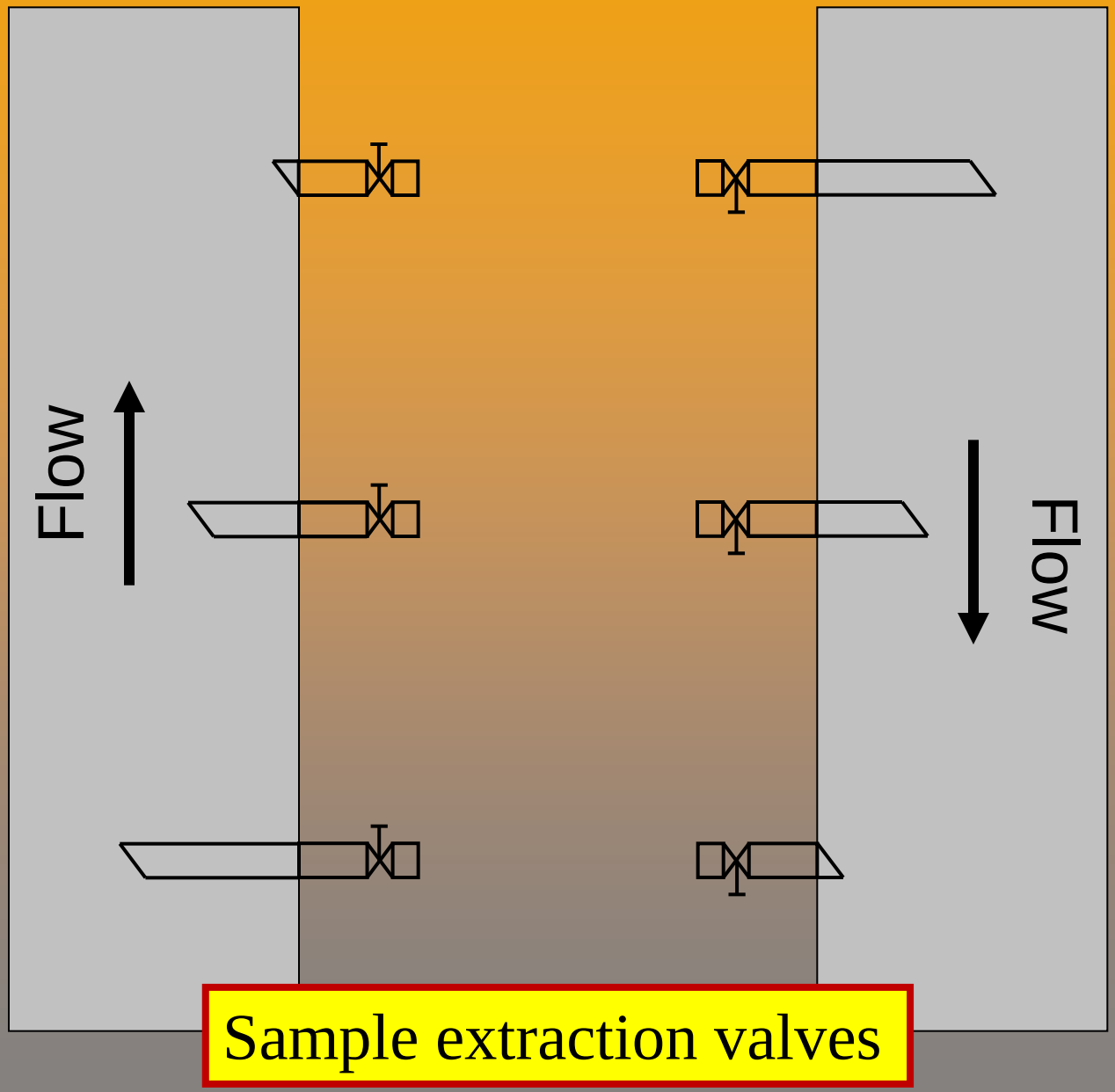


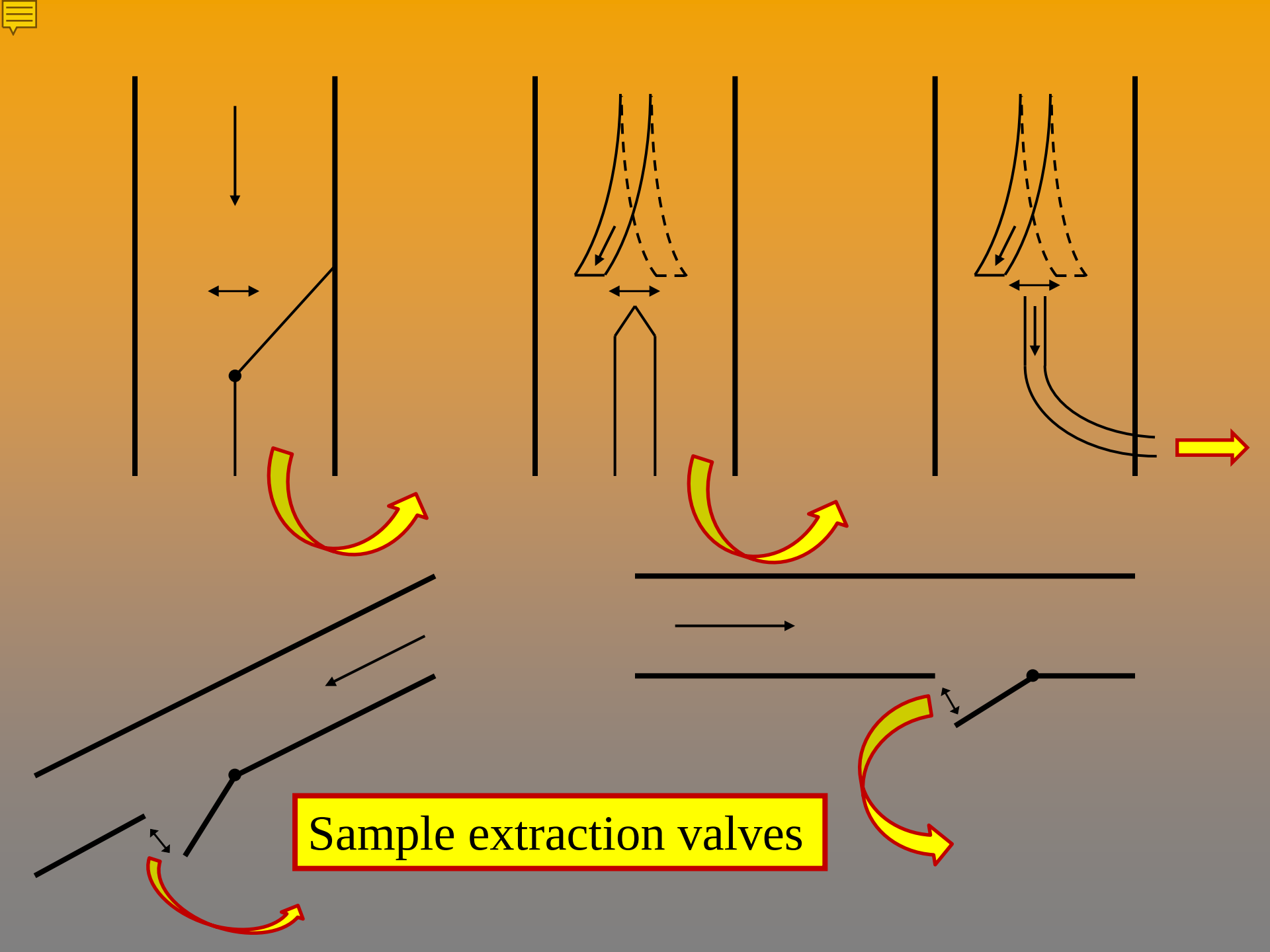

Flow
Fiber optics
Fiber optics
Fiber optics


Flow
Fiber optics
Fiber optics
Fiber optics



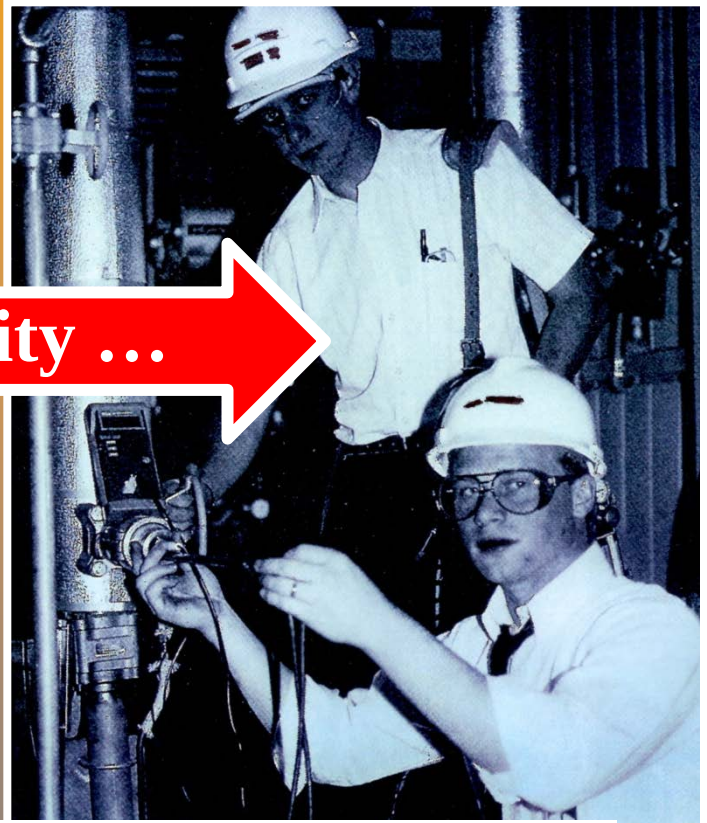
Sample extraction valves


Flow
Flow
Sample extraction valves

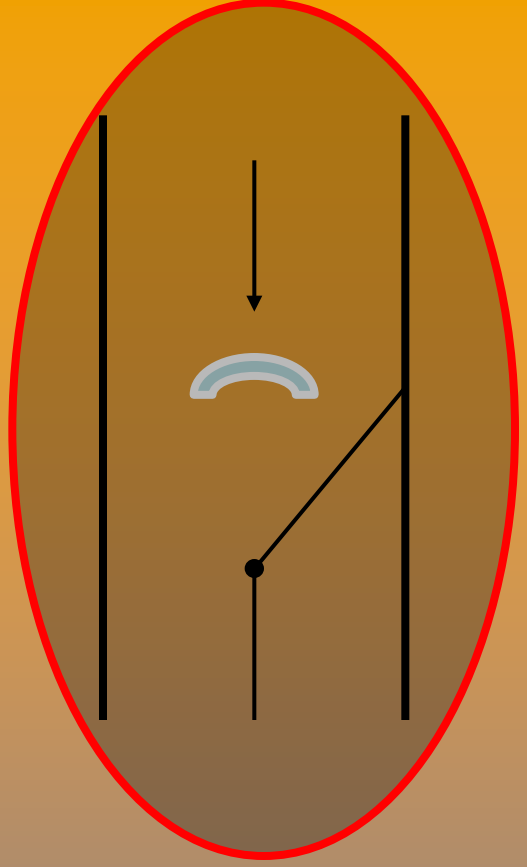

Sample extraction valves

This lecture's agenda: DUALITY

Duality …

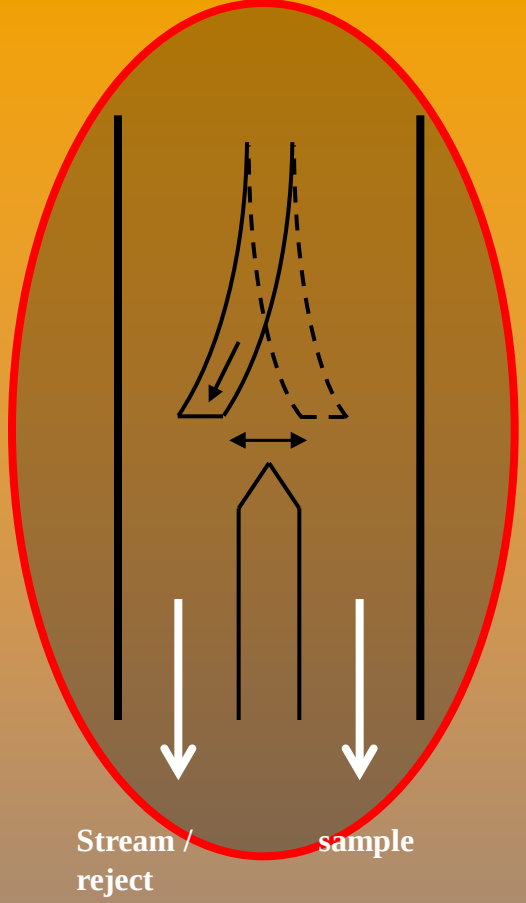


N.B. Identical "sampling issues" with, as without sensor technologies (PAT) ..



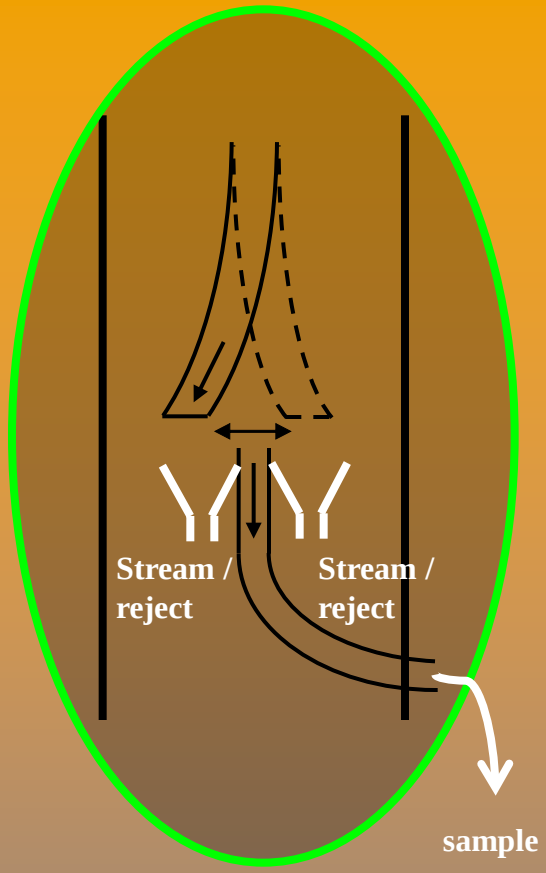
Flap valve design

1.



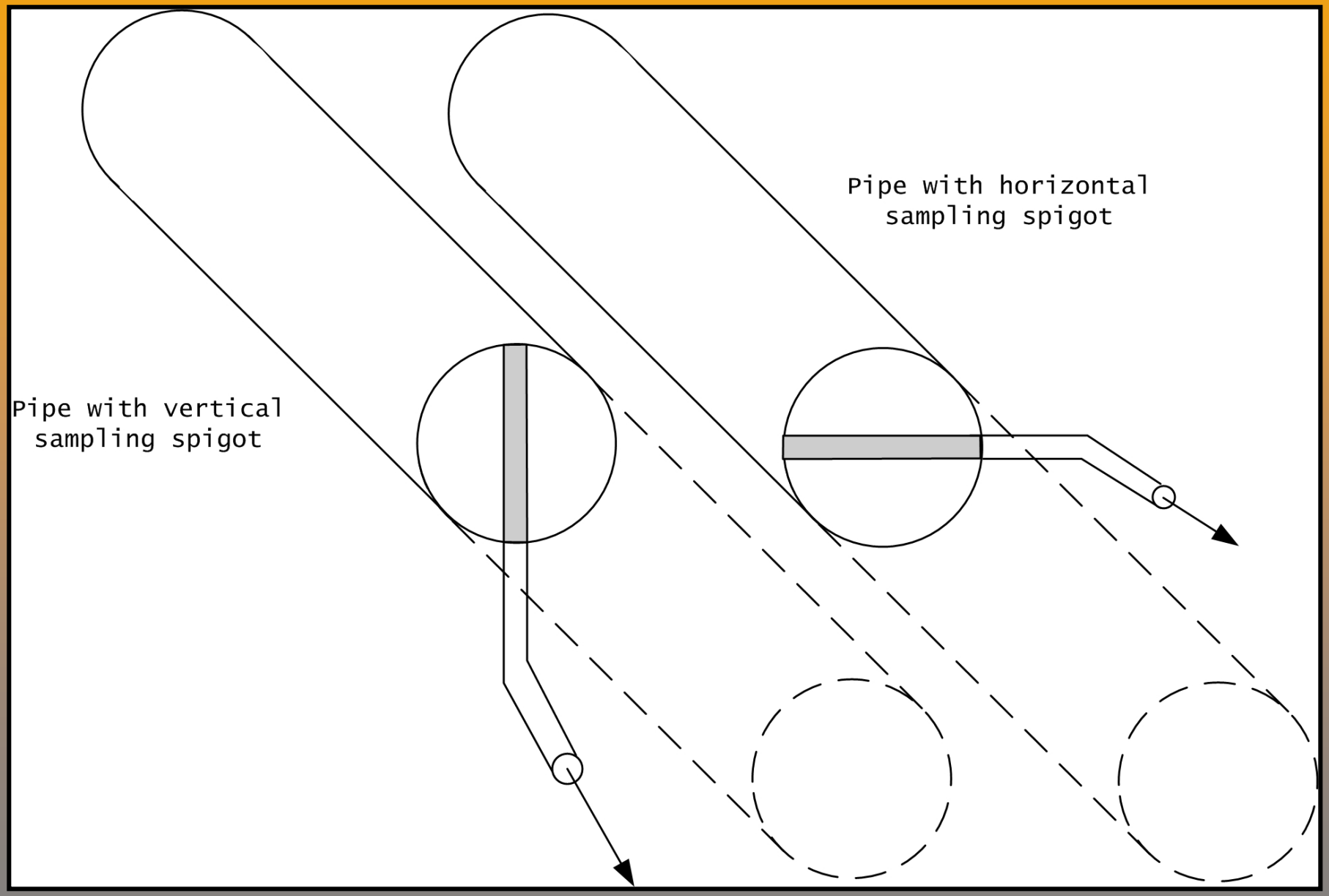
”Inversed” flap valve

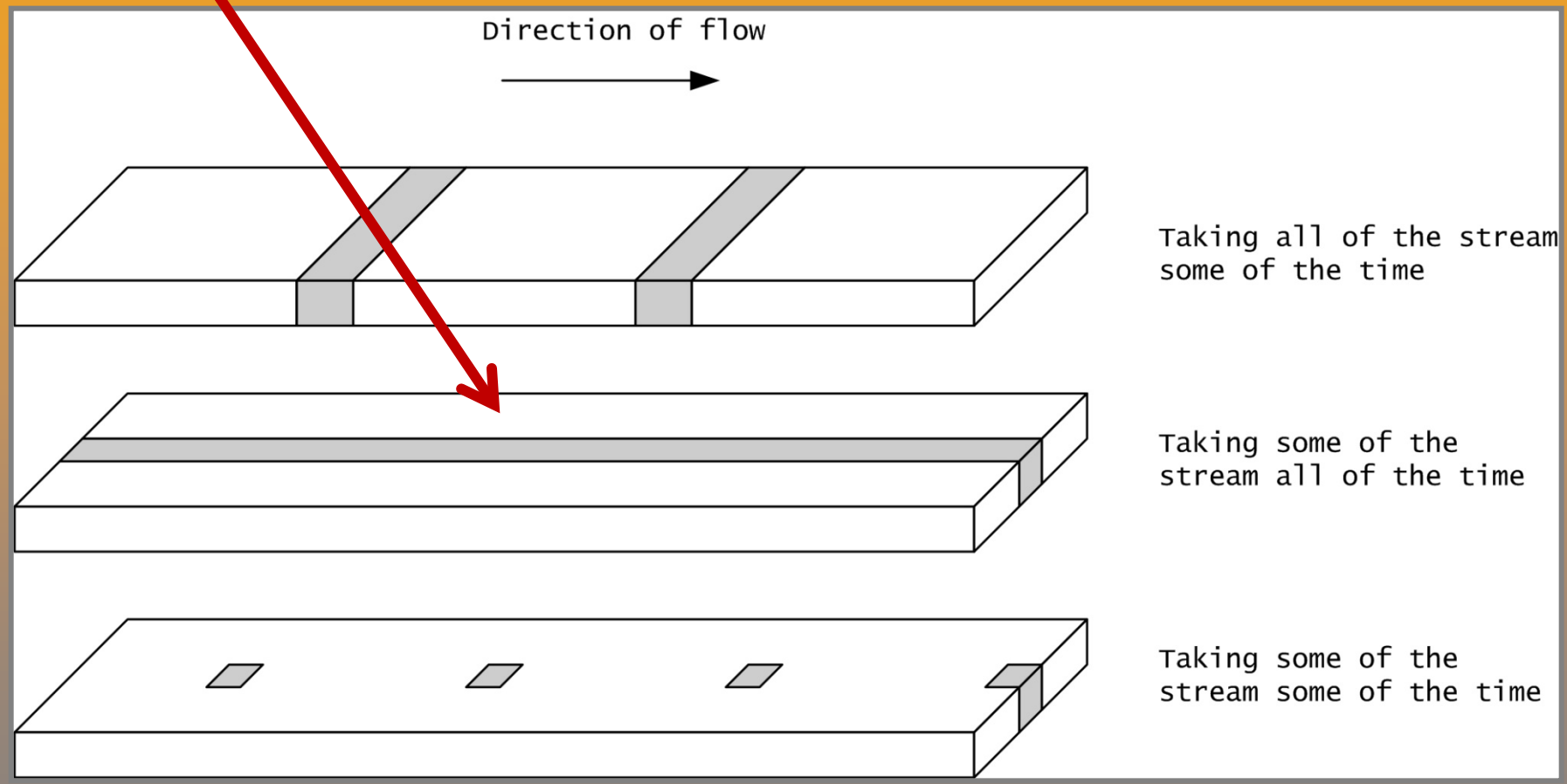
2.

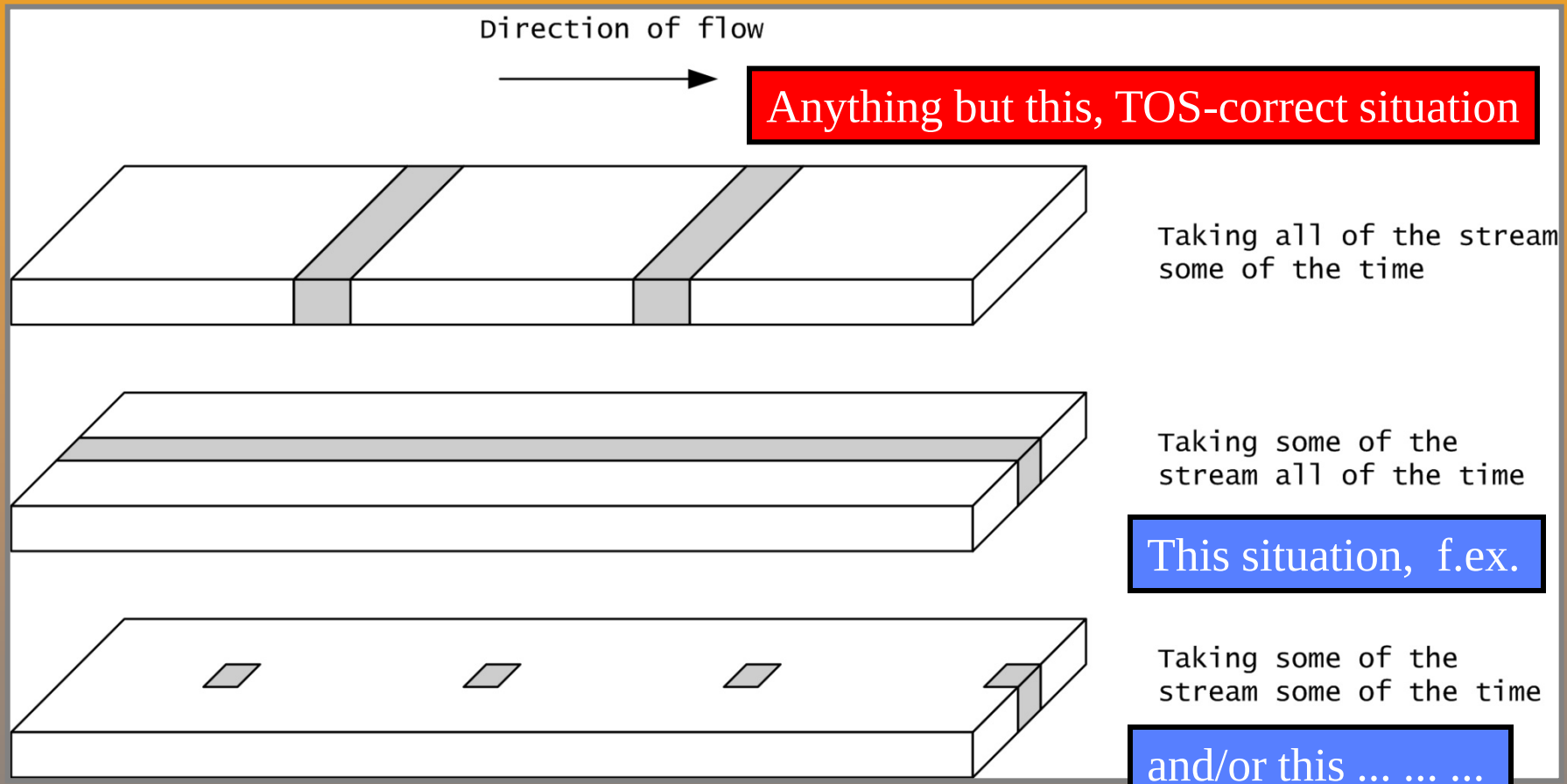


In-line outtake valve

3.

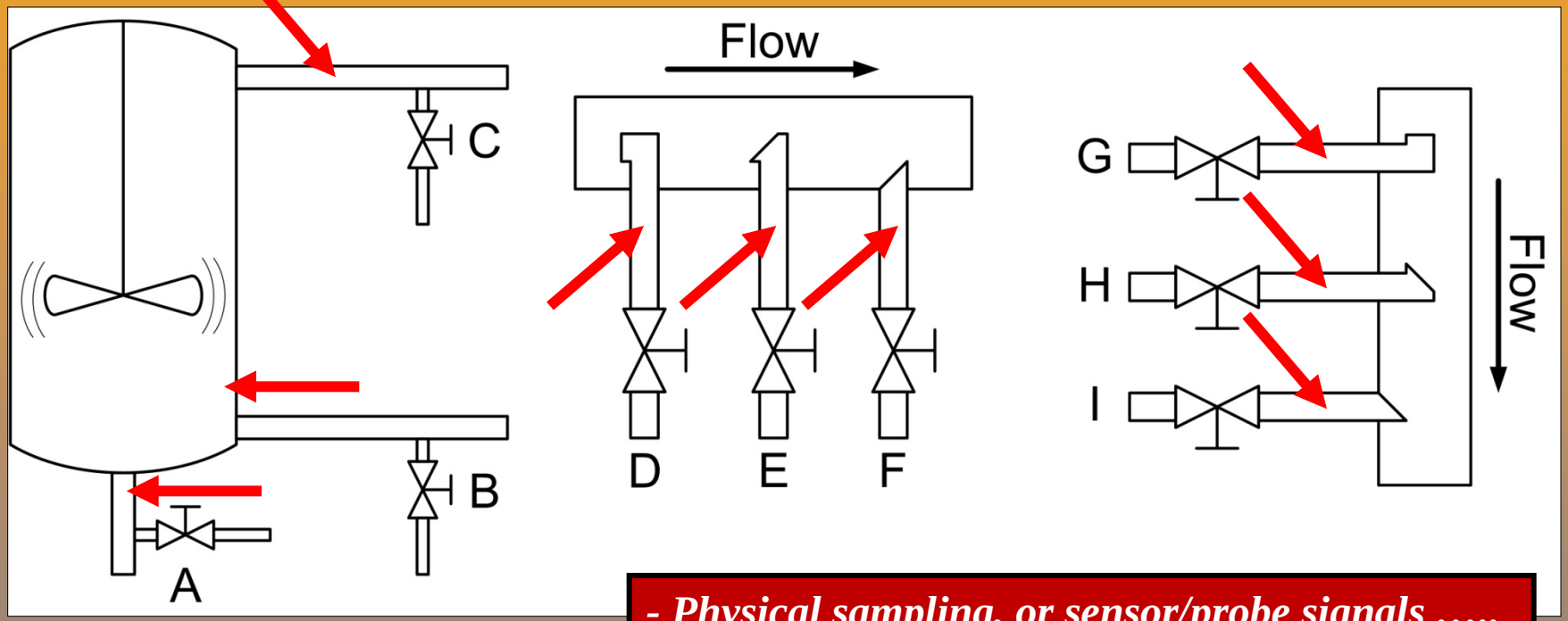

Pipe with horizontal
sampling spigot
Pipe with vertical
sampling spigot


Direction of flow
Taking all of the stream some of the time
Taking some of the stream all of the time
Taking some of the stream some of the time


Direction of flow
Anything but this, TOS-correct situation
Taking all of the stream some of the time
Taking some of the stream all of the time
This situation, f.ex.
Taking some of the stream some of the time
and/or this ... ...


Flow
Flow
Flow

Quick overview of 99 %-ile of PAT ”process sampling design”…



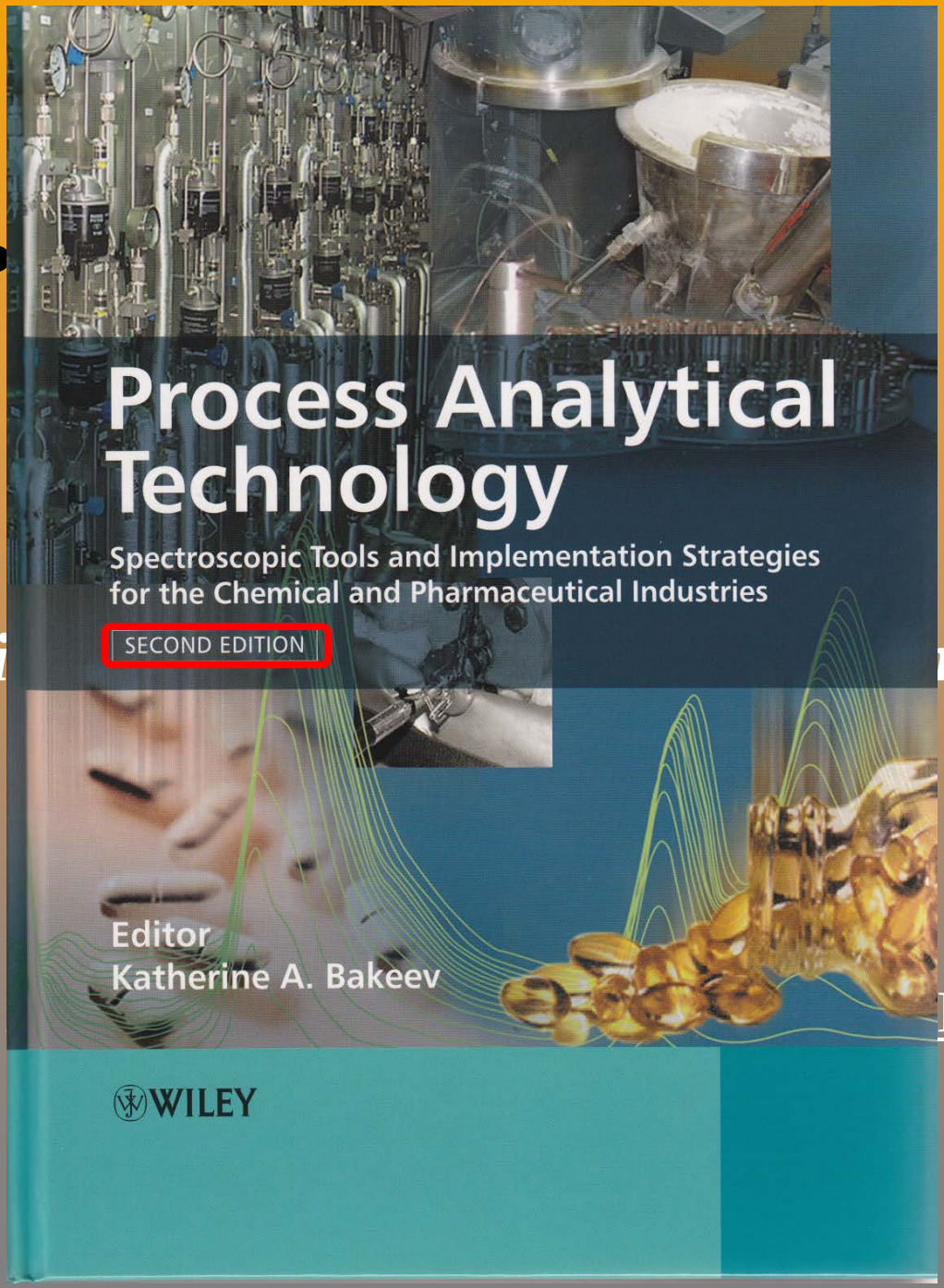
- Physical sampling, or sensor/probe signals …..

<u>All</u> these designs are ”*<u>incorrect</u>*” – sampling process is *non-representative* !

TOS terminology (Theory of Sampling)

P

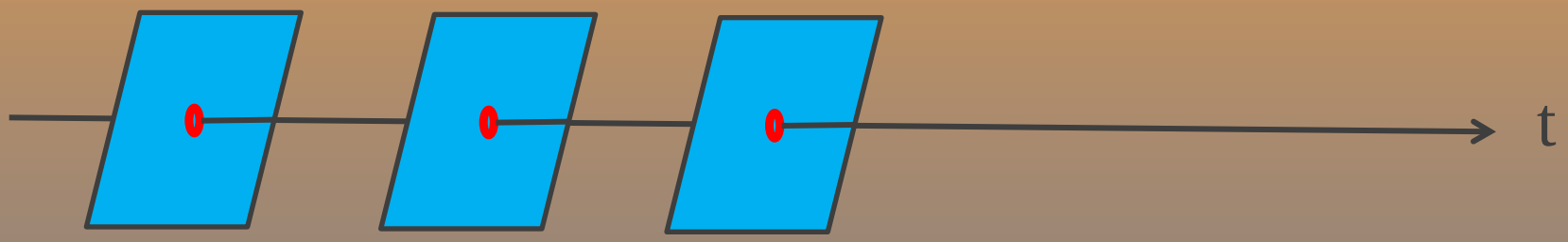
Ki *n*



PROCESS Ed.(*2009*)

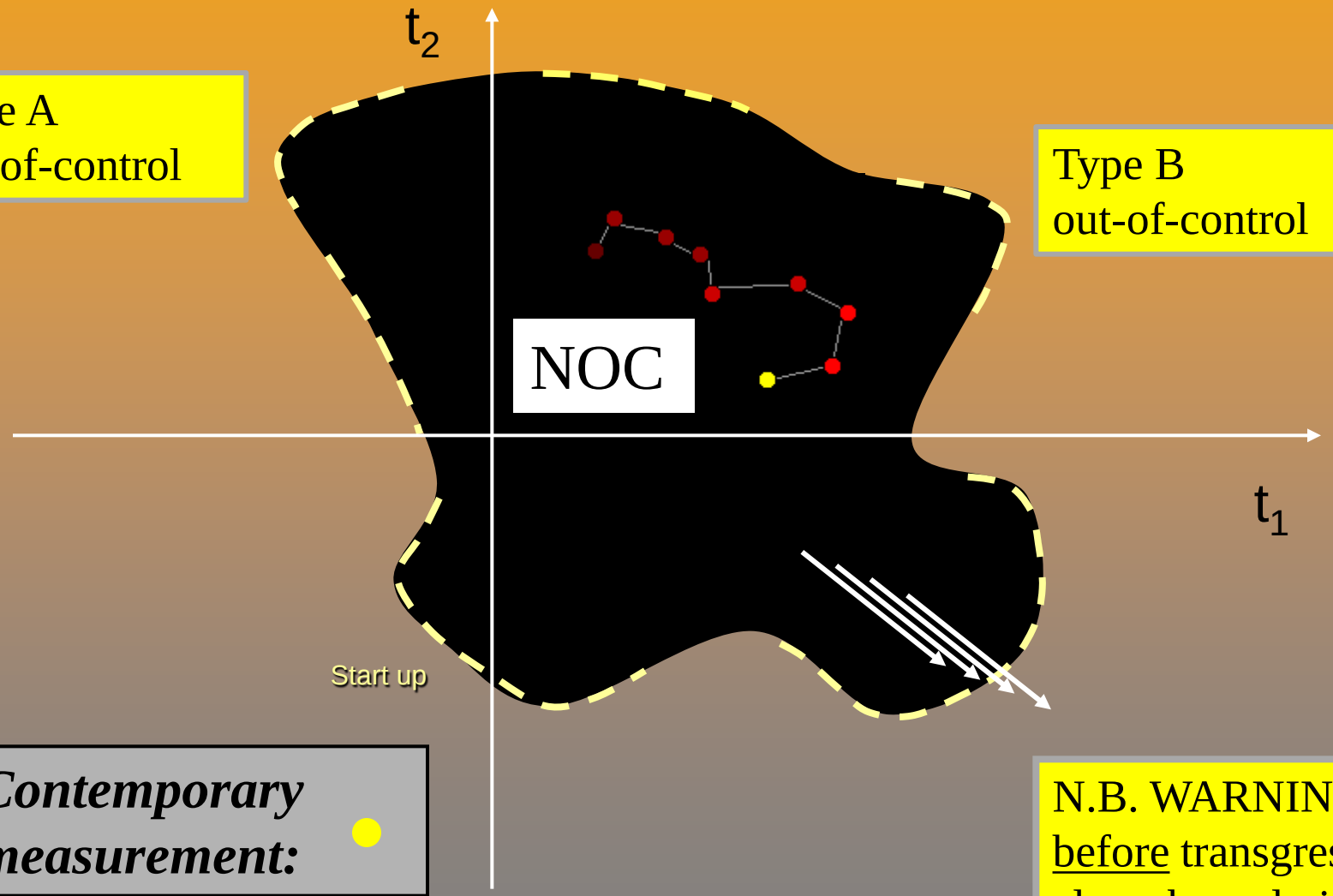
Process Chemometrics (PAC, PAT):

⇨ ”time” (process time / chrolonological time) … t



- essentially chemomrtics *following along* process time:
- PCA, PLS, R-way: MVDA data analysis/modeling
- e.g. t-t, p’-p’, w’-w’, T^2, Q, residuals, outlier, upsets .. …

Generic on-line process monitoring score plot



Type A
out-of-control

Type B
out-of-control

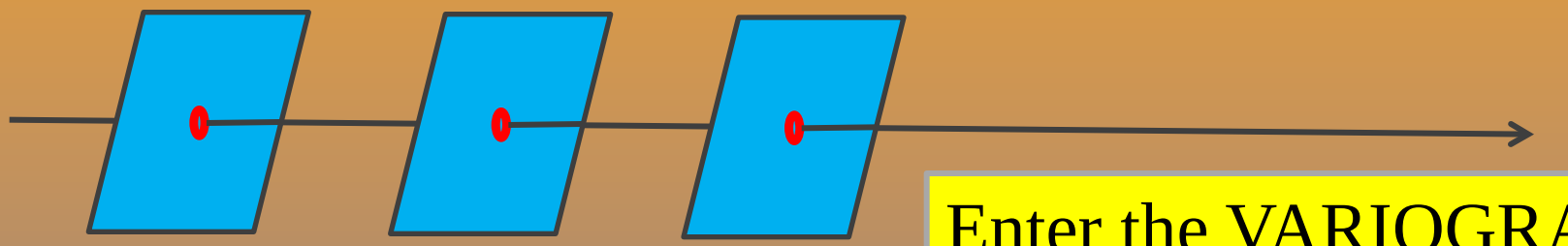
Contemporary measurement:

N.B. WARNING(s) <u>before</u> transgressing alarm boundaries !!

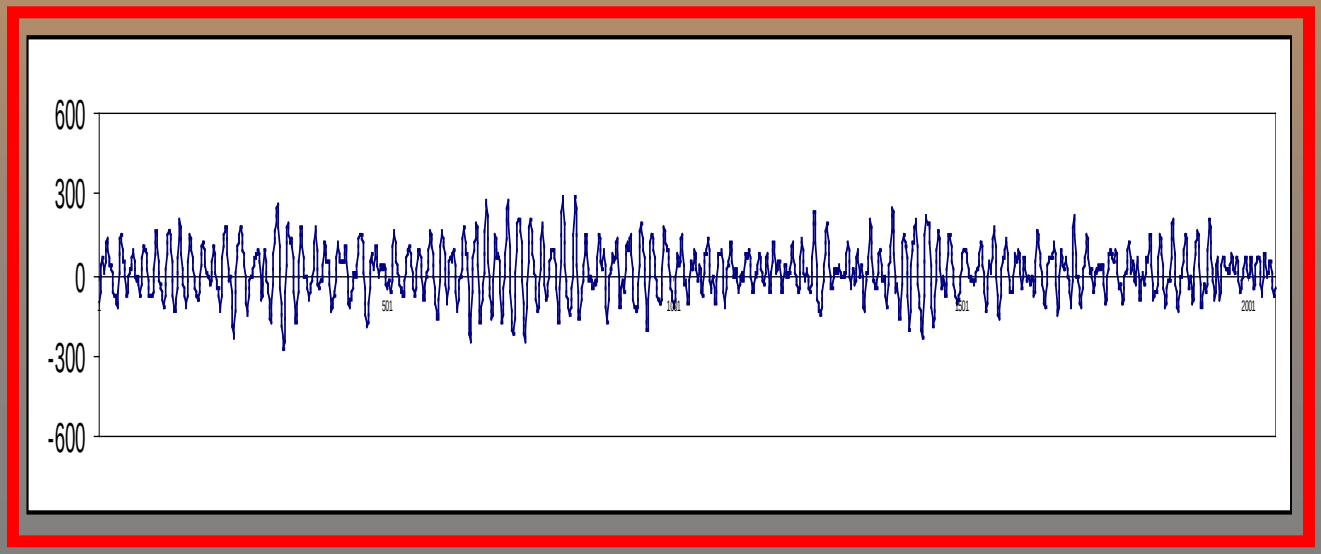
The thrust of this overview presentation:

Is there information in “small-scale variability”?

HOW TO characterise, and utilize ,TS signals?



Enter the VARIOGRAM



Process Sampling Principles: (1-D lots)

Variography (variographics)

1-D lots: process data, stationary piles, ordered series

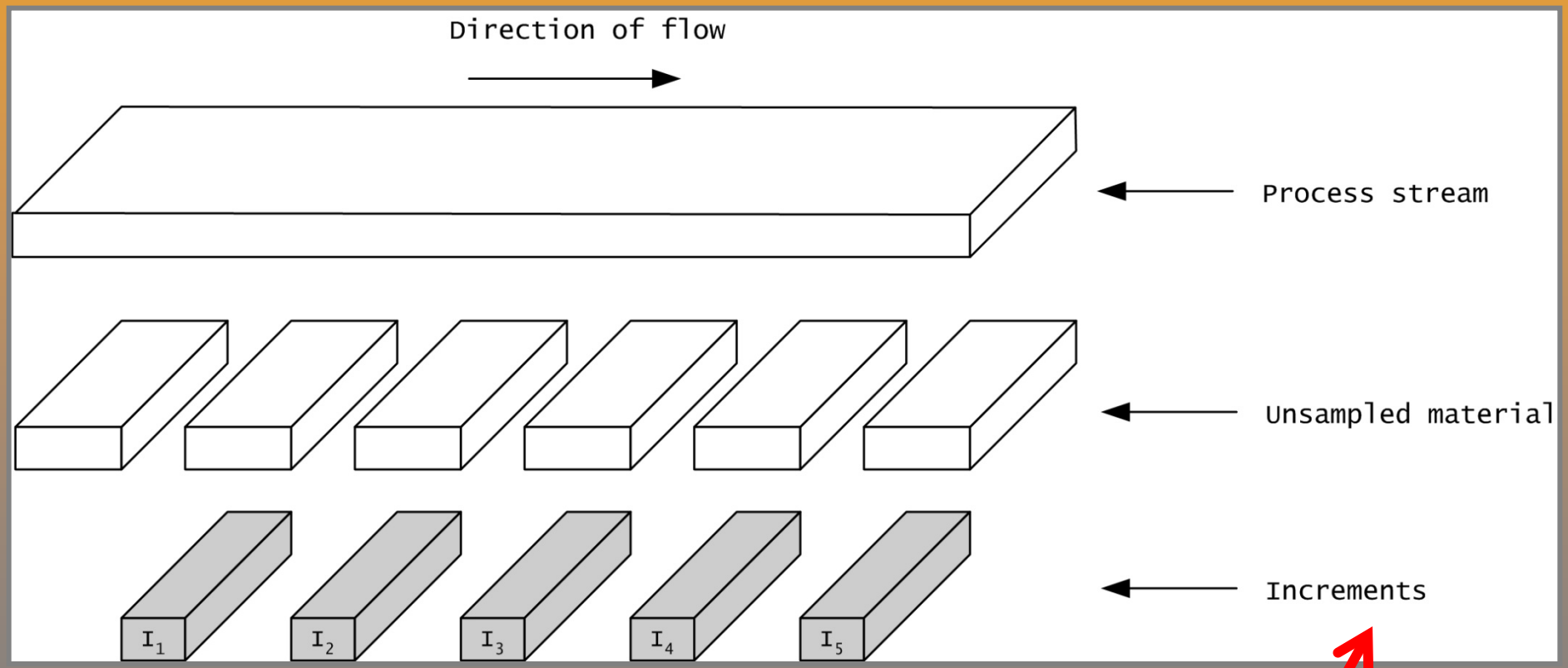
1-dimensional heterogeneity characterization

Valuable information about process *variation*

both major trends, upsets, periodic phenomena …

as well as small-scale variability

Some time, or other – the primary process sampling gets going .....



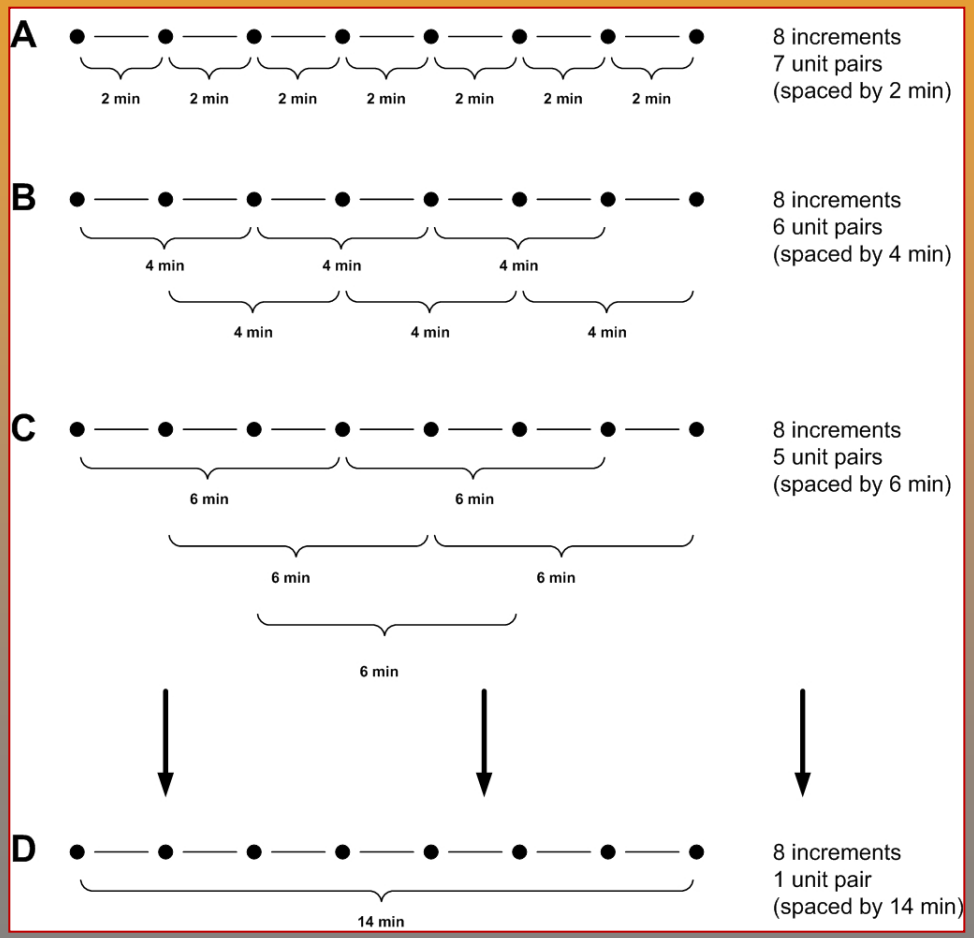
TOS terminology (Theory of Sampling)

Process Sampling Principles: (1-D case)

Variography basics:

"Lag" is the distance between samples *along* the time or the spatial dimension …

A "variogram" displays the total "variation" as a function of the "lag"



Lag = 1

Lag = 2

Lag = 3

Lag = 7

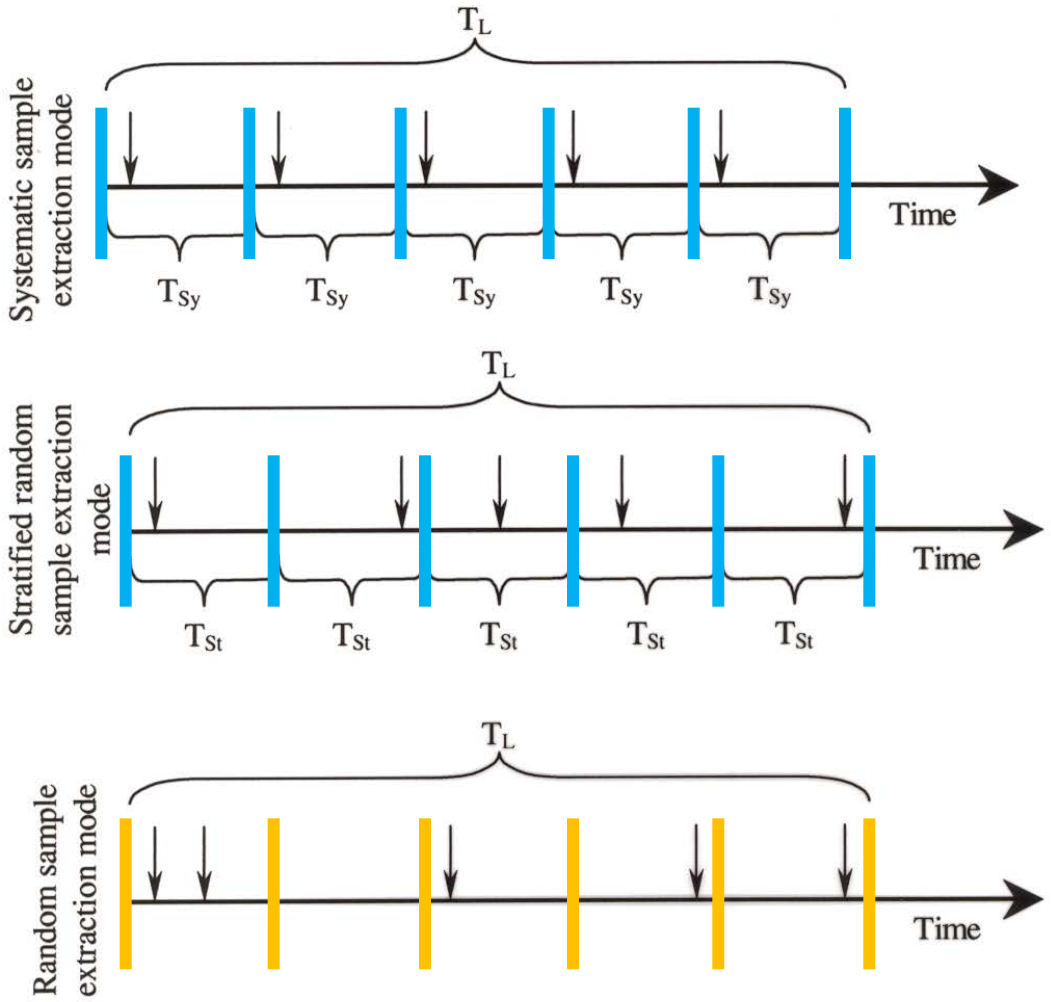
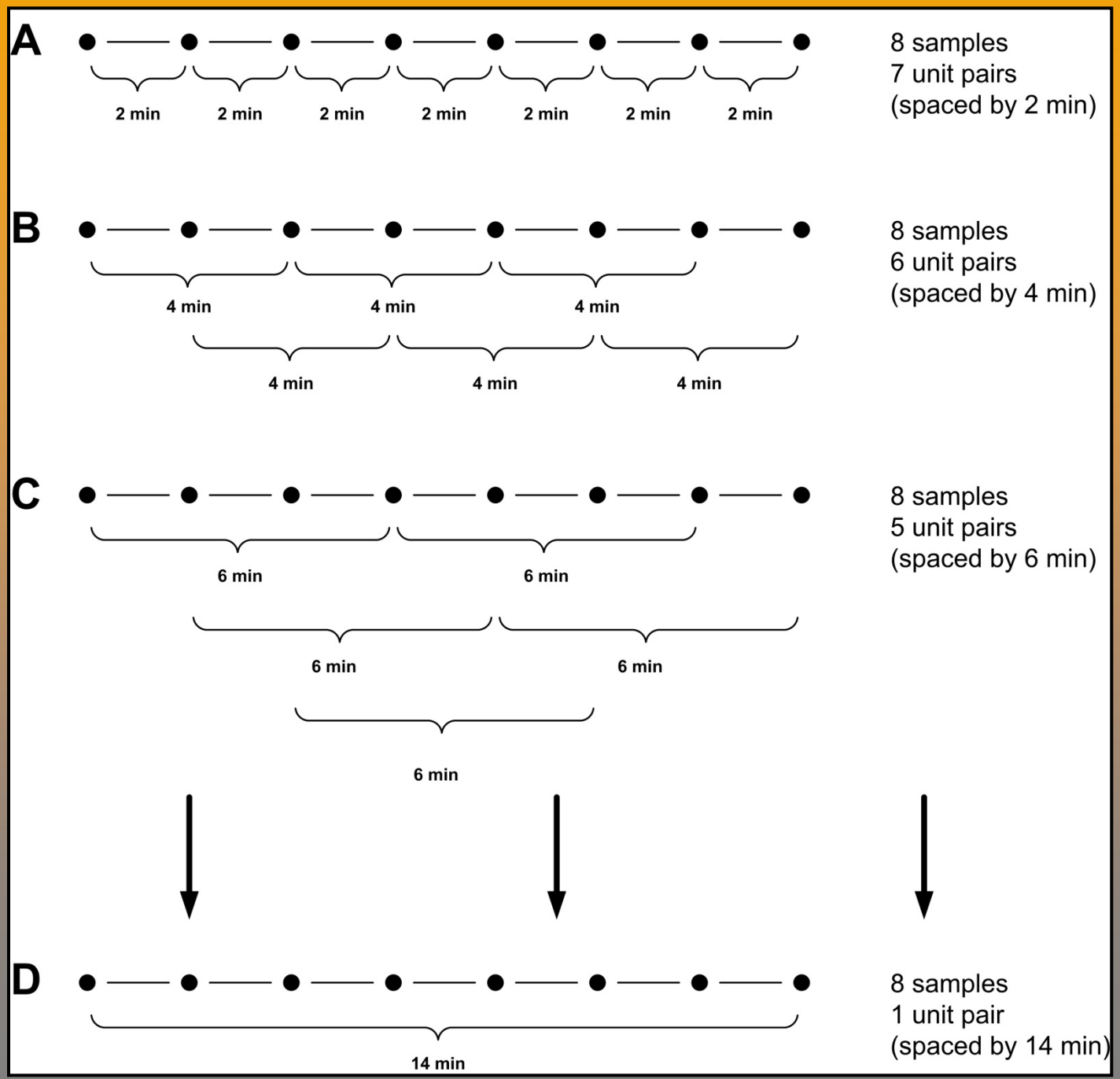


Illustration of the different sample extraction modes. The vertically arrows (↓) represent sample extractions. Notice that the number of sample extractions (Q) for systematic sample extraction mode is equal to T_L/T_{Sy}. The number of sample extractions (Q) for the stratified random sample extraction mode is equal to = T_L/T_{St}.



$$j = \frac{\theta}{\theta_{min}}$$

N_U = total number of data points

$$V(j) = \frac{1}{2(N_U - j)} \sum_m (h_{m+j} - h_m)^2$$

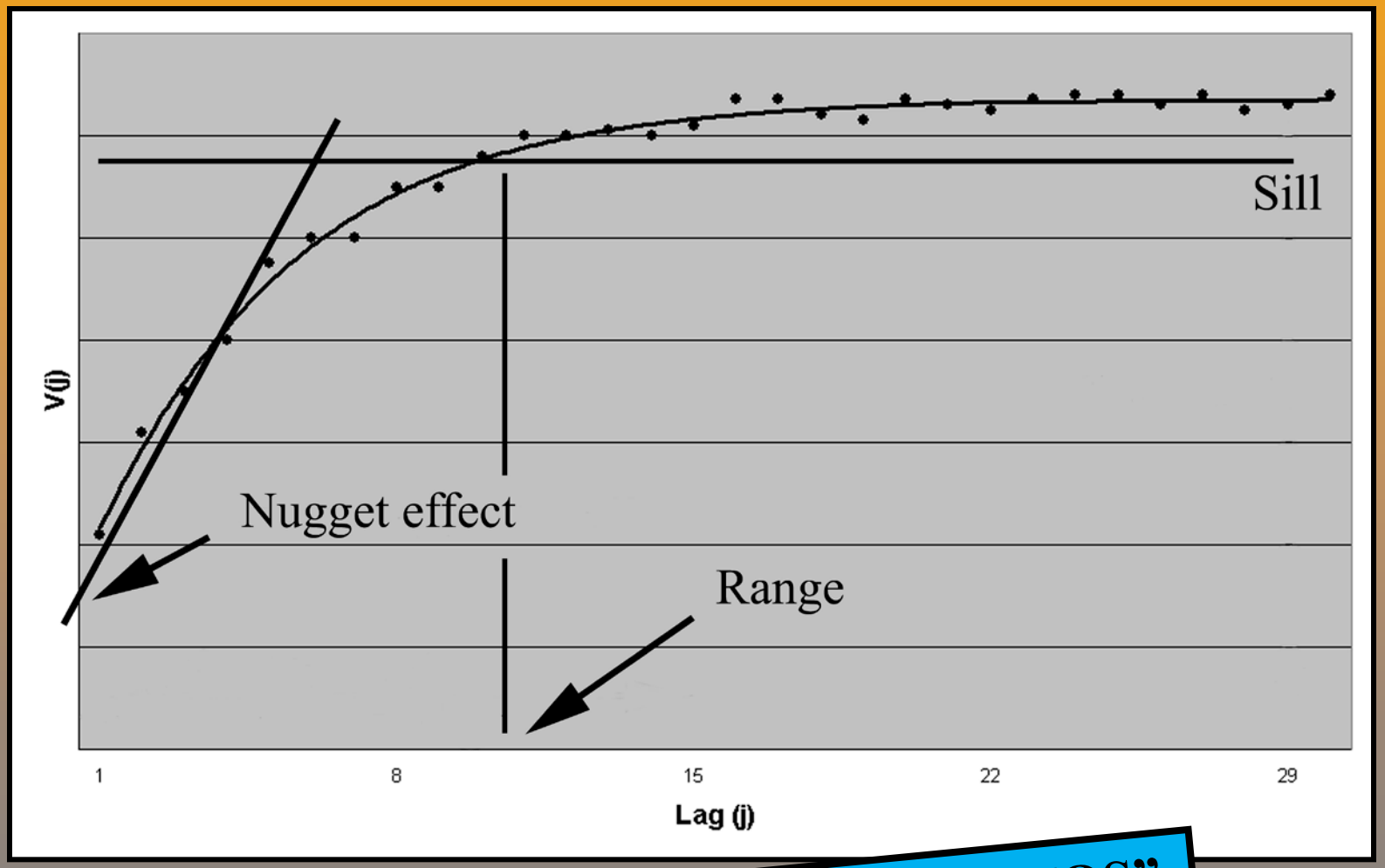
$$V(j) = \frac{1}{2(N_U - j)a_L^2} \sum_m [a_{m+j} - a_m]^2$$

V(j) = Variogram function [relative (h_m) or absolute (a_m)

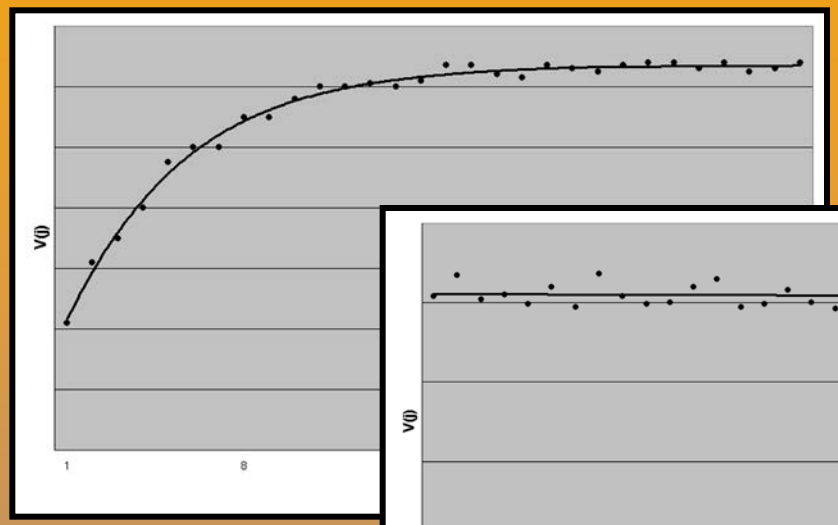
h_m = Sample heterogeneity contribution (mass prop. a_n)

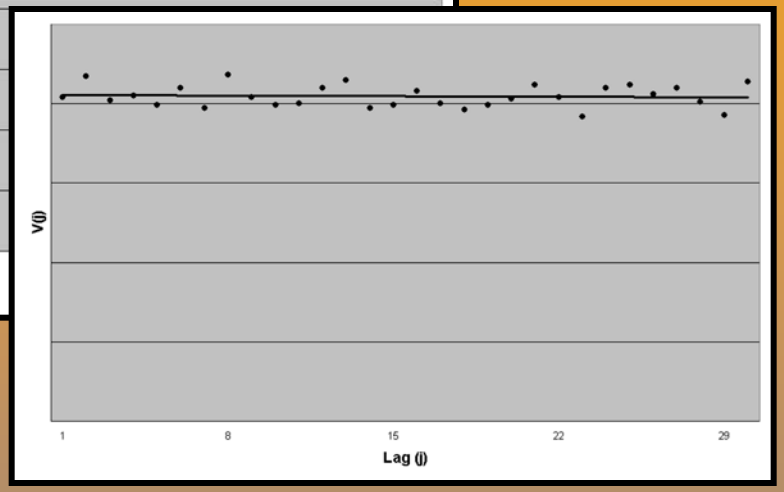
$$h_n = \frac{a_n - a_L}{a_L} \times \frac{M_n}{M_n}$$

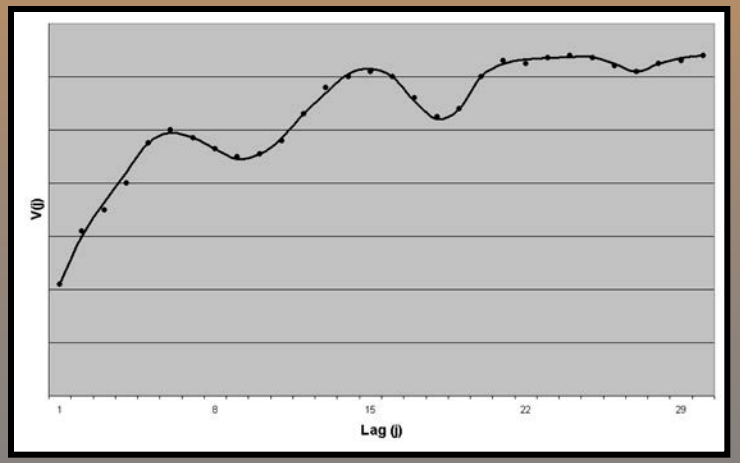
a_n: sample concentration
a_L: lot grade (process average)
M_n: sample mass

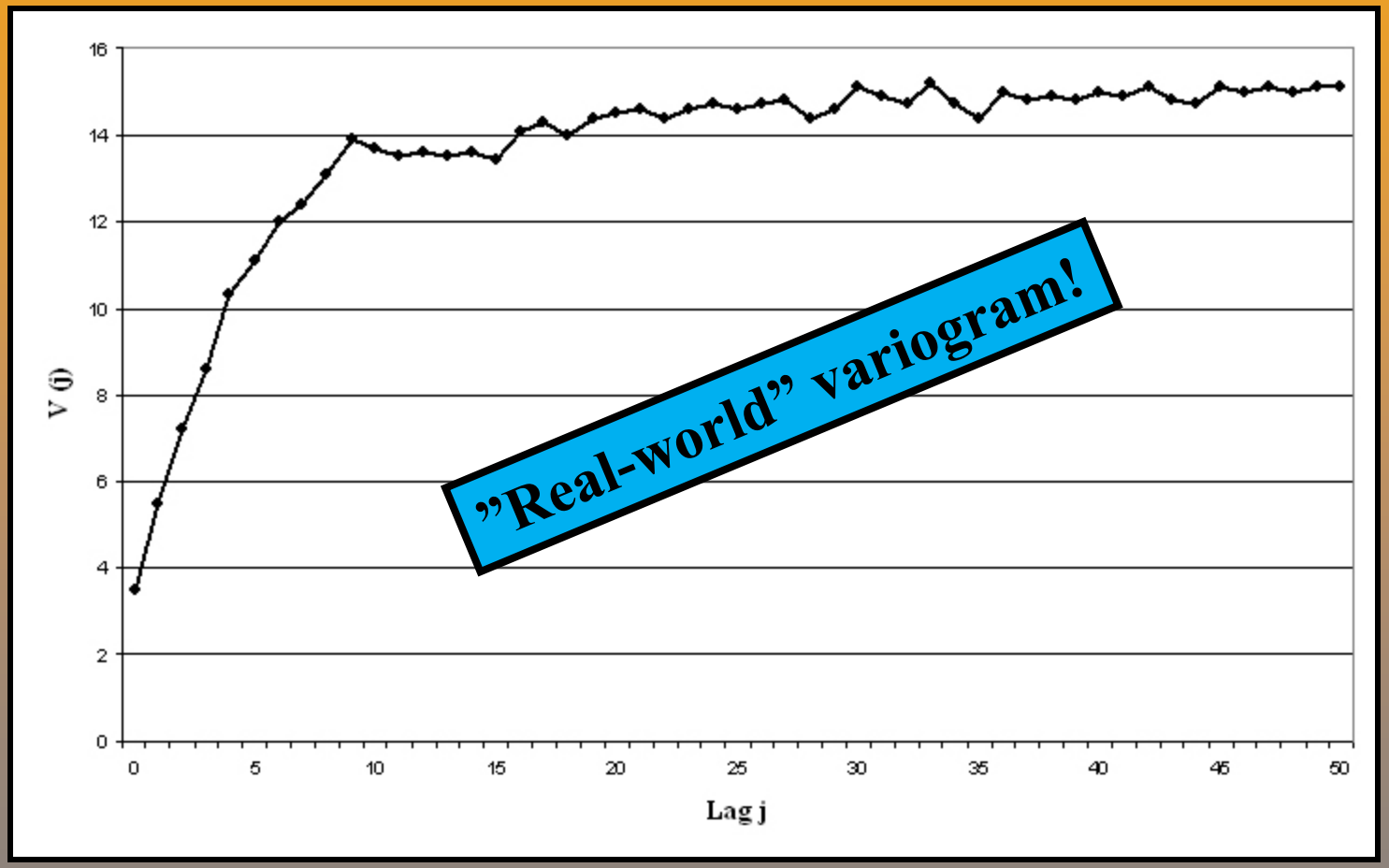


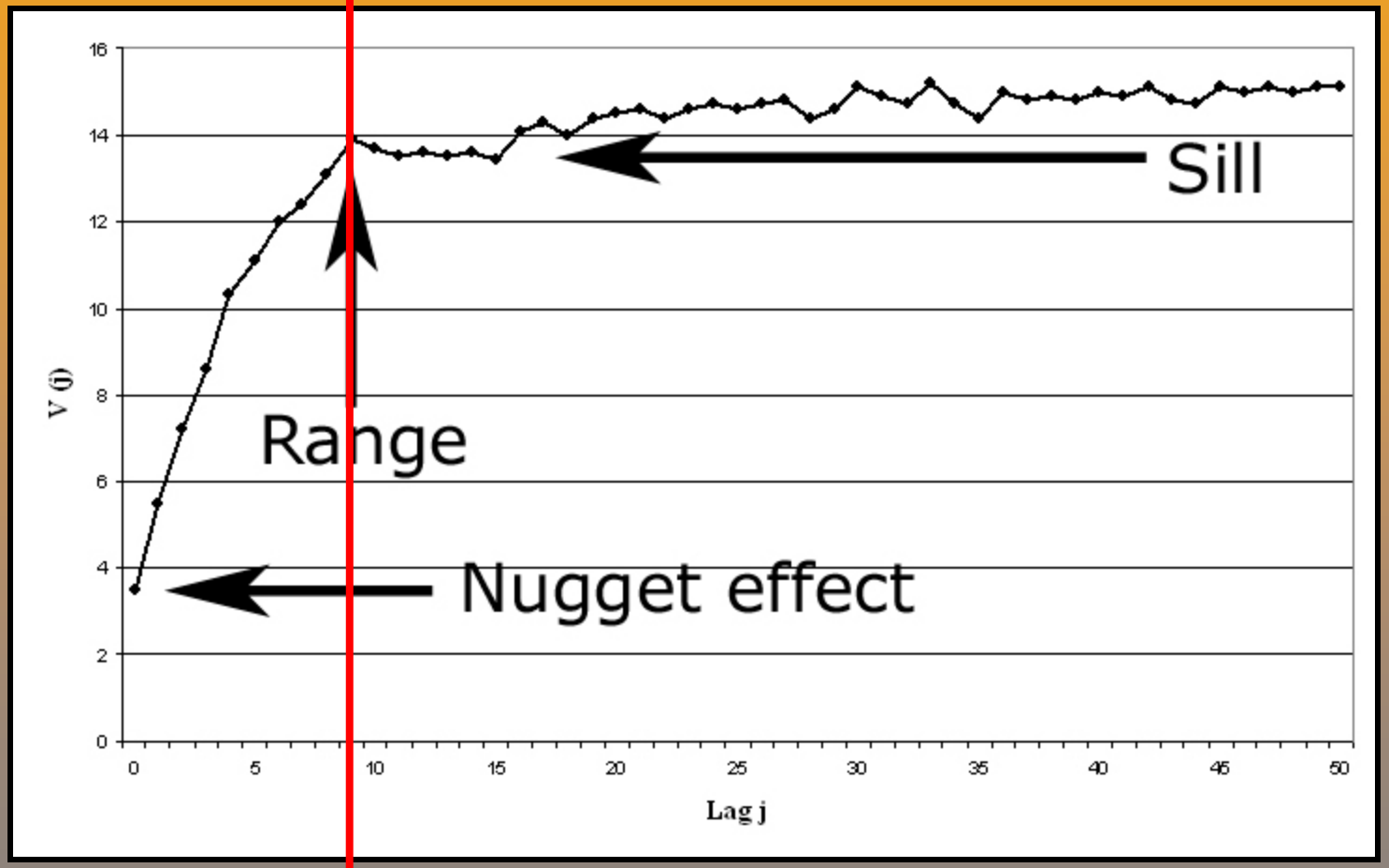
"Geostatistics *vs.* Process TOS"

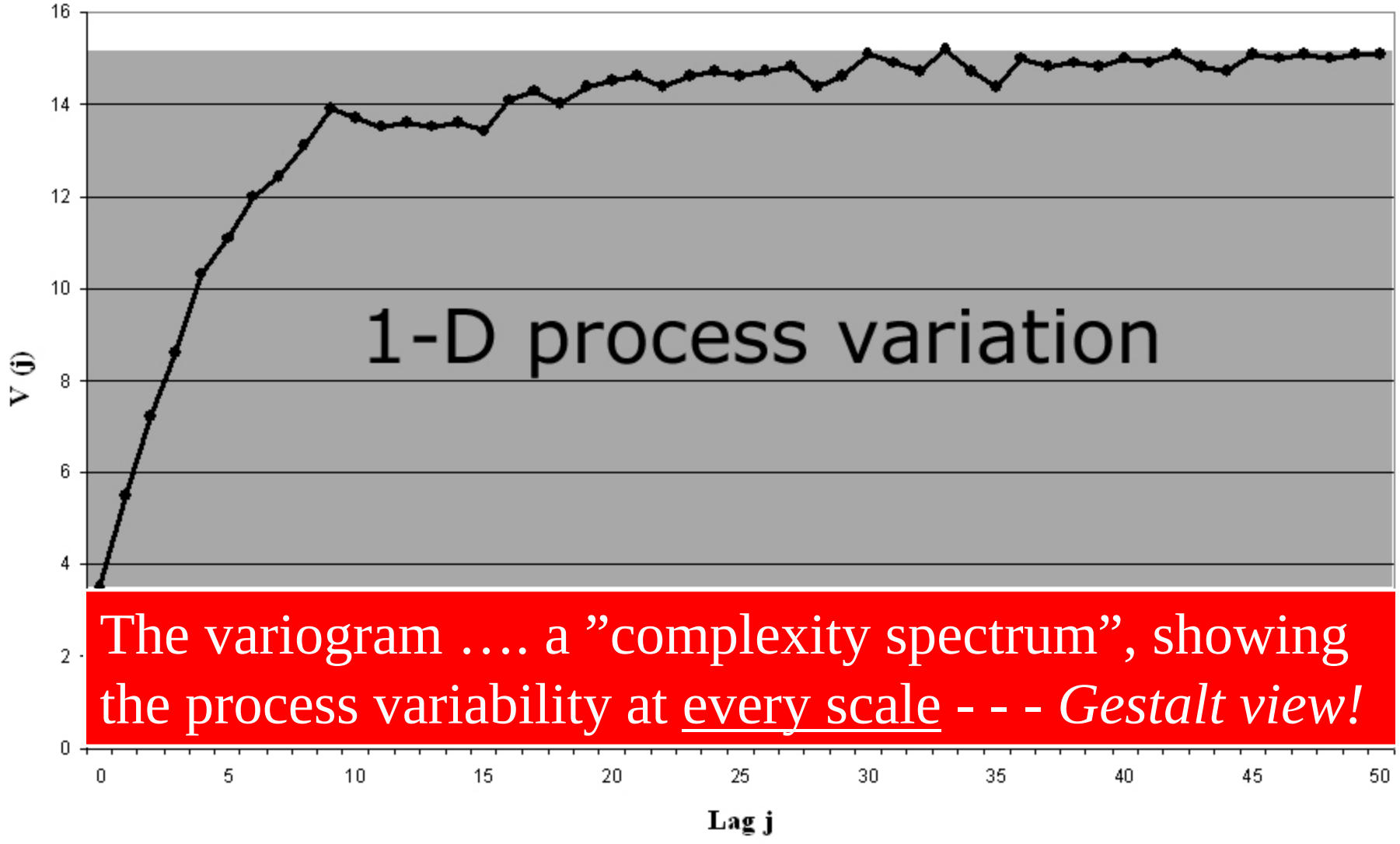

V(j)
1
8


V(j)
1
8
15
22
29
Lag (j)

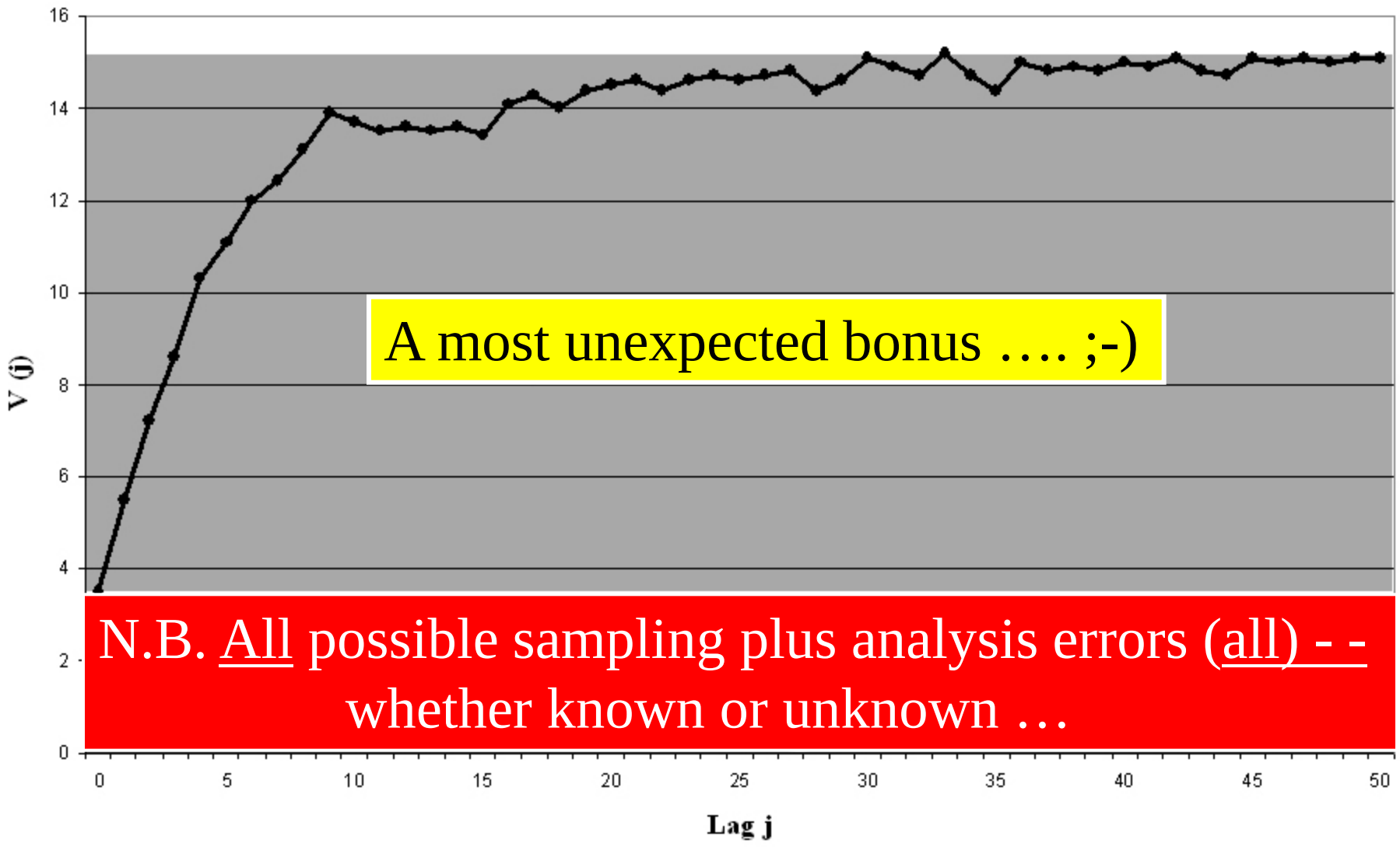

V(j)
1
8
15
22
29
Lag (j)

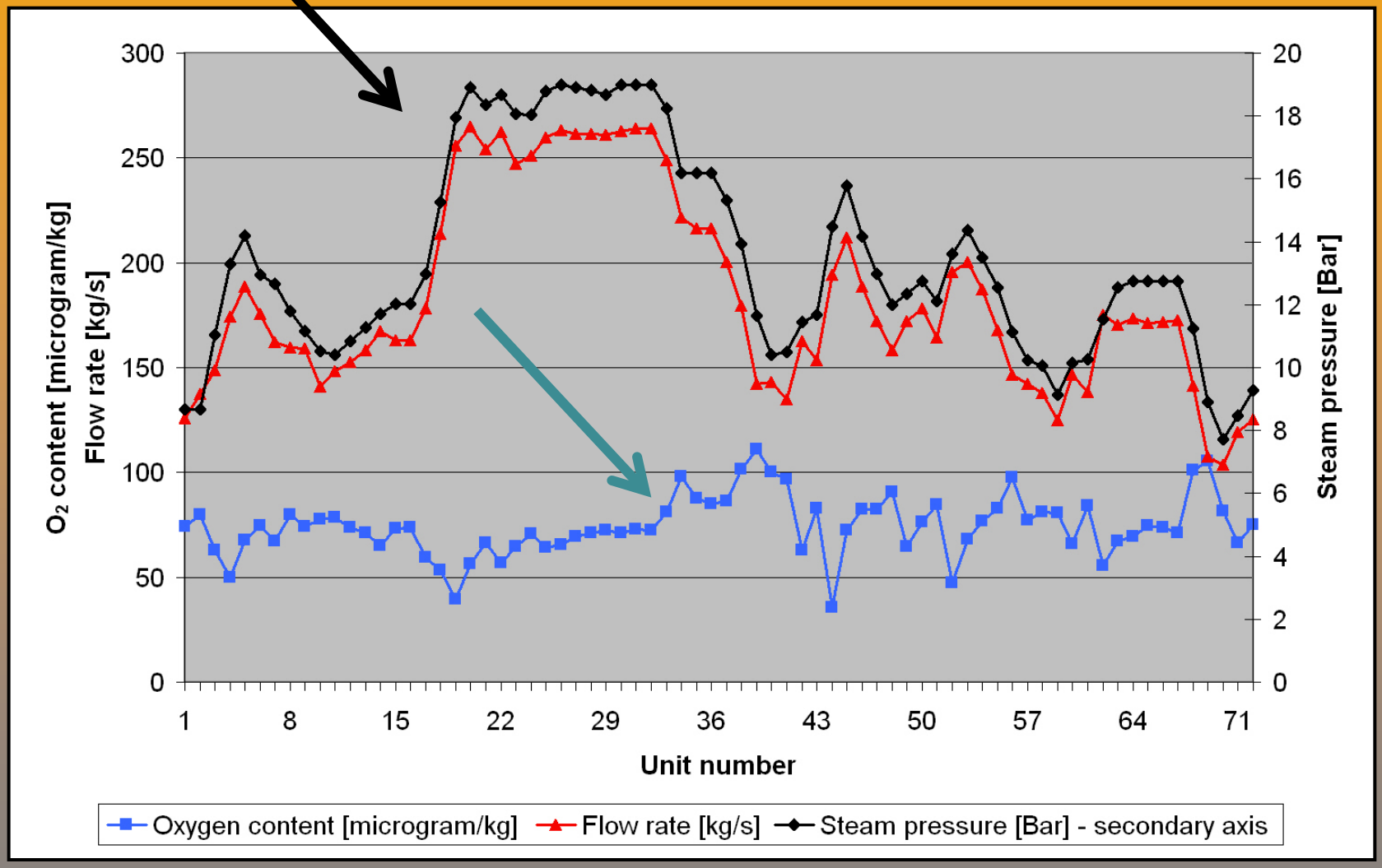

16
14
12
10
8
6
4
2
0
V (j)
0
5
10
15
20
25
30
35
40
45
50
Lag j
”Real-world” variogram!


Sill
Range
Nugget effect
V (j)
Lag j
0
2
4
6
8
10
12
14
16
0
5
10
15
20
25
30
35
40
45
50

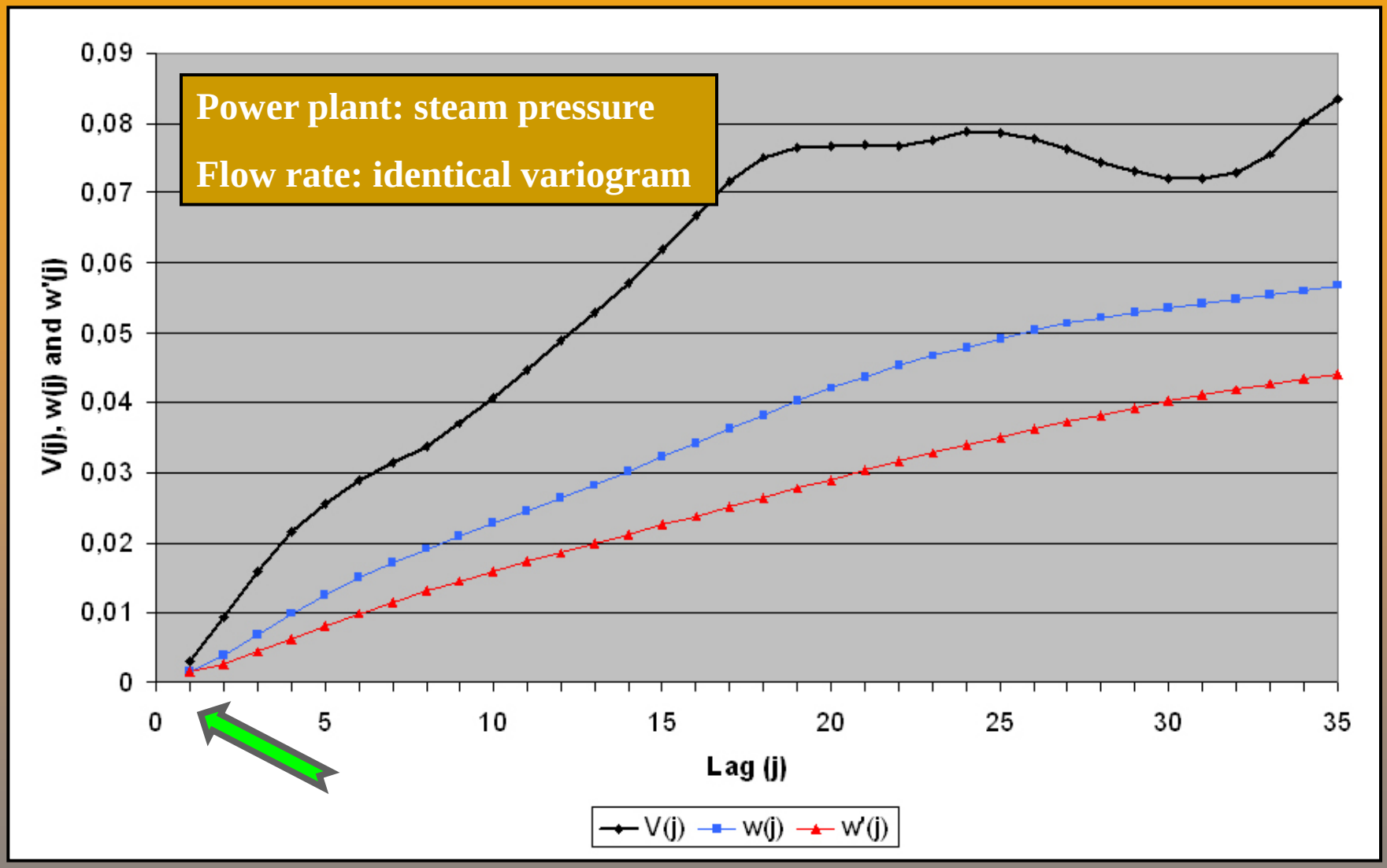


The variogram …. a ”complexity spectrum”, showing the process variability at every scale - - - *Gestalt view!*

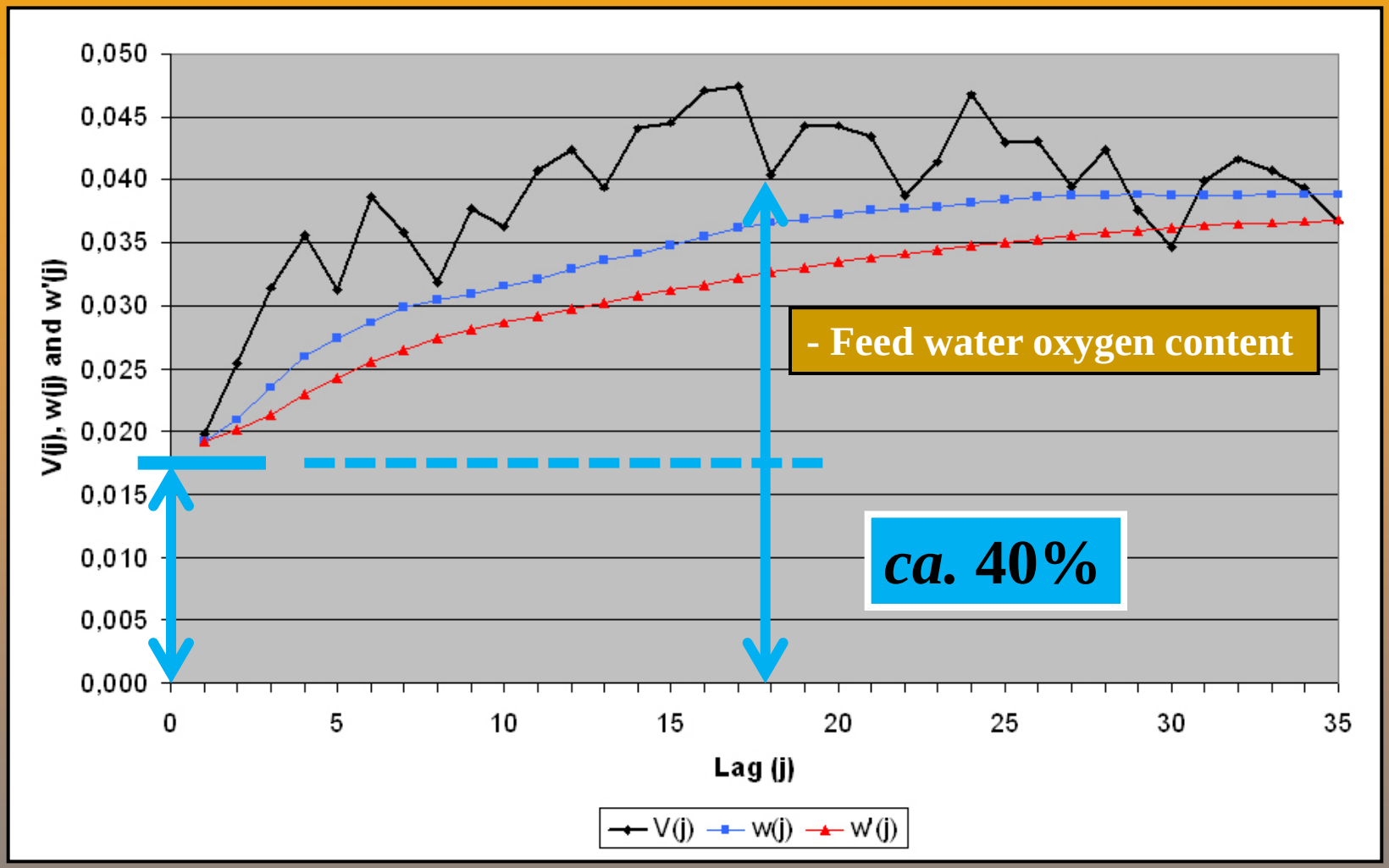

16
14
12
10
8
6
4
2
0
V (j)
0
5
10
15
20
25
30
35
40
45
50
Lag j
A most unexpected bonus …. ;-)
N.B. All possible sampling plus analysis errors (all) - -
whether known or unknown …



- Power plant example: Denmark

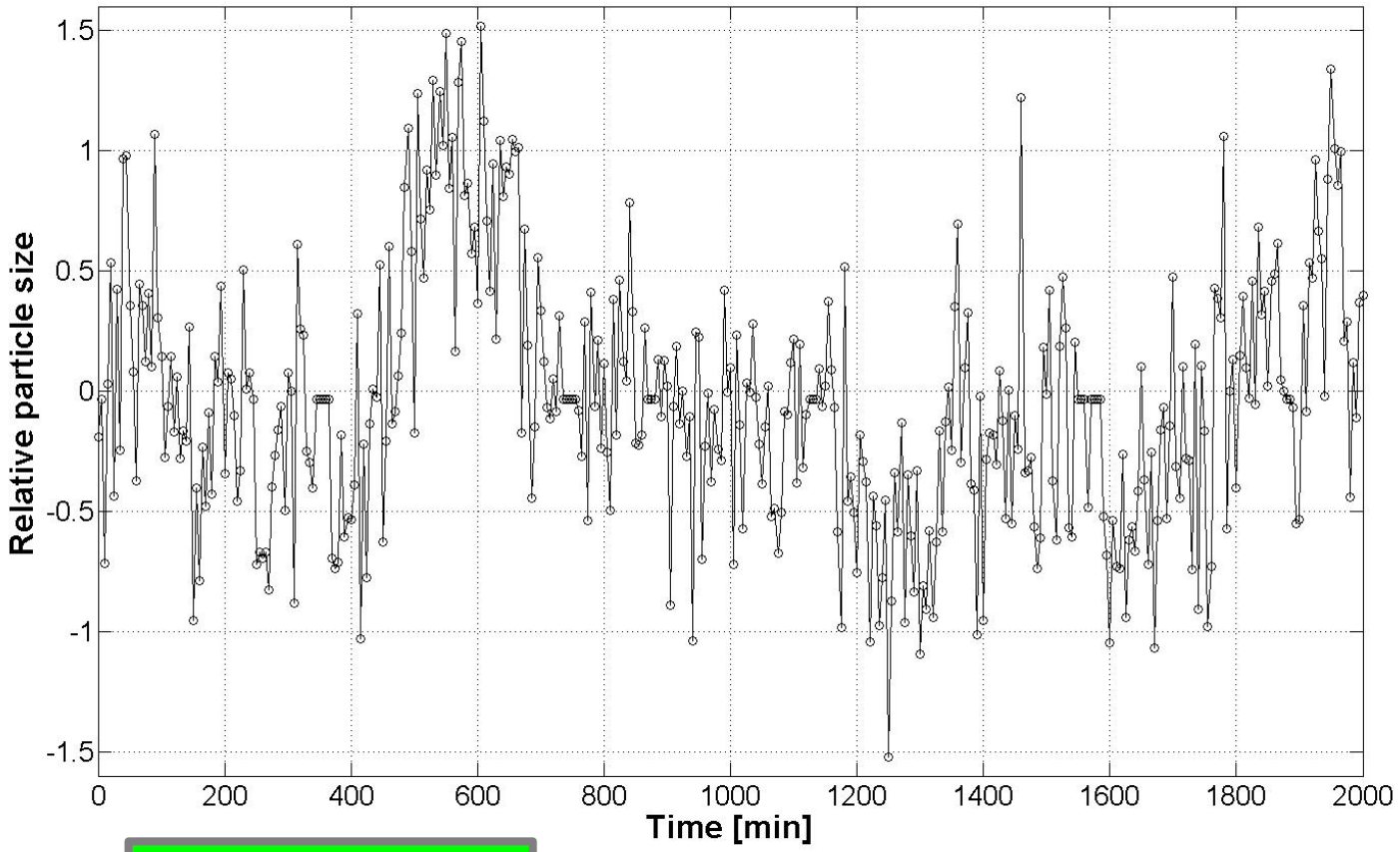


- Power plant example: Denmark



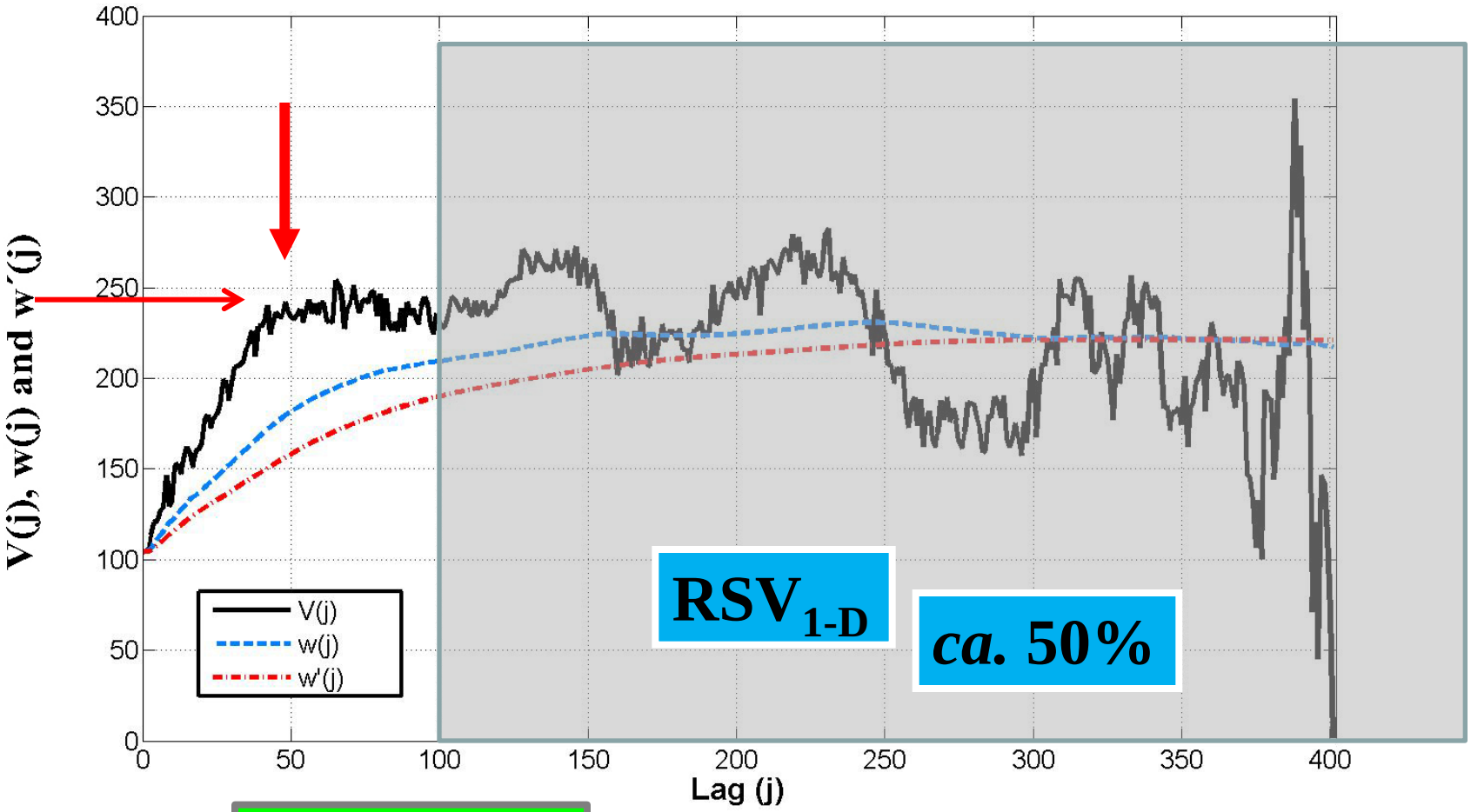
- Power plant example: Denmark

~Peter Paasch-Mortensen



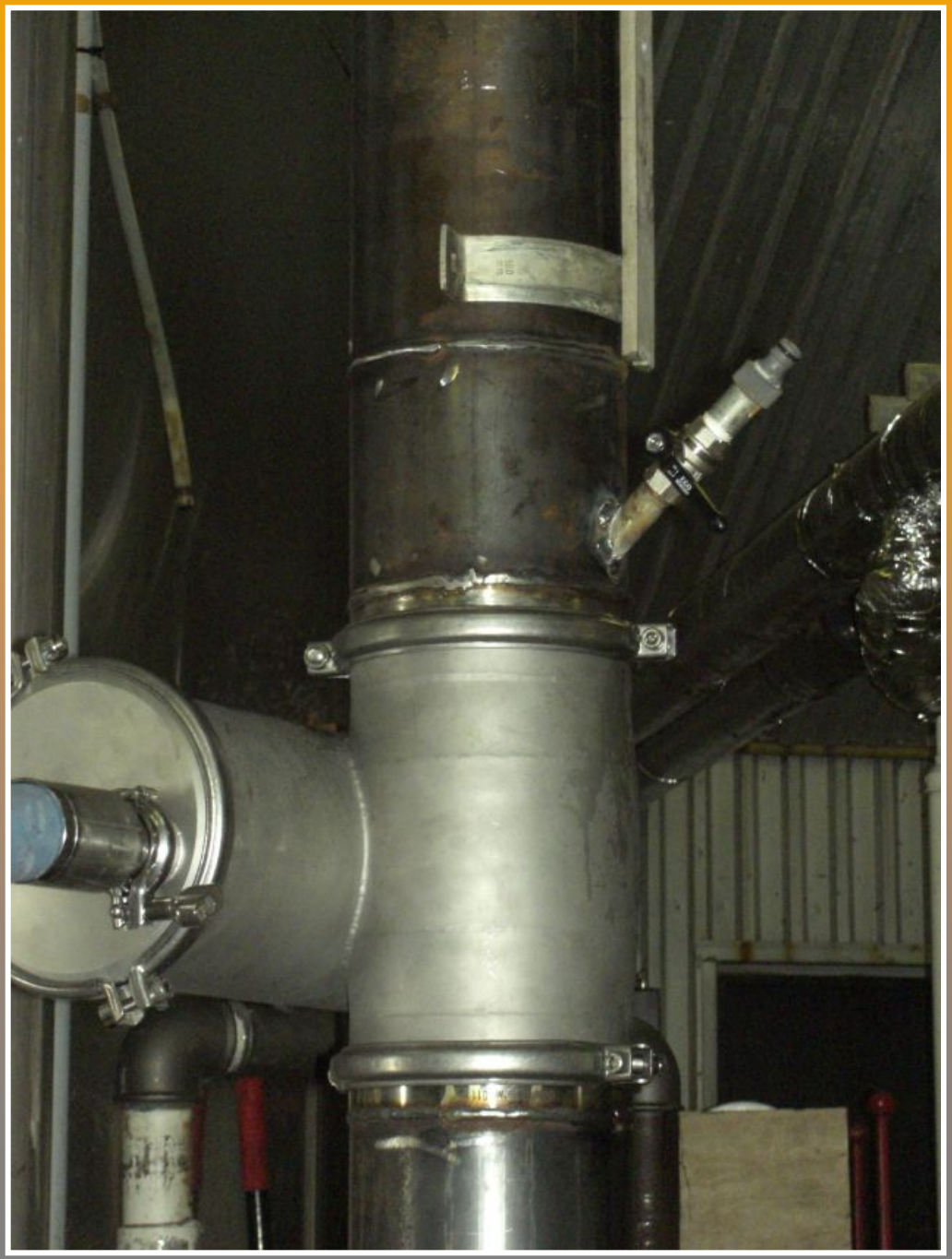
Sampling rate, **r**

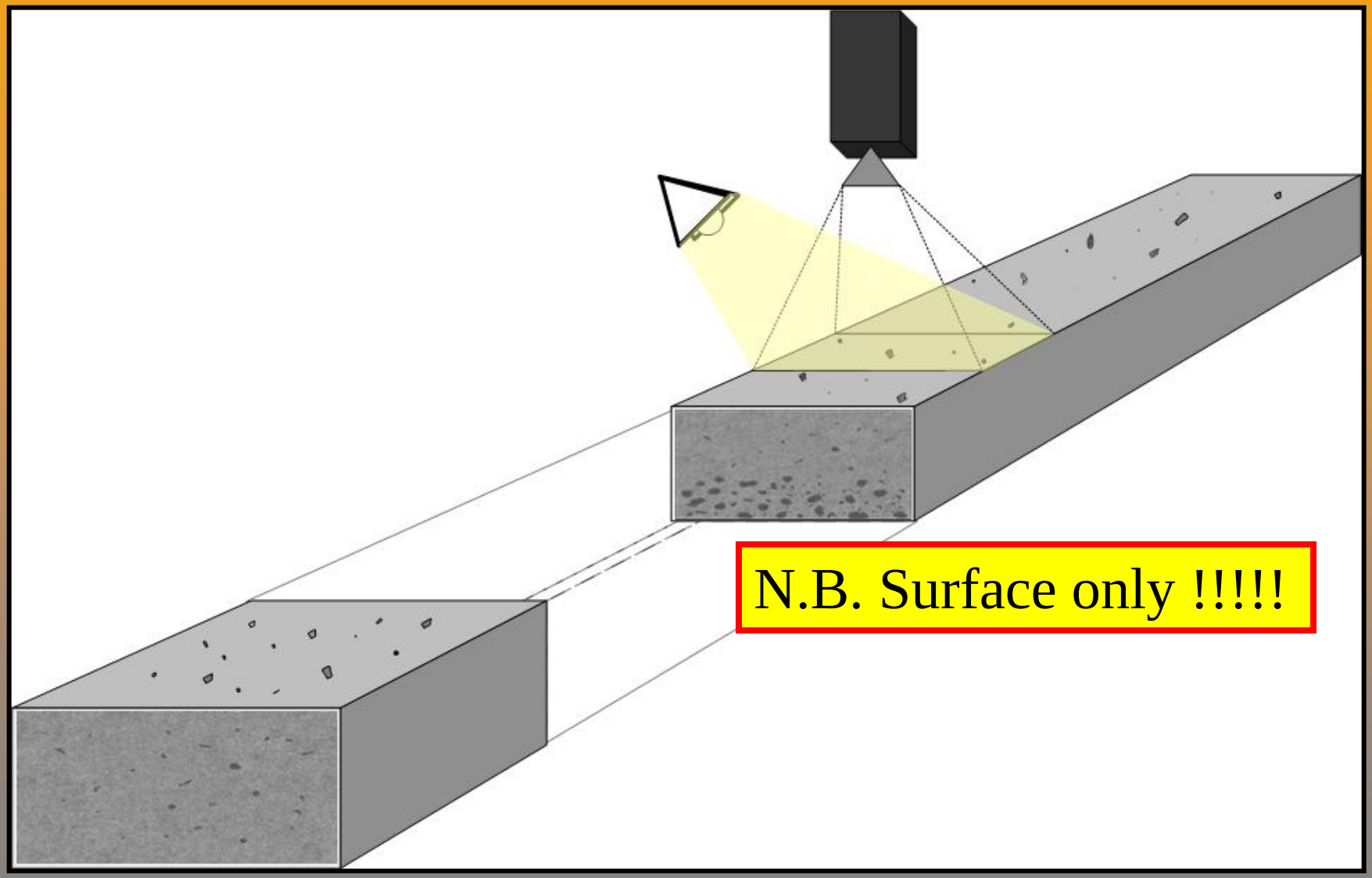
Particle size predictions (estimates) from acoustic chemometrics on-line measurements in a granulated product.
Approximately 2000 measurements were performed every 5 minutes for 34 hours.
All measurements pertain to the same formulation production campaign.



Variogram and auxiliary functions for on-line measurement of particle sizes.
The variogram has a range of j = 50, signifying that units spaced by more than 250 minutes are no longer correlated with each other.

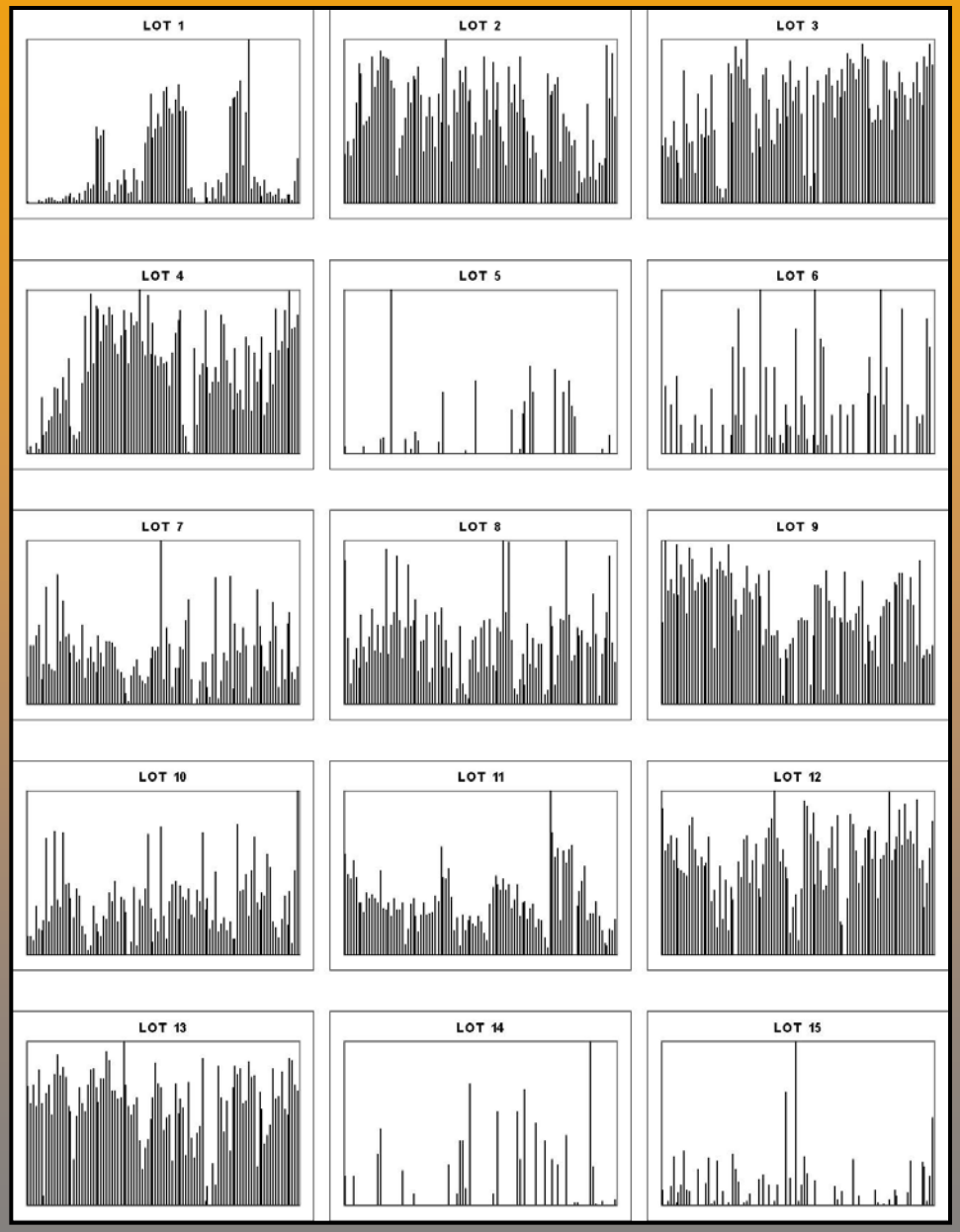
Full-scale biogas plant trials, 2007 - 2008
NIR on-line PAT monitoring (fermenter 3): 2400 m^3

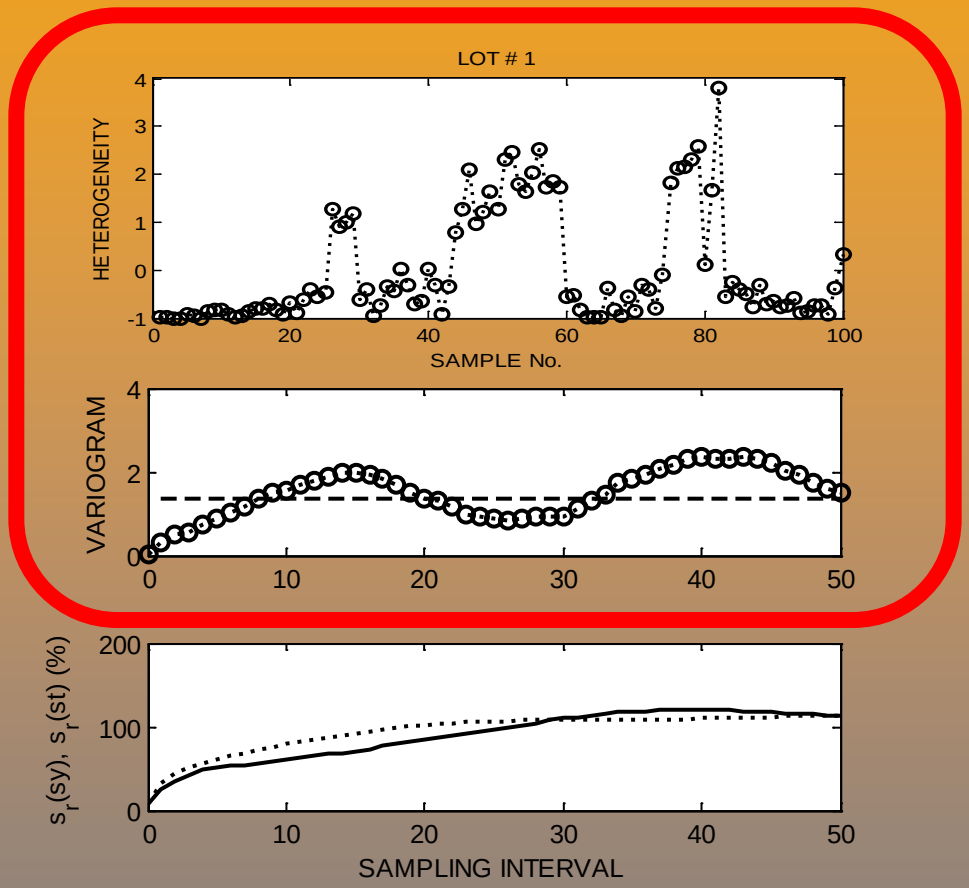


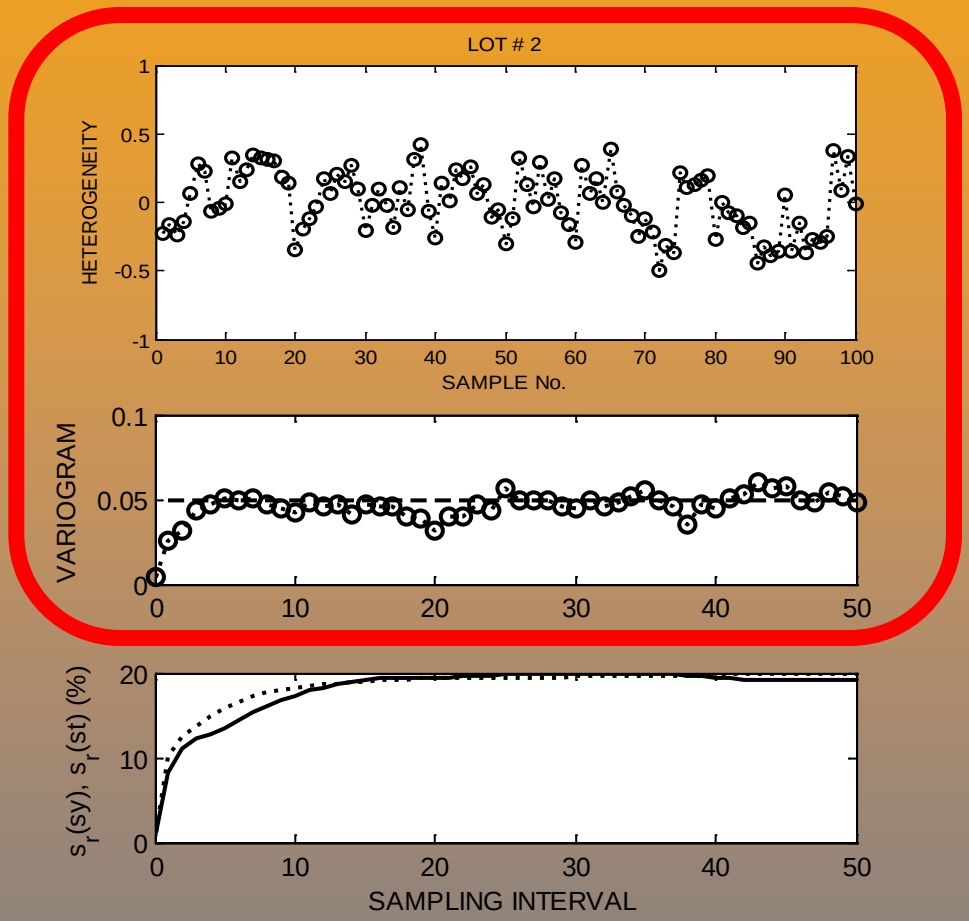

N.B. Surface only !!!!!

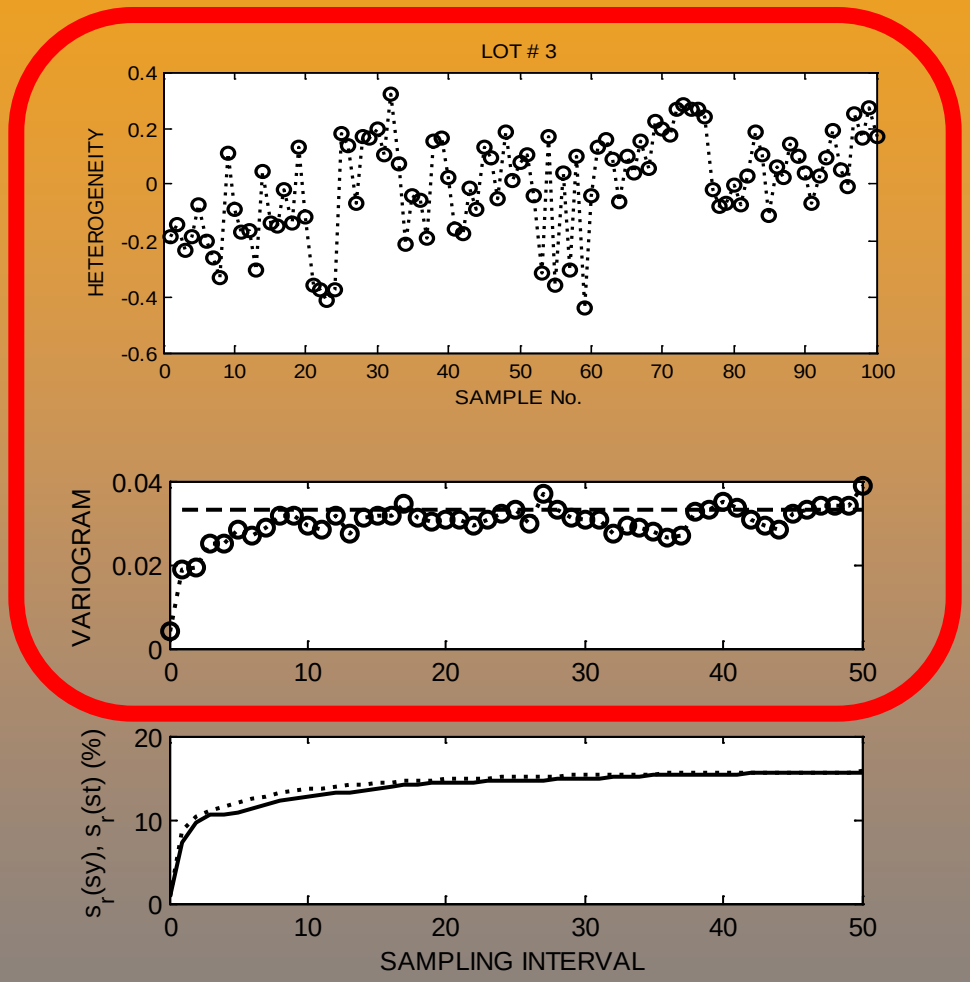
15 industrial process lots: 3D, 1D

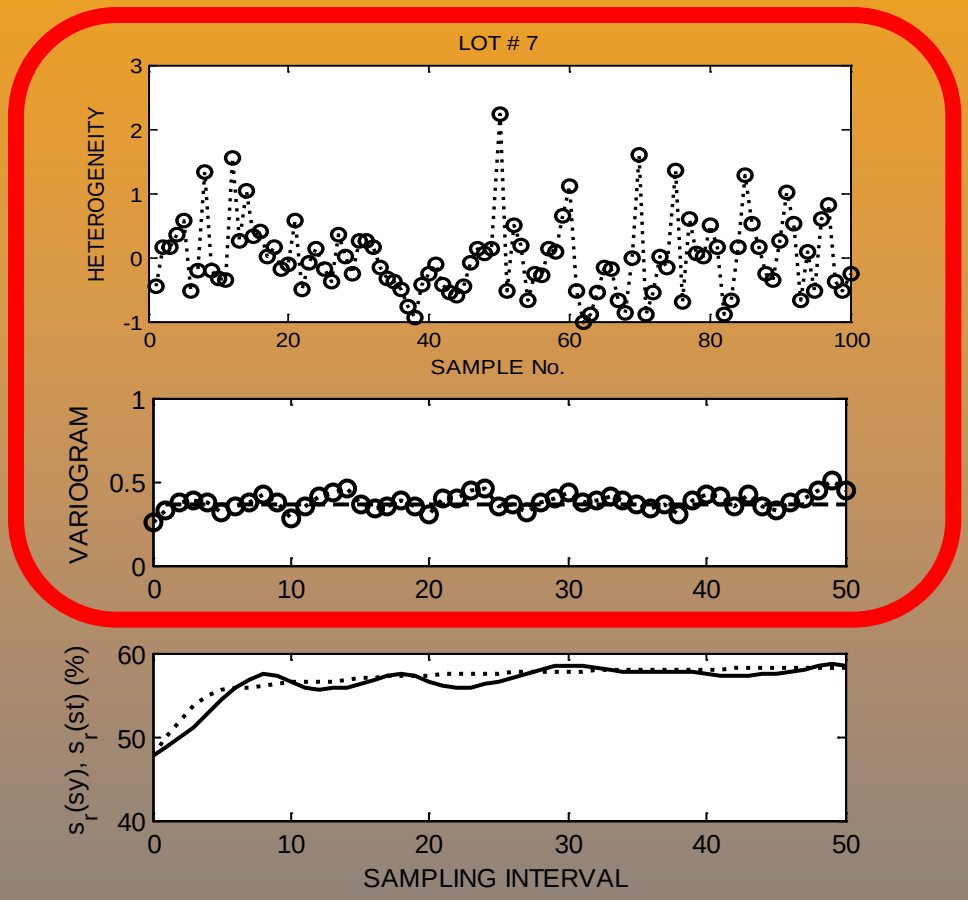
15 ship cargos being off-loaded




LOT # 1
HETEROGENEITY
SAMPLE No.
VARIOGRAM
$s_r(sy)$, $s_r(st)$ (%)
SAMPLING INTERVAL

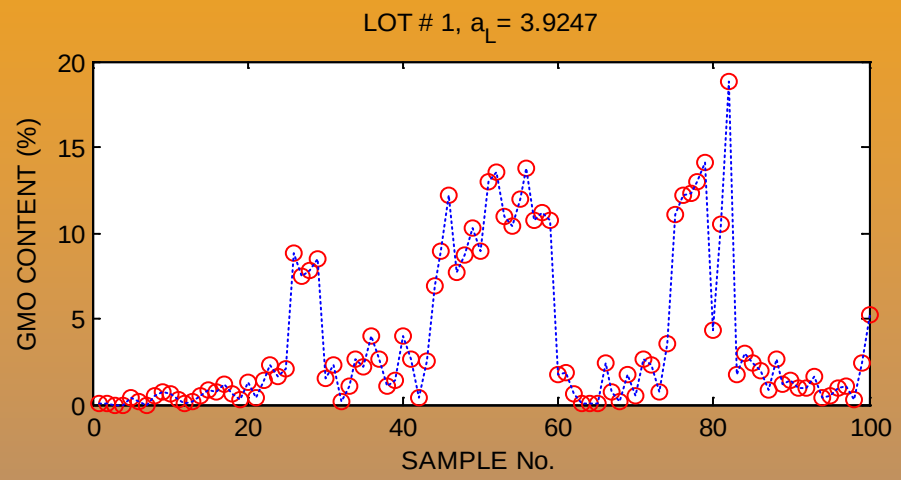

LOT # 2
HETEROGENEITY
1
0.5
0
-0.5
-1
0
10
20
30
40
50
60
70
80
90
100
SAMPLE No.
VARIOGRAM
0.1
0.05
0
0
10
20
30
40
50
s_r(sy), s_r(st) (%)
20
10
0
0
10
20
30
40
50
SAMPLING INTERVAL

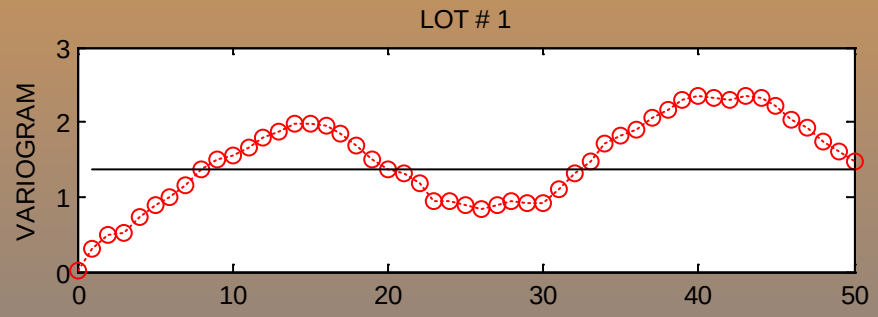

LOT # 3
HETEROGENEITY
0.4
0.2
0
-0.2
-0.4
-0.6
0
10
20
30
40
50
60
70
80
90
100
SAMPLE No.
VARIOGRAM
0.04
0.02
0
0
10
20
30
40
50
s_r(sy), s_r(st) (%)
20
10
0
0
10
20
30
40
50
SAMPLING INTERVAL

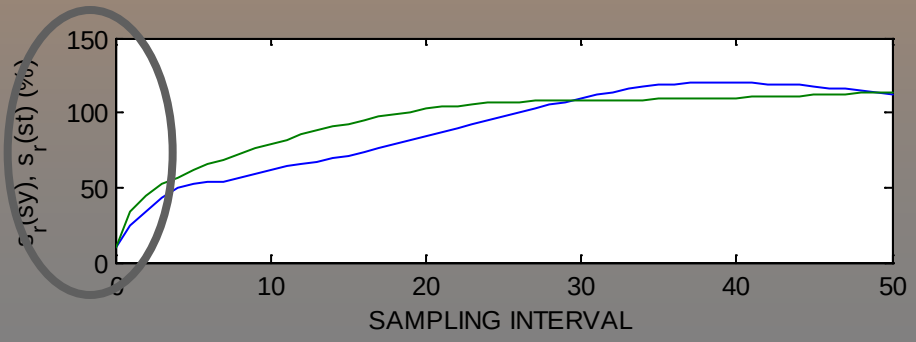

LOT # 7
HETEROGENEITY
3
2
1
0
-1
0
20
40
60
80
100
SAMPLE No.
VARIOGRAM
1
0.5
0
0
10
20
30
40
50
$s_r(sy)$, $s_r(st)$ (%)
60
50
40
0
10
20
30
40
50
SAMPLING INTERVAL

KeLDA variographic analysis

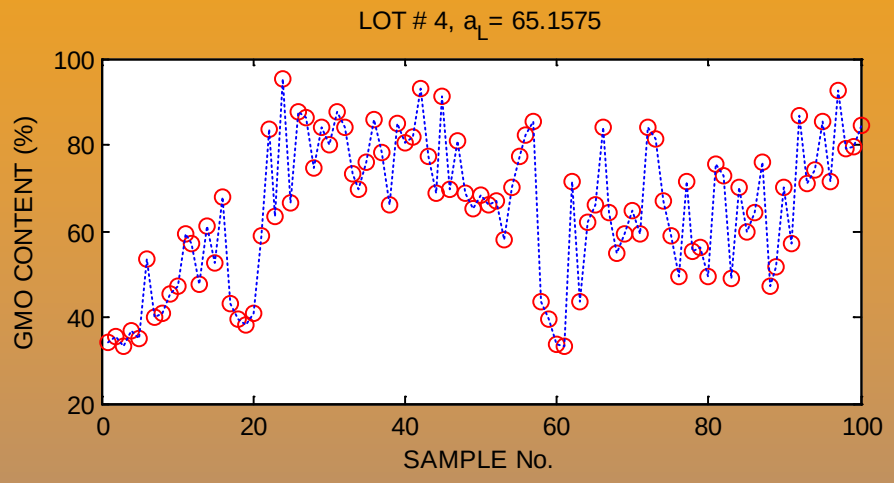
Extreme heterogeneity

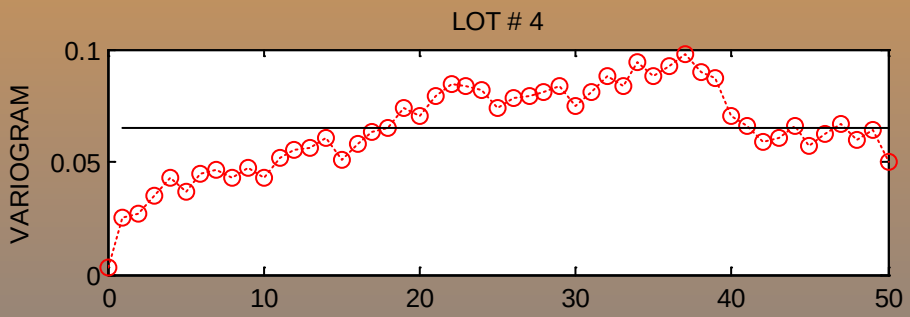


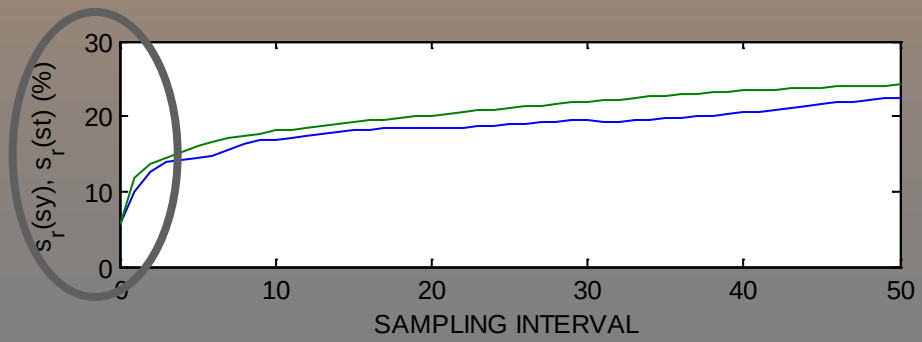




Intermediate heterogeneity

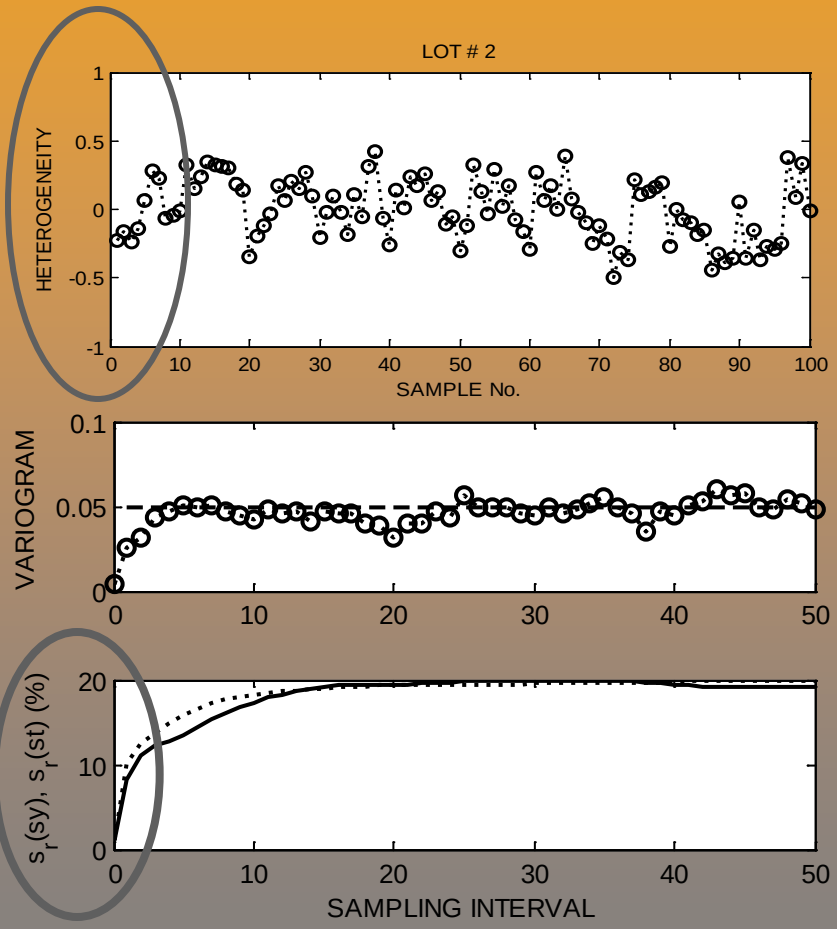




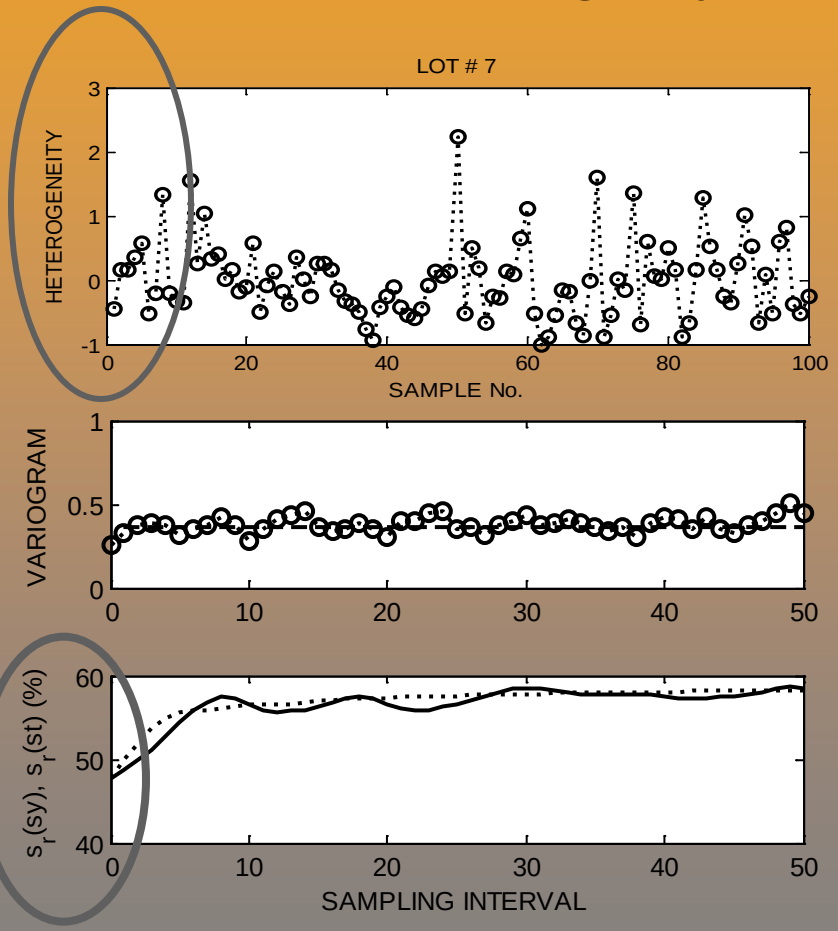


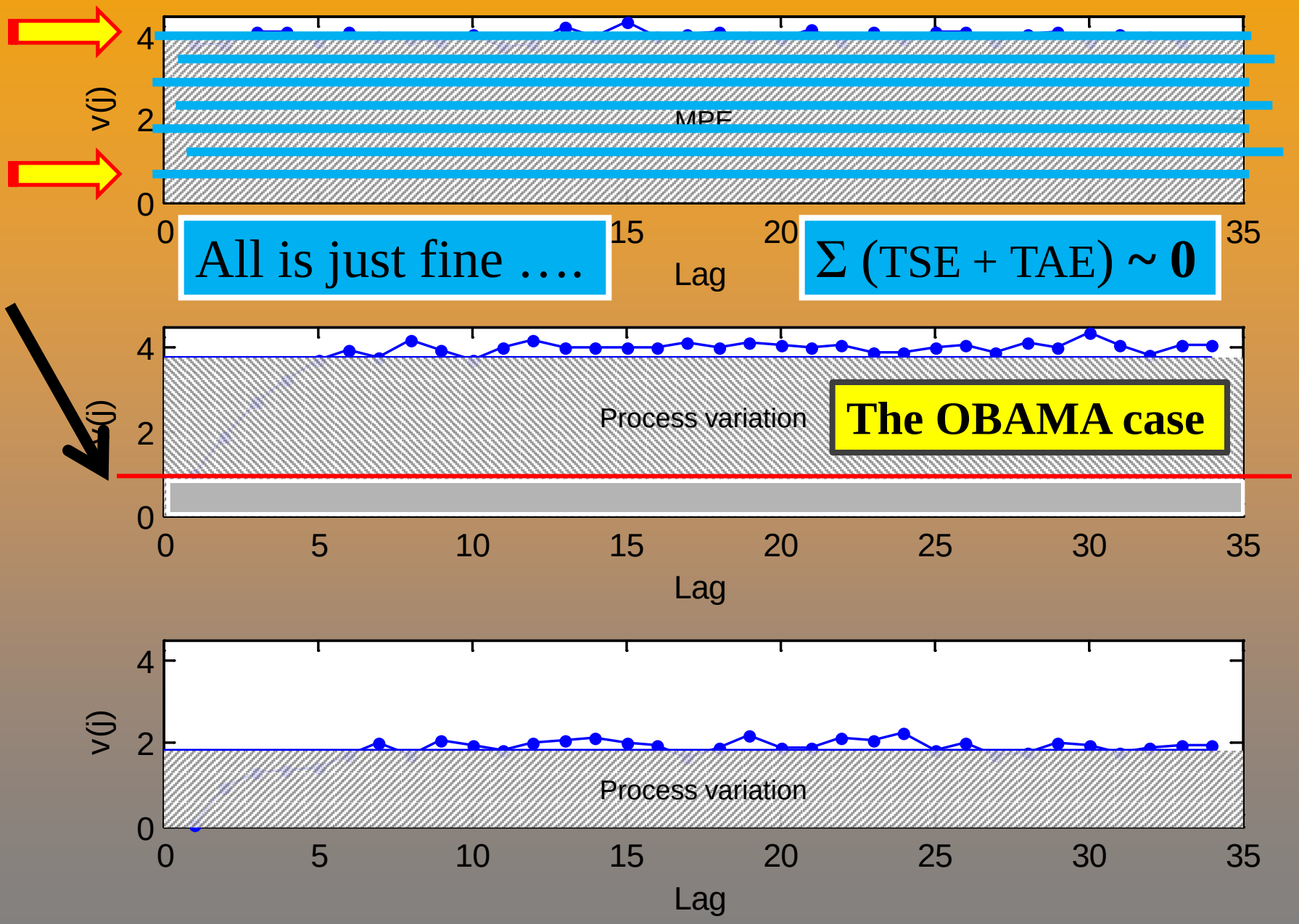
KeLDA variographic analysis

Rel low heterogeneity



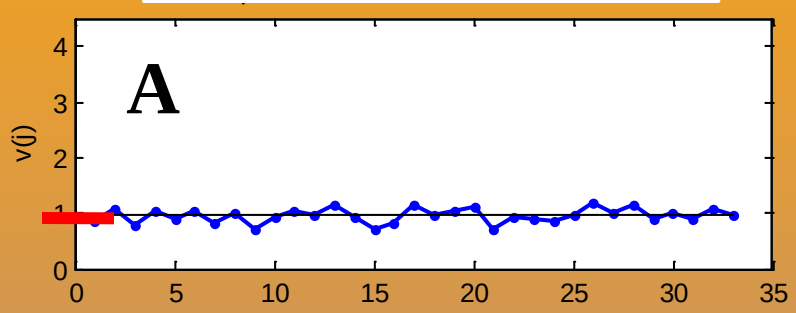
Intermediate heterogeneity



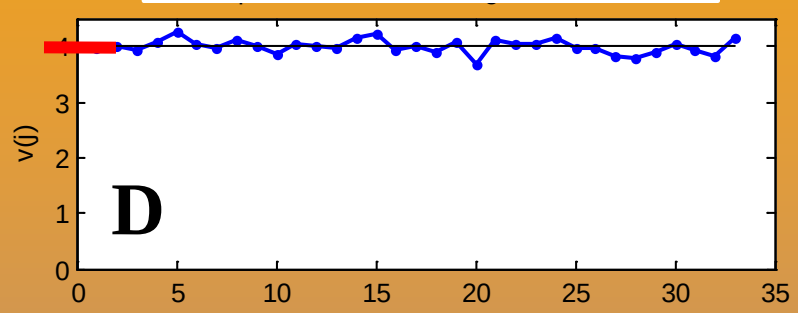

v(j)
4
2
0
MPE
0
15
20
35
Lag
All is just fine ….
Σ (TSE + TAE) ~ 0
v(j)
4
2
0
Process variation
The OBAMA case
0
5
10
15
20
25
30
35
Lag
v(j)
4
2
0
Process variation
0
5
10
15
20
25
30
35
Lag

Variogram: Unparalelled corporate QC / QA tool !!!

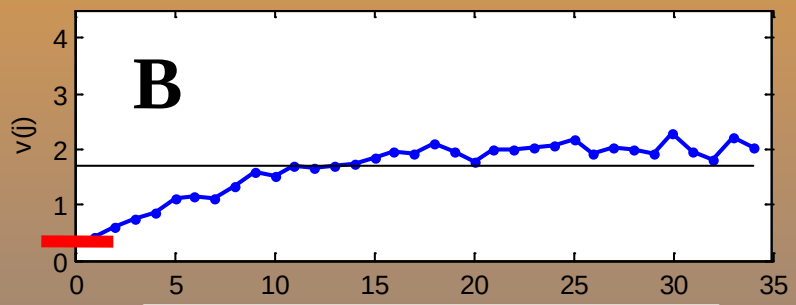
LOW SILL – Process stable & OK



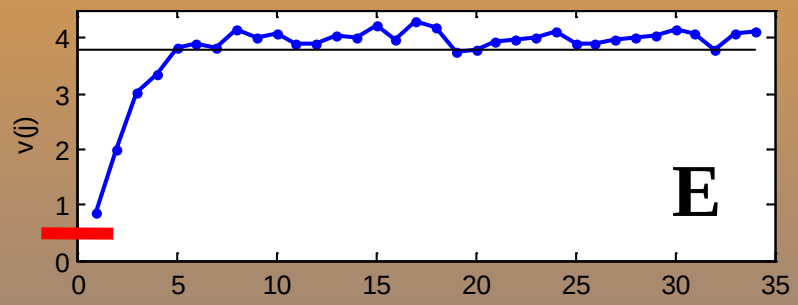
EXTREMELY HIGH SILL – ;-(;-(



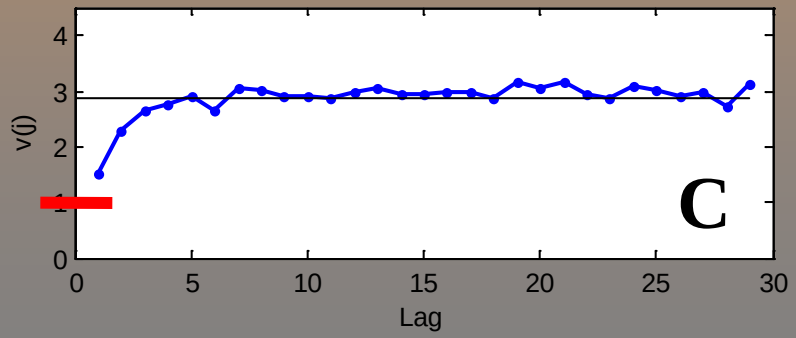
INTERM./LOW SILL – low n,e,



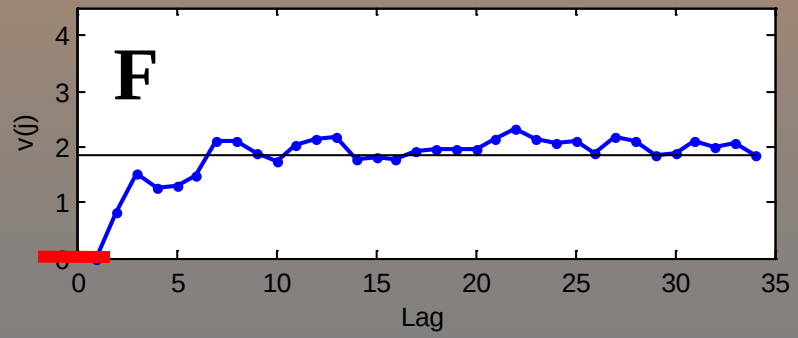
VERY HIGH SILL – very low n.e.



INTERMEDIATE SILL – low n,e,

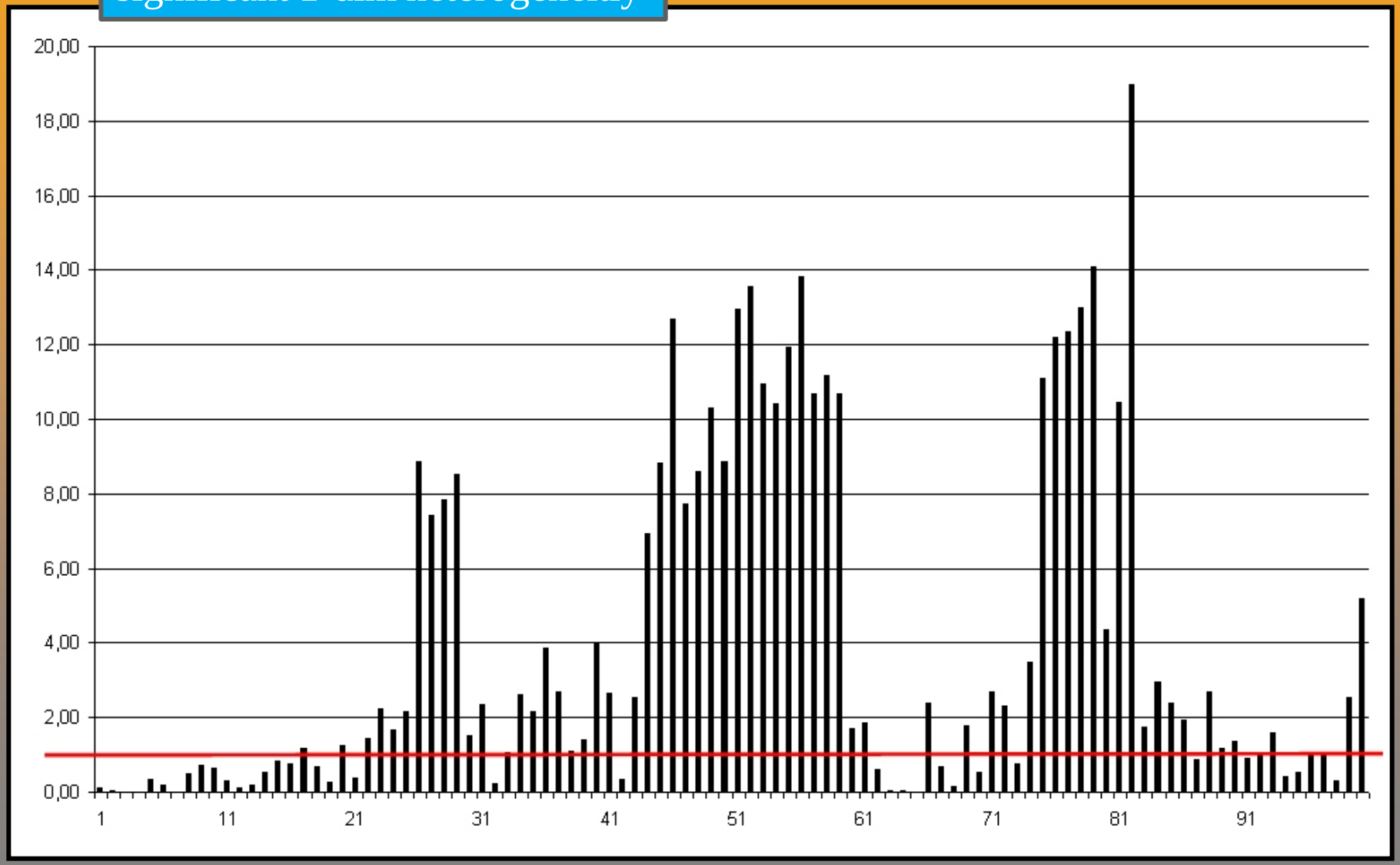


INTERM./LOW SILL – low n,e,

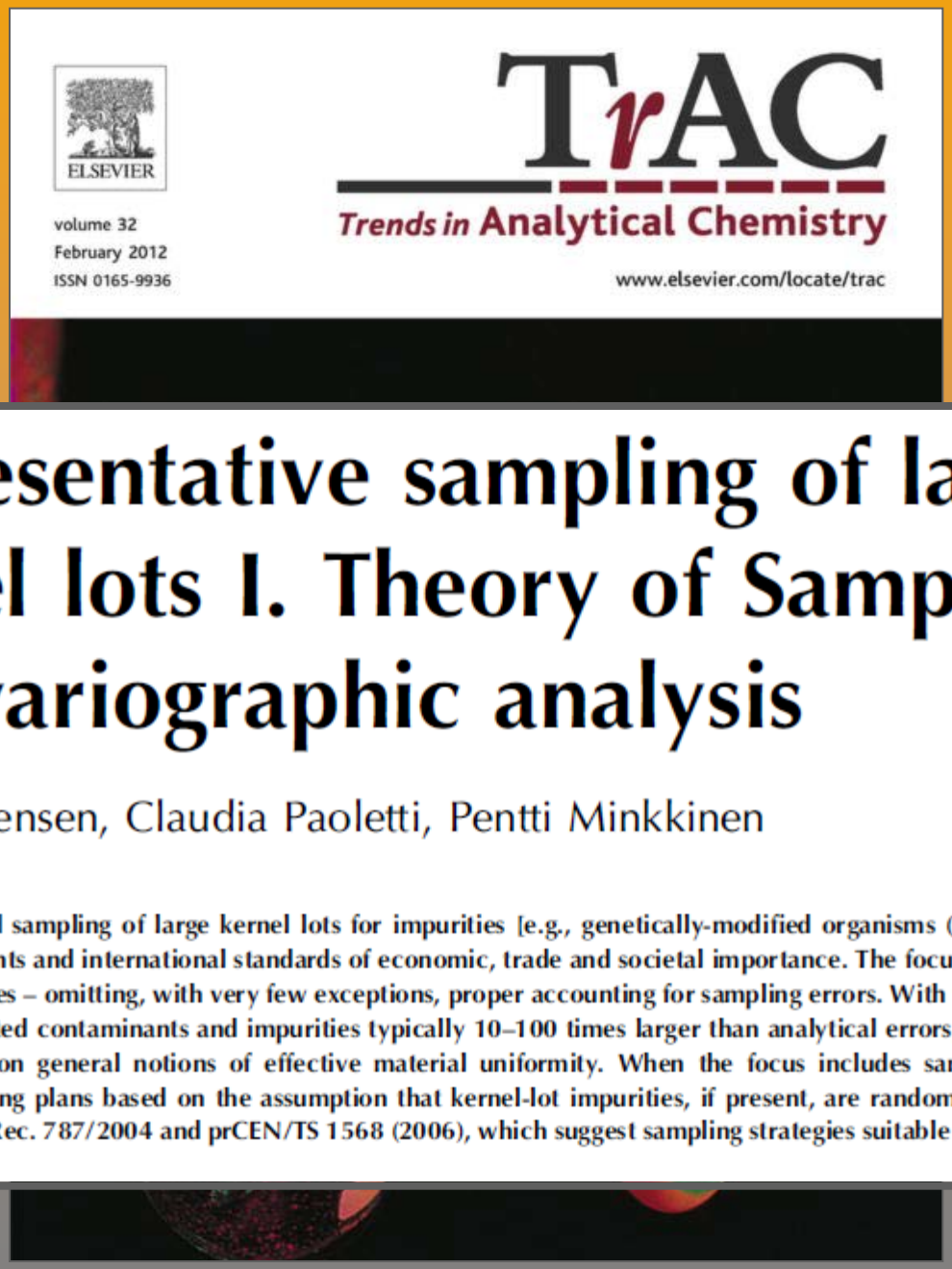


Generic process characterisation of all kinds of materials with significant 1-dim heterogeneitiy

This lot can be used as a general *exemplar* : extreme heterogeneity



ACABS Research Group



GEUS

Representative sampling of large kernel lots I. Theory of Sampling and variographic analysis

Kim H. Esbensen, Claudia Paoletti, Pentti Minkkinen

Official testing and sampling of large kernel lots for impurities [e.g., genetically-modified organisms (GMOs)] is regulated by normative documents and international standards of economic, trade and societal importance. The focus nearly always includes only analytical issues – omitting, with very few exceptions, proper accounting for sampling errors. With total sampling errors for irregularly distributed contaminants and impurities typically 10–100 times larger than analytical errors, this issue is critical for procedures based on general notions of effective material uniformity. When the focus includes sampling, most guidelines recommend sampling plans based on the assumption that kernel-lot impurities, if present, are randomly distributed. The only exceptions are EC Rec. 787/2004 and prCEN/TS 1568 (2006), which suggest sampling strategies suitable for more heterogeneous situations.

ACABS
Research Group

GEUS

Representative sampling of large kernel lots II. Application to soybean sampling for GMO control

Pentti Minkkinen, Kim H. Esbensen, Claudia Paoletti

Official testing and sampling of large kernel lots for impurities [e.g., genetically modified organisms (GMOs)] is regulated by normative documents and international standards of economic, trade and societal importance.

In Part I, we reviewed current official guides and standards for sampling large contaminated kernel lots and the basic concepts of the Theory of Sampling (TOS) for chemical analysis. Here, we re-interpret the data collected in a recent field study (KeLDA) from a stringent TOS perspective, focusing on representative process sampling and variographic analysis in order to characterize the heterogeneities of large kernel lots and to estimate both Total Sampling Error (*TSE*) and Total Analytical Error (*TAE*). This is used as a basis for developing a general approach for optimization of kernel sampling protocols that are "fit for purpose" i.e. robust to heterogeneity and sufficiently accurate also to detect critically low levels of concentration.

We demonstrate that both *TSE* and *TAE* are significantly large for GMO quantitation, but that *TSE* still can be up two orders of magnitude larger than *TAE*, depending on heterogeneity, sampling mode and GMO concentration, signifying that efforts to reduce uncertainties should focus on sampling plans and not on further refinements of analytical precision.

For GMO testing based on the current labeling threshold (0.9%) in European Union regulations, we show that 42 is the absolute minimum number of increments needed for reliable characterization of all lots with a heterogeneity comparable to the most severely heterogeneous KeLDA lots (Lot #1).

ACABS Research Group

GEUS

Representative sampling of large kernel lots III. General considerations on sampling heterogeneous fo

Not just food/feed !!! !!!

Kim H. Esbensen, Claudia Paoletti, Pentti Minkkinen

All materials with ~ *similar* concentration heterogeneities

Part I reviewed the Theory of Sampling (TOS) as applied to quantitation of g analyzed KeLDA data from a variographic analysis perspective and estimated Error (*TAE*).

Results from this analysis are here used as a basis for developing a general approach to optimization of kernel-sampling protocols that are fit for purpose (i.e. scaled with respect to the effective heterogeneity while simultaneously sufficiently accurate to detect critically low concentration levels). While *TAE* is significantly large for GMO quantitation, *TSE* can still be up two orders of magnitude larger, signifying that efforts to reduce GMO-analysis uncertainties should focus on improving or optimizing sampling plans and not on further refinements of analytical precision.

Variogram: Many applications/implementations

i) One variable
ii) One component (P variables) (scores)

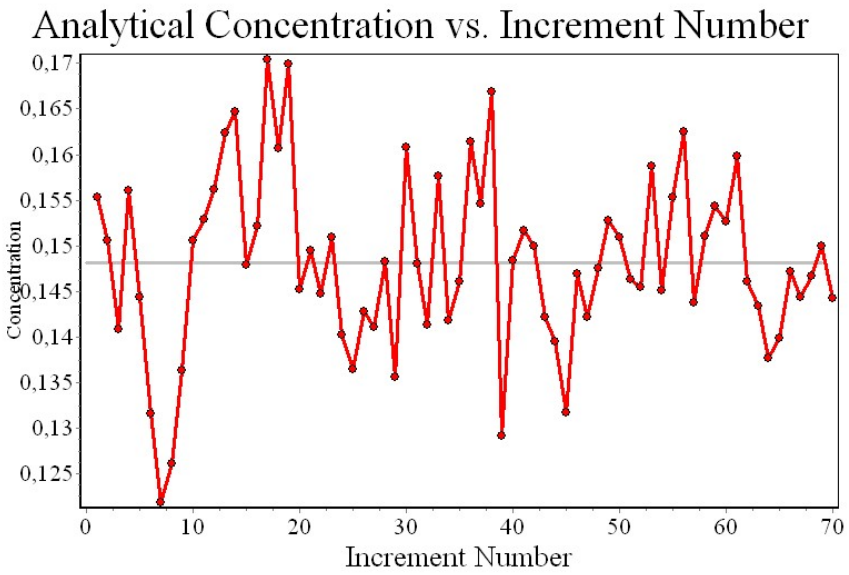
iii) One process deployment location
iv) Several deployment locations

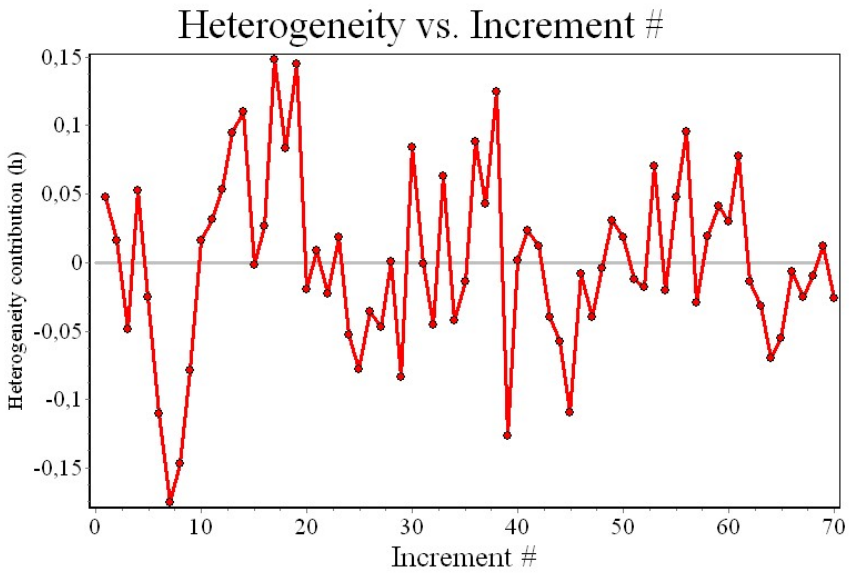
v) Single time location
vi) Several time locations

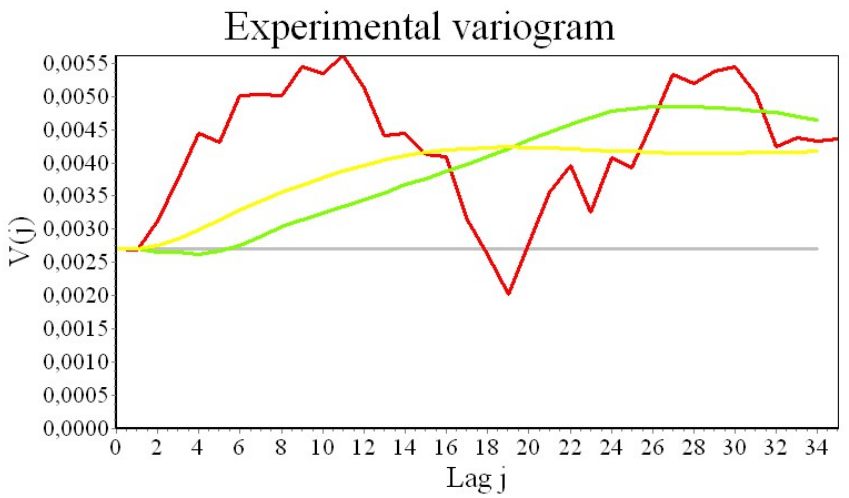
} Temporal analysis …

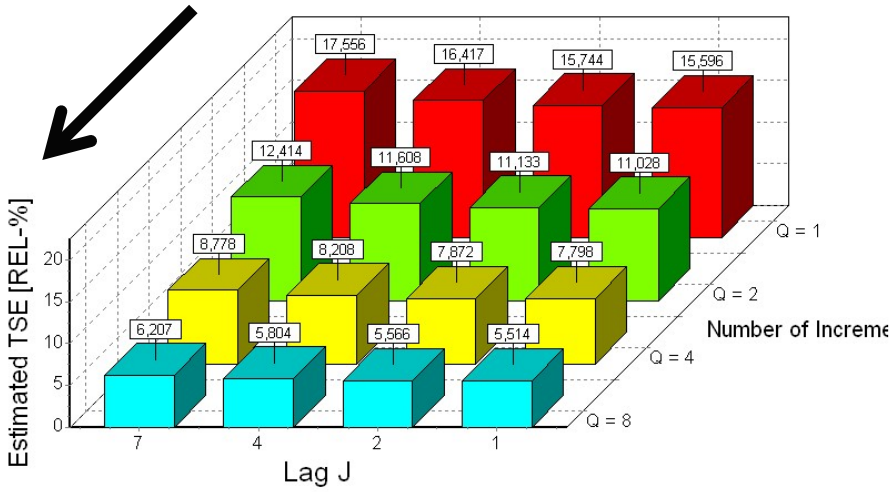
Quality control of the total measurement system:

Nugget effect / Sill (<30%)









TOS' primary use of variographic *analysis*: Prediction of the Total sampling Error (TSE)

DS F

Horizontal – Representative Sampling

Danish Standard 3077, *2013*

Variographic analysis as quality assurance, codified in international sampling standard

Quality control of the total measurement system:

Nugget effect / Sill (<30%)

DS | Dansk standard

DS 3077

2. udgave

2013-08-26

Repræsentativ prøvetagning – Horisontal standard

Representative sampling – Horizontal standard

www.ds.dk
ke@geus.dk

DS F

Horizontal – Representative Sampling

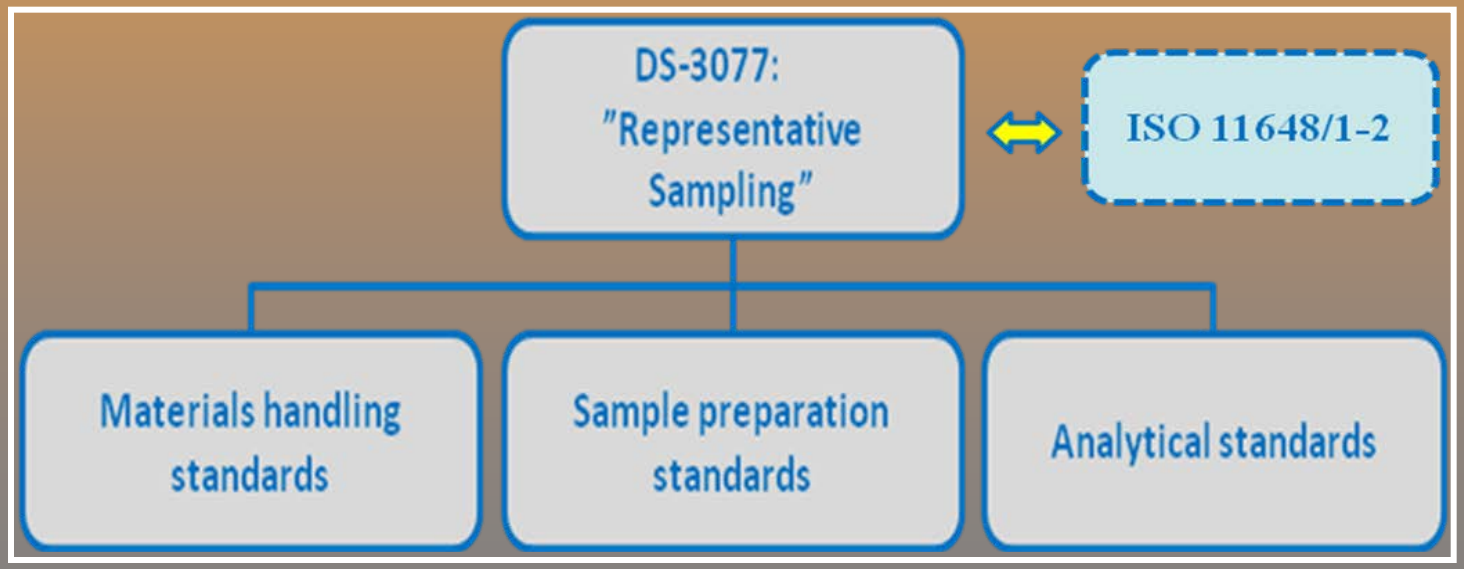
DS-3077

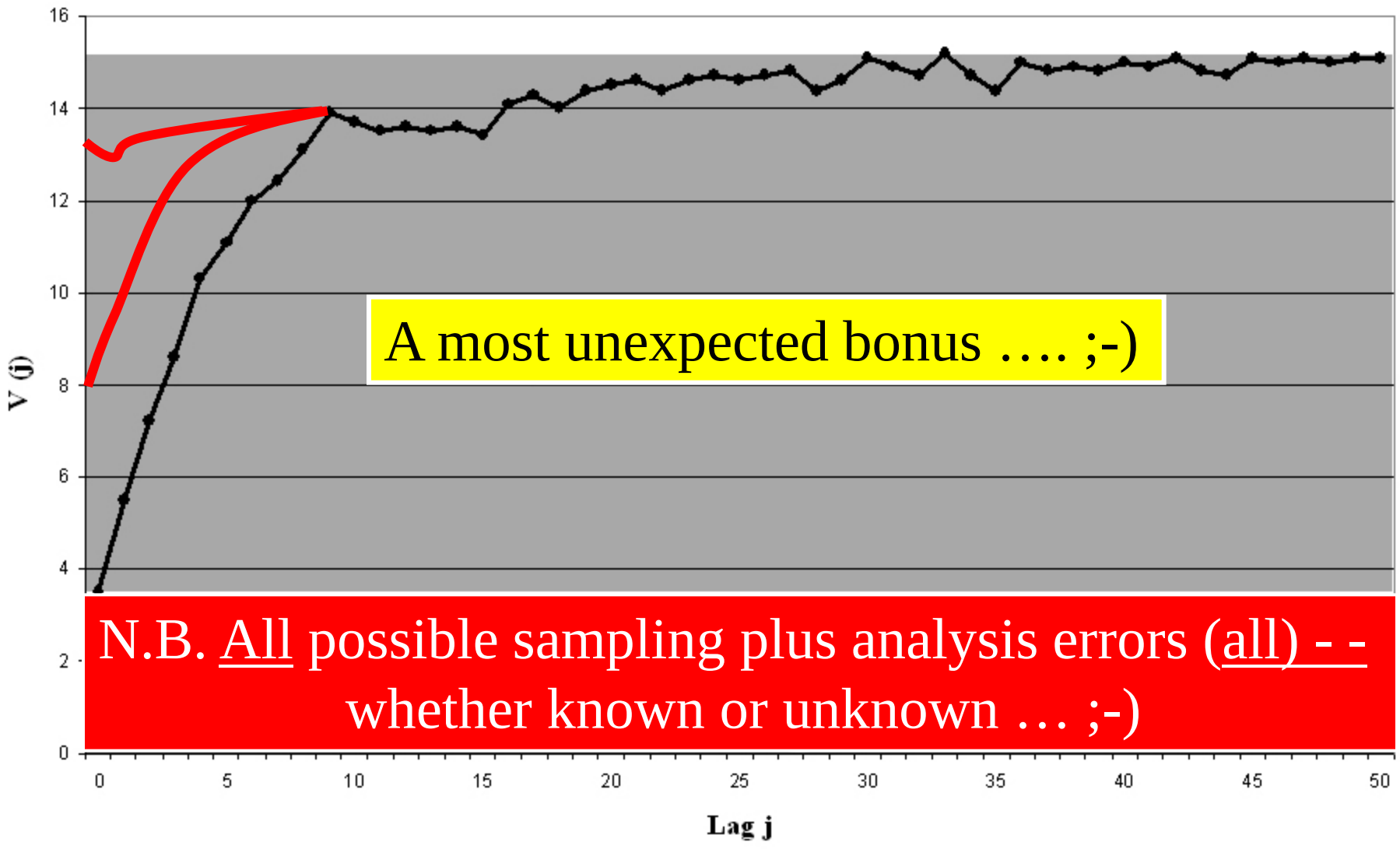
This standard outlines a practical, iterative, self-controlling approach with minimal complexity, based on the Theory of Sampling (TOS). The generic sampling process described and all elements involved are sufficient and necessary for the stated objective, with the consequence that no exceptions can be allowed in order to be able to document the intended sampling representativity. It is necessary to consider the full pathway from primary sampling to analytical results in order to be able to guarantee a reliable and valid analytical outcome. This standard, including normative references, annexes (and further, optional references) constitute a complete and sufficient competence basis for this purpose. The present approach will ensure appropriate levels of accuracy and precision for both primary sampling as well as for all sub-sampling procedures and mass-reduction systems at the subsequent laboratory stages before analysis.

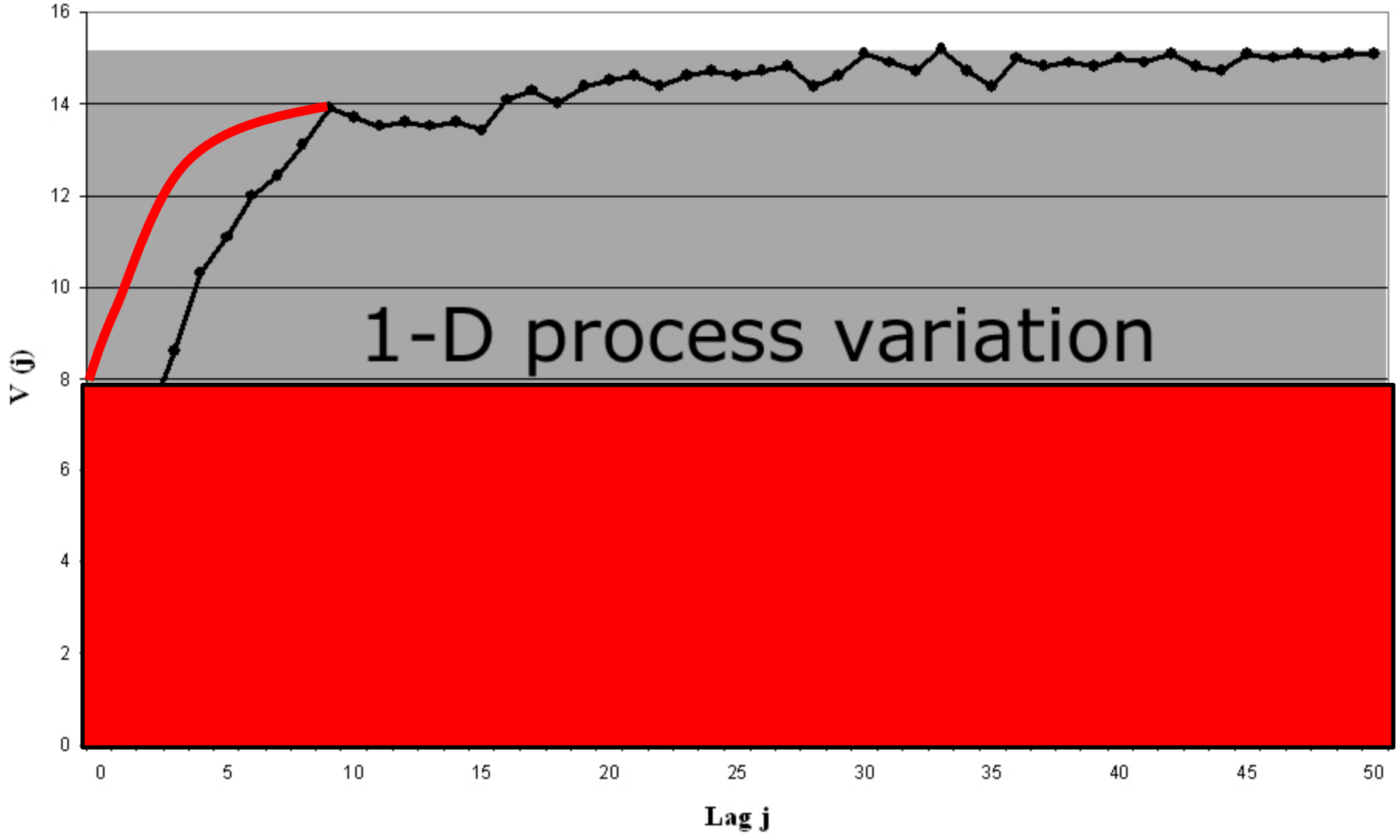
DS F

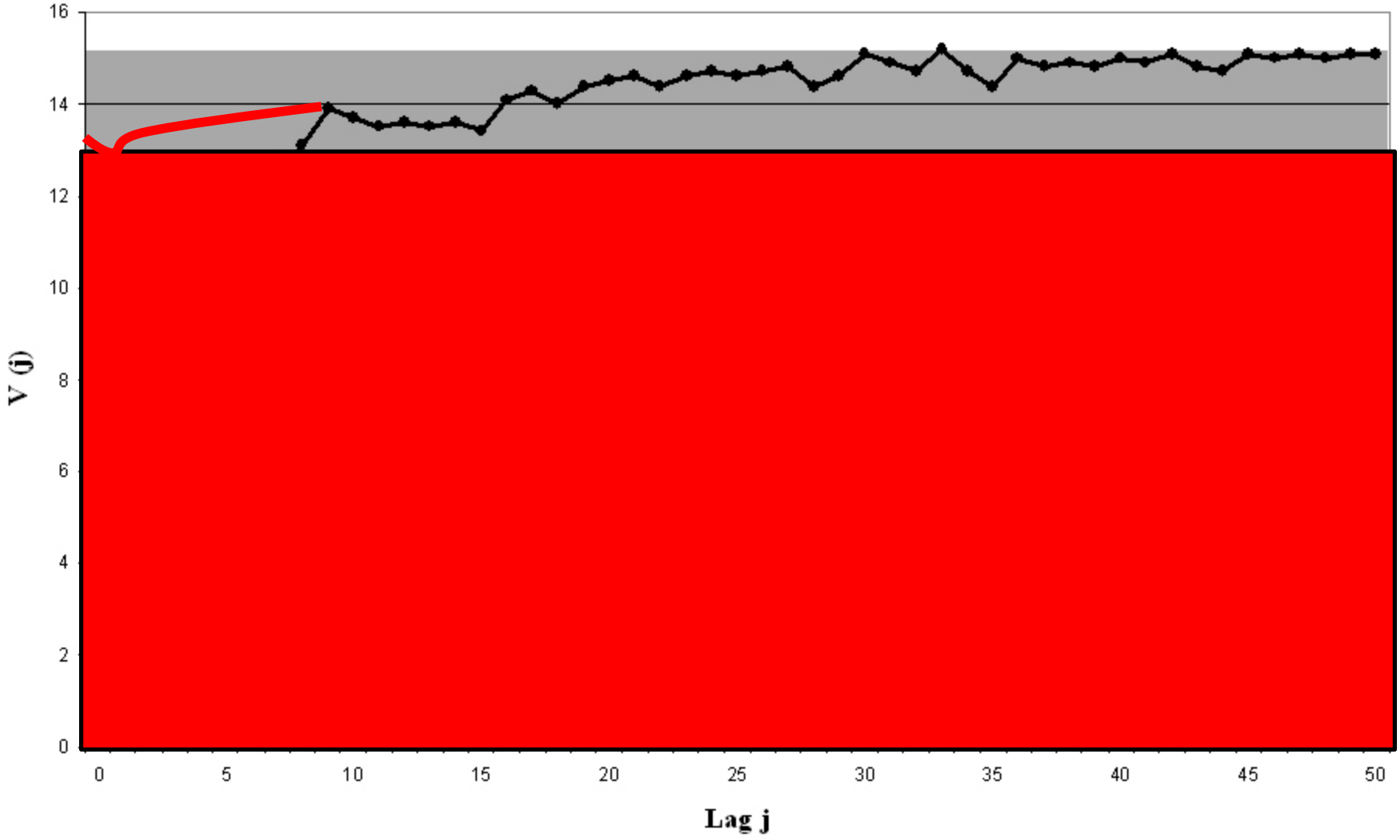
Horizontal – Representative Sampling

DS-3077




16
14
12
10
8
6
4
2
0
V (j)
0
5
10
15
20
25
30
35
40
45
50
Lag j
A most unexpected bonus …. ;-)
N.B. All possible sampling plus analysis errors (all) - -
whether known or unknown … ;-)


16
14
12
10
8
6
4
2
0
V (j)
1-D process variation
0
5
10
15
20
25
30
35
40
45
50
Lag j


16
14
12
10
8
6
4
2
0
V (j)
0
5
10
15
20
25
30
35
40
45
50
Lag j

DS | Dansk standard

DS 3077

2. udgave

2013-08-26

Repræsentativ prøvetagning – Horisontal standard

Representative sampling – Horizontal standard

www.ds.dk
ke@geus.dk

Representative sampling, data quality, validation – a necessary trinity in chemometrics

Kim H. Esbensen & Lars Petersen Julius

in Brown, S, Tauler, R, Walczak, B. (Eds.)
COMPREHENSIVE CHEMOMETRICS
Wiley Major Reference Works, vol. 4, pp.1-20

REPRESENTATIVE PROCESS SAMPLING FOR RELIABLE DATA ANALYSIS

Lars Petersen & Kim H. Esbensen

Aalborg University Esbjerg, Denmark

Journal of Chemometrics (2005)

REPRESENTATIVE PROCESS SAMPLING - in practice

Variographic analysis and estimation of total sampling errors (TSE)

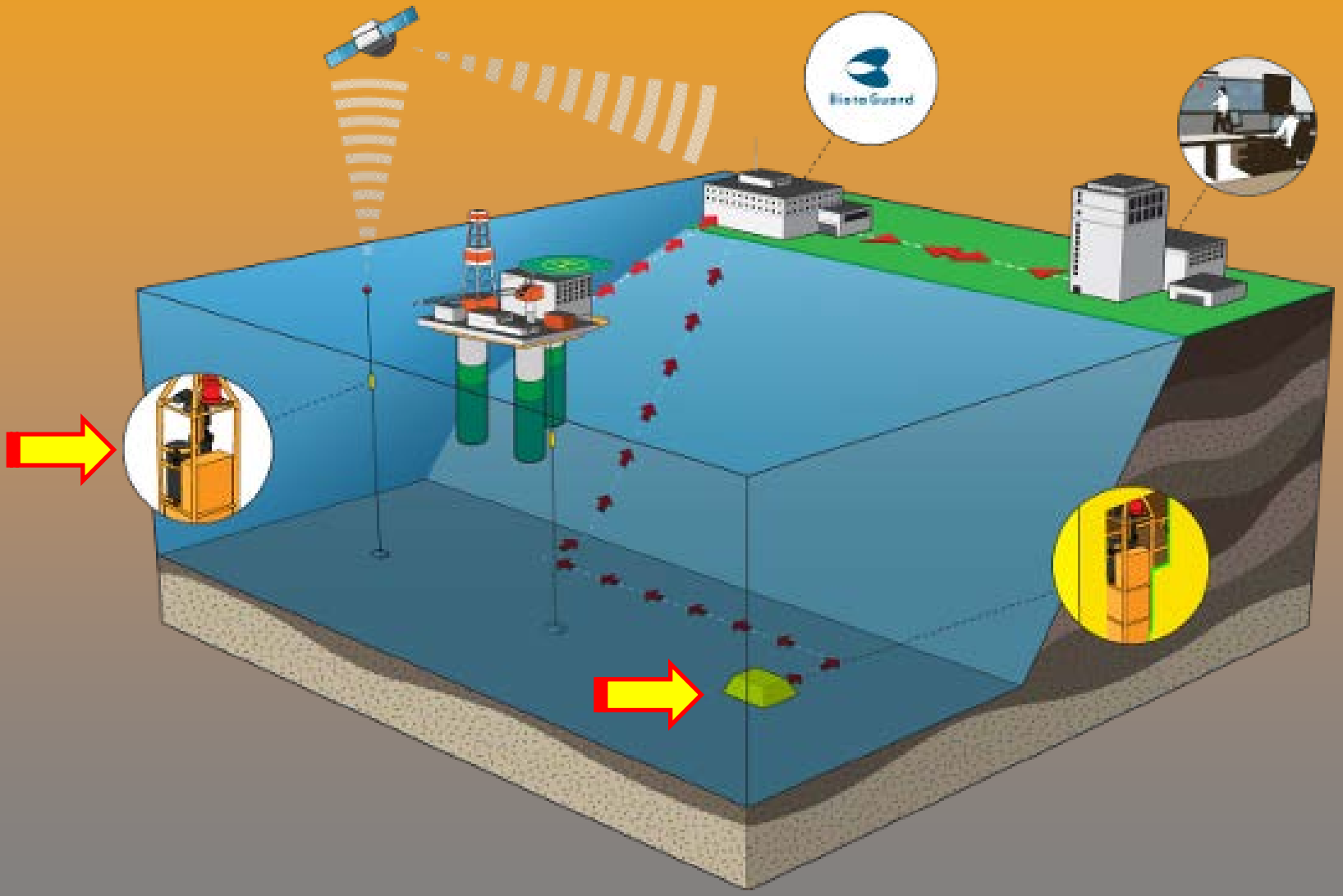
Kim H. Esbensen*, Hans Henrik Friis-Petersen, Lars Petersen, Jens Bo Holm-Nielsen, Peter P. Mortensen

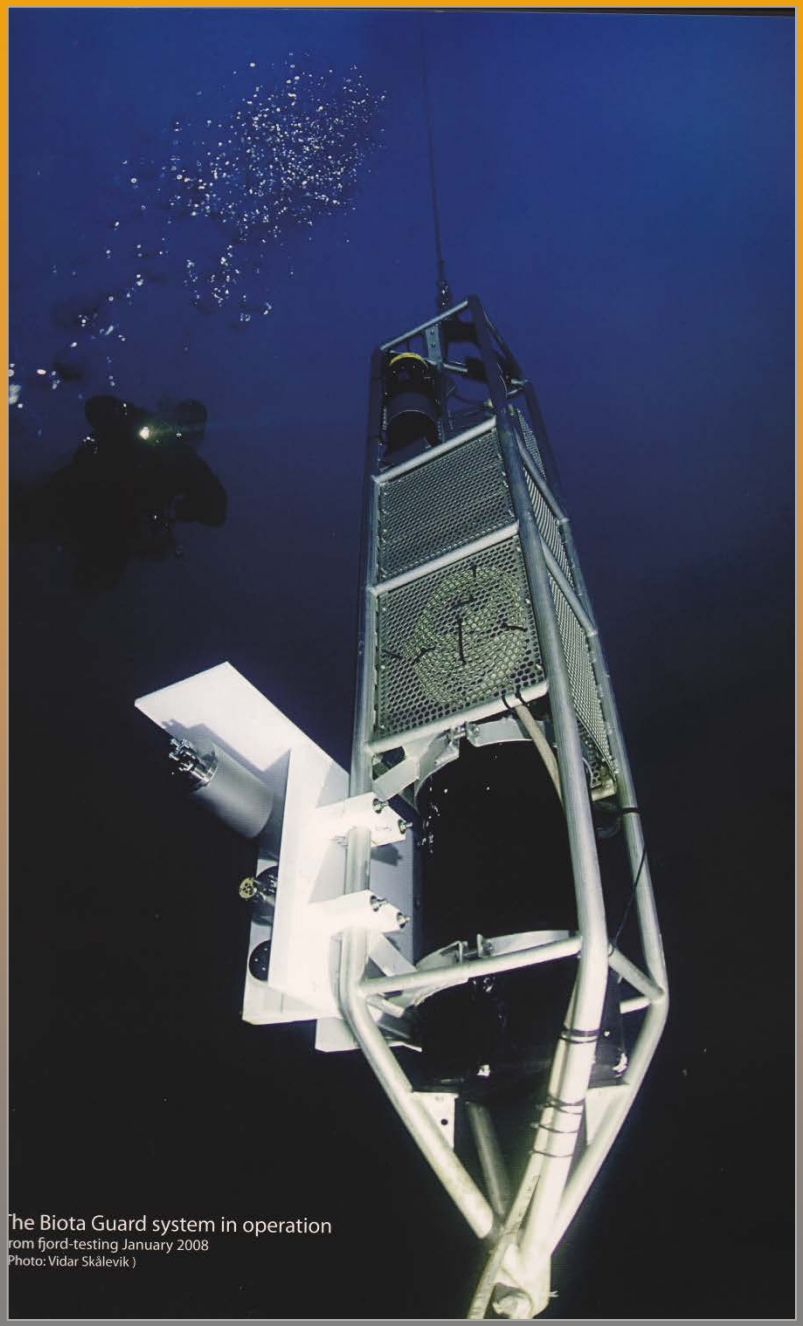
Chemometrics and Intelligent Laboratory Systems (*2007*)

Thank you for your attention!

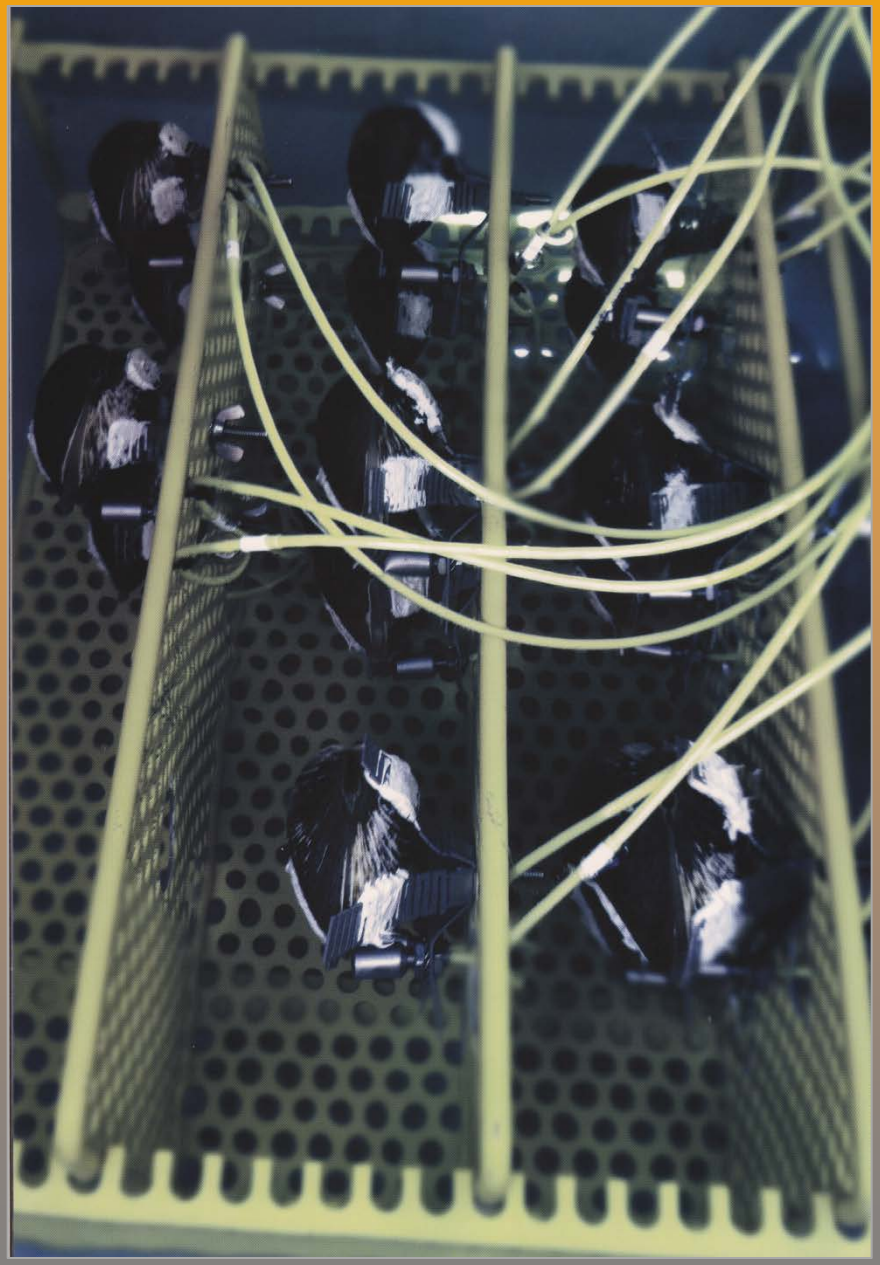
ke@geus.dk

Novel application – Oil pollution monitoring



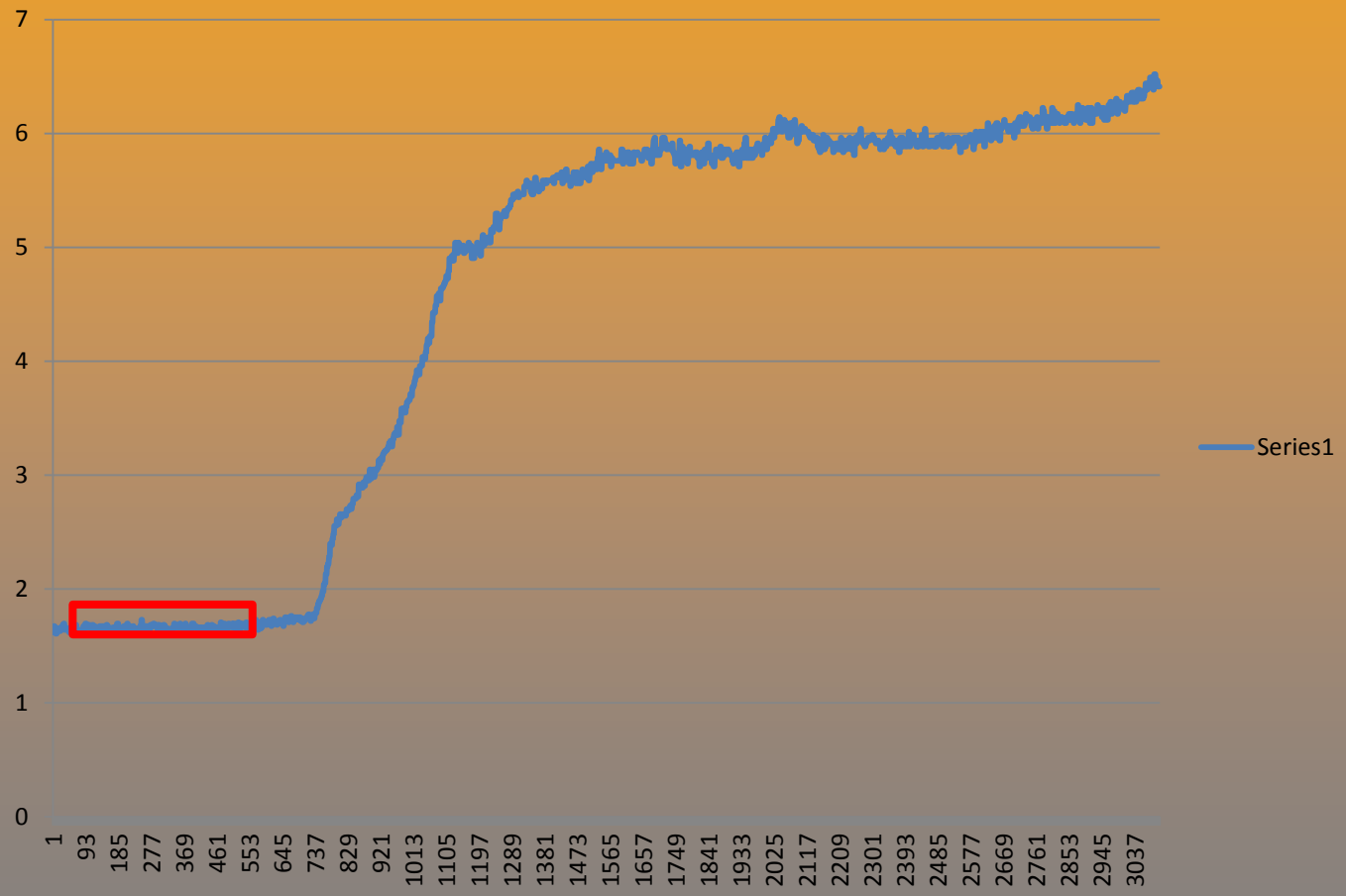


The Biota Guard system in operation
rom fjord-testing January 2008
Photo: Vidar Skålevik)

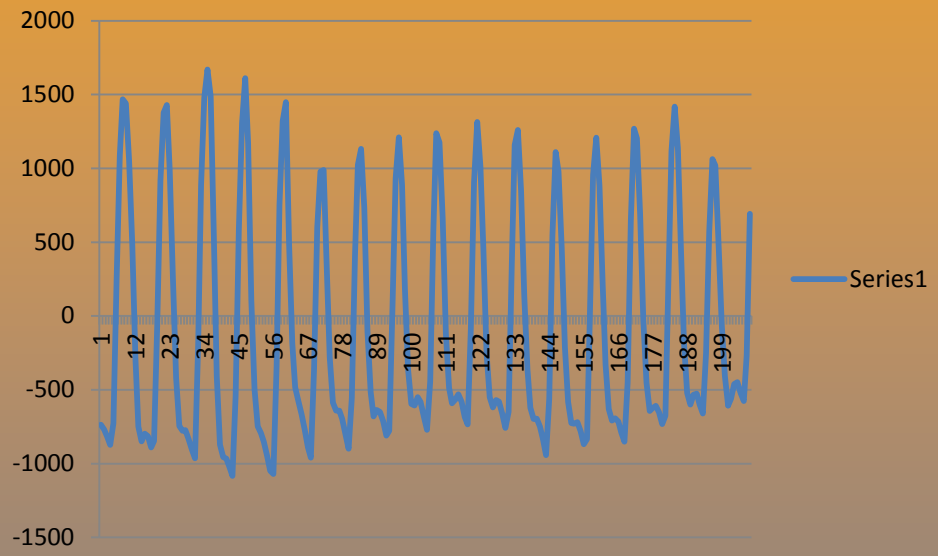


NIR sensor

Biosensor time series

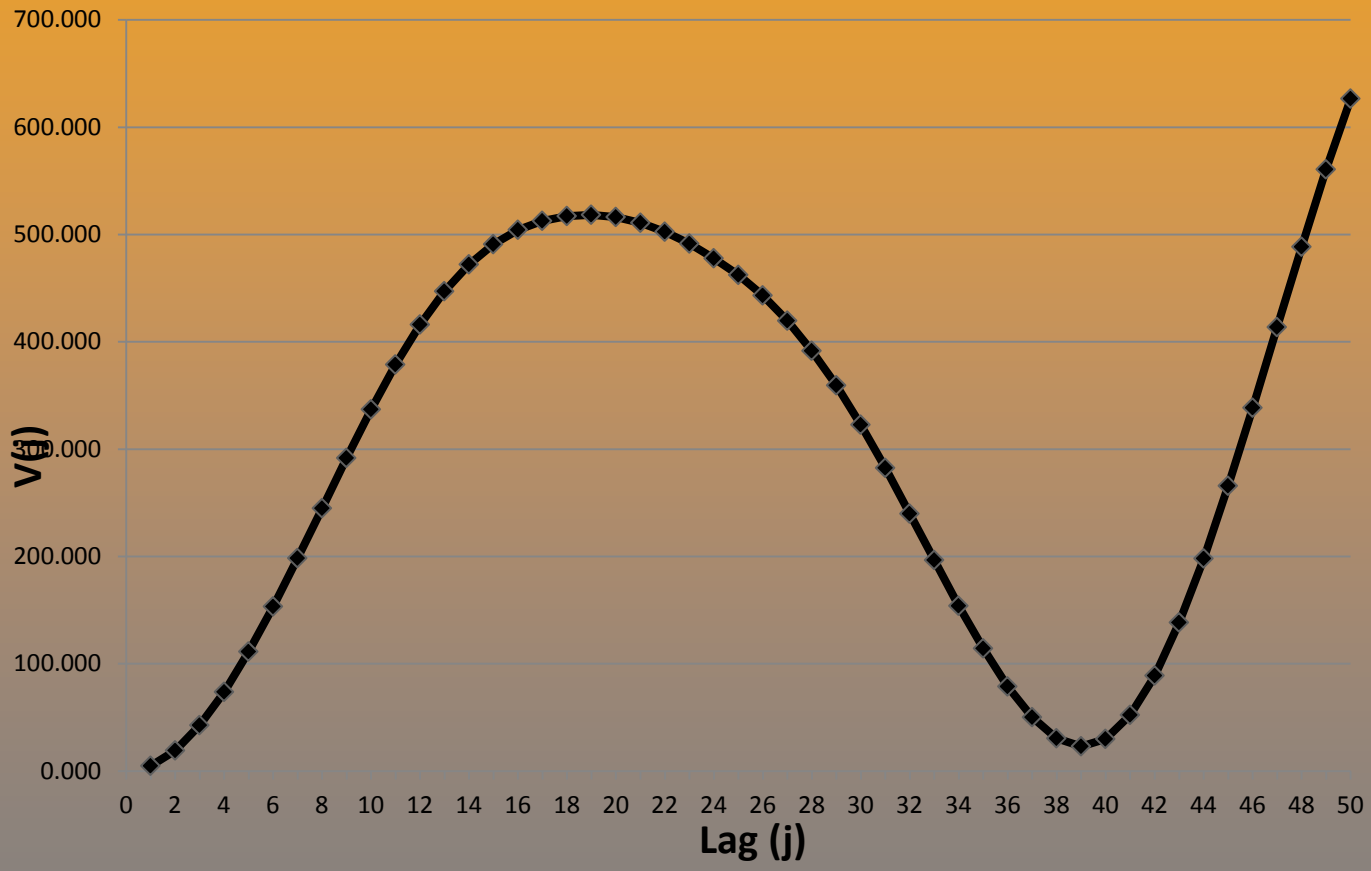


Biosensor time series



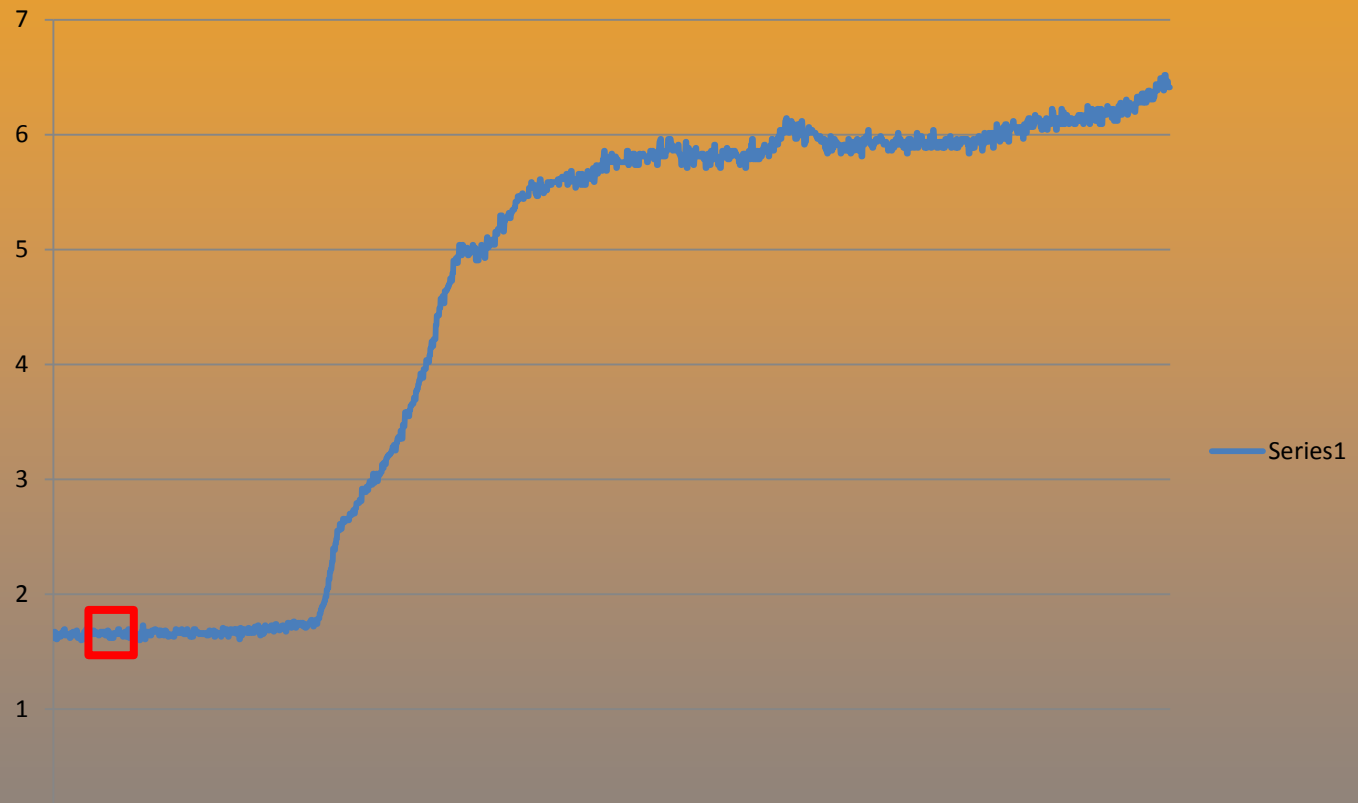
Measurement variable: Heart beat

Biosensor variogram characteristic



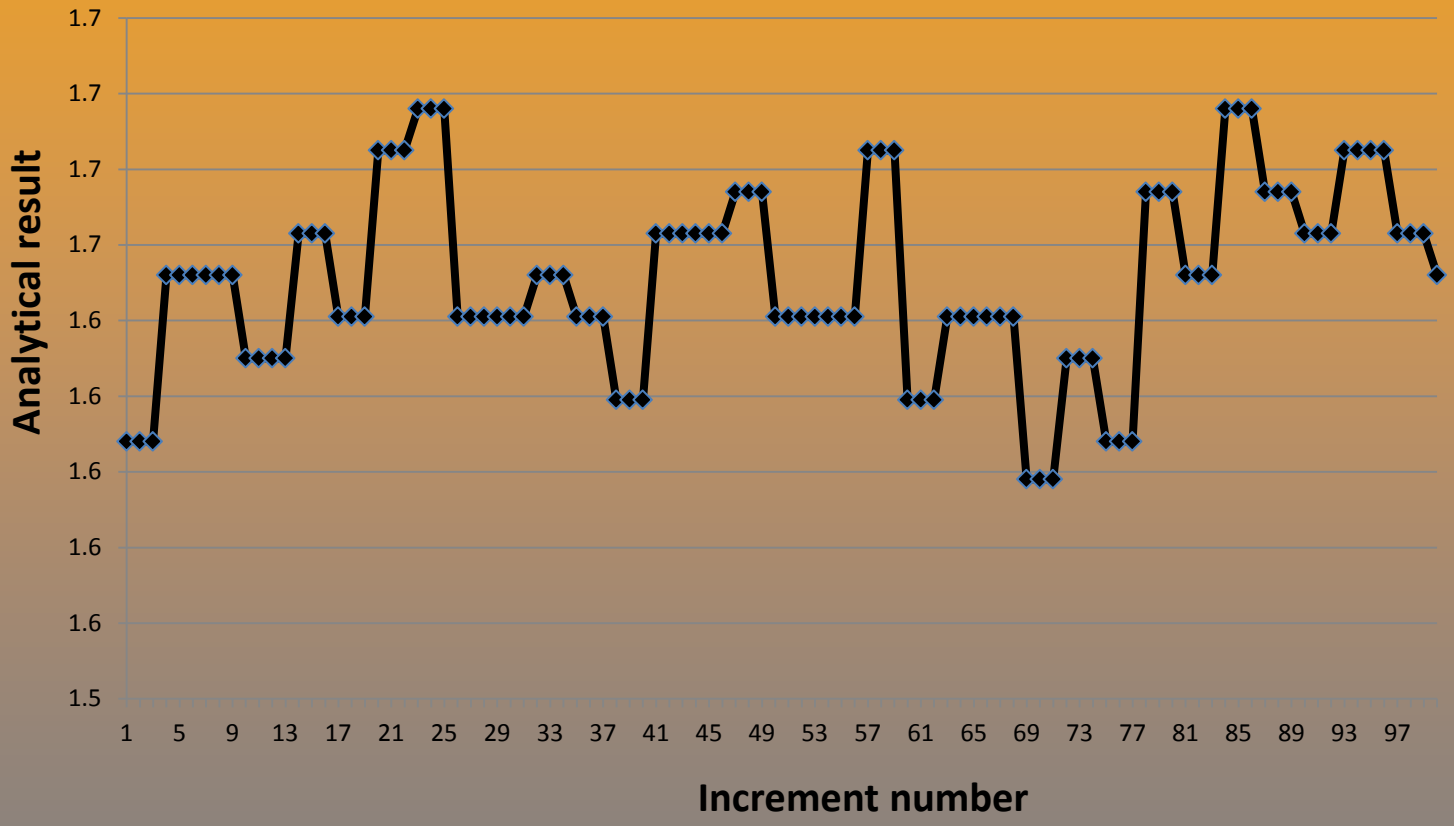
Measurement variable: Heart beat

Biosensor time series



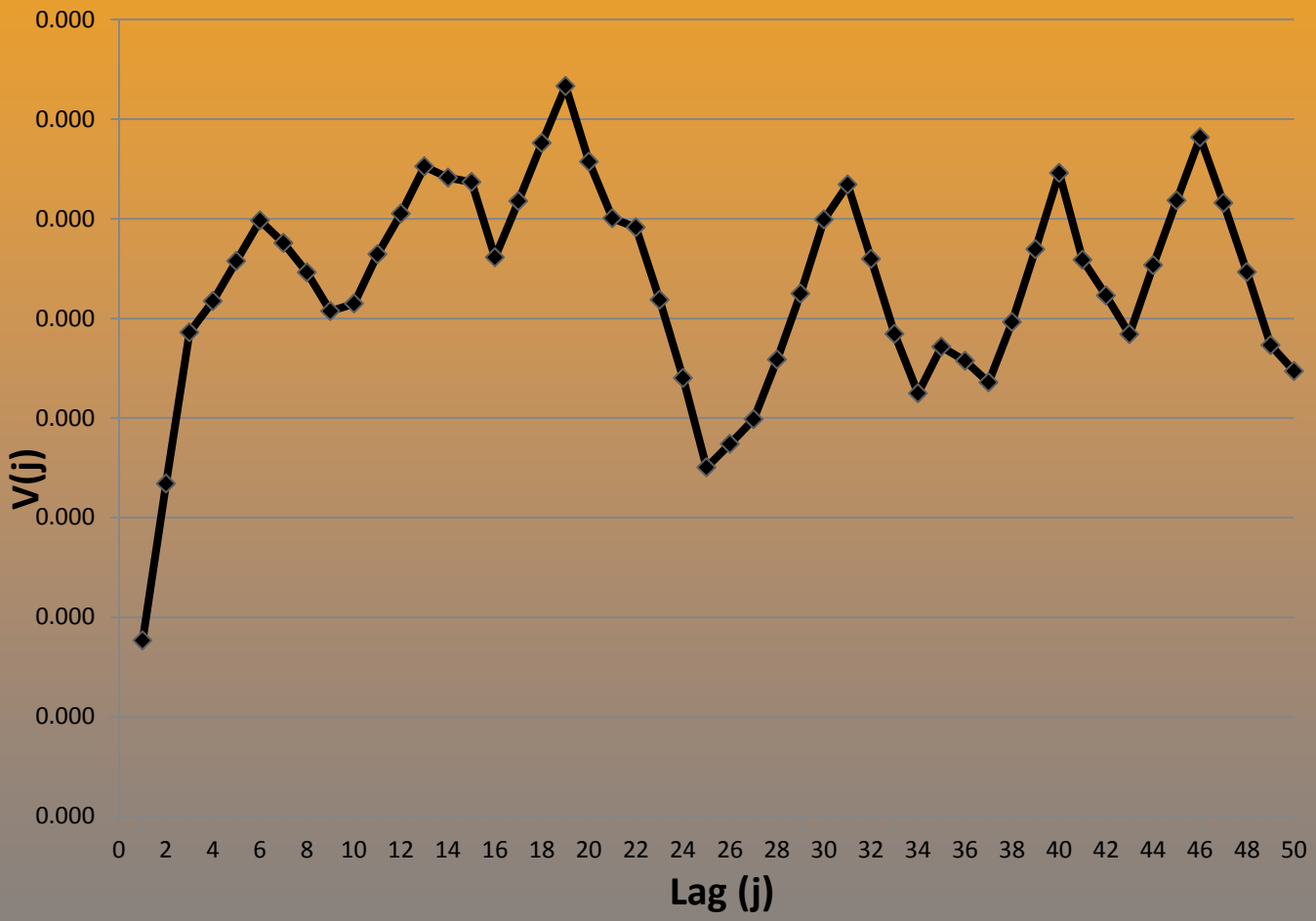
Measurement variable: Bivalve gap opening

Measurement series (biosensor)



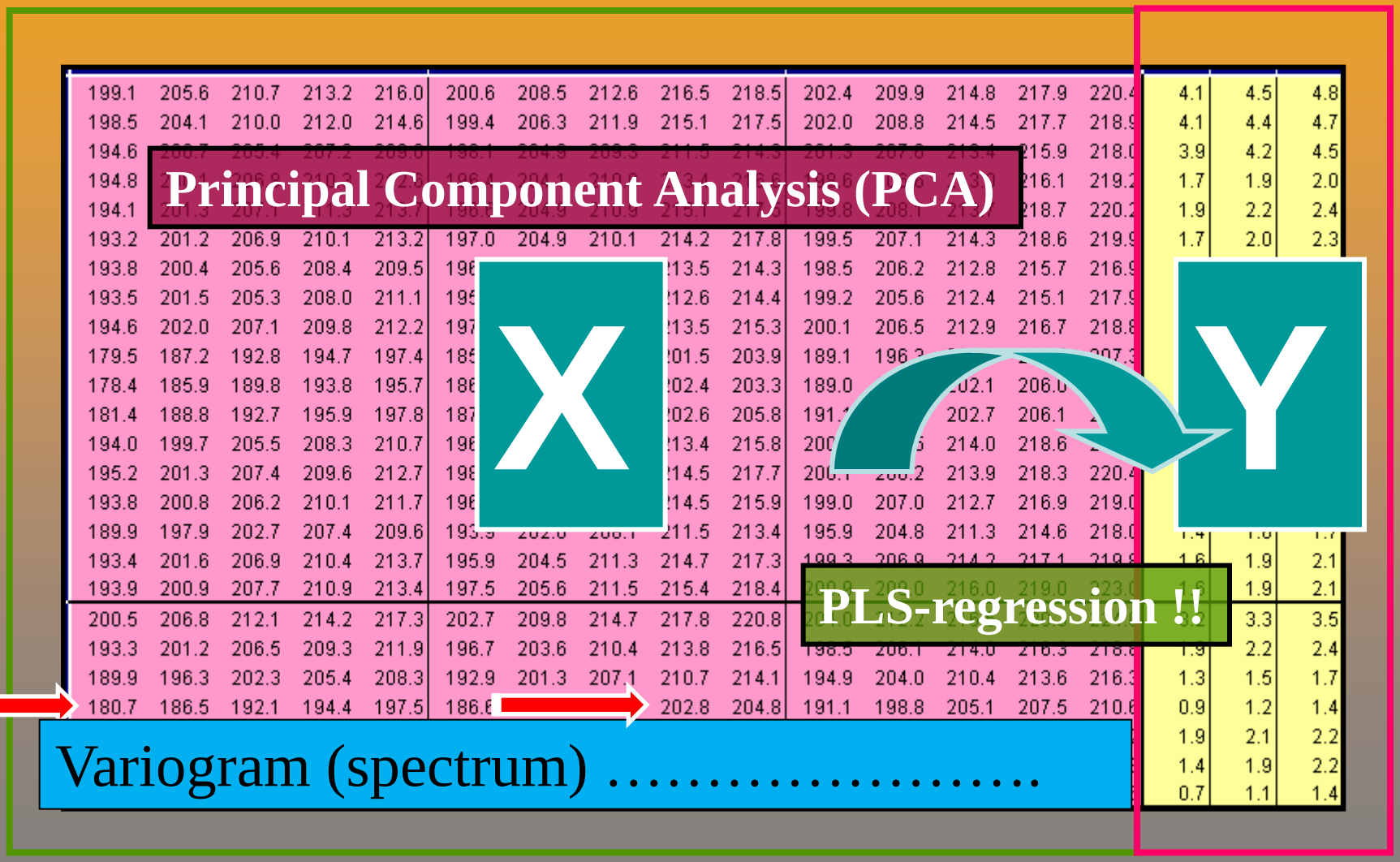
Measurement variable: Bivalve gap opening

Biosensor time series



Measurement variable: Bivalve gap opening

Soft modelling: chemometrics



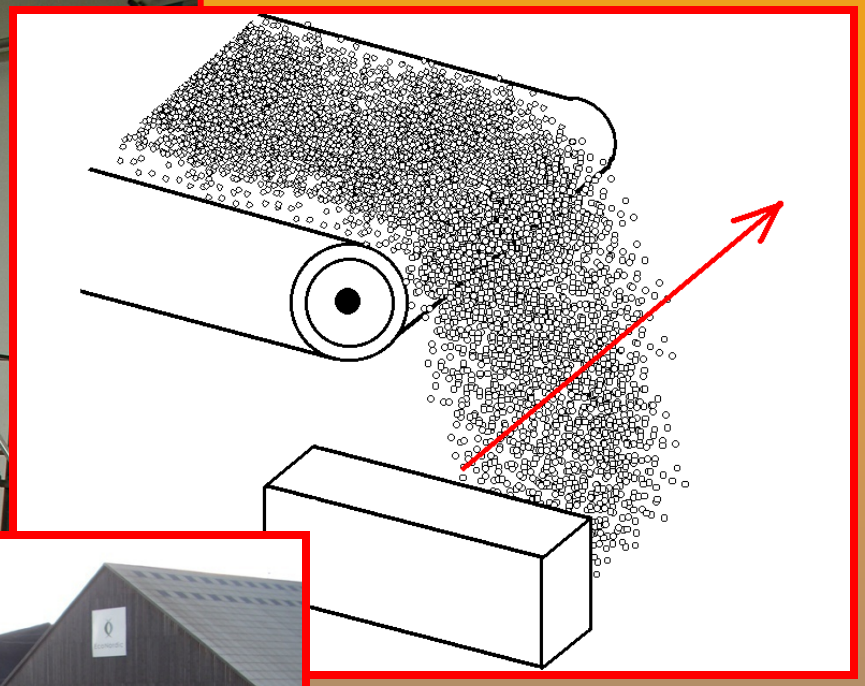
Agenda

- Variogram: complexity spectrum; f(scale)
- Variogram: row in X-matrix
- PCA, PLS … … [Y: functional param.]
- Ample possibilities for PaRC, SIMCA … Ample pos. for process chemometrics

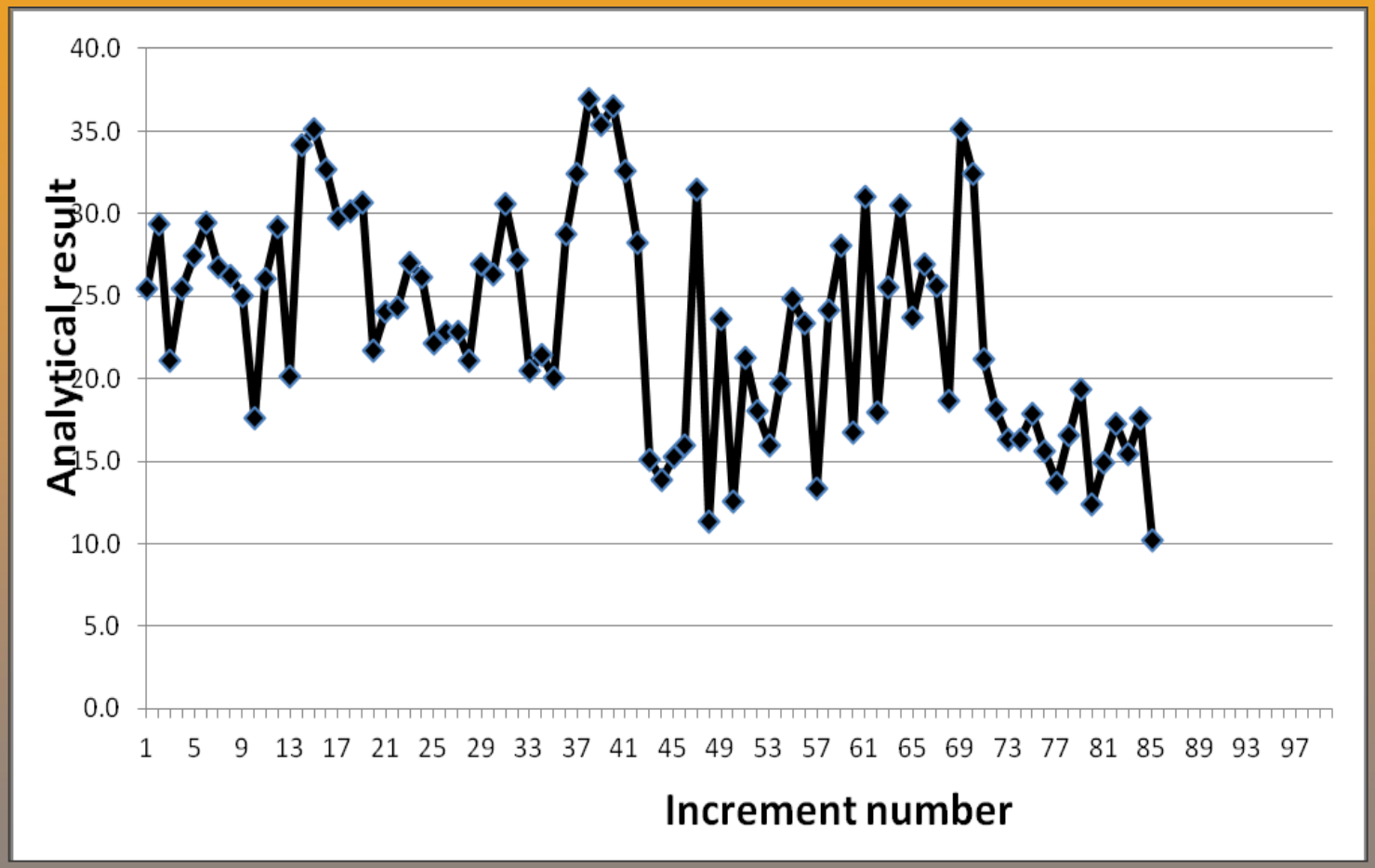
- Only lack of imagination sets limits …

MYTHBUSTERS in Chemometrics

<u>You</u> are welcome as a guest-buster ..



Variographic transects: Unsurpassed soil complexity characteristics (1-/2-D)


40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
Analytical result
1 5 9 13 17 21 25 29 33 37 41 45 49 53 57 61 65 69 73 77 81 85 89 93 97
Increment number

”Stage-less” mass reduction equipment