KHE Consulting

Theory and Practice of Representative Sampling

Proper sampling is the critical success factor before analysis and decision making

Who / What / When ….

Kim H. Esbensen

Geological Survey of Denmark & Greenland (GEUS):
Research professor (*sampling, chemometrics)* 2010-2015

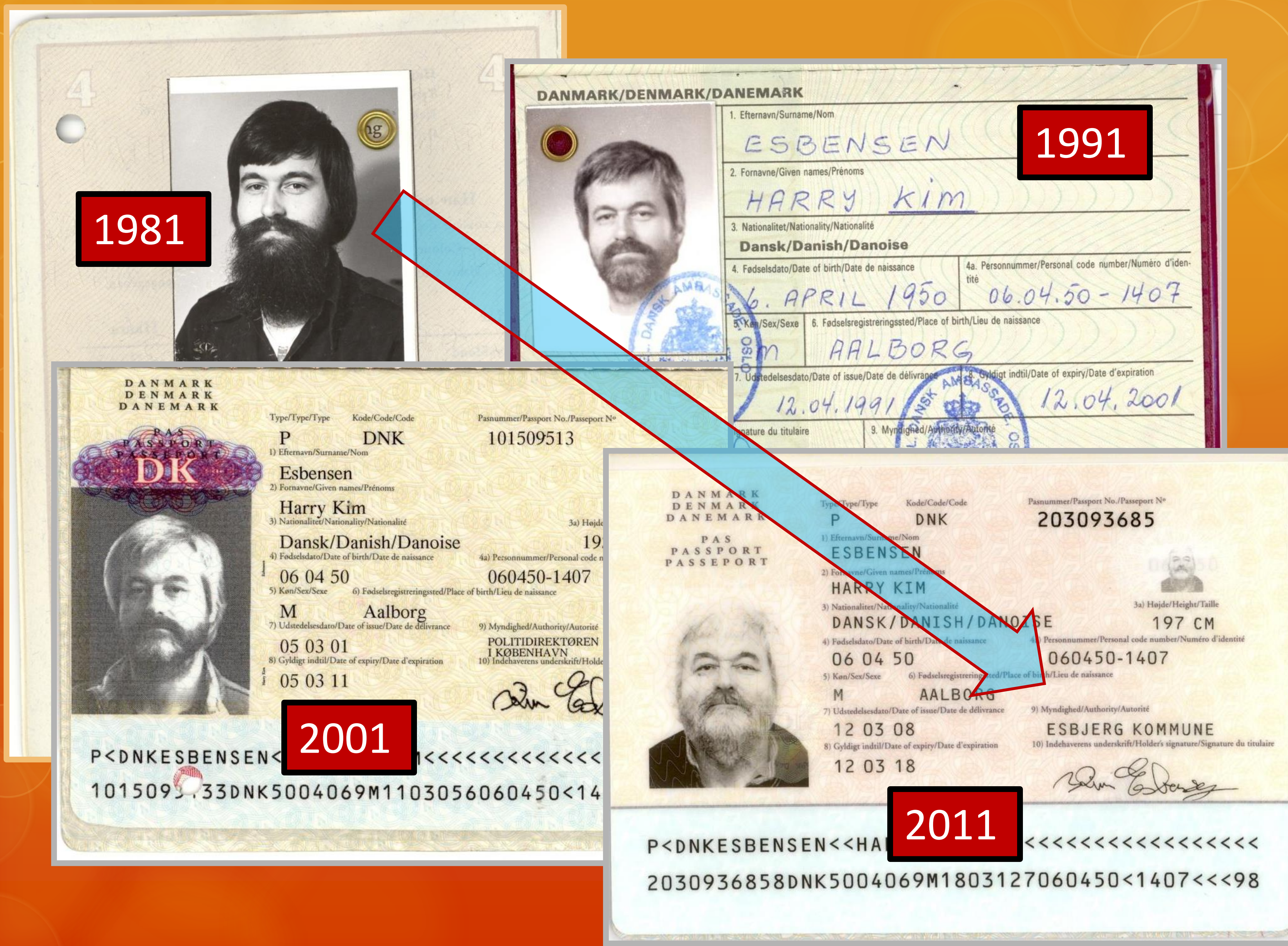
Aalborg University (*chemometrics, sampling*): professor 2001-

Telemark University of Process Technology (HIT): professor 1991-

Norwegian Computing Center & SINTEF: research scientist 1985-

Terra Swede (exploration): geochemical data analyst 1982-

Technical University of Denmark (DTH), Ph.D. (metallurgy, data analysis) 1981
University of Aarhus, Denmark: M.Sc. (geology) 1979


1981
1991
DANMARK/DENMARK/DANEMARK
1. Efternavn/Surname/Nom
ESBENSEN
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
9. Myndighed/Authority/Autorité
2001
DANMARK
DENMARK
DANEMARK
PAS
PASSPORT
PASSEPORT
DK
Type/Type/Type
P
Kode/Code/Code
DNK
Pasnummer/Passport No./Passeport N°
101509513
1) Efternavn/Surname/Nom
Esbensen
2) Fornavne/Given names/Prénoms
Harry Kim
3) Nationalitet/Nationality/Nationalité
Dansk/Danish/Danoise
4) Fødselsdato/Date of birth/Date de naissance
06 04 50
4a) Personnummer/Personal code
060450-1407
5) Køn/Sex/Sexe
M
6) Fødselsregistreringssted/Place of birth/Lieu de naissance
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/Holder
P<DNKESBENSEN<
2011
DANMARK
DENMARK
DANEMARK
PAS
PASSPORT
PASSEPORT
Type/Type/Type
P
Kode/Code/Code
DNK
Pasnummer/Passport No./Passeport N°
203093685
1) Efternavn/Surname/Nom
ESBENSEN
2) Fornavne/Given names/Prénoms
HARRY KIM
3) Nationalitet/Nationality/Nationalité
DANSK/DANISH/DANOISE
3a) Højde/Height/Taille
197 CM
4) Fødselsdato/Date of birth/Date de naissance
06 04 50
4a) Personnummer/Personal code number/Numéro d'identité
060450-1407
5) Køn/Sex/Sexe
M
6) Fødselsregistreringssted/Place of birth/Lieu de naissance
AALBORG
7) Udstedelsesdato/Date of issue/Date de délivrance
12 03 08
9) Myndighed/Authority/Autorité
ESBJERG 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
2030936858DNK5004069M1803127060450<1407<<<98

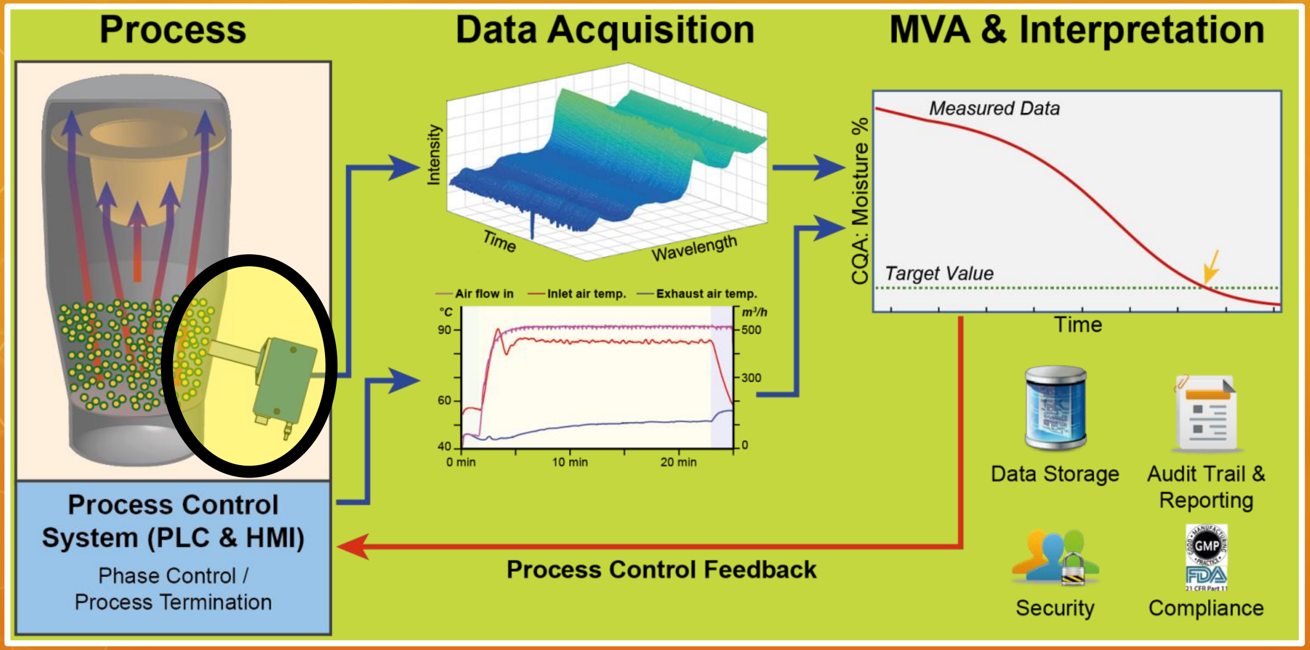
KHE CONSULTING

KHE Consulting

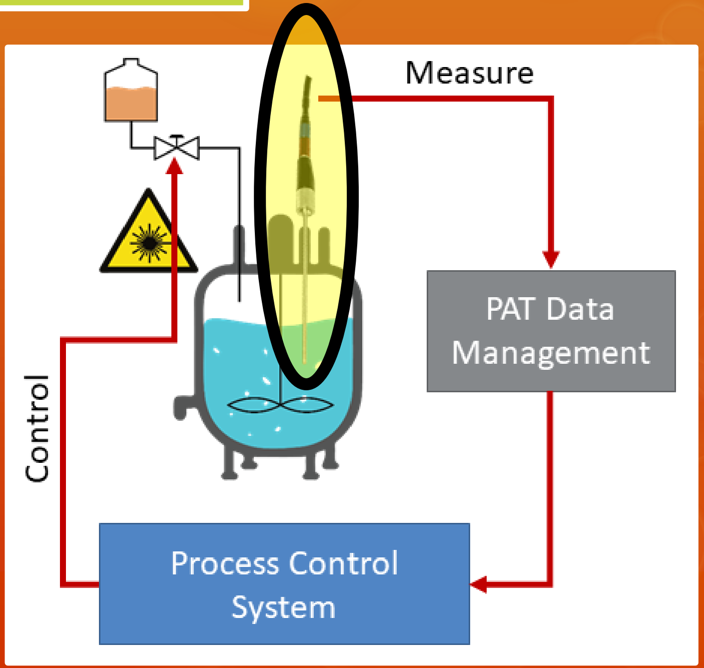
Theory and Practice of Representative Sampling

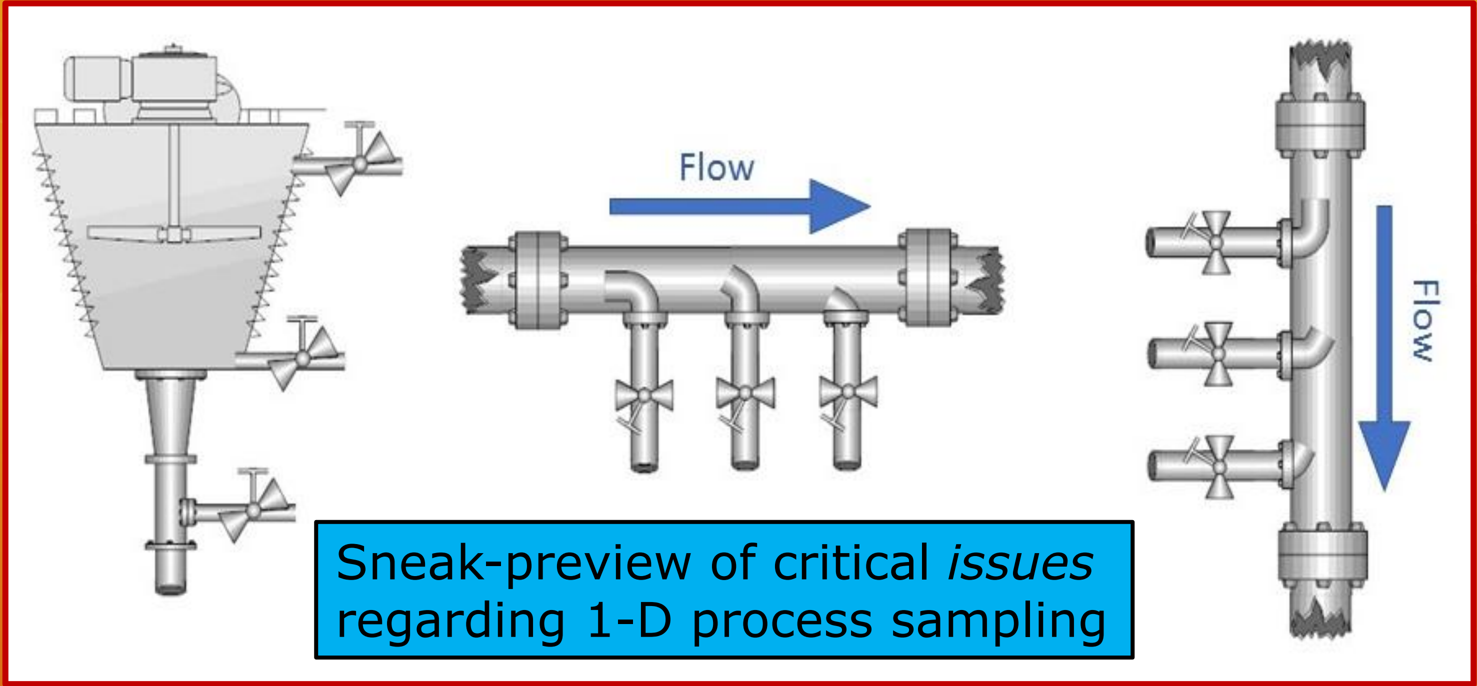
Proper sampling <u>the</u> critical success factor before analysis and decision making:

- for the company; for the customer; for the scientist; for the technician
- in science, technology, industry; in the analytical laboratory
- for compliance; for safety; for society
- for the seller; for the buyer; for the middleman, for the arbiter
- for stationary lots and for <u>moving lots</u> (PAT process monitoring & control)

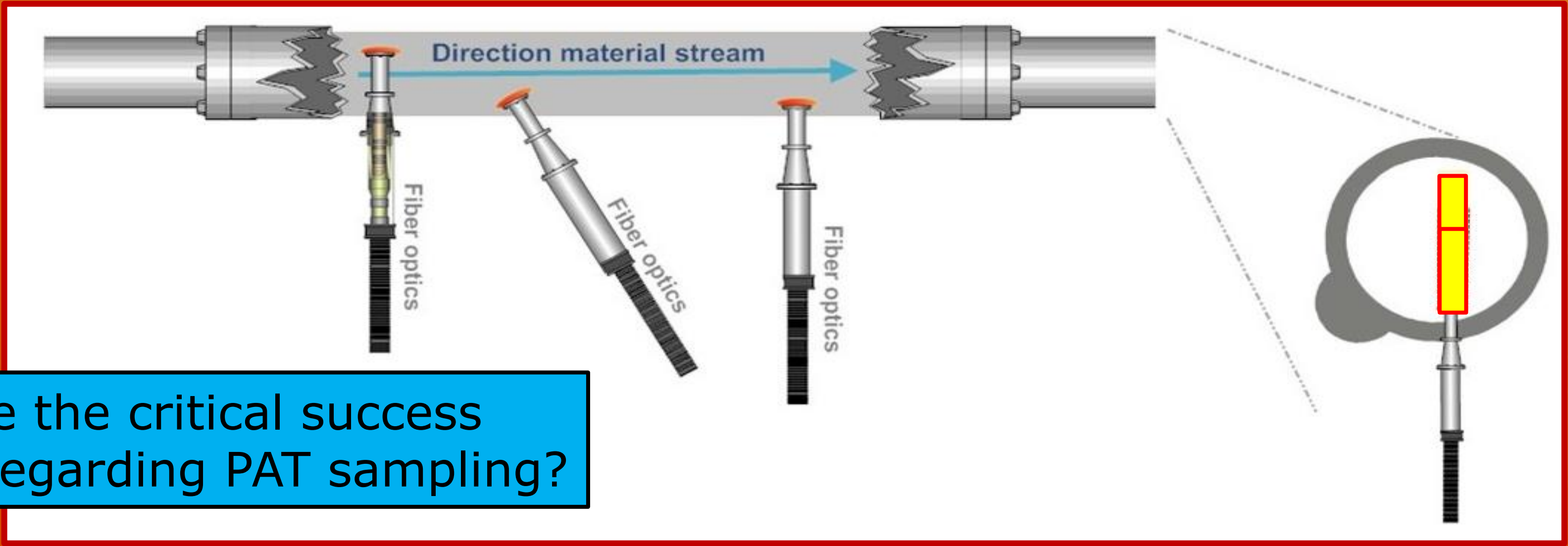


Examples of PAT integration incl. PAT Data Management





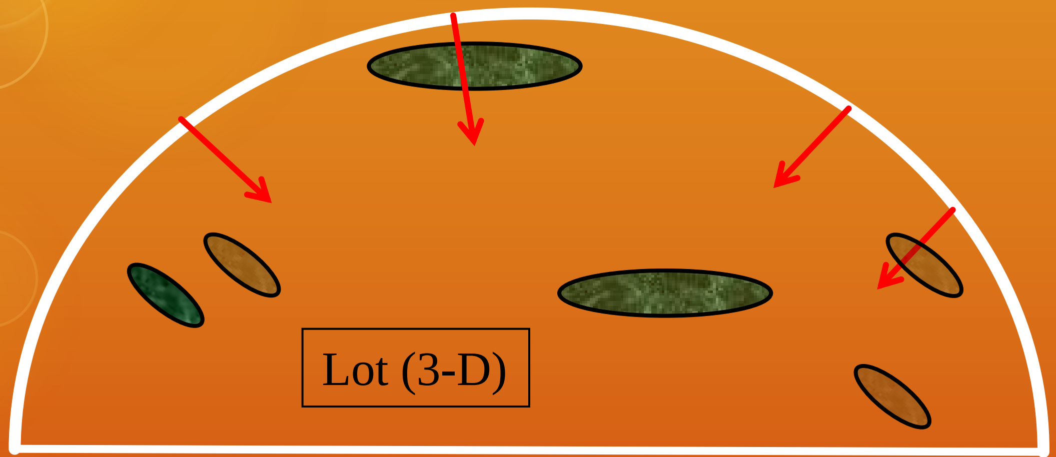
Sneak-preview of critical *issues* regarding 1-D process sampling



What are the critical success factors regarding PAT sampling?

Sampling tool: spear sampler <u>or</u> PAT sensor

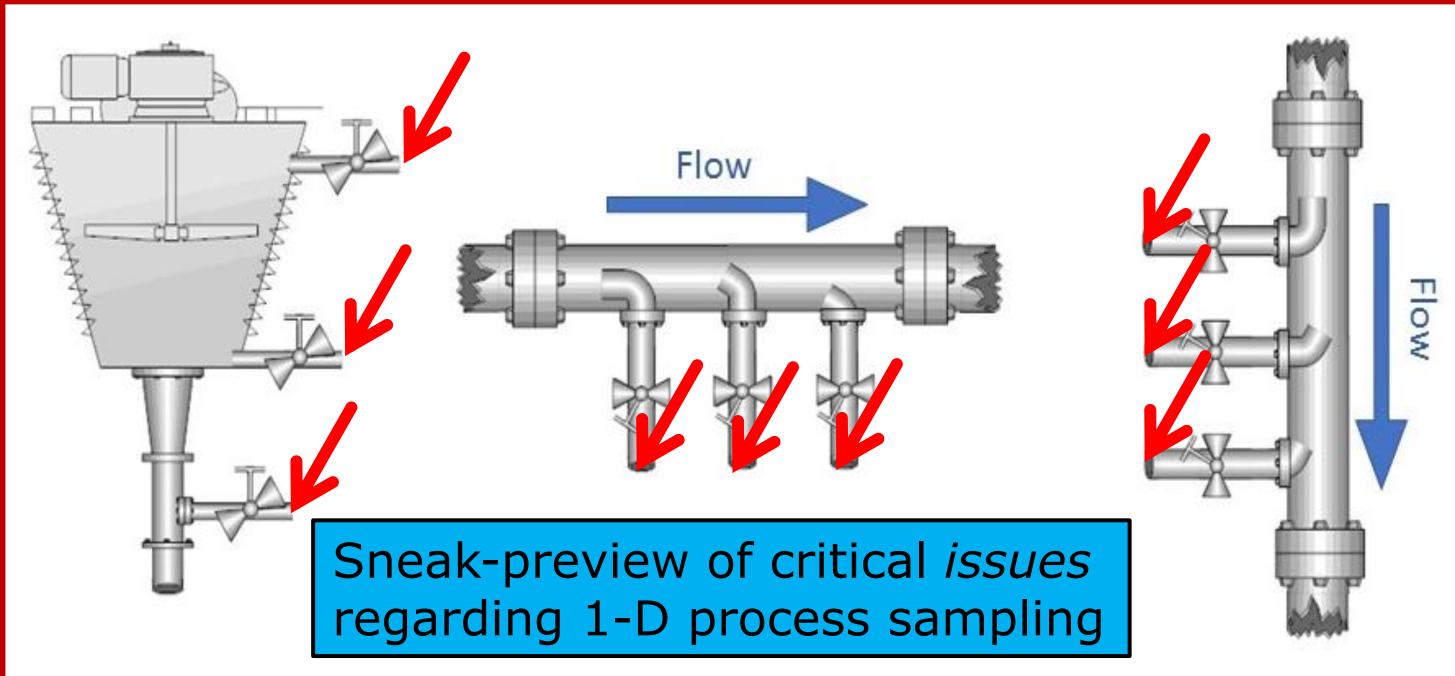
Heterogeneity (hidden)



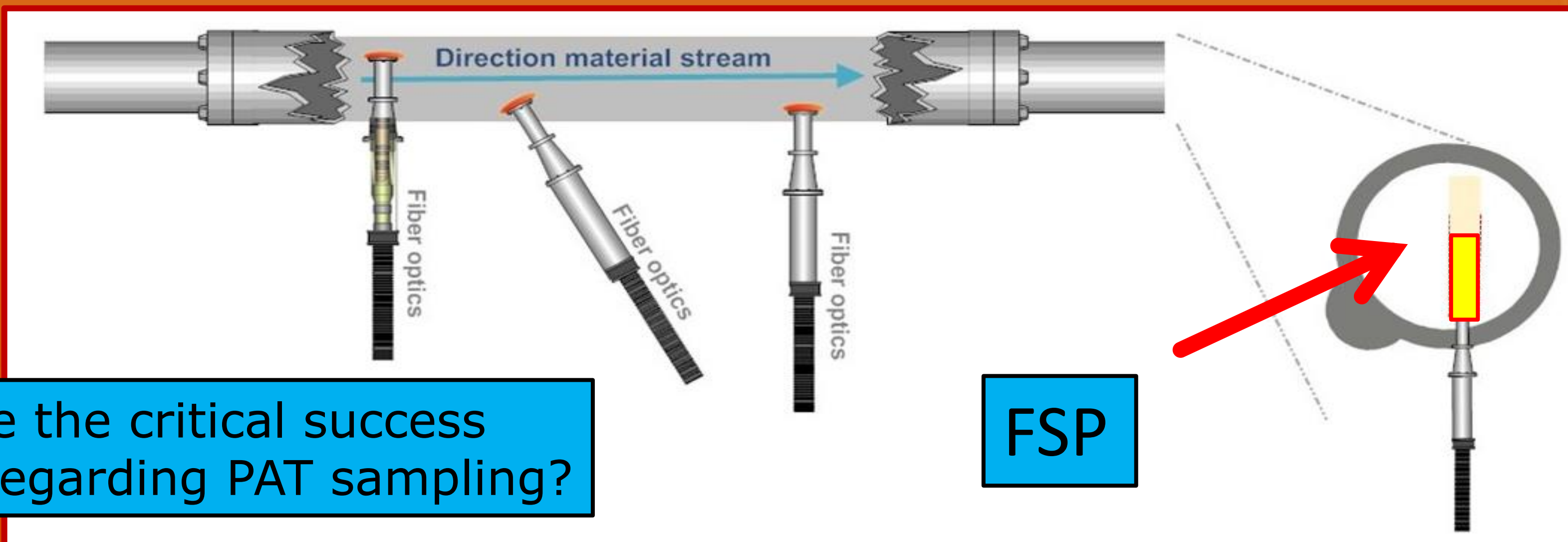
FSP

Fundamental Sampling Principle (FSP):

All increments must have the *same* (non-zero) probability of ending up in the sample

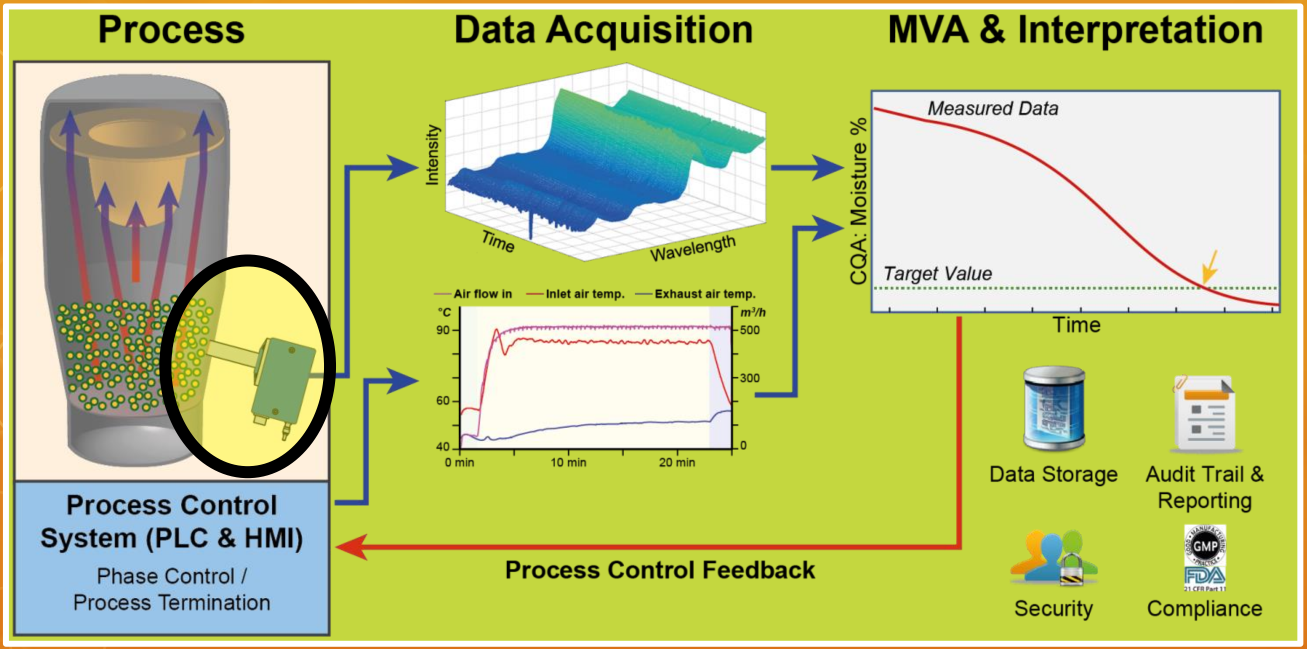


Sneak-preview of critical *issues* regarding 1-D process sampling



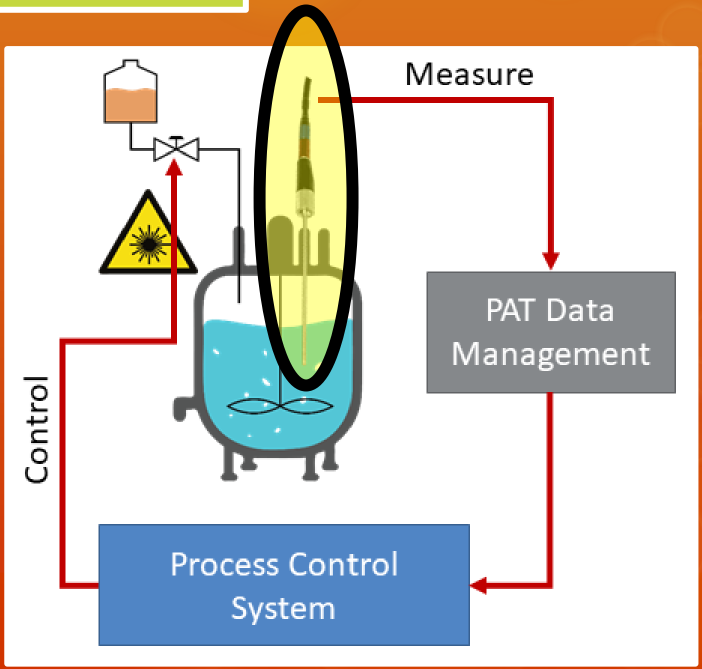
FSP

What are the critical success factors regarding PAT sampling?

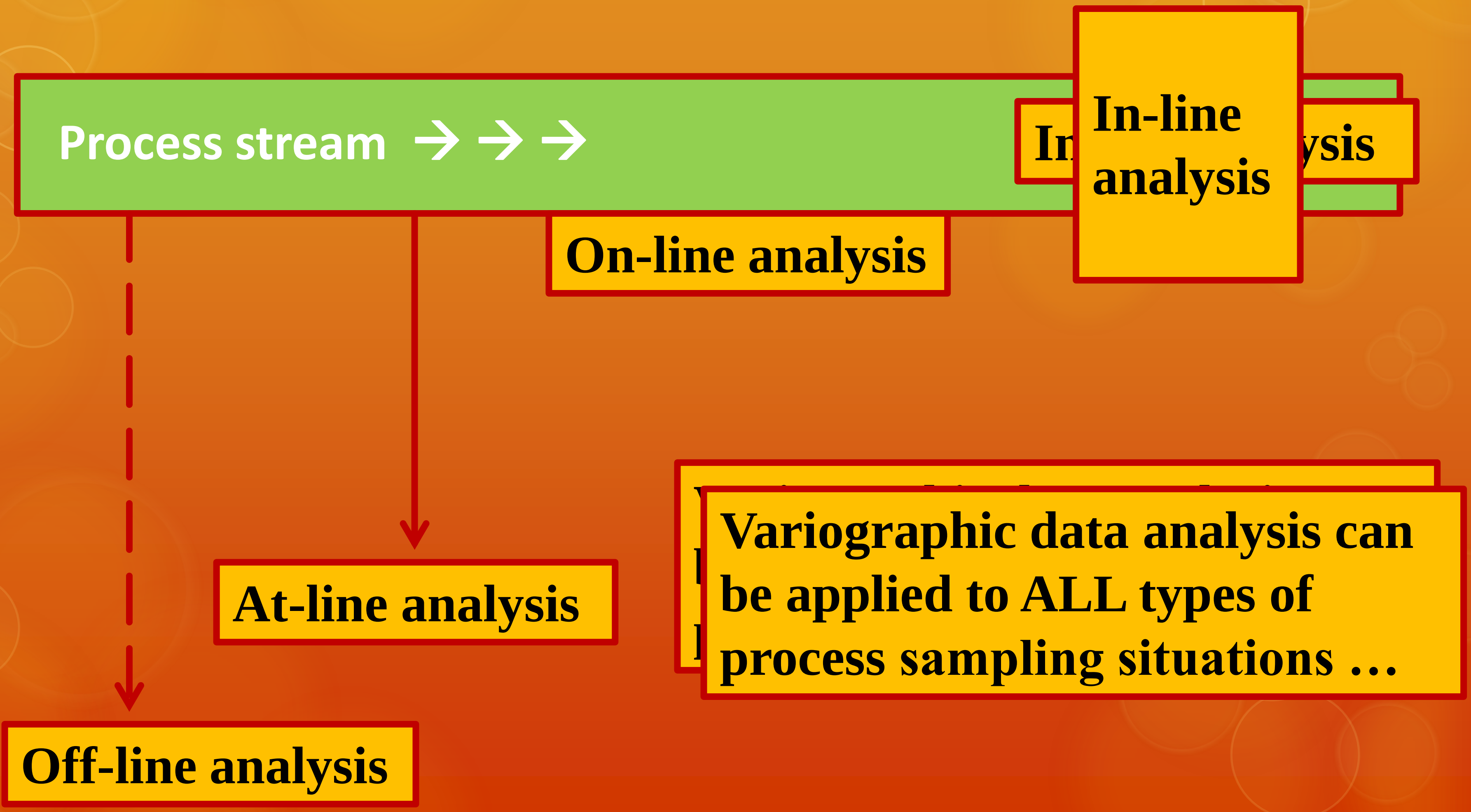


Examples of PAT integration incl. PAT Data Management

Primary sampling: spectral acquisition – where data quality is established

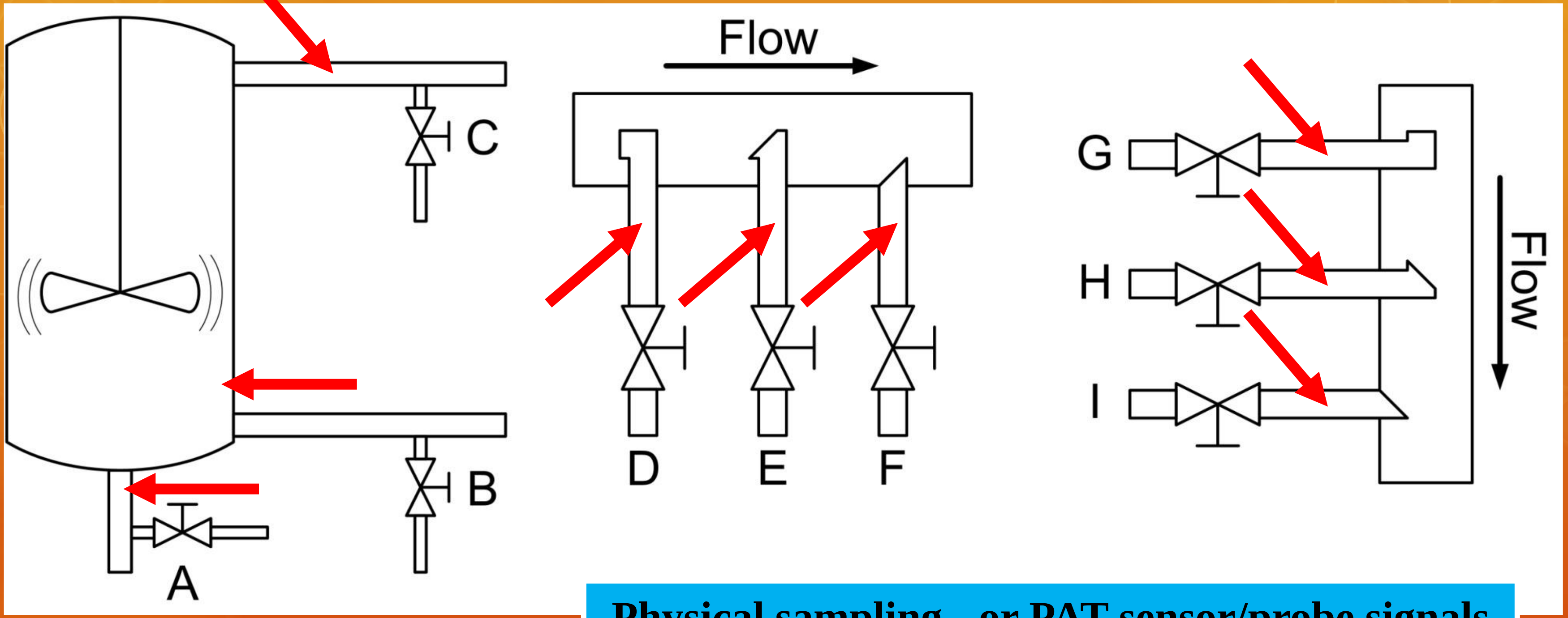


”Process sampling – generic overview”



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

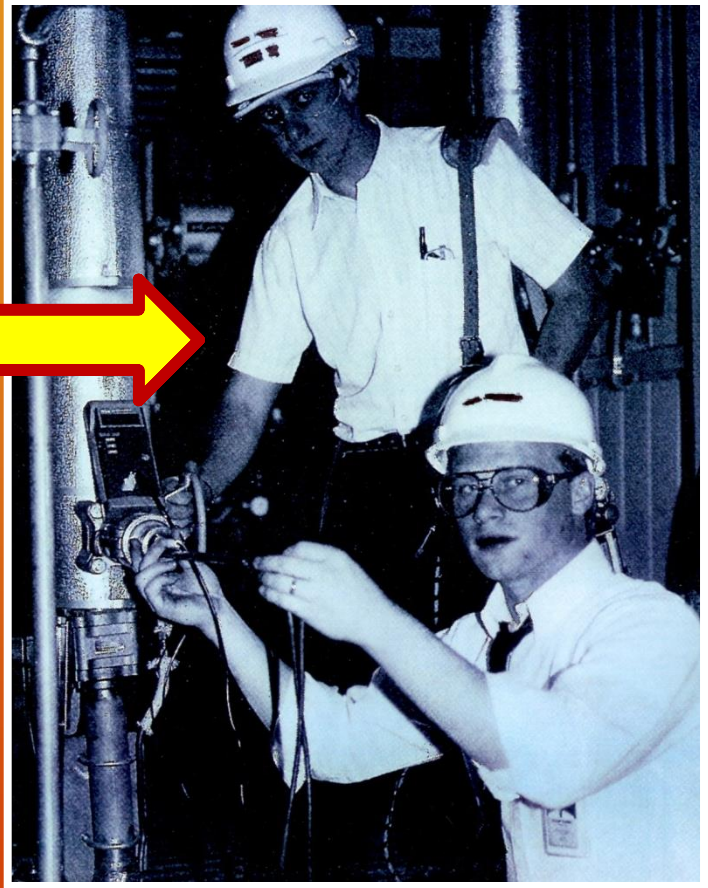
Plant sampling



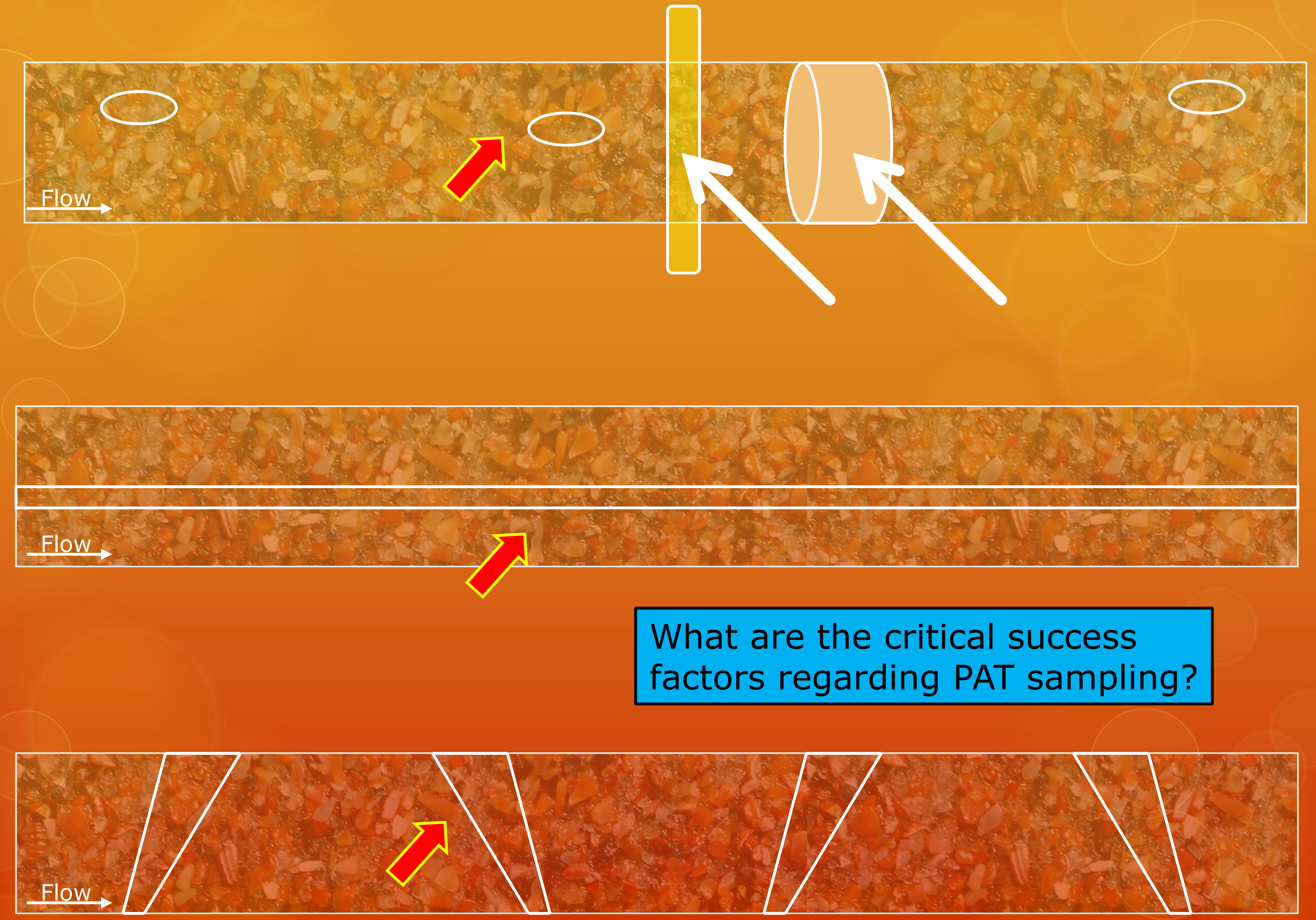
Physical sampling - <u>or</u> PAT sensor/probe signals

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

TOS agenda: complete DUALITY

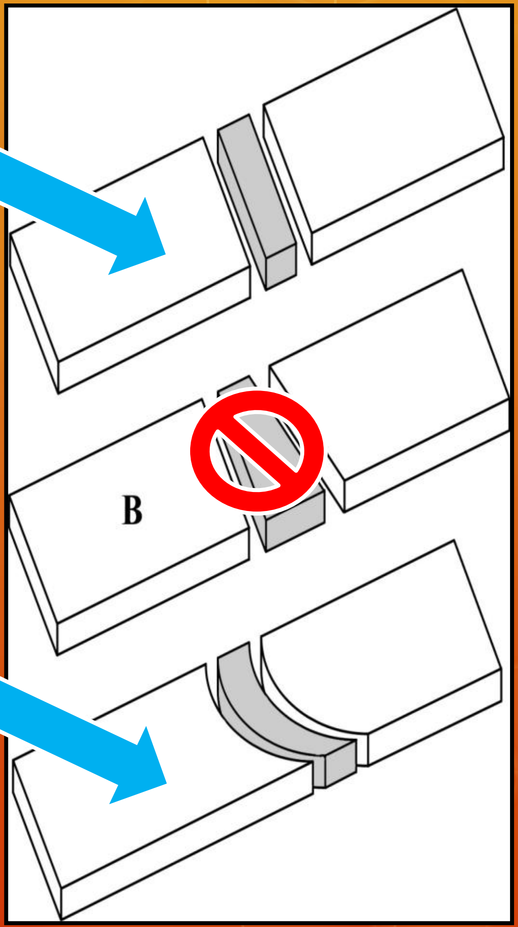


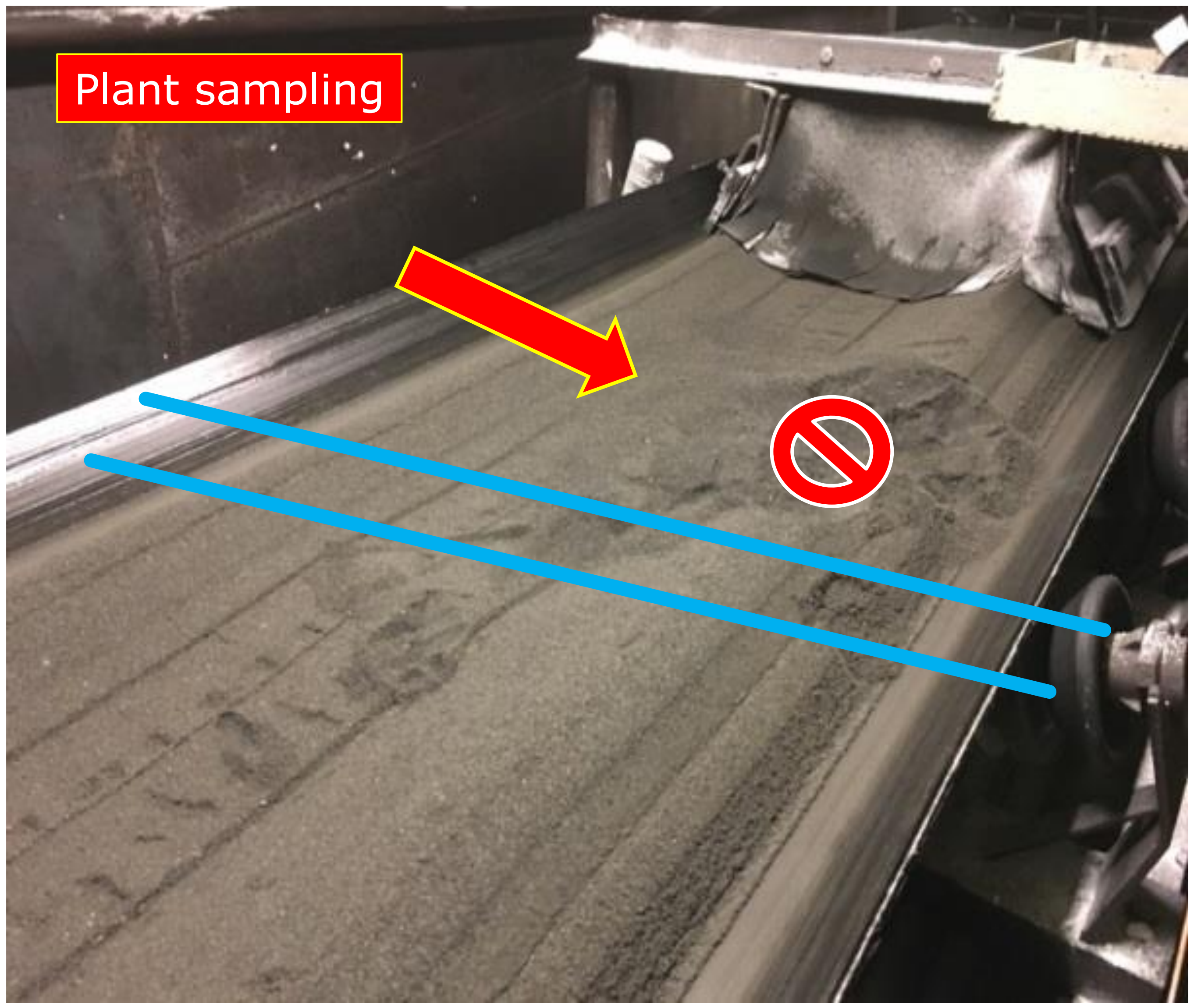
Identical ”sampling issues” with or without sensor technologies (PAT) ..

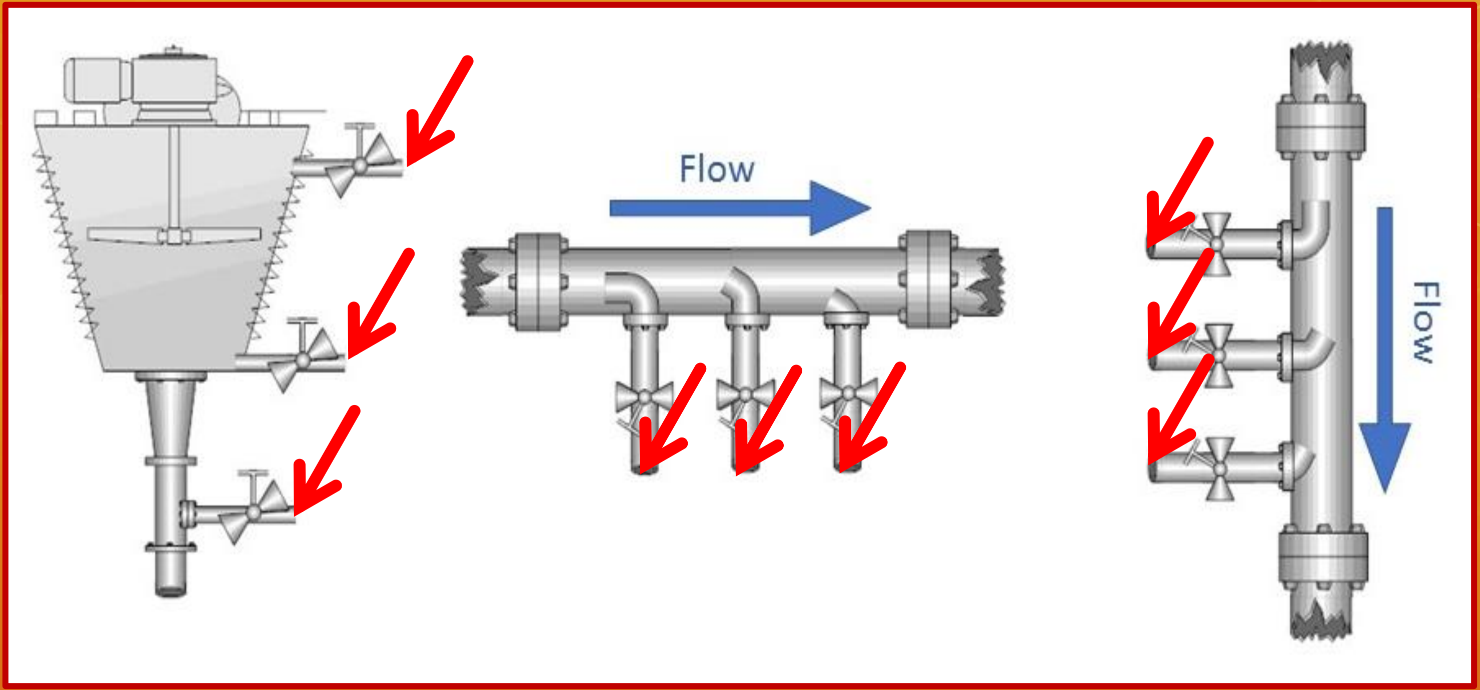


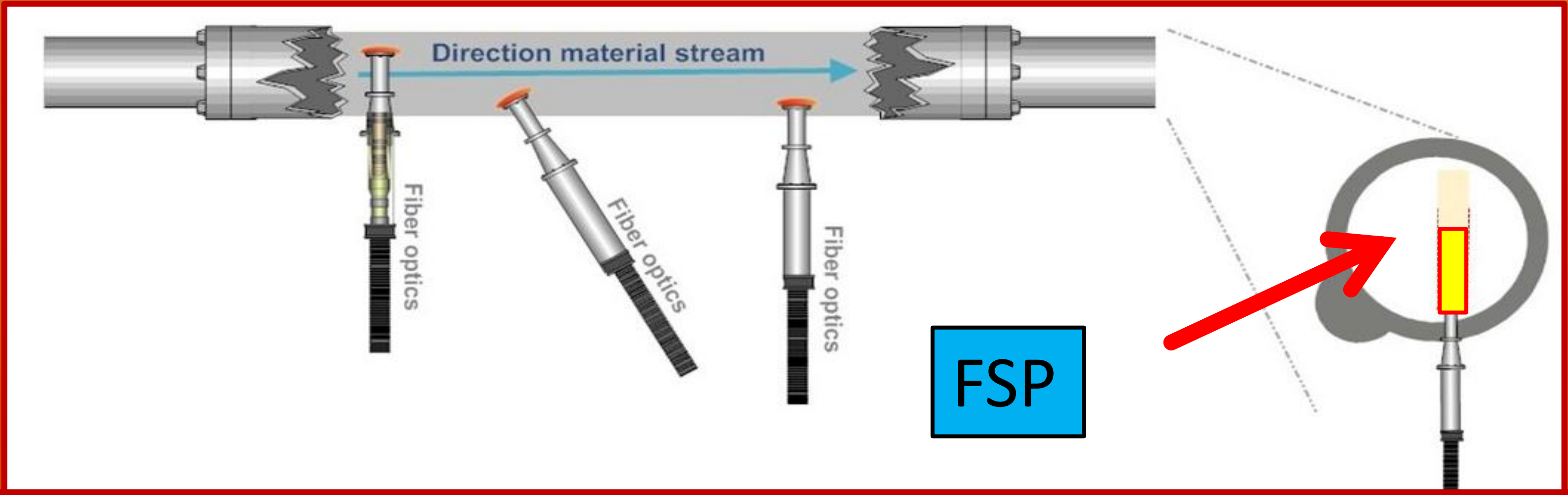
What are the critical success factors regarding PAT sampling?

Plant sampling




Plant sampling



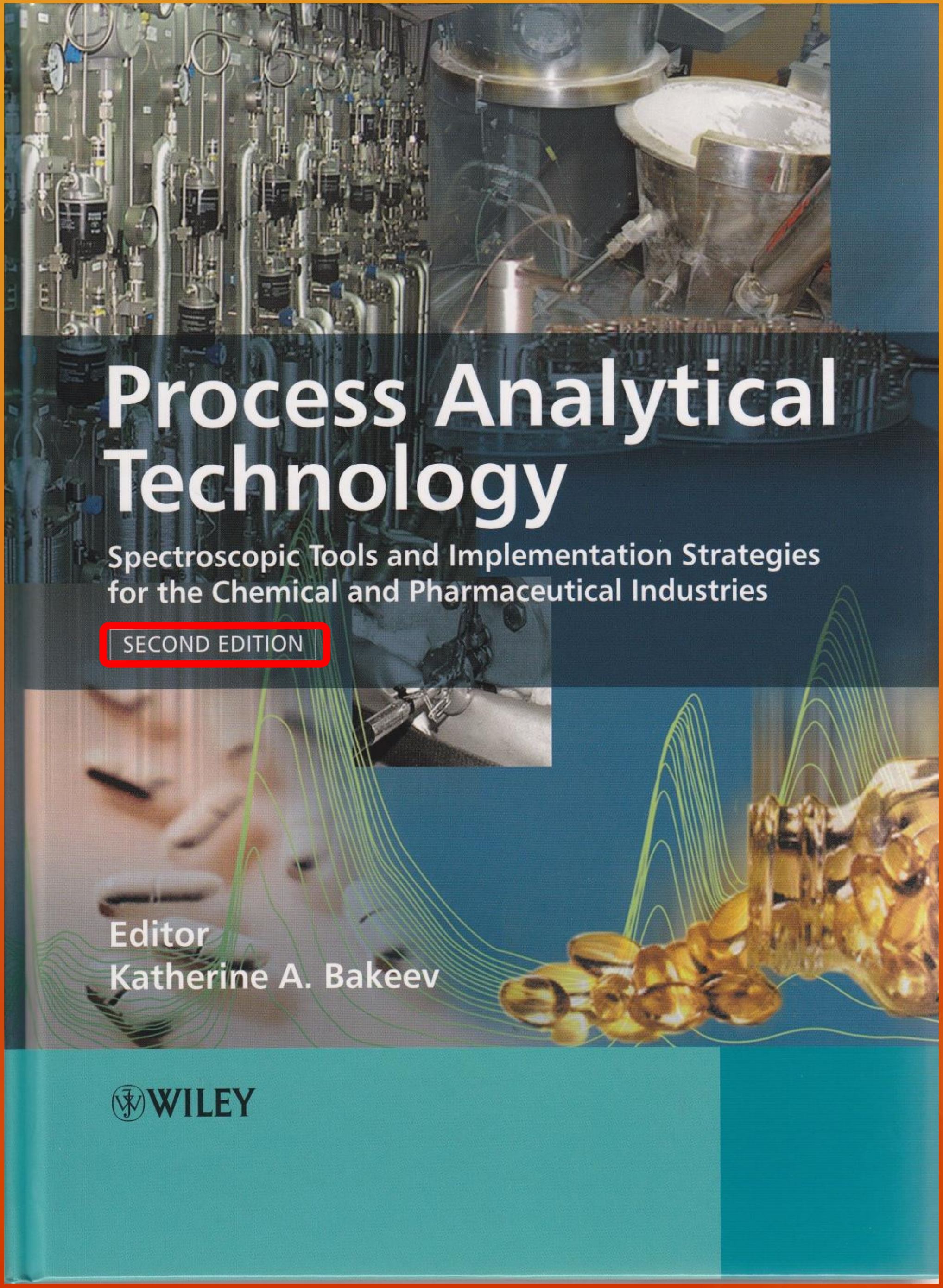


This sensor is principally "*incorrect*" – sampling process is *non-representative* !

“PROCESS SAMPLING (TOS) – the missing link in PAT”

Kim H. Esbensen & P. Paasch-Mortensen

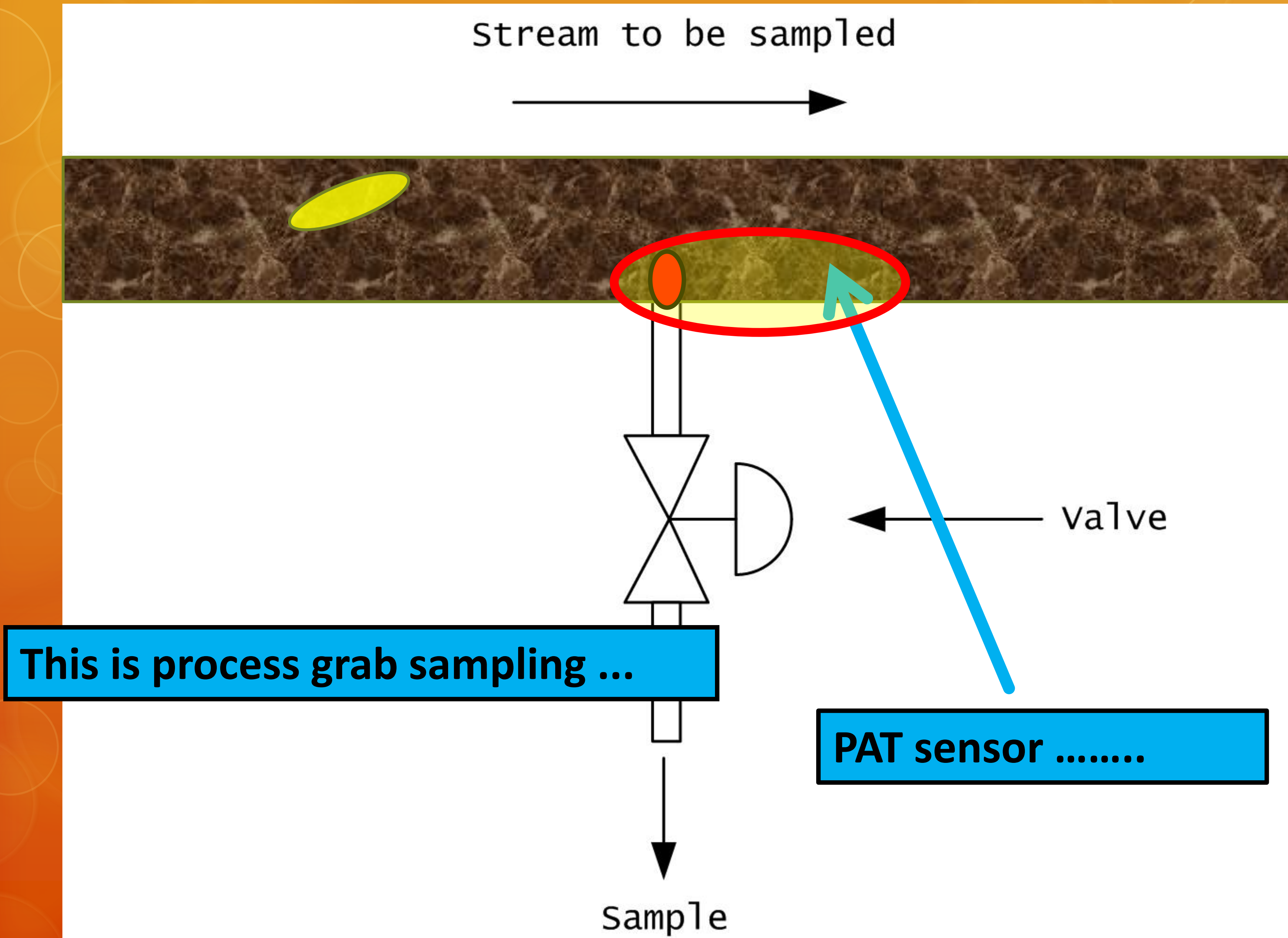
Bakeev, K. (Ed.)
PROCESS ANALYTICAL TECHNOLOGY, 2.nd Ed.(*2009*)
(chapter 3)


Process Analytical Technology
Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries
SECOND EDITION
Editor
Katherine A. Bakeev
WILEY

Sampling - looking for *unifying* principles & actions

PAT priorities (incl. TOS):

1. Representativity <u>rules</u>!
2. Equipment (<u>TOS correct</u>)
3. Data acquisition ...
4. Data management ...
5. Multivariate calibration
6. Proper prediction model validation



Stream to be sampled
Valve
This is process grab sampling ...
PAT sensor ....
Sample

KHE Consulting

Theory and Practice of Representative Sampling

Basic principles and unit operations – an overview

Theory of Sampling (TOS)

This is a representative sample ;-)

What are the criteria for representativity?

TOS

”SAMPLING – should not be gambling!”

Attributed to: **Pierre Gy**, founder of Theory of Sampling (TOS)

”TOS-sampling is not relevant for my:

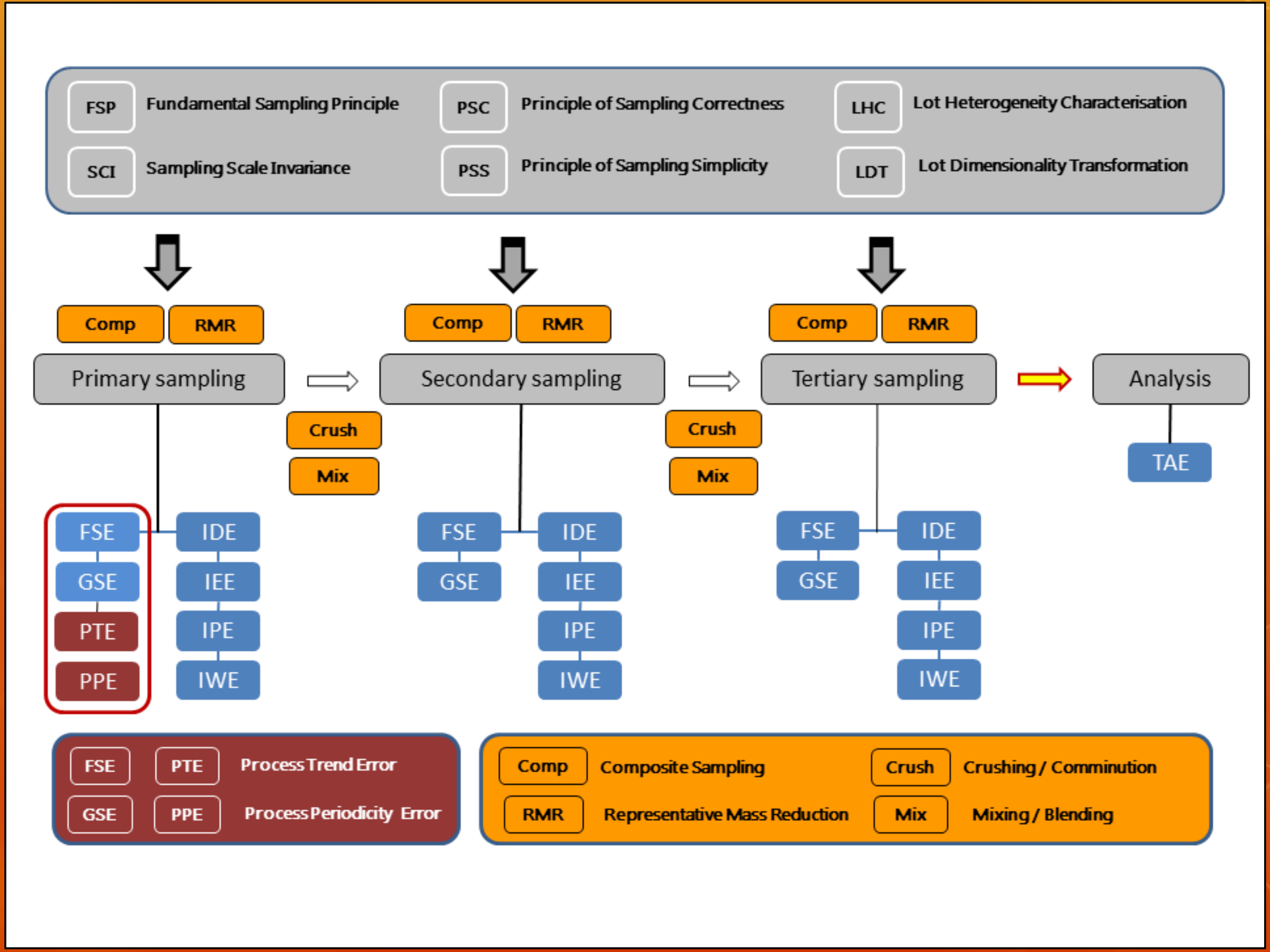
- type of material ...

- type of samples ...

- type of data ... ... ”

???

Theory of Sampling (TOS) – everything in a glance



A most important insight … where all sampling starts

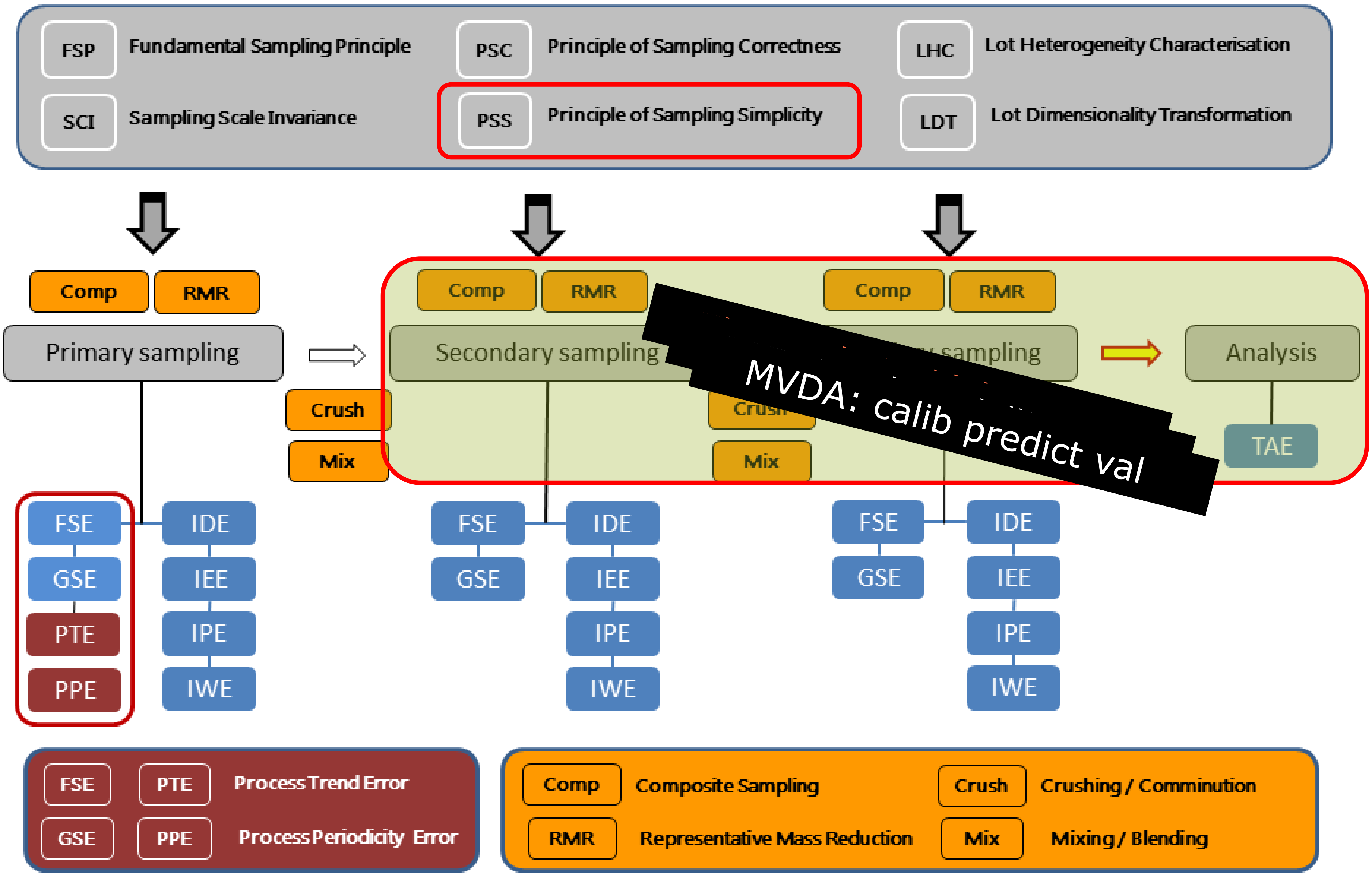
1. Sampling is <u>never</u> a one-shot operation

2. Sampling is <u>always</u> a multi-stage process:

   i) Primary sampling +

   ii) Representative Mass-reduction

Princple of Sampling Simplicity (PSI)


FSP
Fundamental Sampling Principle
PSC
Principle of Sampling Correctness
LHC
Lot Heterogeneity Characterisation
SCI
Sampling Scale Invariance
PSS
Principle of Sampling Simplicity
LDT
Lot Dimensionality Transformation
Comp
RMR
Primary sampling
Crush
Mix
Comp
RMR
Secondary sampling
Comp
RMR
Mix
MVDA: calib predict val
Analysis
TAE
FSE
GSE
PTE
PPE
IDE
IEE
IPE
IWE
FSE
GSE
IDE
IEE
IPE
IWE
FSE
GSE
IDE
IEE
IPE
IWE
FSE
PTE
Process Trend Error
GSE
PPE
Process Periodicity Error
Comp
Composite Sampling
Crush
Crushing / Comminution
RMR
Representative Mass Reduction
Mix
Mixing / Blending

Representative Sampling: Theory of Sampling (TOS)

TOS - Axiomatic exposé

Governing Principles (GP) – Sampling Unit Operations (SUO)

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. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

KHE Consulting

Theory and Practice of Representative Sampling TOS

Material heterogeneity – where it all begins ...

and where it may sometimes end as well (fatally)

Let's get started - HOW to sample?

- How to sample representatively?

- How to DEFINE representativity?

- Sa... vs. bulk materials handling ...

- HOW BIG sh... ample be
  (in order to be repr... ive?)

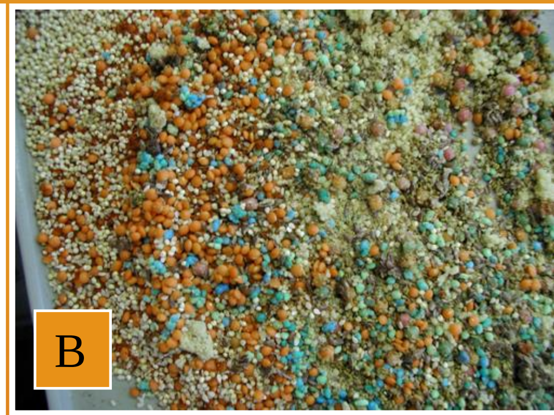
- How to sub-sample (mass redu...
  (in the field? – in the laboratory?)

- How to ... ...

- WHY (all of the above)?

TOS

... It all starts with HETEROGENITY, and how to counteract its effects ...

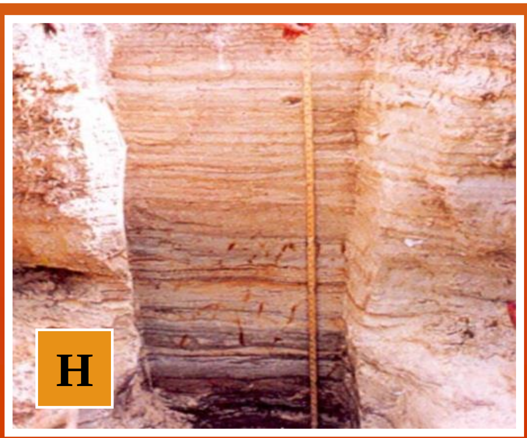
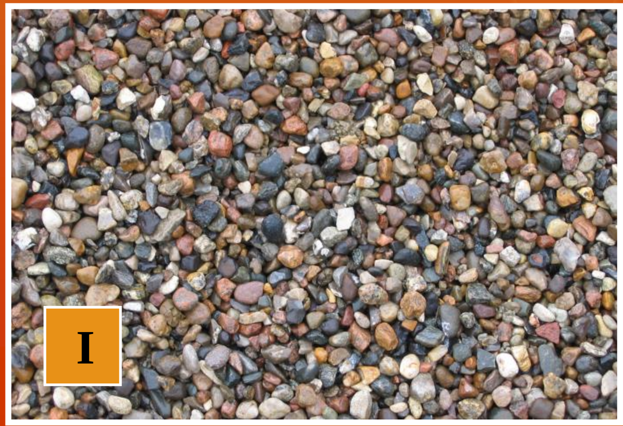

A

B

C

D

E

F

G

H

I

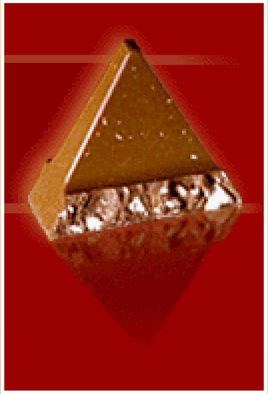
Compositional hetergeneity – CH (CH_{Lot})

Distrubutional hetergeneity – DH (DH_{Lot})

Grain-size hetergeneity → DH (DH_{Lot})

Large diversity of foods/feed products – HETEROGENEITY rules !!!

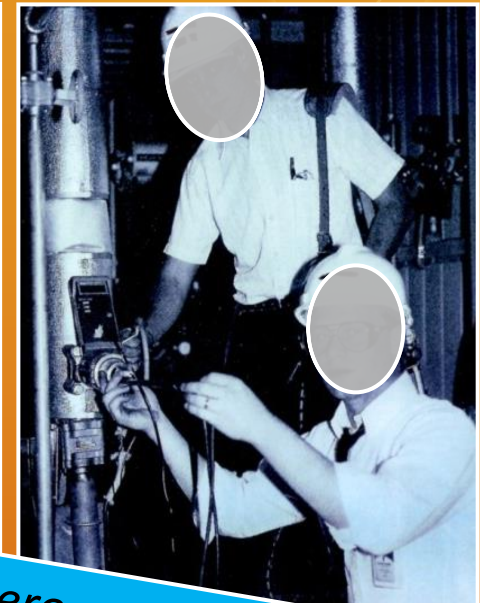
How to ensure representative primary sampling for <u>all</u> types of heterogeneous materials?



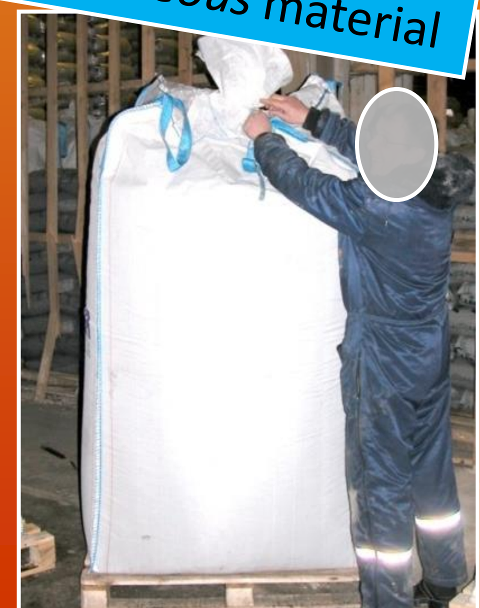
Remember him? Grab sampling!

Well. He is not alone out there in the field!

Grab sampling galore ....



A grab sample can never be representative of a *heterogeneous* material



There is observable heterogeneity – and there is hidden heterogeneity!
This is equally bad w.r.t. the possibility of *proper* representative sampling

A grab sample can never be representative of a *heterogeneous* material

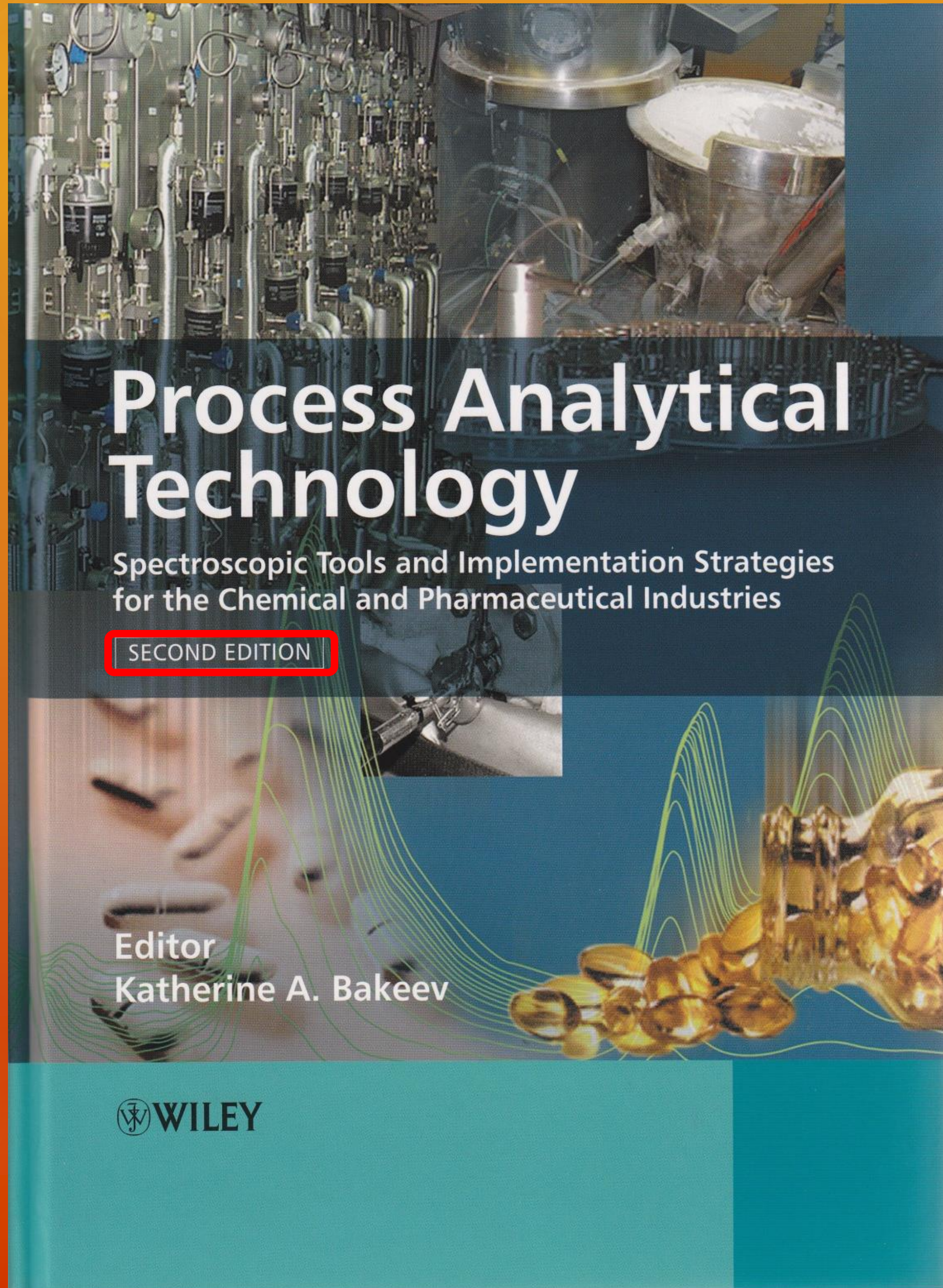
Materials are alwways heterogeneous (only a matter of to which degree)
It is necessary to sample all materials *as if* they were significantly het. !!


But there is a big surprise comming – a feature that will make sampling very, very much easier, despite there being so many very diffrent kinds of materials
34
Sampling is not really so much material dependent --
Sampling is much more up against lot *heterogeneity*

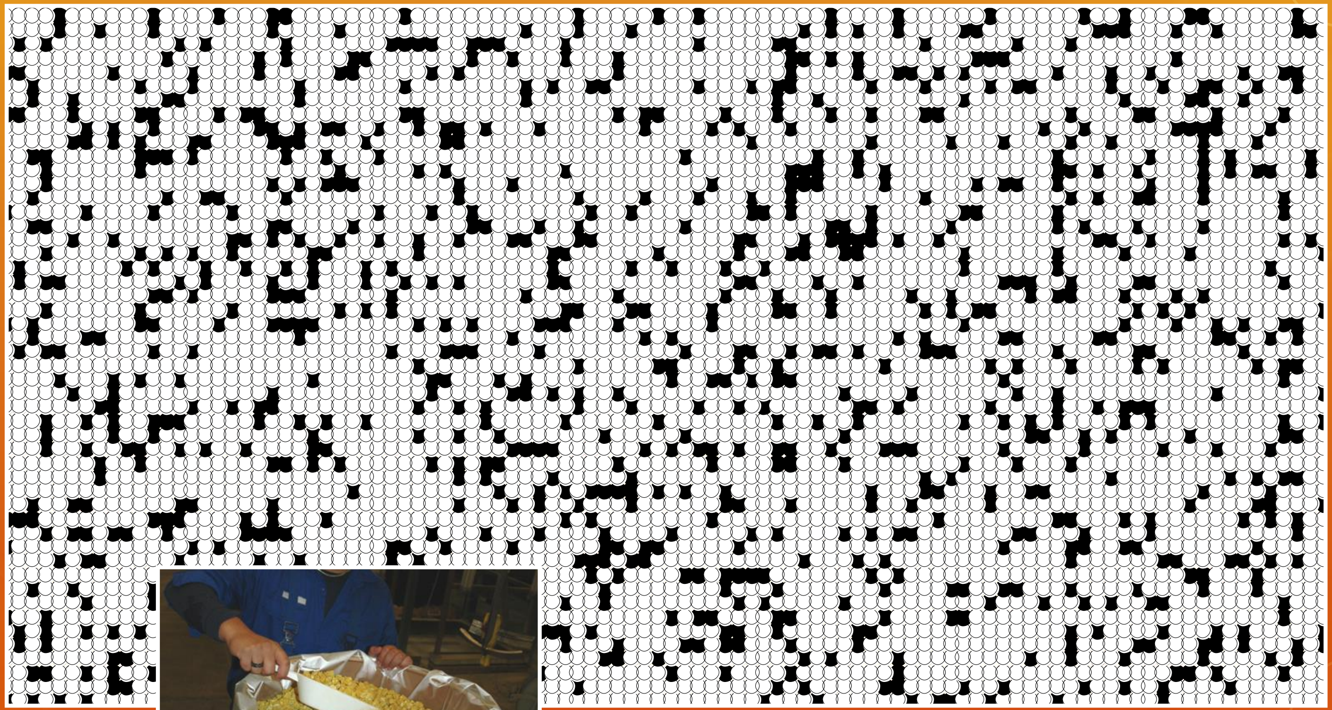
“PROCESS SAMPLING (TOS) – the missing link in PAT”

Kim H. Esbensen & P. Paasch-Mortensen

Bakeev, K. (Ed.)
PROCESS ANALYTICAL TECHNOLOGY, 2.nd Ed.(*2009*)
(chapter 3)


Process Analytical Technology
Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries
SECOND EDITION
Editor
Katherine A. Bakeev
WILEY

HETEROGENEITY – the arch enemy



Grab sampling – the deadly sin !!!

Heterogeneity - the unifying characteristic for all types of material

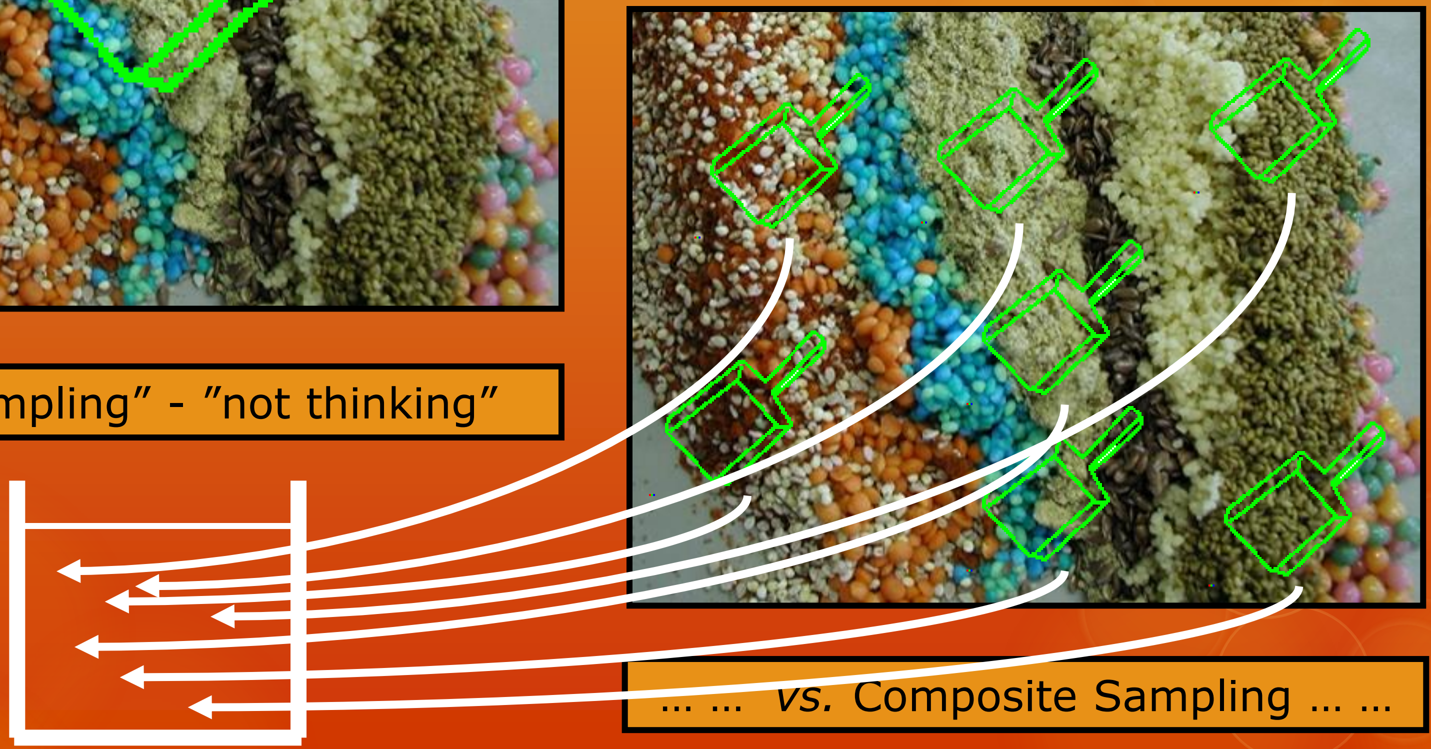
Grab sampling – its a deadly sin !!

TOS

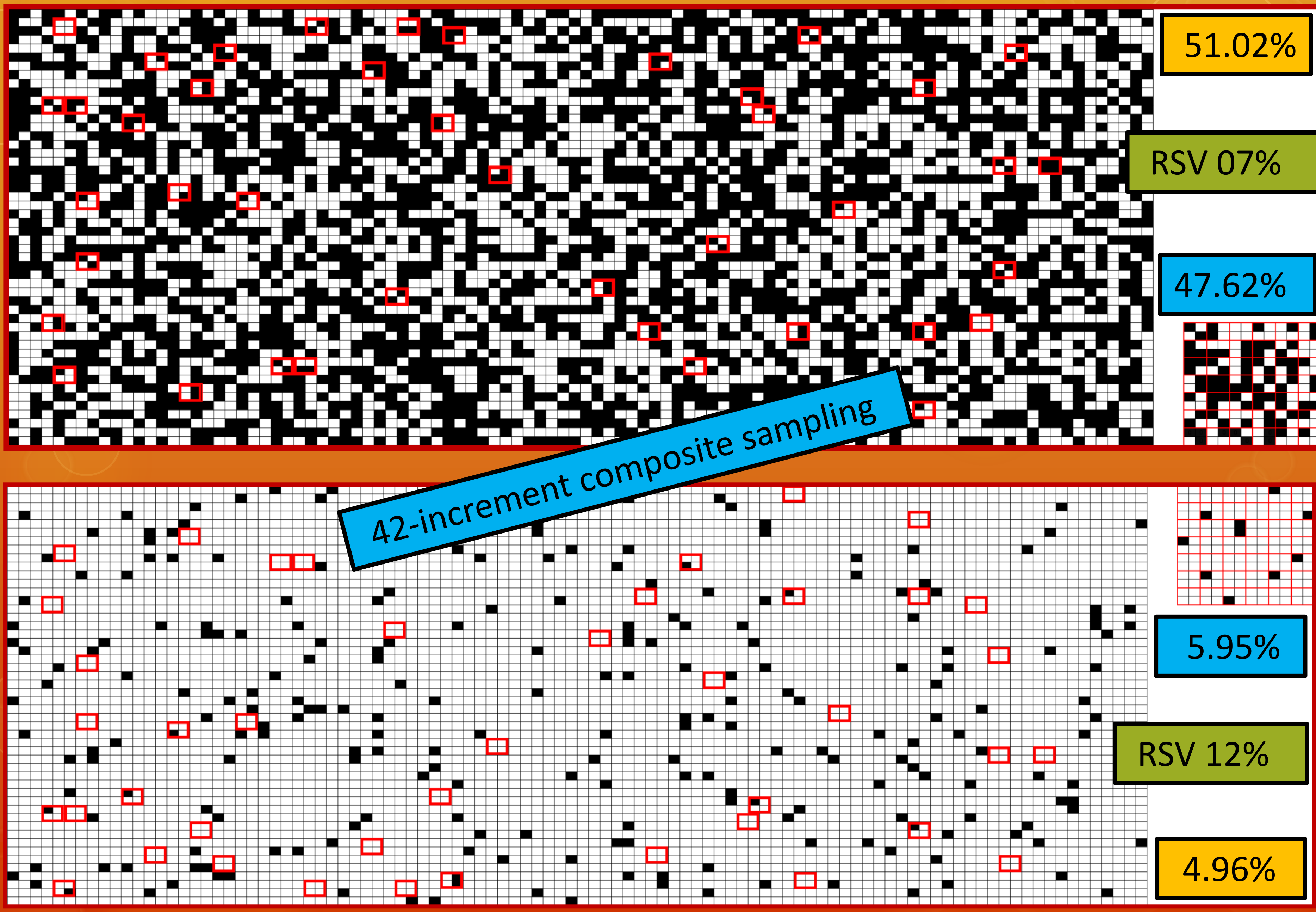
Sampling Unit Operations: Composite Sampling

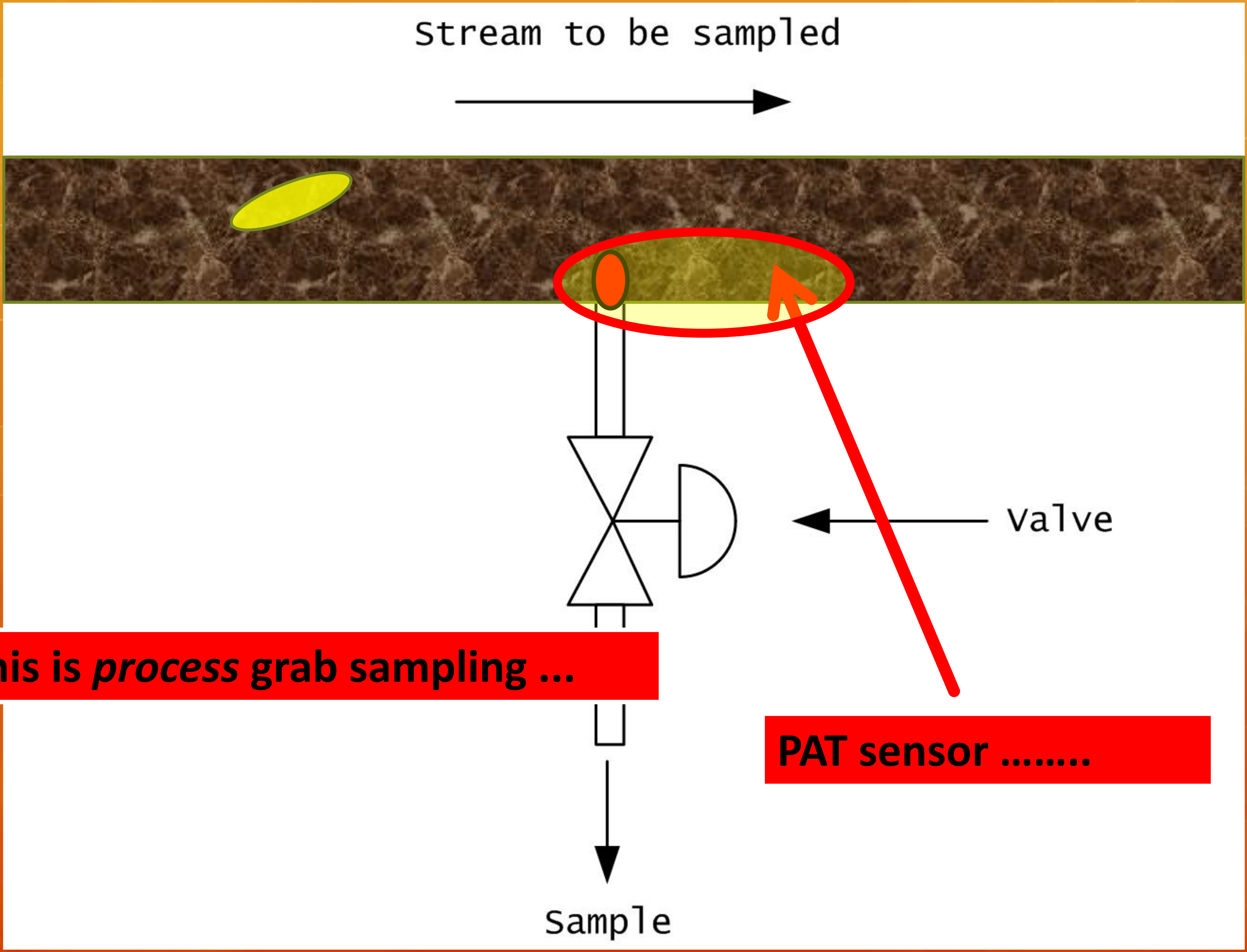
TOS

”Grab Sampling” - ”not thinking”



… … *vs.* Composite Sampling … …


51.02%
RSV 07%
47.62%
42-increment composite sampling
5.95%
RSV 12%
4.96%



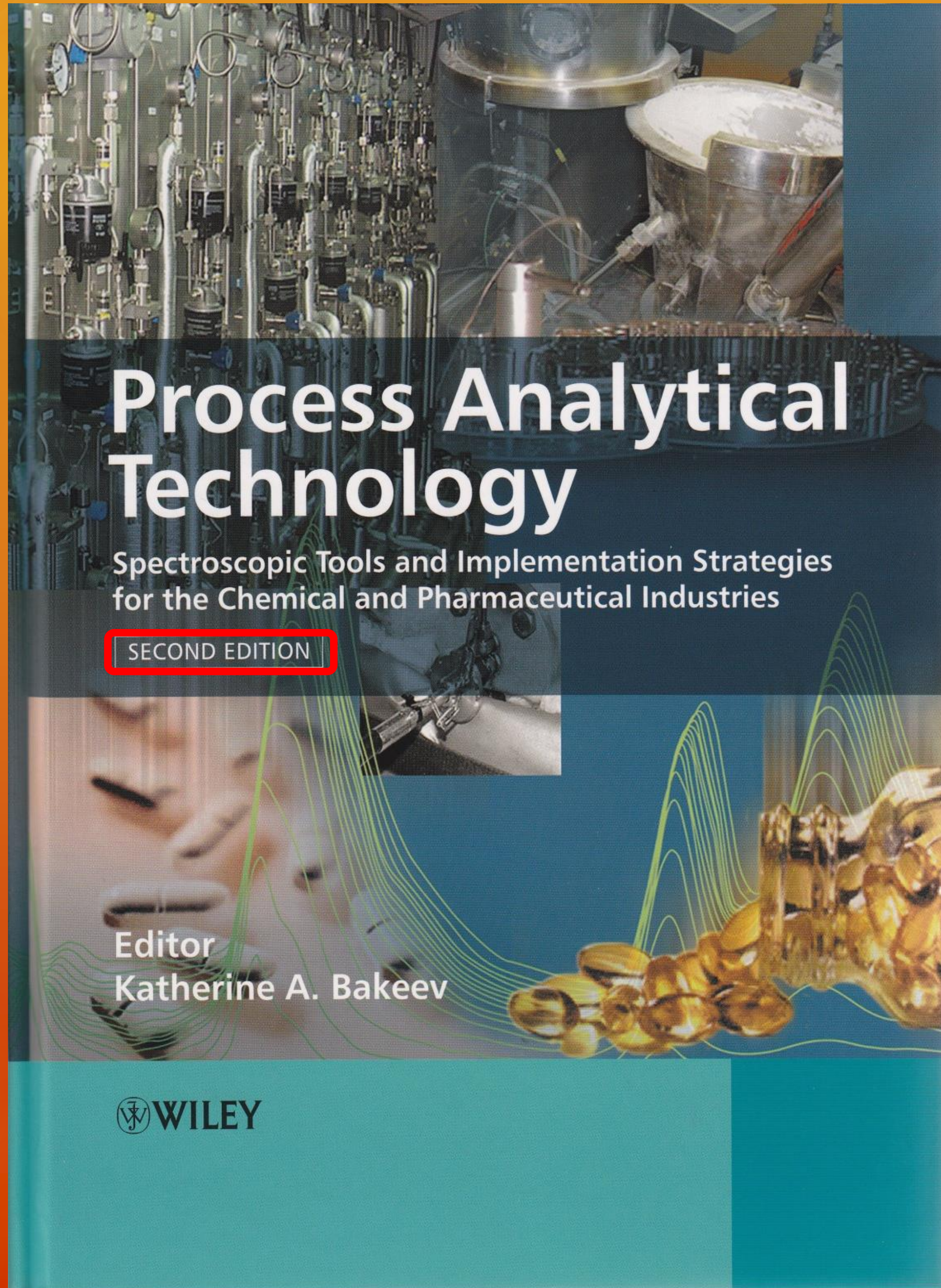
This is *process* grab sampling ...

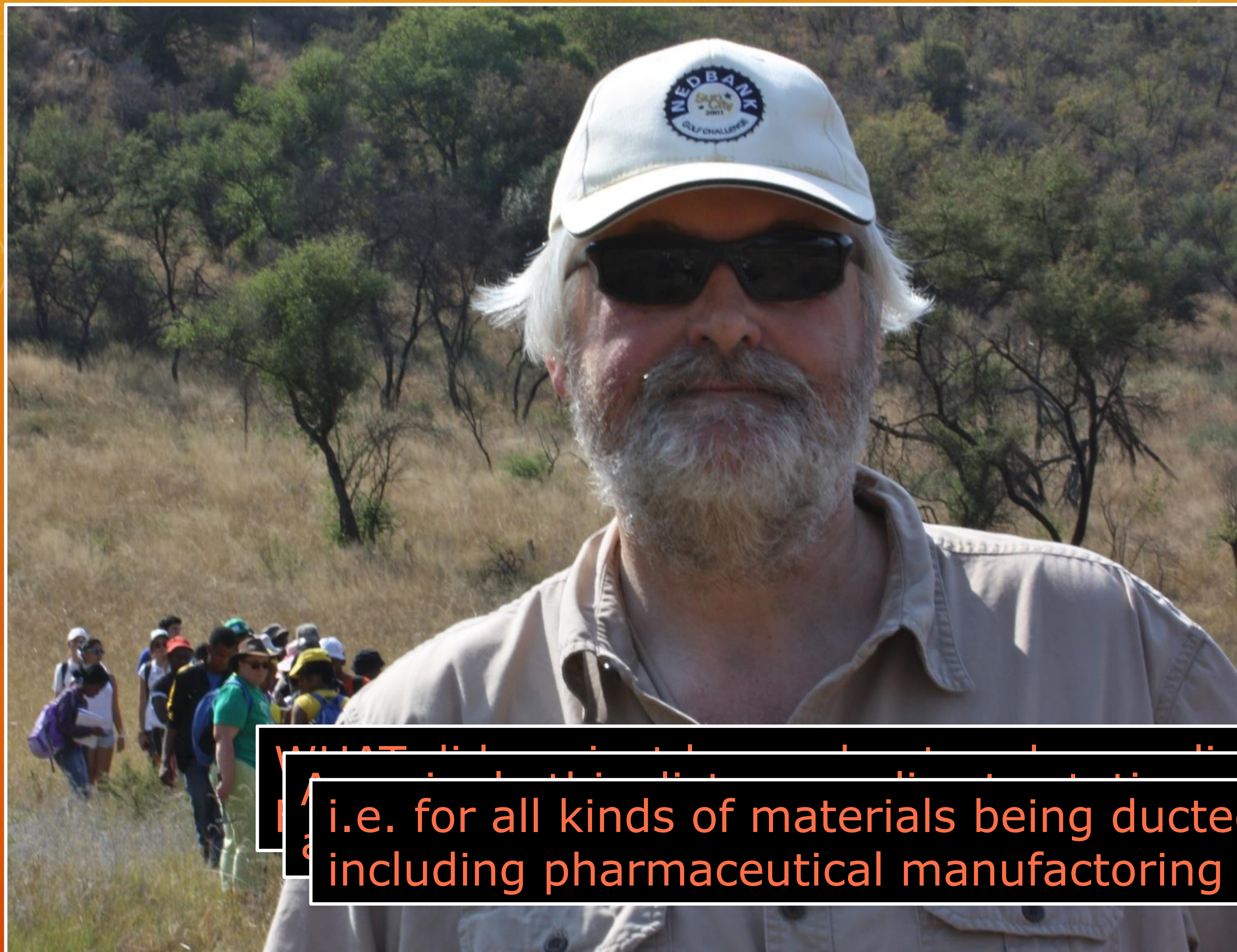
PAT sensor ....

“PROCESS SAMPLING (TOS) – the missing link in PAT”

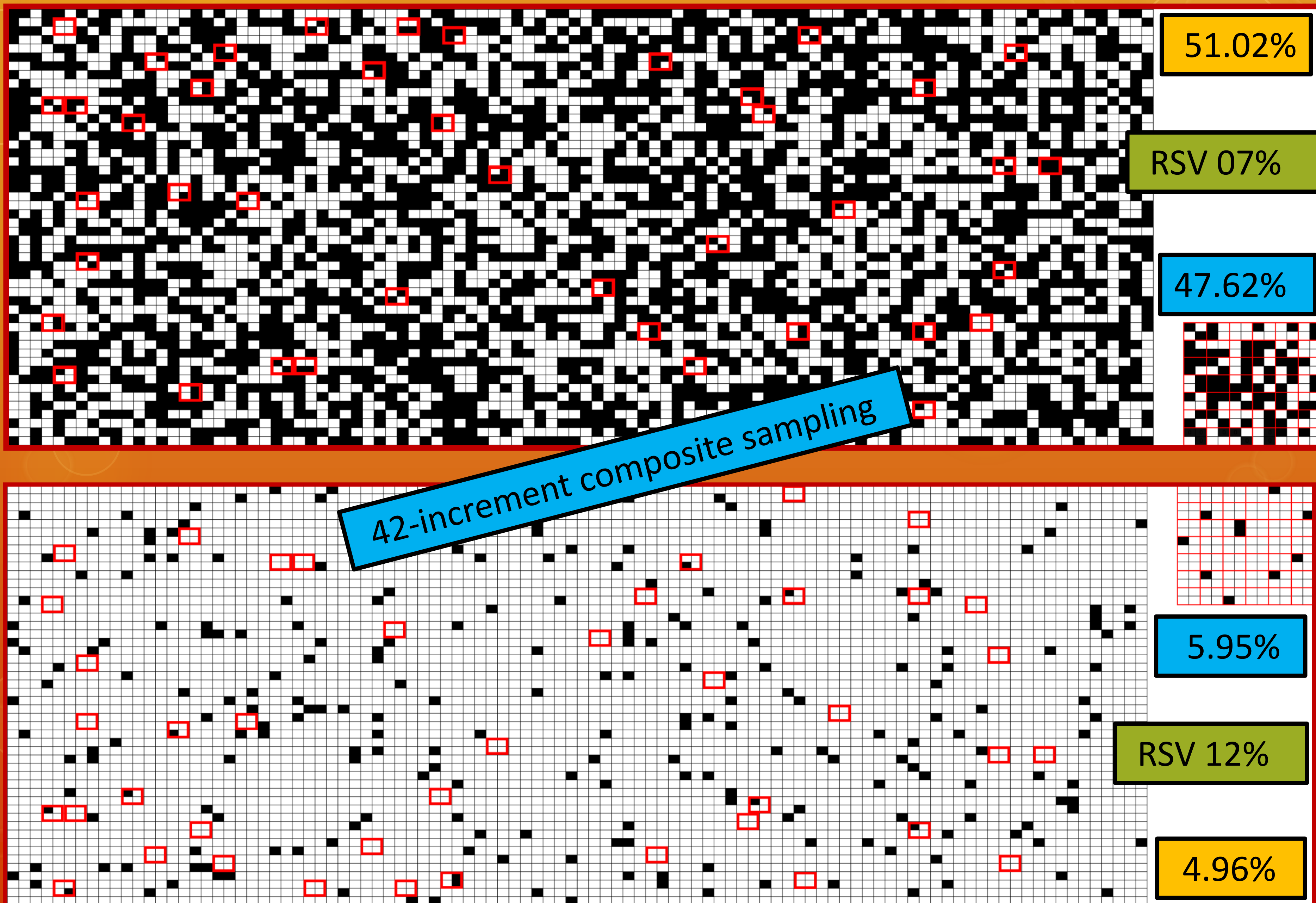
Kim H. Esbensen & P. Paasch-Mortensen

Bakeev, K. (Ed.)
PROCESS ANALYTICAL TECHNOLOGY, 2.nd Ed.(*2009*)
(chapter 3)

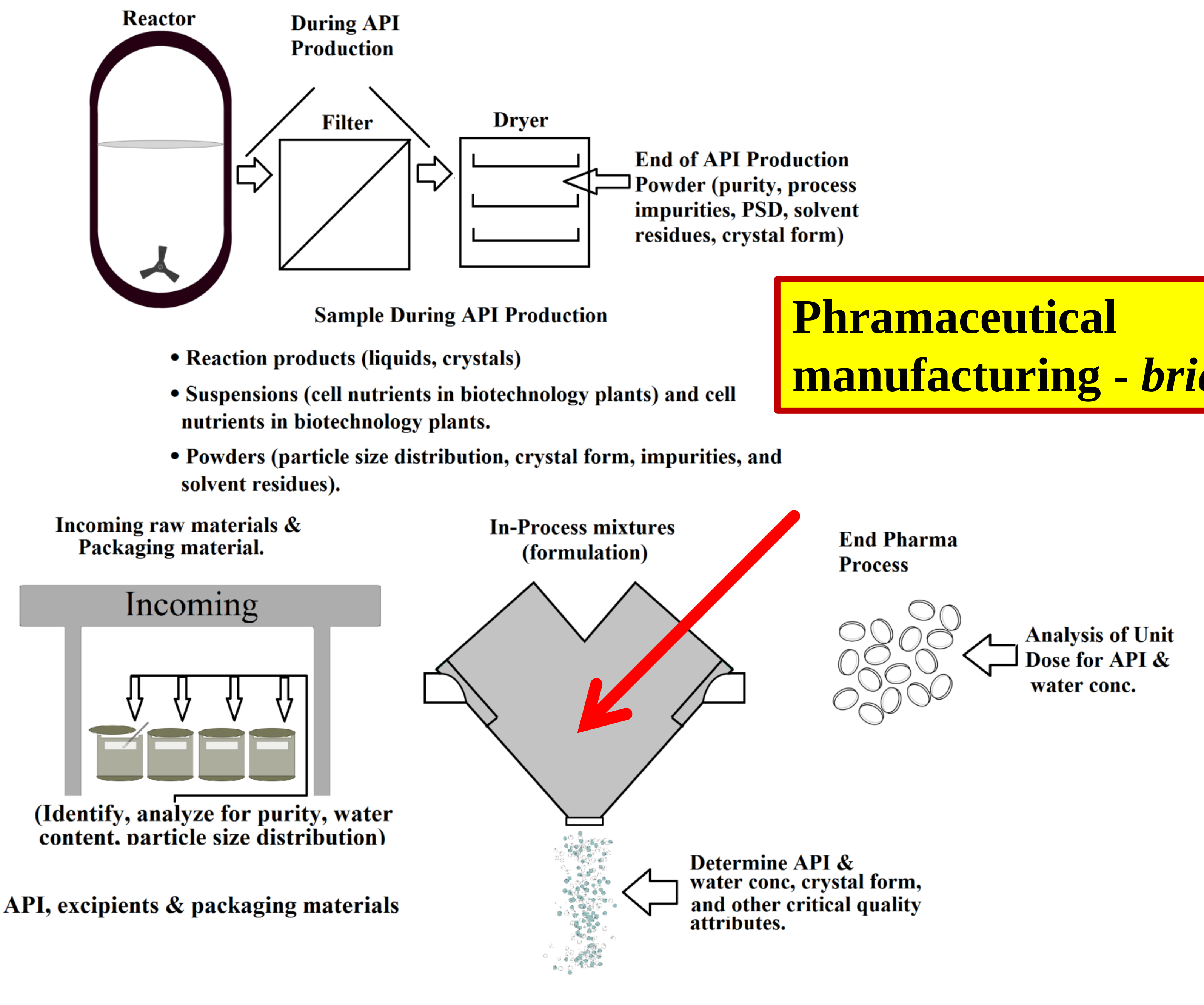

Process Analytical Technology
Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries
SECOND EDITION
Editor
Katherine A. Bakeev
WILEY

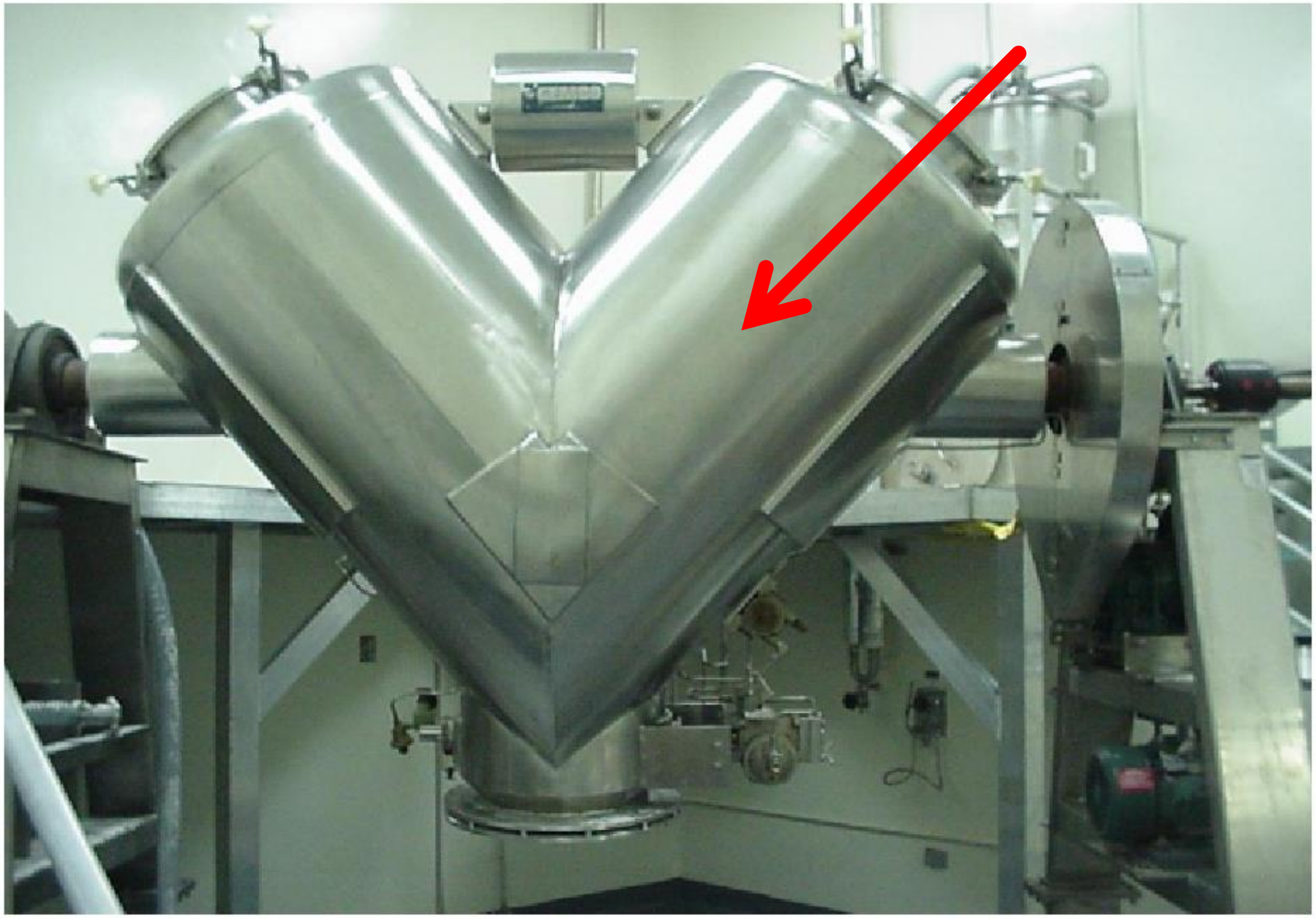


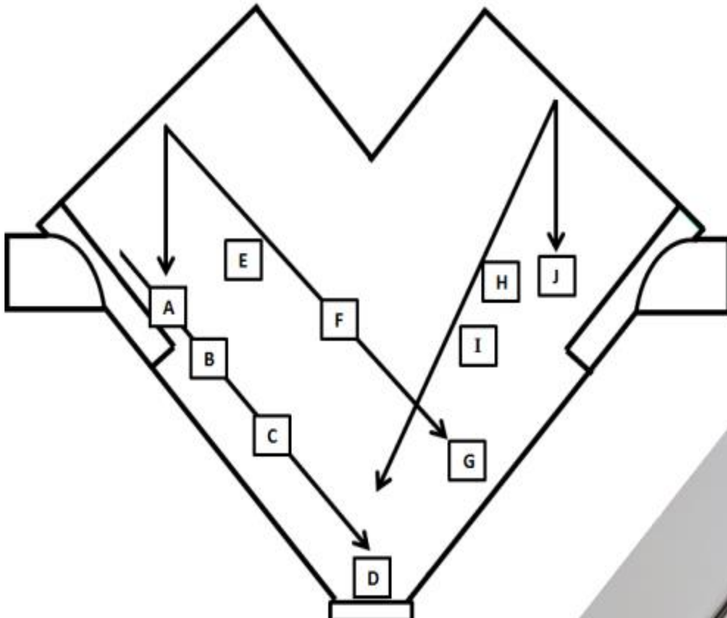
i.e. for all kinds of materials being ducted ... ...
including pharmaceutical manufactoring ... ...

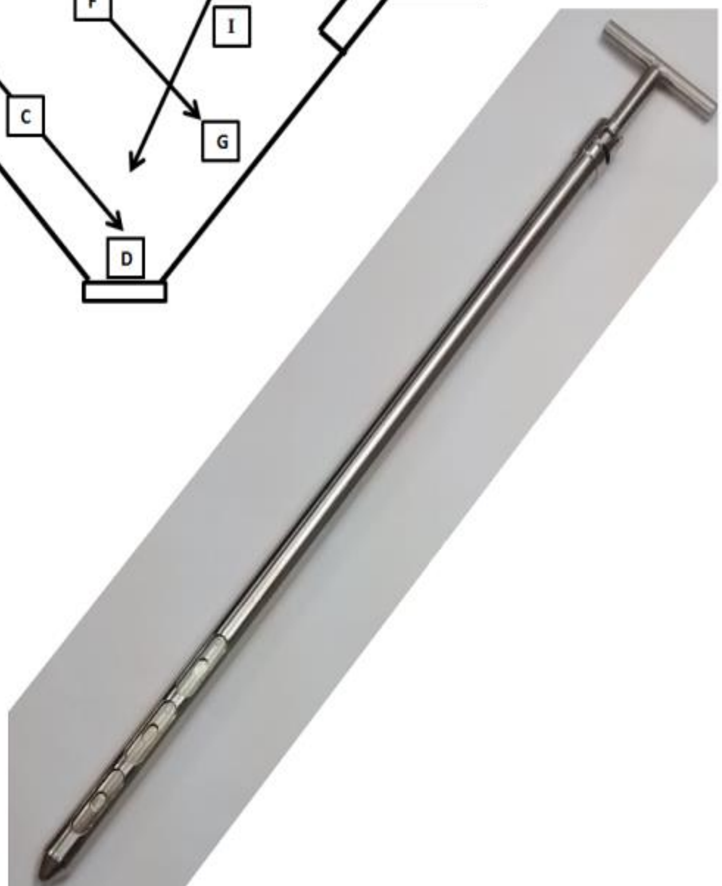

51.02%
RSV 07%
47.62%
42-increment composite sampling
5.95%
RSV 12%
4.96%

Phramaceutical manufacturing - *brief*









1st Sample Set
A

2nd Sample Set
B
C
D

3rd Sample Set
E
F
G

4th Sample set
H
I

5th Sample set
J

Traditional thief sampling (spear sampling) from within pharmaceutical mixing blenders (here a tumbling V-blender) recommends using 10 fixed locations according to a certain order intended to minimize ‘drag down’ of powder from higher locations.

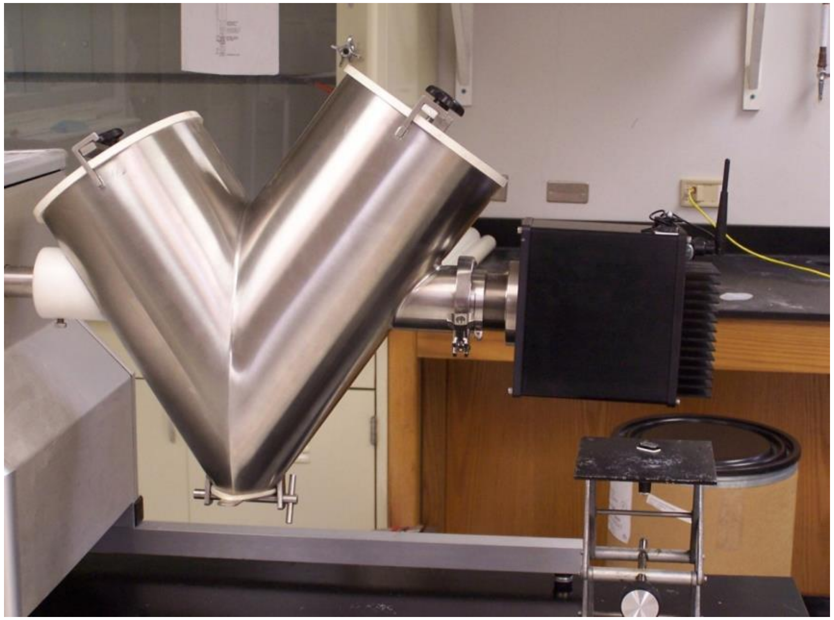
The fundamental assumption is that these locations (including replication at a few locations) represent the “most in-homogenous” parts of any blend, for all types of mixtures, in all types of blenders.

Alas this strong universal assumption is **not** tenable in the industrial practice.

.. In fact, pretty much <u>everything</u> is wrong with this approach:

• Enormous ISE – ditto sampling bias – totally out of control (spears)
• Even CSE are difficult to control: ”one sampling procedure fits all”
• Pharmaceutical mixtures are anything but a similar heterogeneity, or of similar mixing features …
• Confusion, massive confusion regarding heterogeneity vs. sampling


State-of-the-art approach …..
Camera, WIFI, …

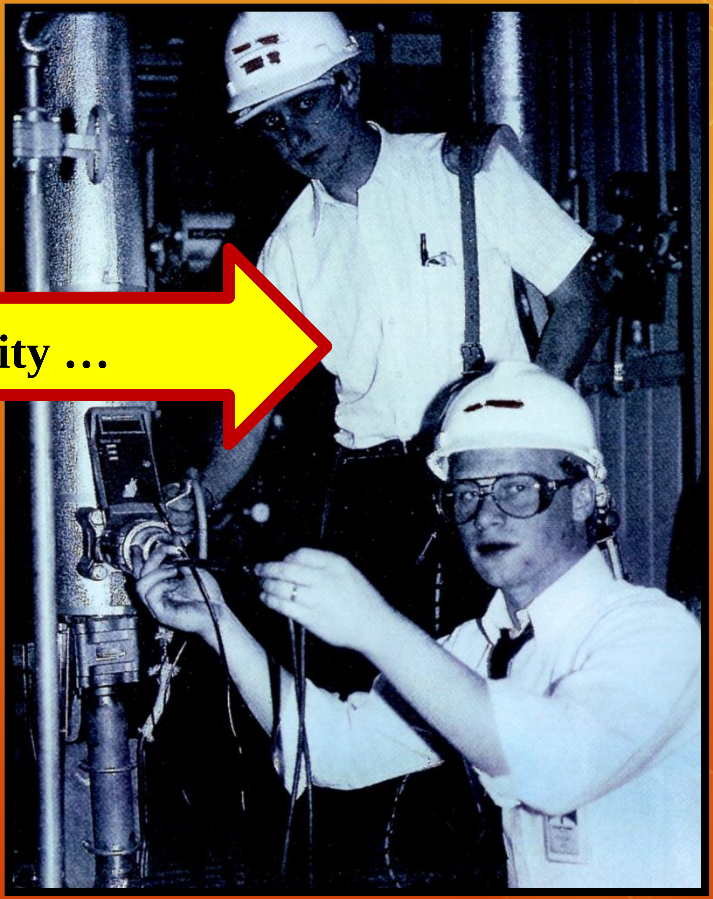


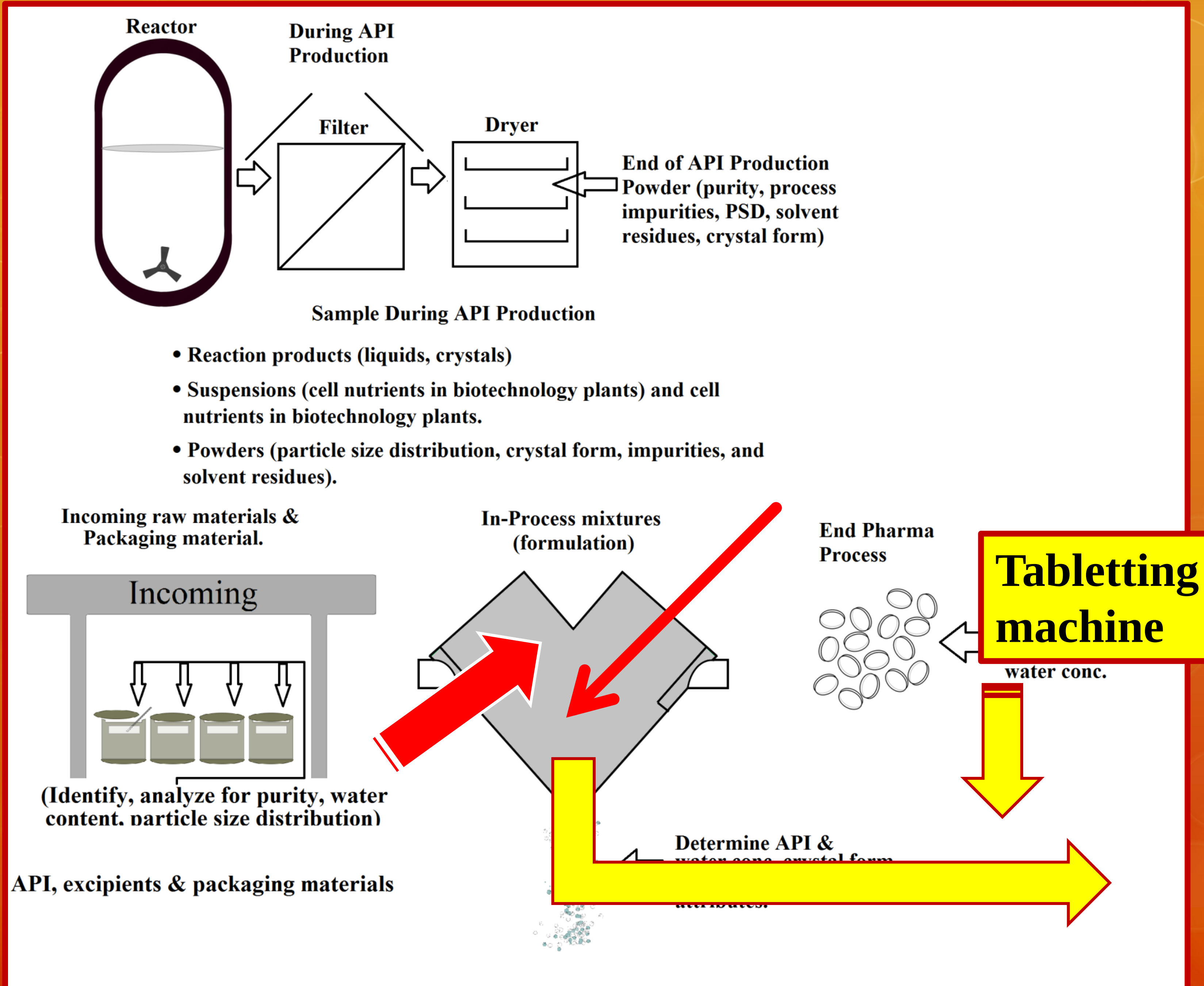
State-of-the-art approach .....

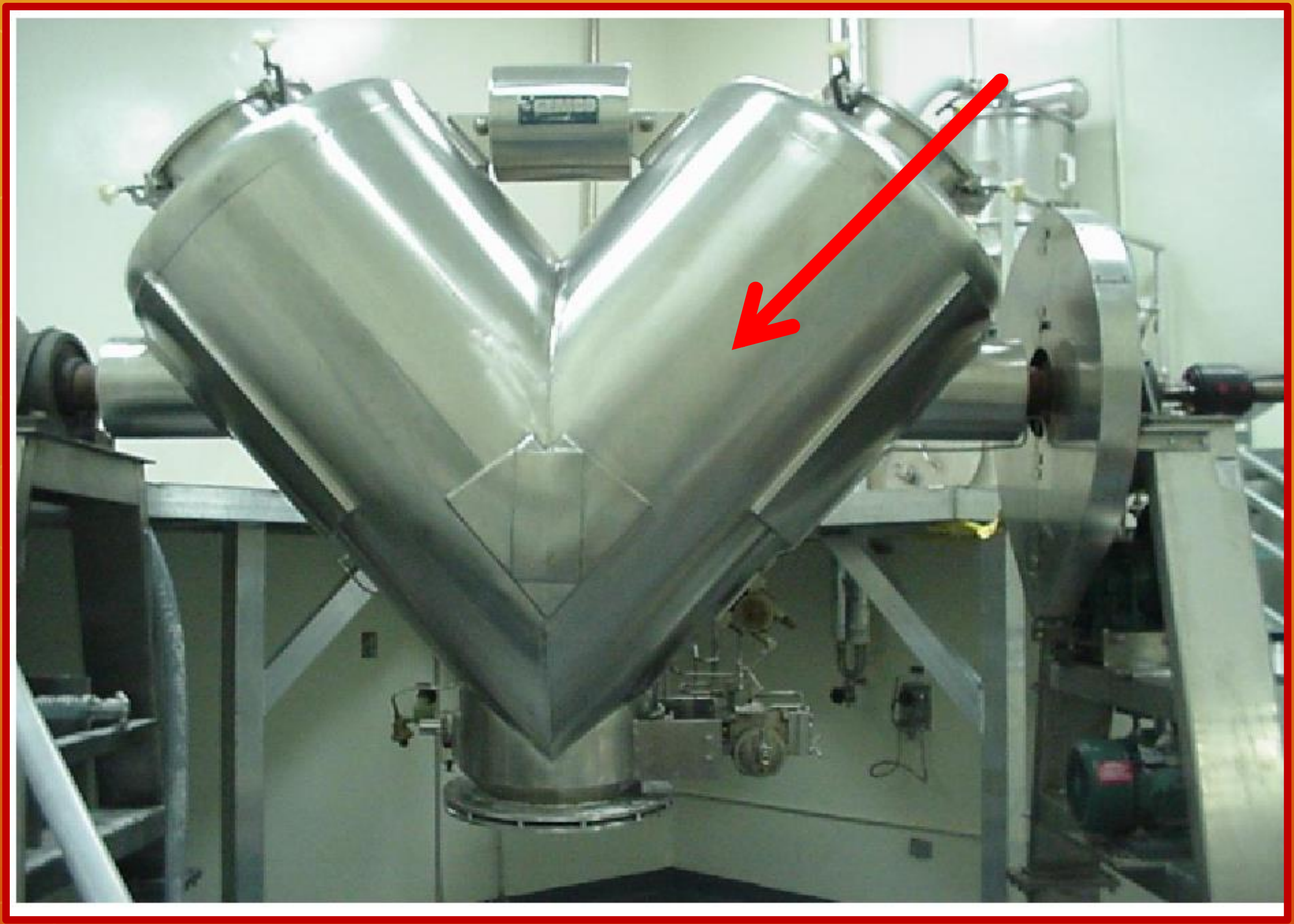
Camera, WIFI, ...


Flow
Flow
Flow

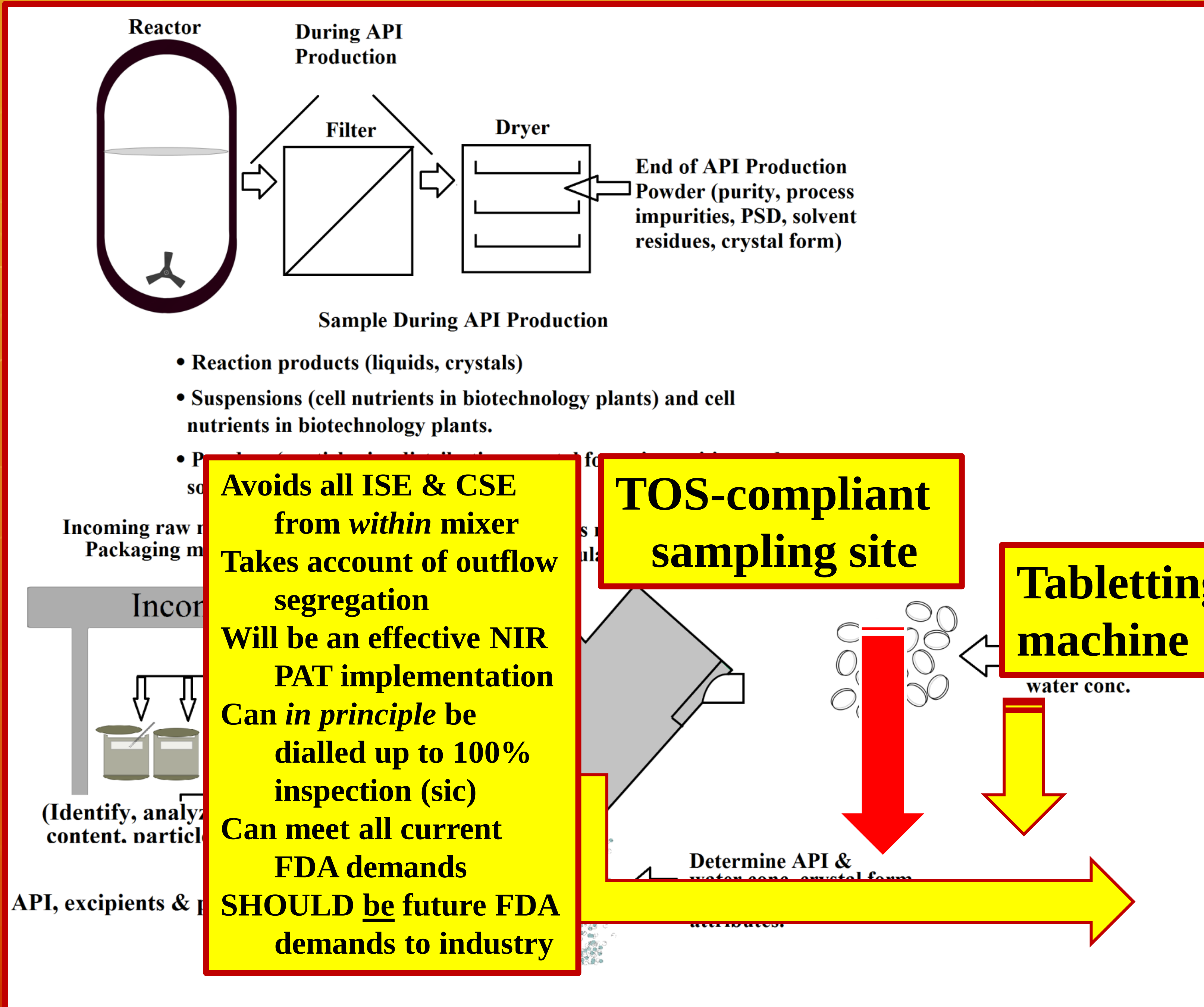
Duality …




Reactor
During API Production
Filter
Dryer
End of API Production Powder (purity, process impurities, PSD, solvent residues, crystal form)
Sample During API Production
• Reaction products (liquids, crystals)
• Suspensions (cell nutrients in biotechnology plants) and cell nutrients in biotechnology plants.
• Powders (particle size distribution, crystal form, impurities, and solvent residues).
Incoming raw materials & Packaging material.
Incoming
(Identify, analyze for purity, water content, particle size distribution)
API, excipients & packaging materials
In-Process mixtures (formulation)
End Pharma Process
Tabletting machine
water conc.
Determine API &



Radical solution: IF you have to sample from within a mixer – DO NOT SAMPLE from within a mixer !!


Reactor
During API
Production
Filter
Dryer
End of API Production
Powder (purity, process
impurities, PSD, solvent
residues, crystal form)
Sample During API Production
• Reaction products (liquids, crystals)
• Suspensions (cell nutrients in biotechnology plants) and cell
nutrients in biotechnology plants.
Avoids all ISE & CSE
from *within* mixer
Takes account of outflow
segregation
Will be an effective NIR
PAT implementation
Can *in principle* be
dialled up to 100%
inspection (sic)
Can meet all current
FDA demands
SHOULD <u>be</u> future FDA
demands to industry
TOS-compliant
sampling site
Tabletting
machine
water conc.
Determine API &

Representative Sampling: Theory of Sampling (TOS)

TOS - Axiomatic exposé

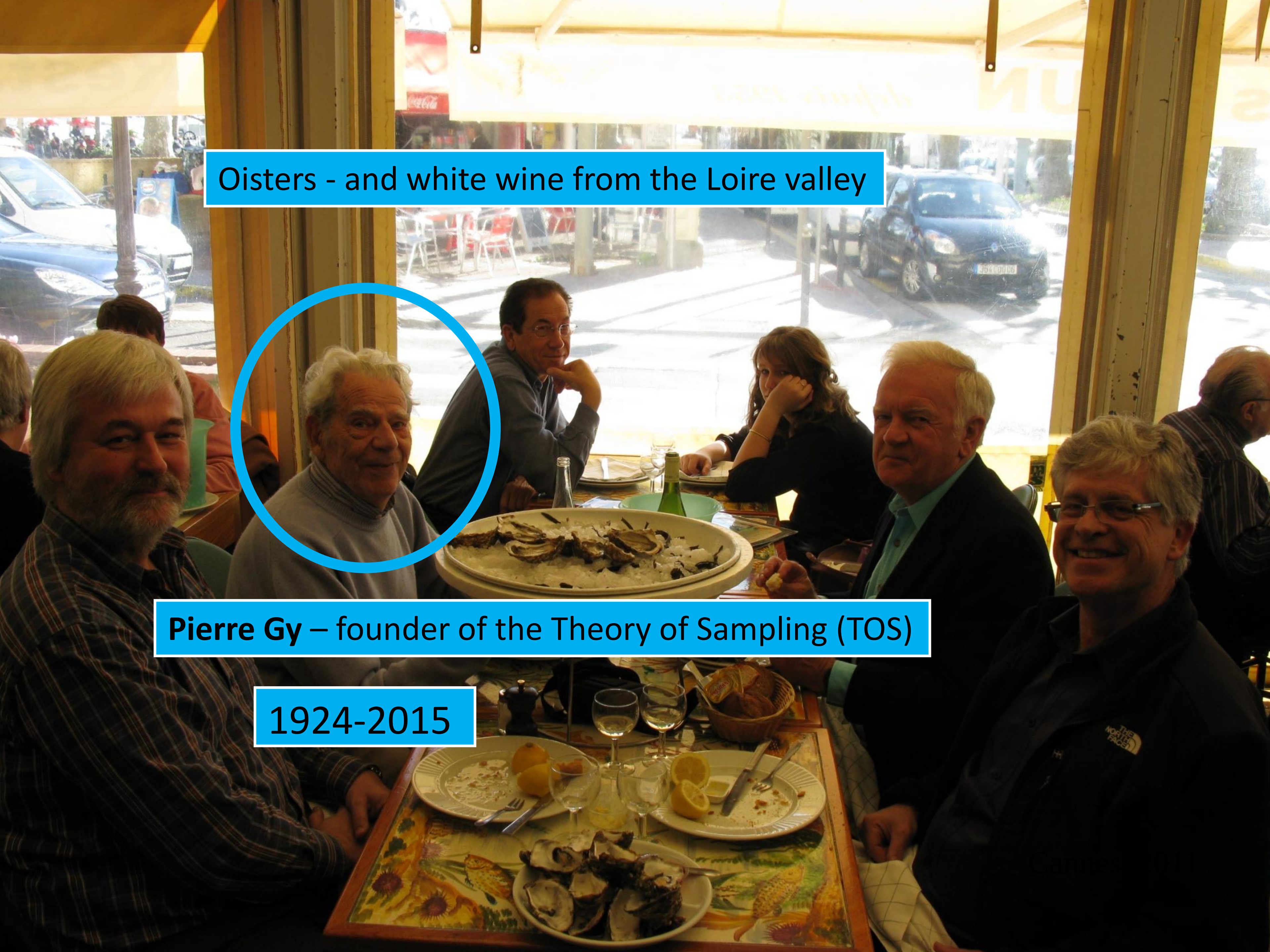
Governing principles (GP) – Sampling unit operations (SUO)

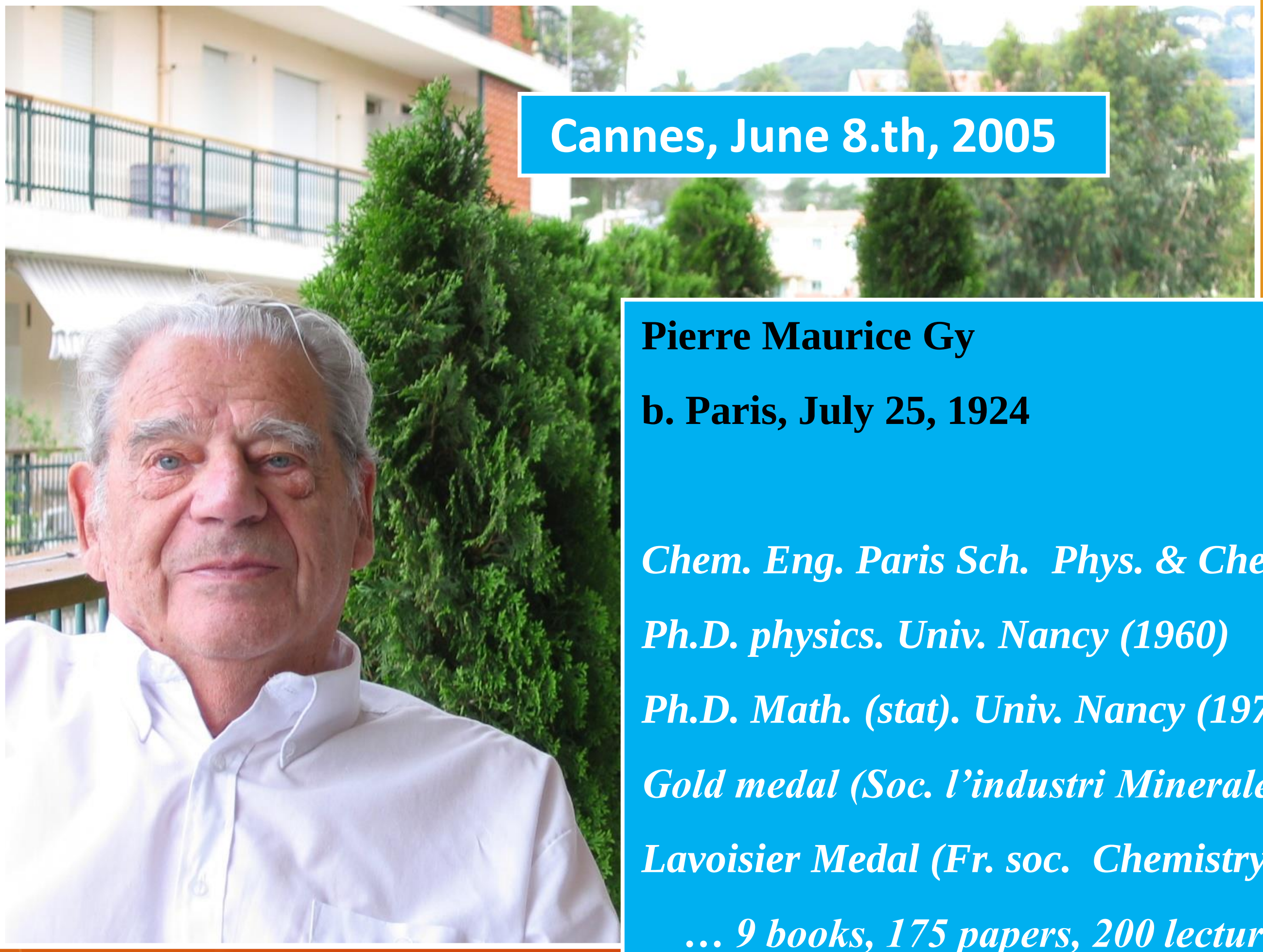
1. FSP: Fundamental Sampling Principle
2. SSI: Sampling Scale Invariance
3. PSC: Sampling Correctness (bias-free sampling)
4. PSS: Sampling Simplicity (primary sampling + mass-reduction)
5. LDT: Lot Dimensionality Transformation
6. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

Exit grab sampling!

The <u>one</u> EXCEPTION from grab sampling

Well there are two exceptions ... ...


Oisters - and white wine from the Loire valley
Pierre Gy – founder of the Theory of Sampling (TOS)
1924-2015



Cannes, June 8.th, 2005

Pierre Maurice Gy

b. Paris, July 25, 1924

Chem. Eng. Paris Sch. Phys. & Chem. (46)

Ph.D. physics. Univ. Nancy (1960)

Ph.D. Math. (stat). Univ. Nancy (1975)

Gold medal (Soc. l'industri Minerale) (63,76)

Lavoisier Medal (Fr. soc. Chemistry) (1995)

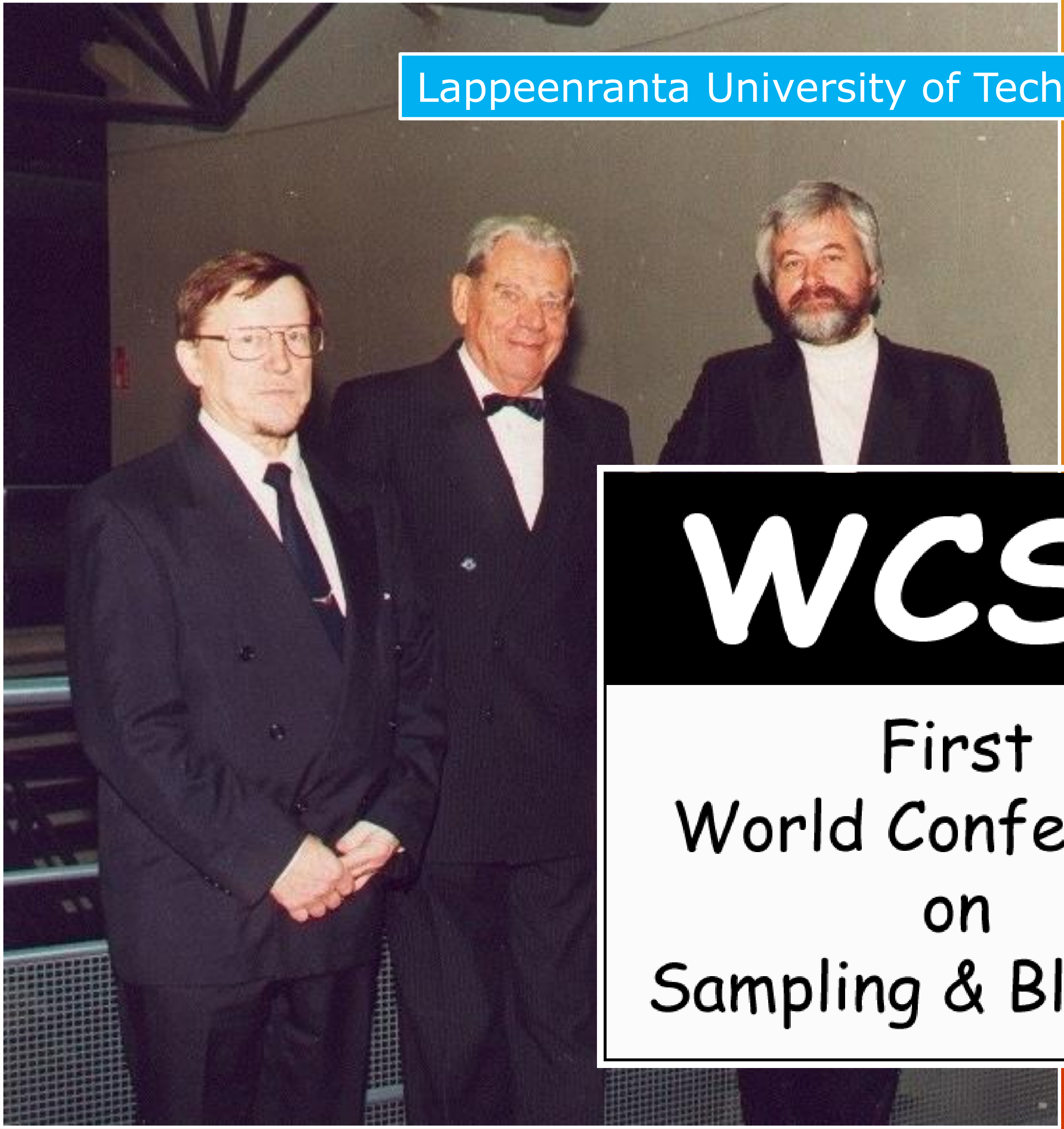
… 9 books, 175 papers, 200 lectures …

Pierre Gy: founder of TOS

A brilliant scientist – a monumental *ouevre* – a gentleman friend to all samplers


From where certain VIKINGS originated
who raided *Le Normandie* 1000 years ago
and later then setteled there:
... The family Pierre Gy ...
The estuary ”Gyland”, Flekkefjord, Norway

Lappeenranta University of Technology (LUT) - 1999



WCSB

First
World Conference
on
Sampling & Blending

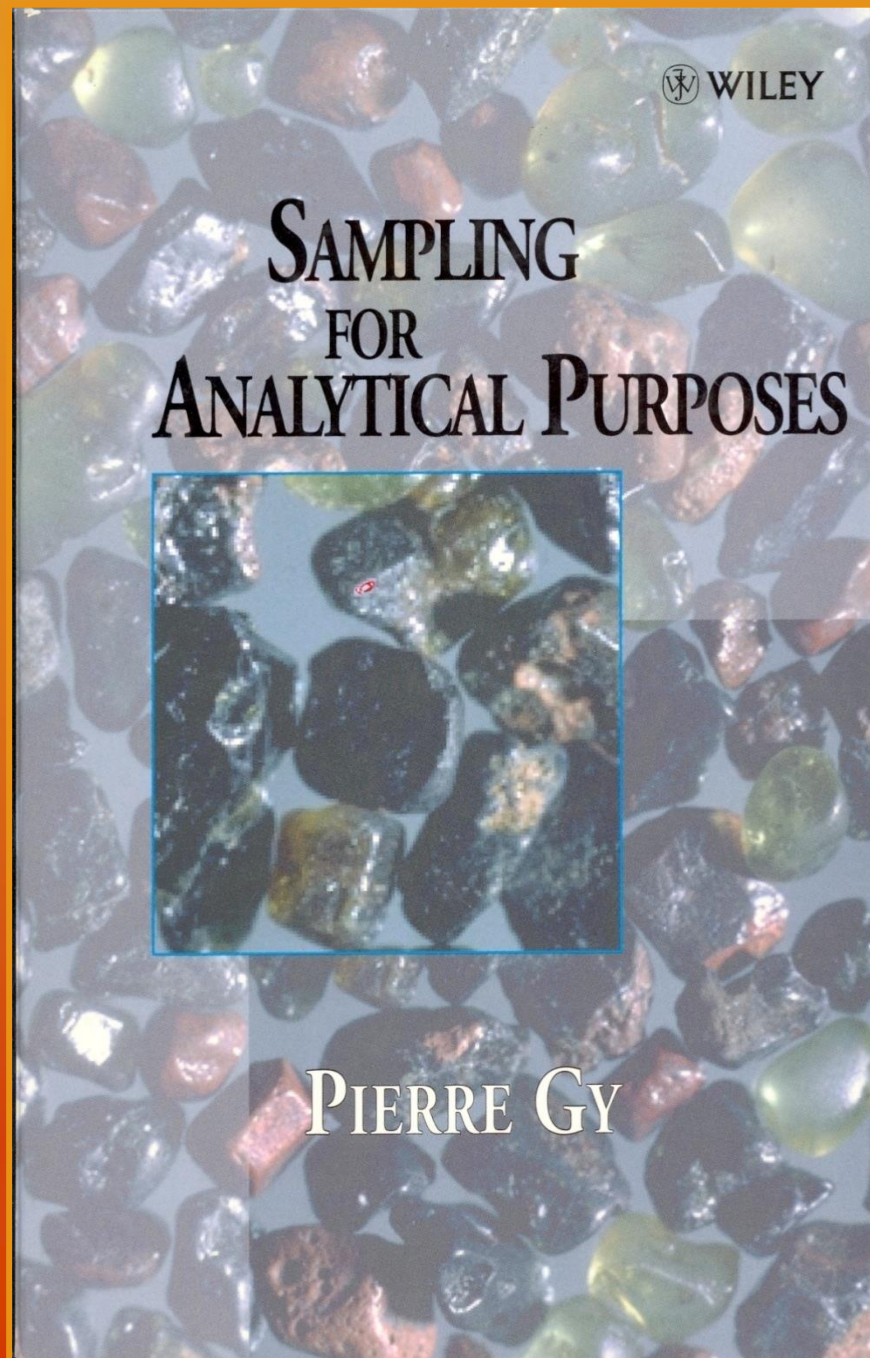
2003

Vol. 74, issue 1, 28 November 2004

ISSN 0169-7439

ELSEVIER

CHEMOMETRICS AND INTELLIGENT LABORATORY SYSTEMS


WILEY
SAMPLING
FOR
ANALYTICAL PURPOSES
PIERRE GY

KHE Consulting

Theory and Practice of TOS Representative Sampling

A minimum understanding of governing principles and sampling unit operations

- for all types of materials (all degrees of heterogeneity:low / intermediate / high
- at all scales (for all lot sizes: small / intermediate / big / extreme)
- unifying principles of representative sampling: field/plant/laboratory

مقام معظم رهبری:
دهه فجر
تجدید عهد
با انقلاب است.
اداره کل مدیریت بحران استانداری گیلان
والفجر و لیال عشر

The THEME of today's short course: WHAT comes BEFORE analysis?

Measurement Uncertainty (MU)

Theory of Sampling (TOS)

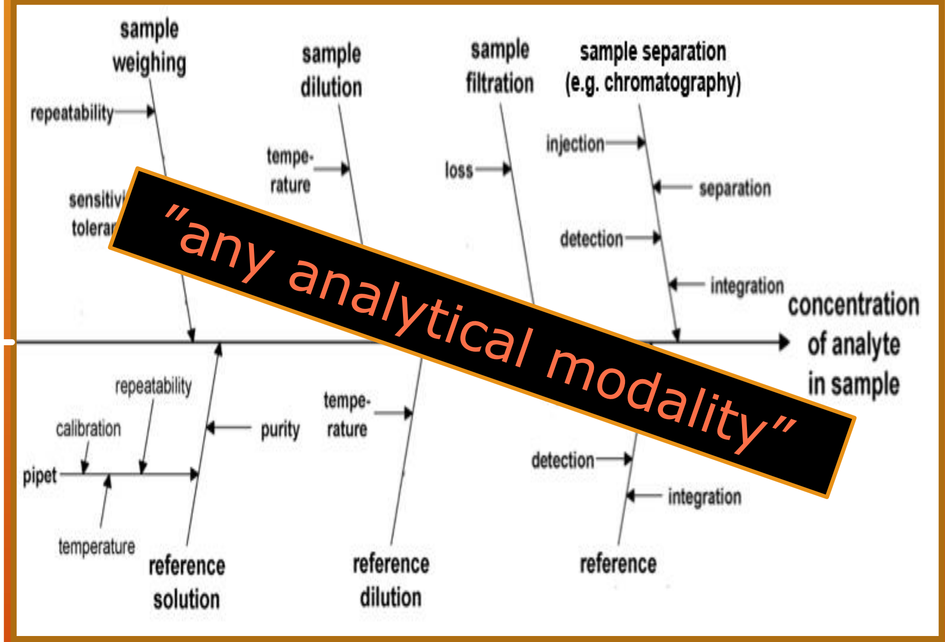
APGC + MS/MS
(single) MS
(universal) MS
LC-MS + APGC + CI + (tandem) MS
GC –GC
HPLC / HRMS .. (data independent MS/MS)
Tandem Ionization
UHPLC / QT [MS/MS libraries]
GC – TF-MS
.....
.....

Total Sampling Error (TSE)

Total Analytical Error (TAE)

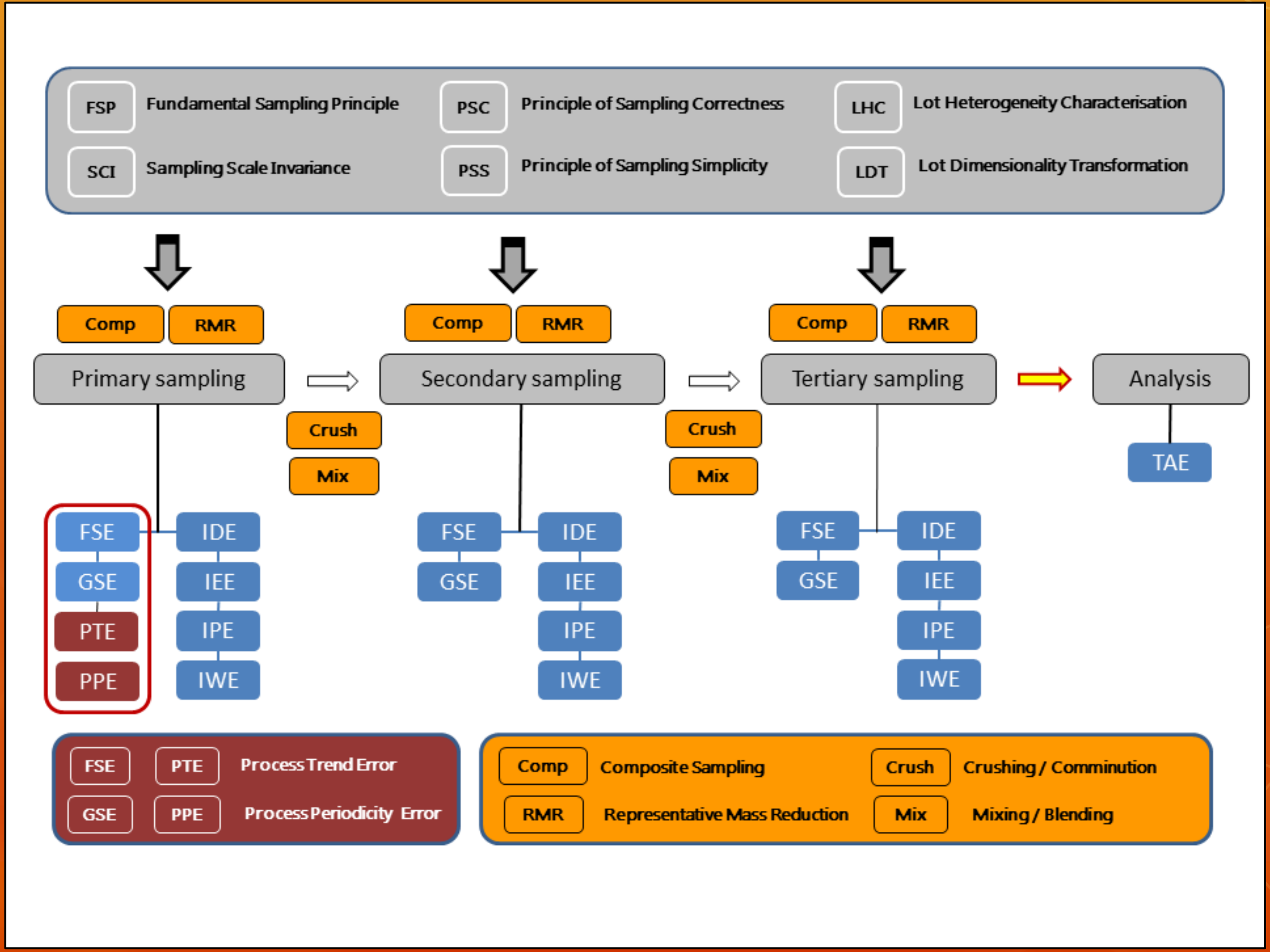
Theory of Sampling (TOS)

Measurement Uncertainty (MU)



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

Theory of Sampling (TOS) – everything in a glance

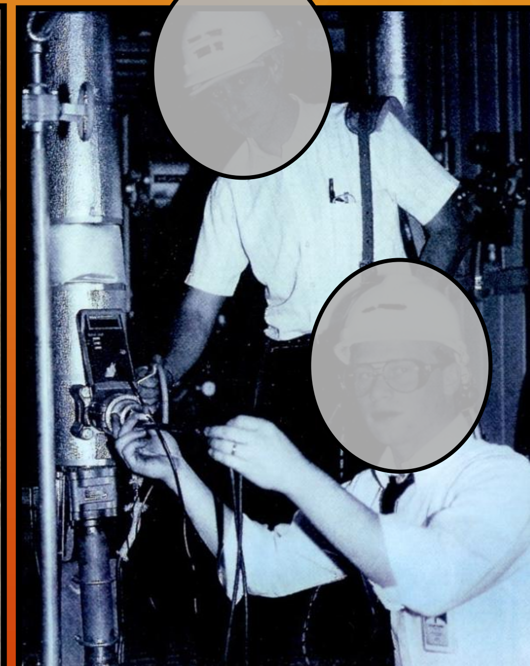


Sampling *competence*

The sampling *fool*

So, let us all become *fully* TOS competent!

The sampling *fool*



Sampling is actually only dependent upon three things

Representative Sampling: Theory of Sampling (TOS)

TOS - Axiomatic exposé

Governing principles (GP) – Sampling unit operations (SUO)

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. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

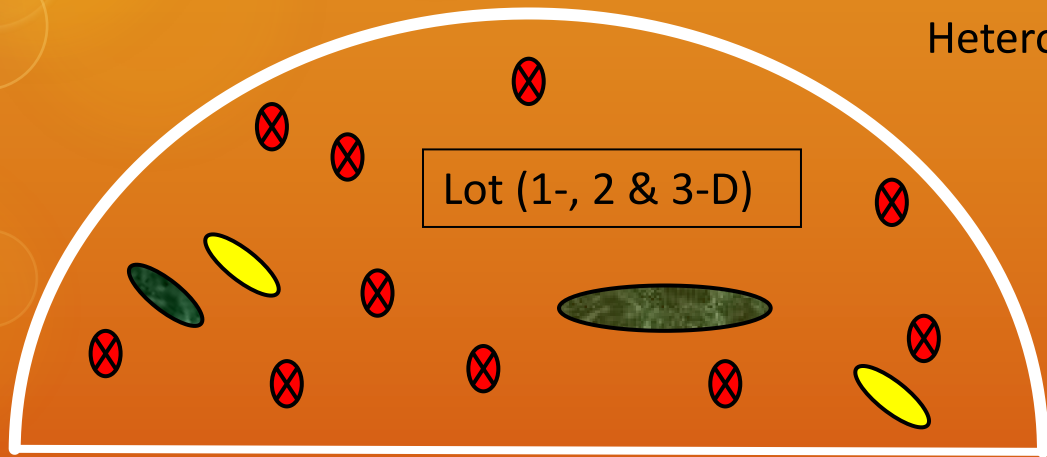
A grab sample ...

Fundamental Sampling Error (FSE) [+ GSE]

Several increments ... ...

Trying (very well) to deal with (GSE) [+ *FSE*]

Heterogeneity (hidden)



Fundamental Sampling Principle (FSP):

All increments must have the *same* (non-zero) probability of ending up in the sample (non-neg)

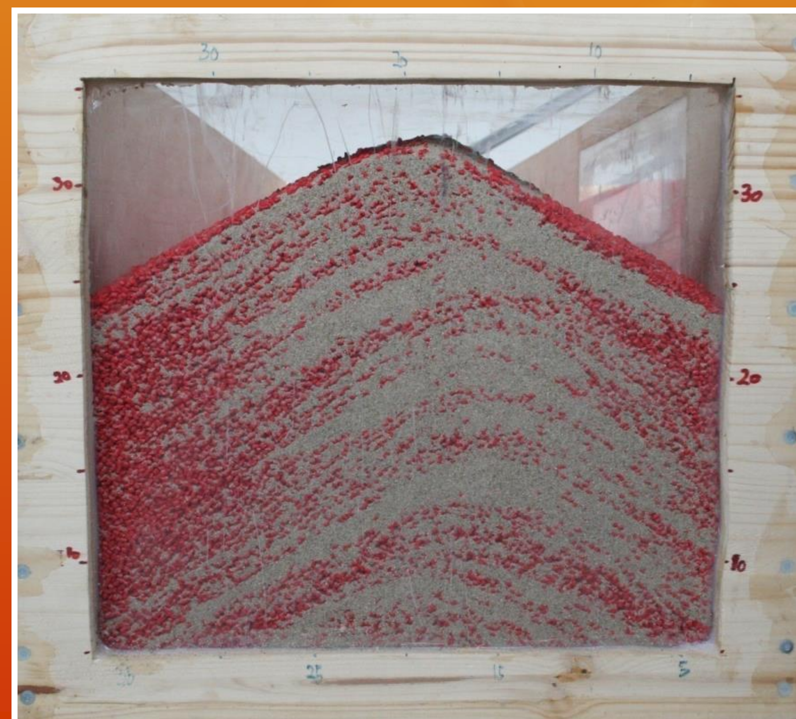
FSP

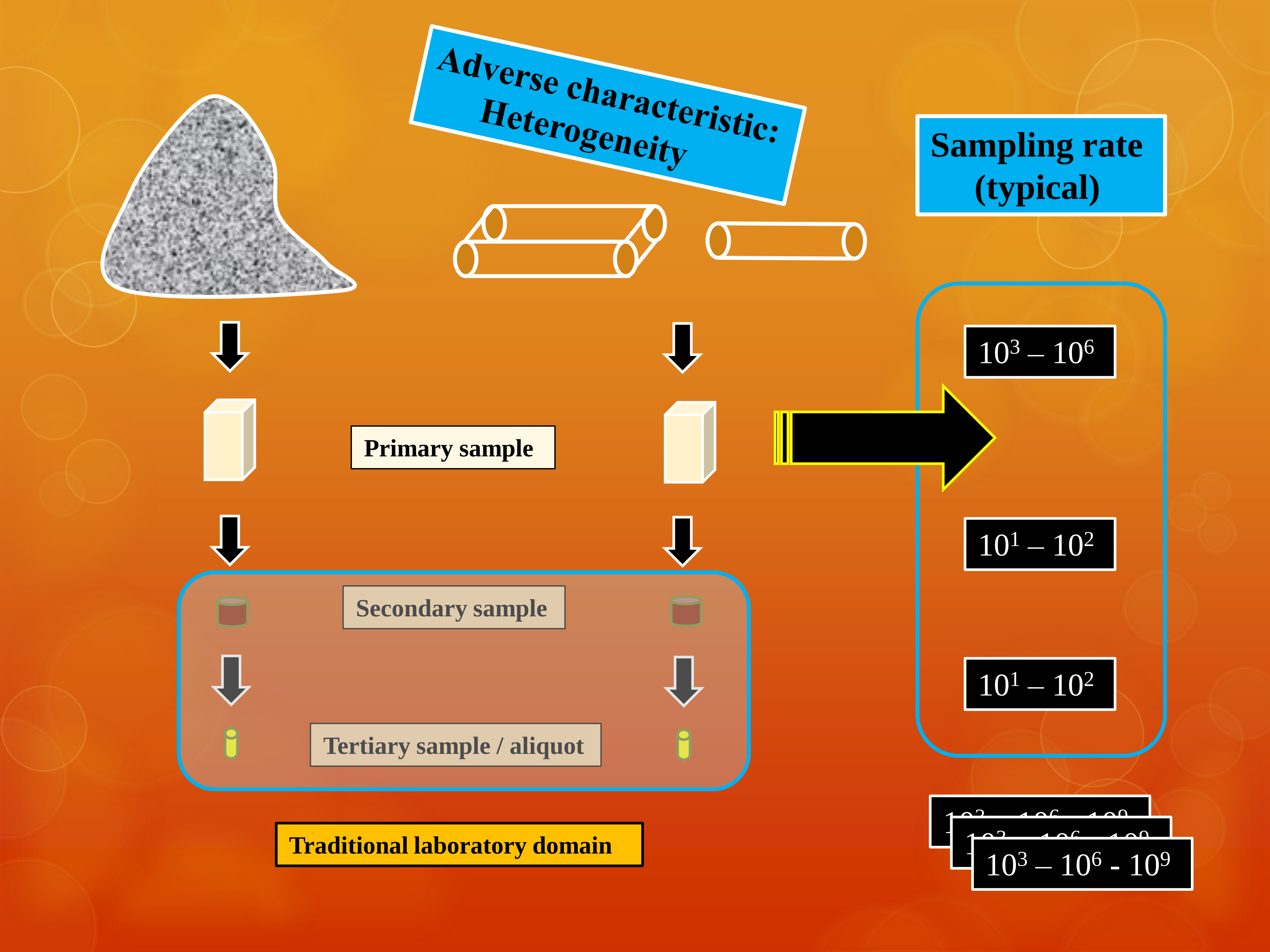
Fundamental Sampling Principle (FSP) 3-D lot

Fundamental Sampling Principle (FSP) 3-D lot

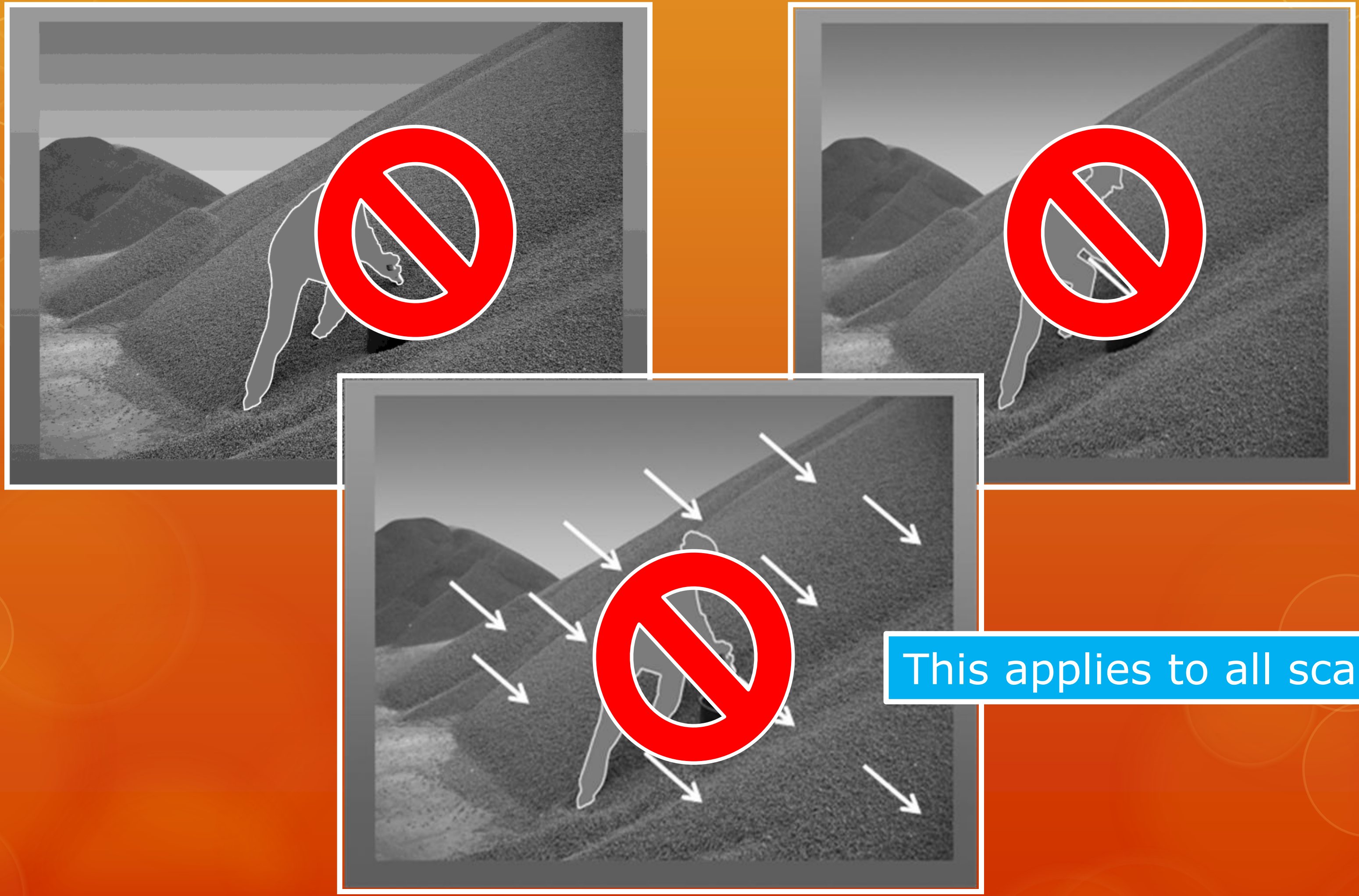
Fundamental Sampling Principle (FSP) 3-D lot

FSP



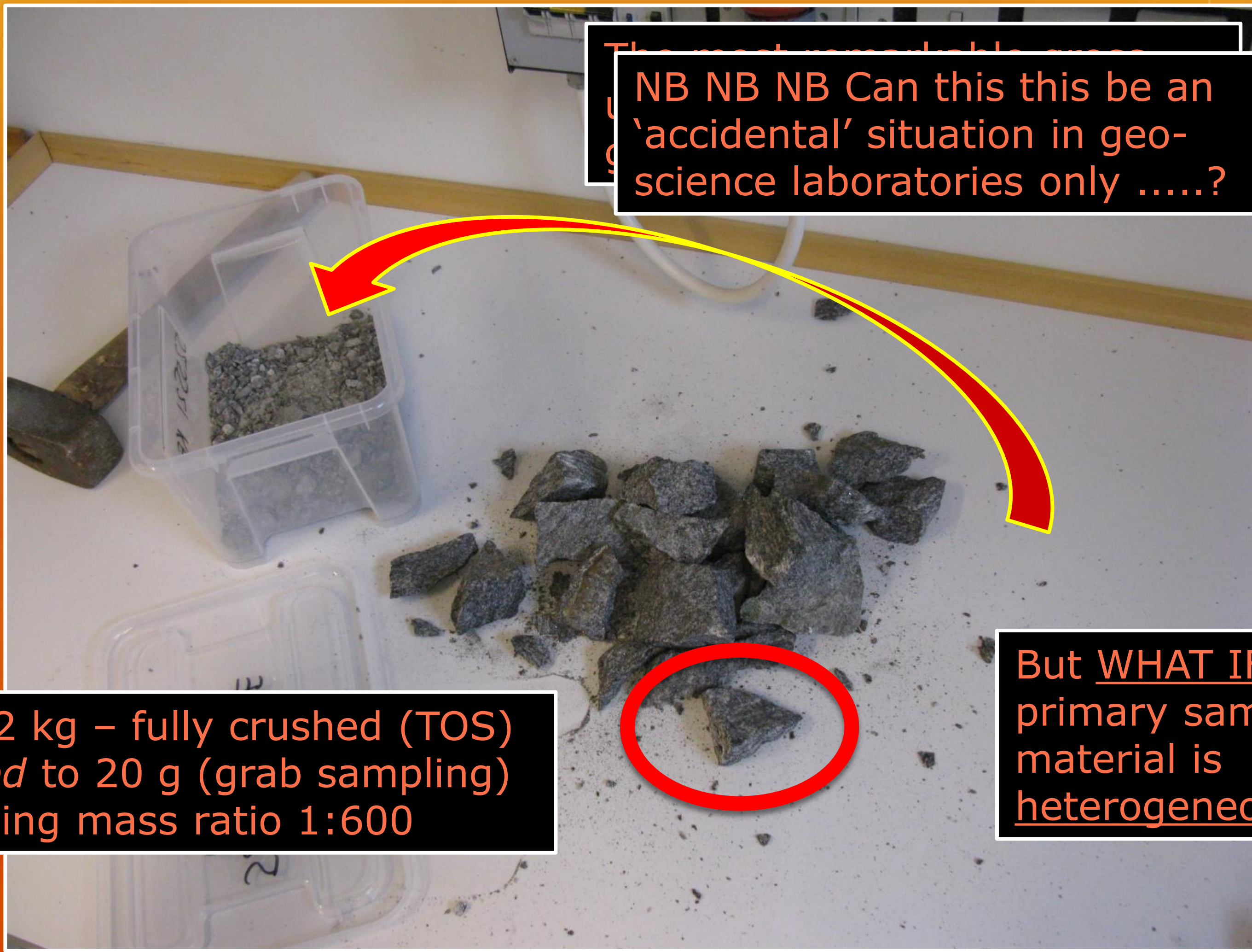

Adverse characteristic:
Heterogeneity
Sampling rate
(typical)
Primary sample
Secondary sample
Tertiary sample / aliquot
Traditional laboratory domain
$10^3 – 10^6$
$10^1 – 10^2$
$10^1 – 10^2$
$10^3 – 10^6 - 10^9$

Fundamental Sampling Principle (FSP) 3-D lot



This applies to all scales!

- the analytical result <u>depends</u> on the sampling procedures used ....



NB NB NB Can this this be an 'accidental' situation in geo-science laboratories only .....?

Whole 12 kg – fully crushed (TOS) *compared* to 20 g (grab sampling) – missing mass ratio 1:600

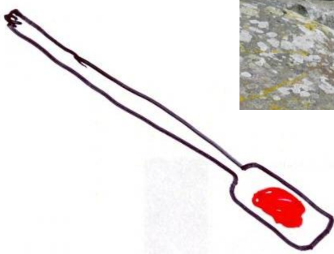
But <u>WHAT IF</u> the primary sample material is <u>heterogeneous</u>?

Representative Sampling: Theory of Sampling (TOS)

TOS - Axiomatic exposé

Governing principles (GP) – Sampling unit operations (SUO)

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. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)



...... only the SCALE !!!

Sampling <u>in the laboratory</u>: What’s the difference w.r.t. the field/plant etc?

<u>Identical</u> sampling issues and problems - at all scales ..

TOS' Simplifying, Governing Principles …

Principle of Sampling Scale Invariance (SSI)

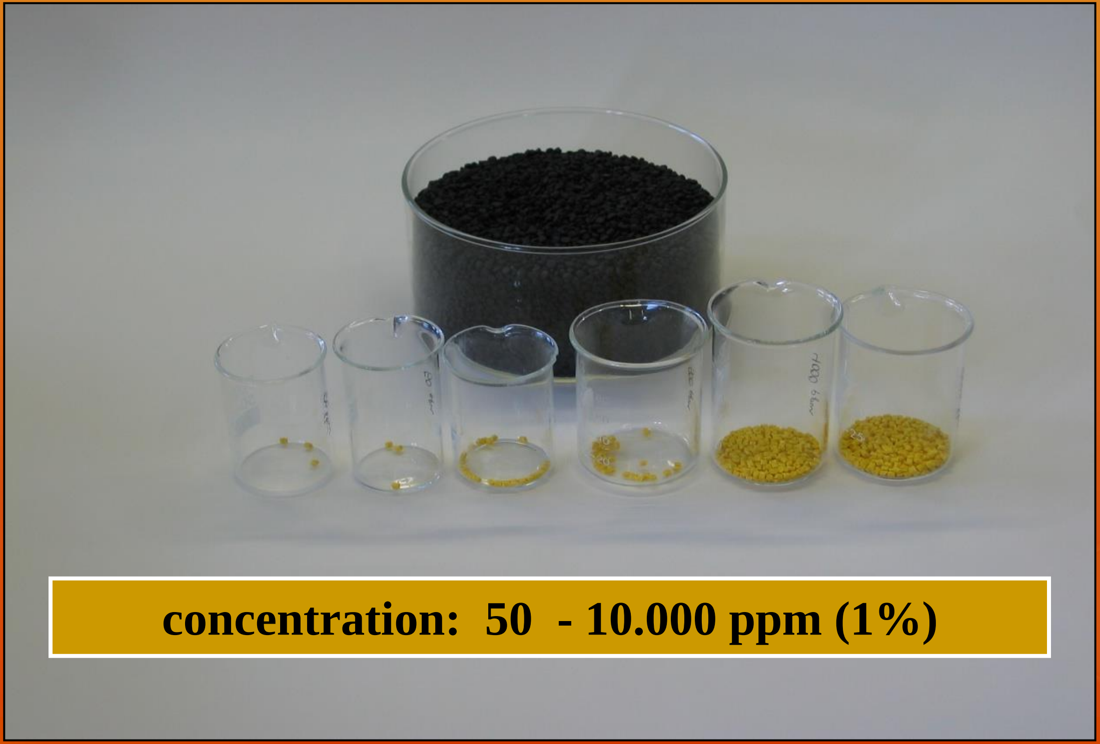
”1 gram for analysis -----”

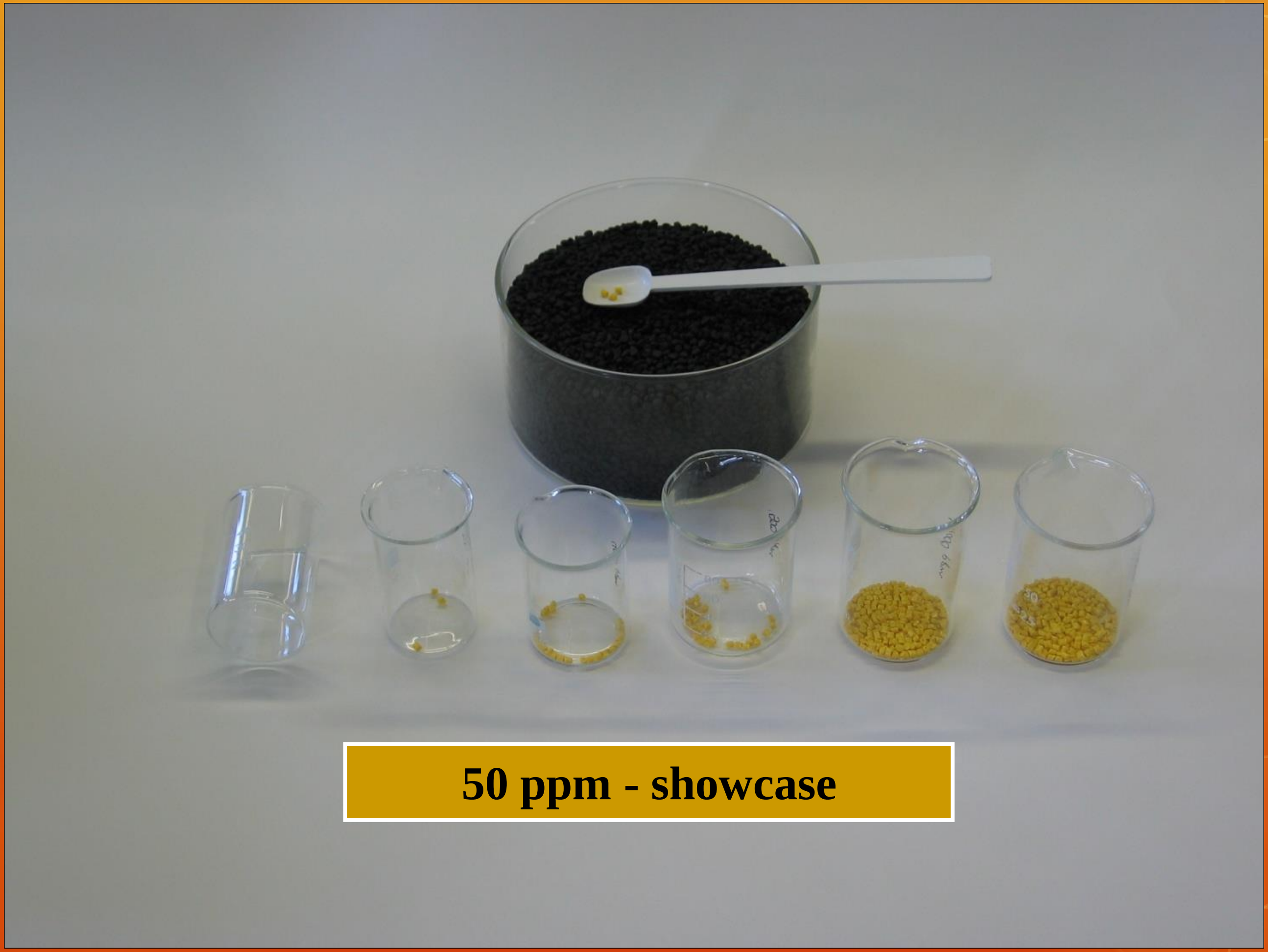
How big is the original lot, which shall be characterised by the analytical result?

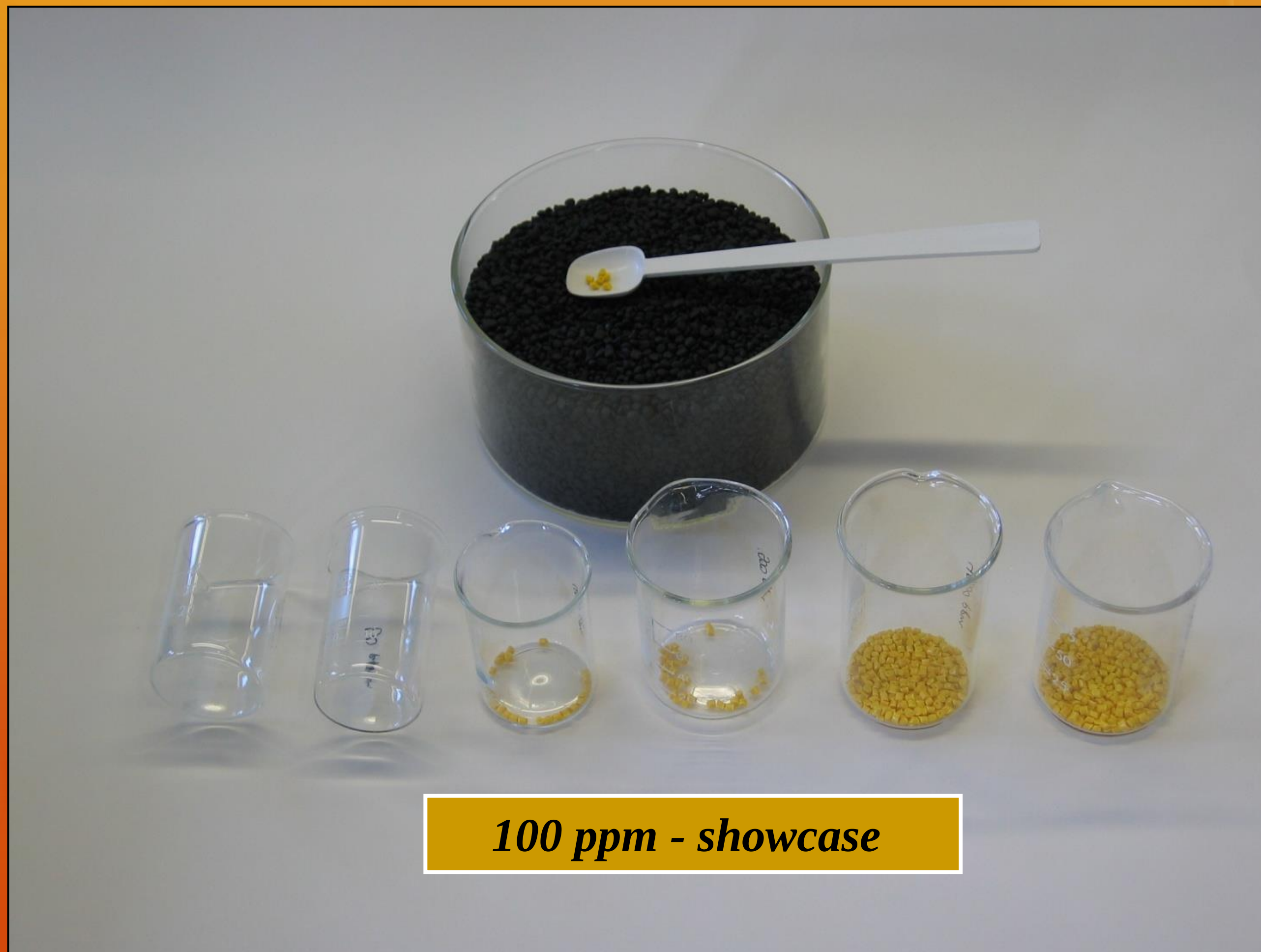
	Sampling rate:
1 kg	$1 / 10^3$
10 kg	$1 / 10^4$
100 kg	$1 / 10^5$
1.000 kg (1 ton)	$1 / 10^6$
10.000 kg (5 ton)	$1 / 10^7$
......	....

GRAPHIC illustration (trace concentrations)

- trace concentration: below 1%

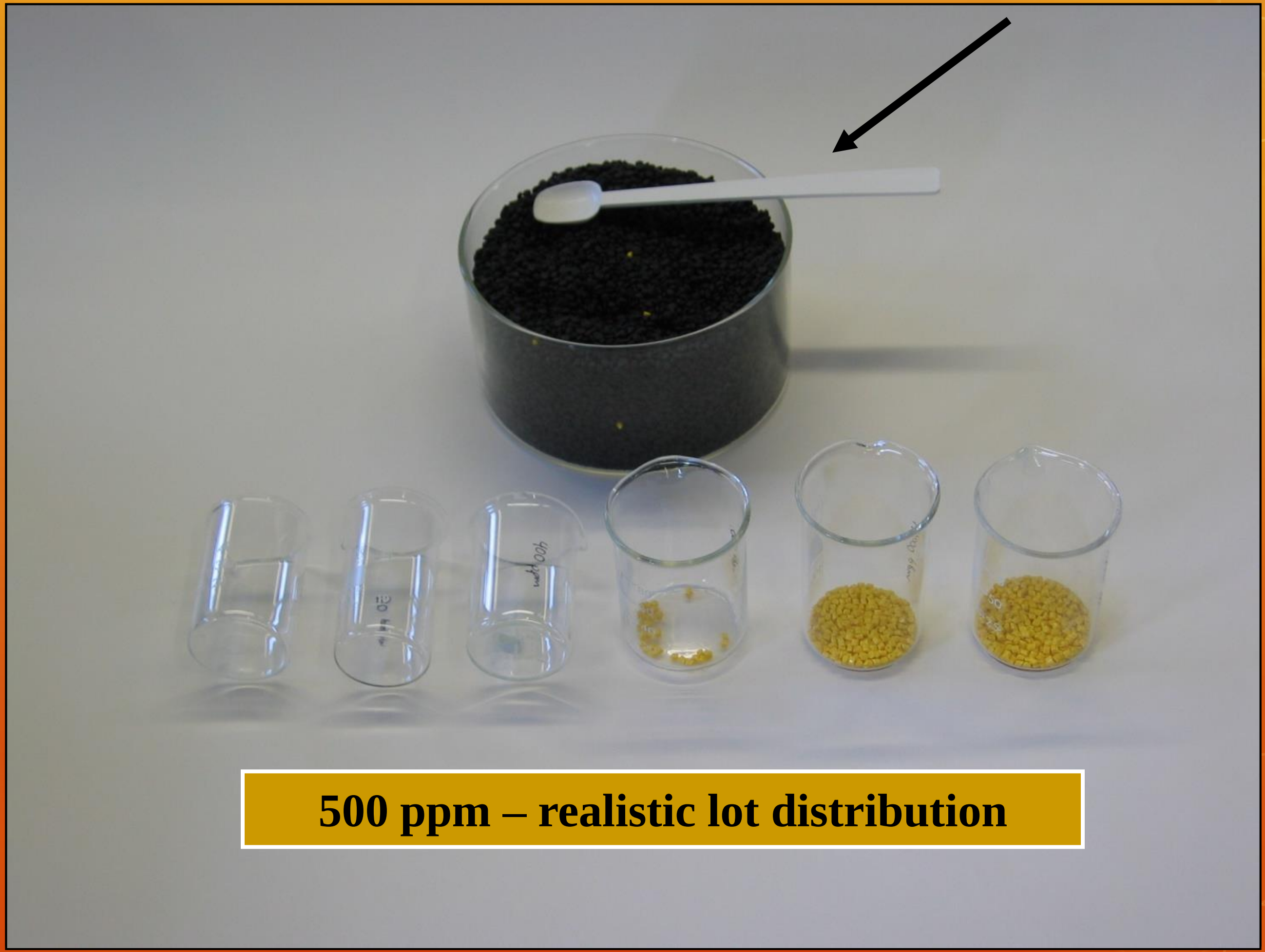



50 ppm - showcase


100 ppm - showcase

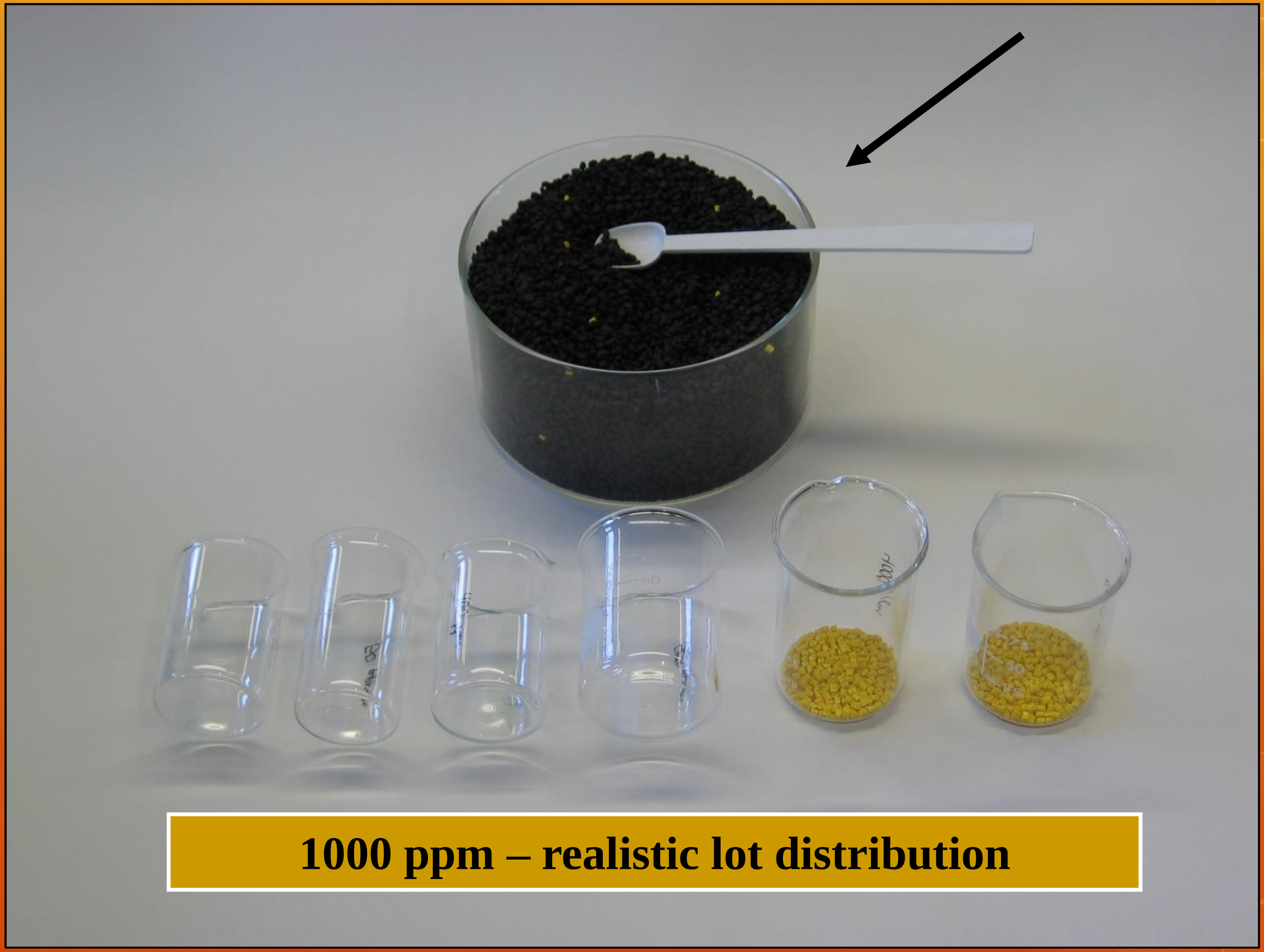
500 ppm - showcase

beware of spoon size

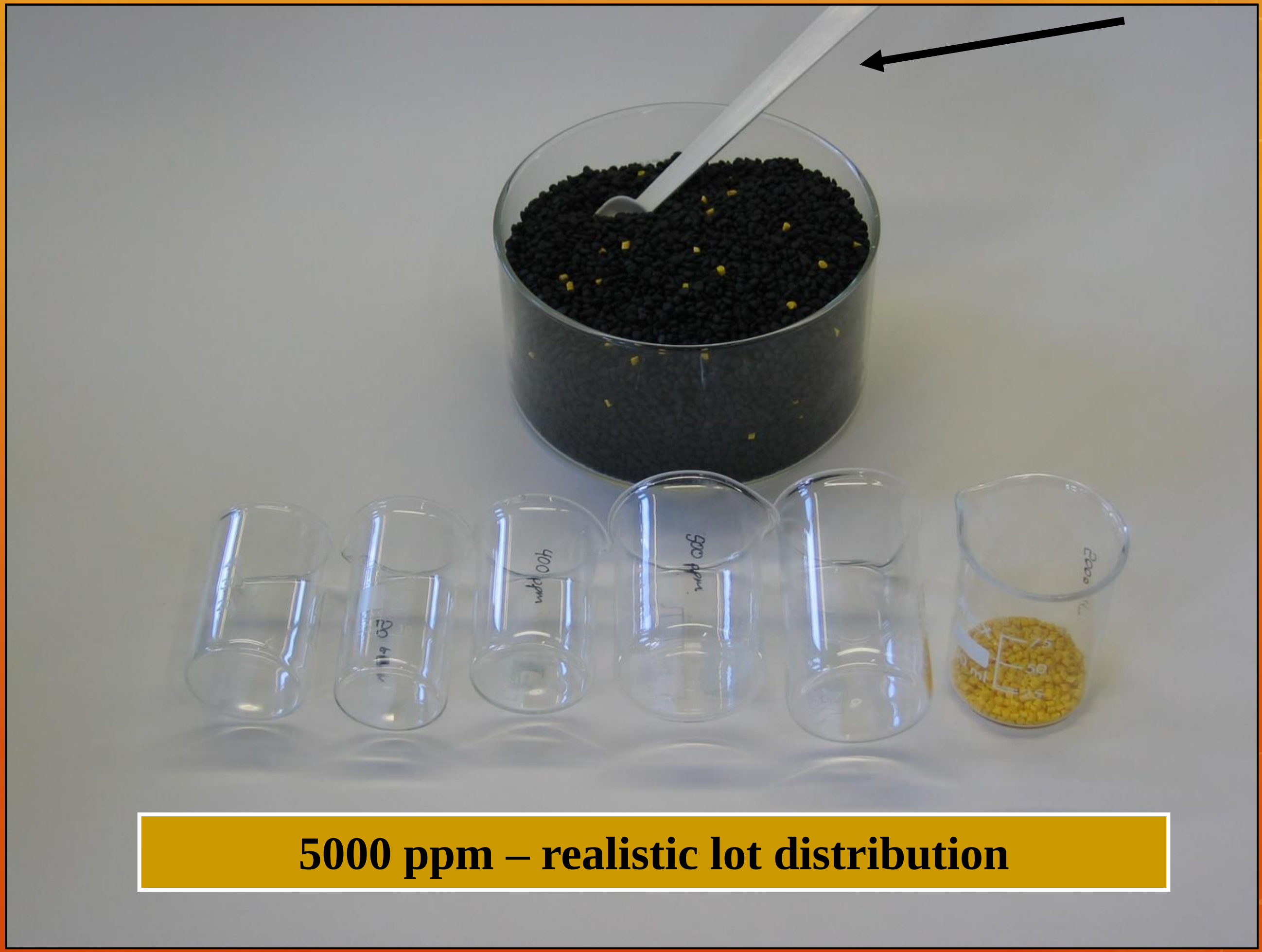


500 ppm – realistic lot distribution

beware of spoon size



beware of spoon size

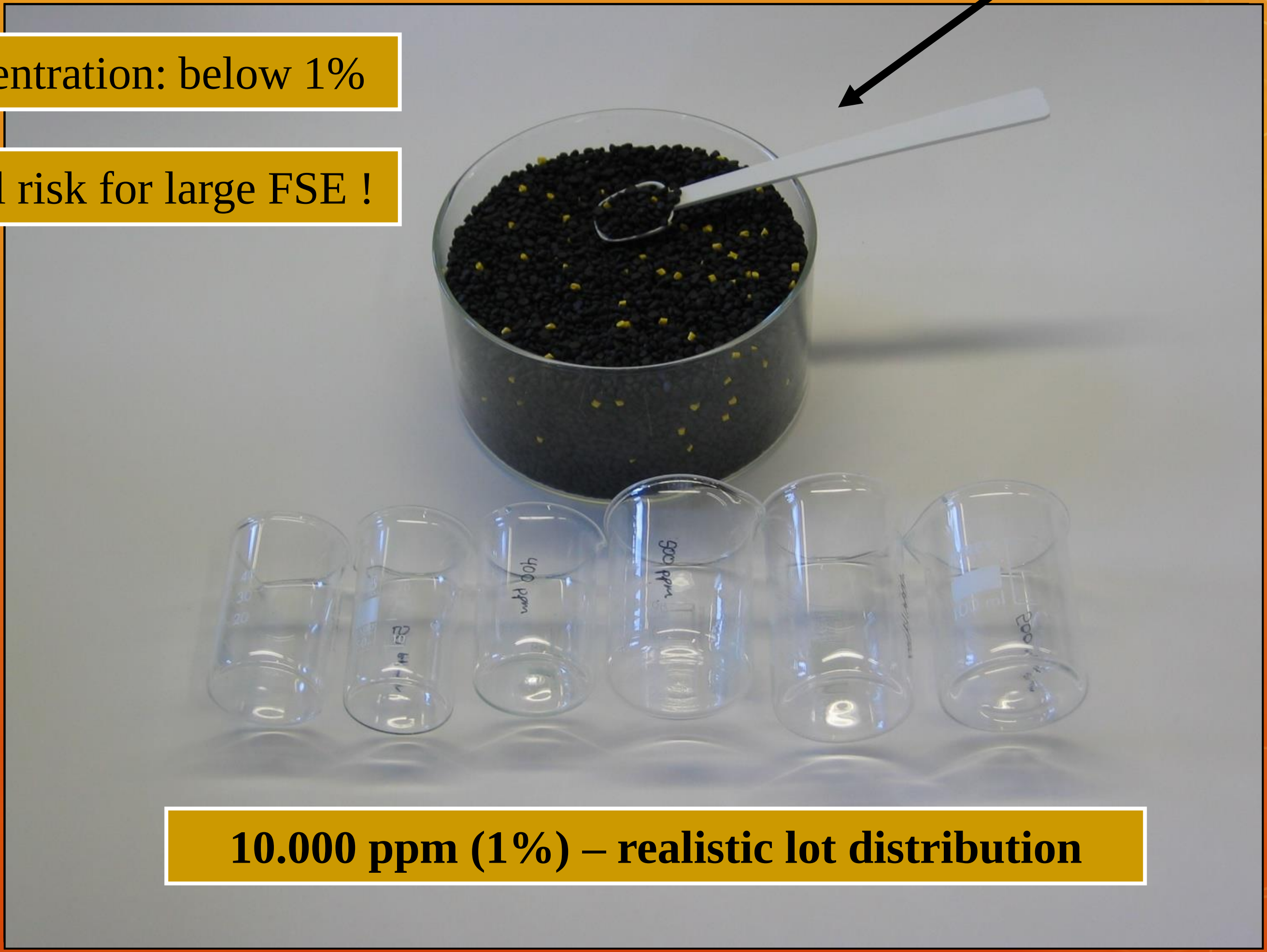


5000 ppm – realistic lot distribution

beware of spoon size

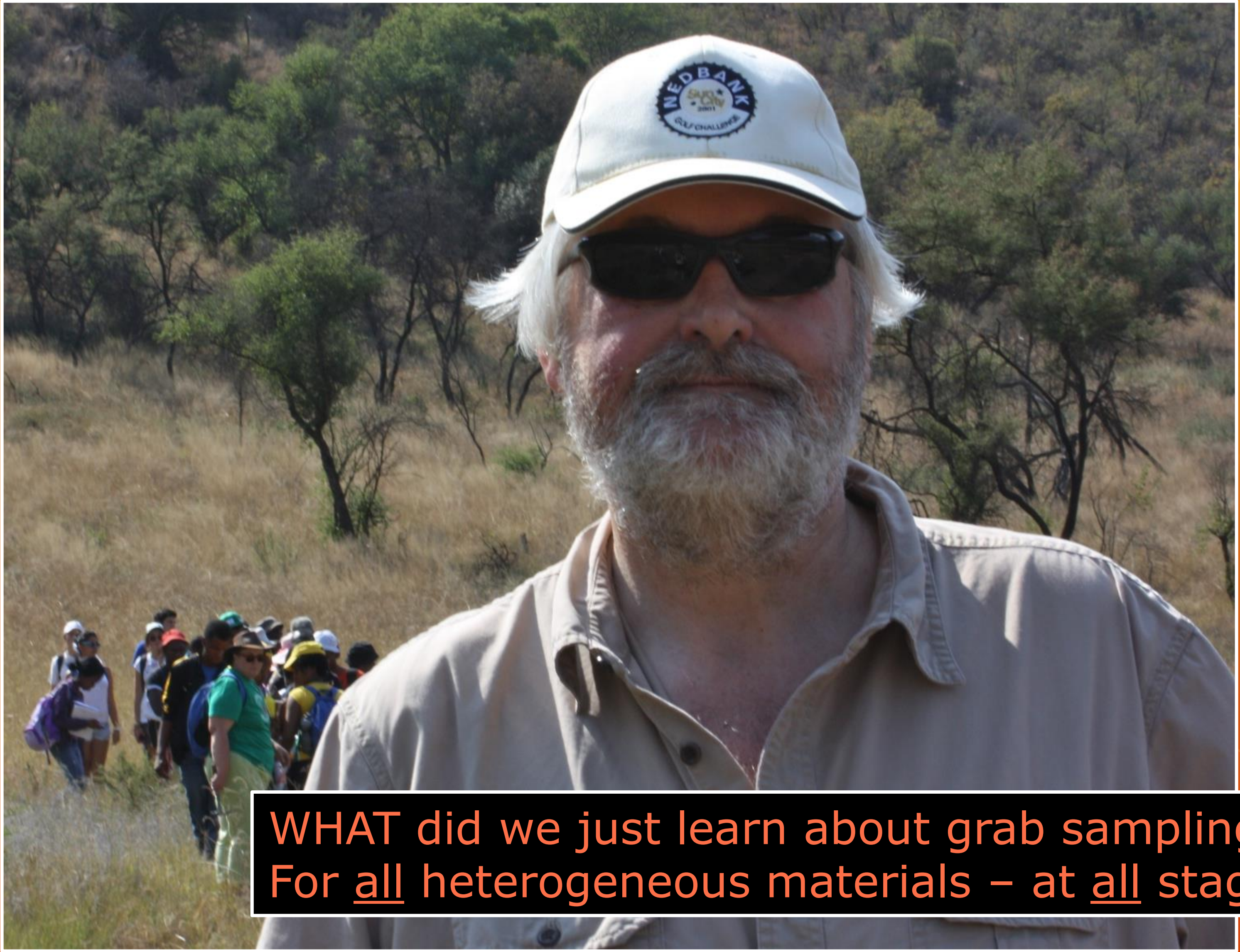
- trace concentration: below 1%

- Substabtial risk for large FSE !



10.000 ppm (1%) – realistic lot distribution

Experiment exemplum - credits: Suzanne Roy & Lars Beck

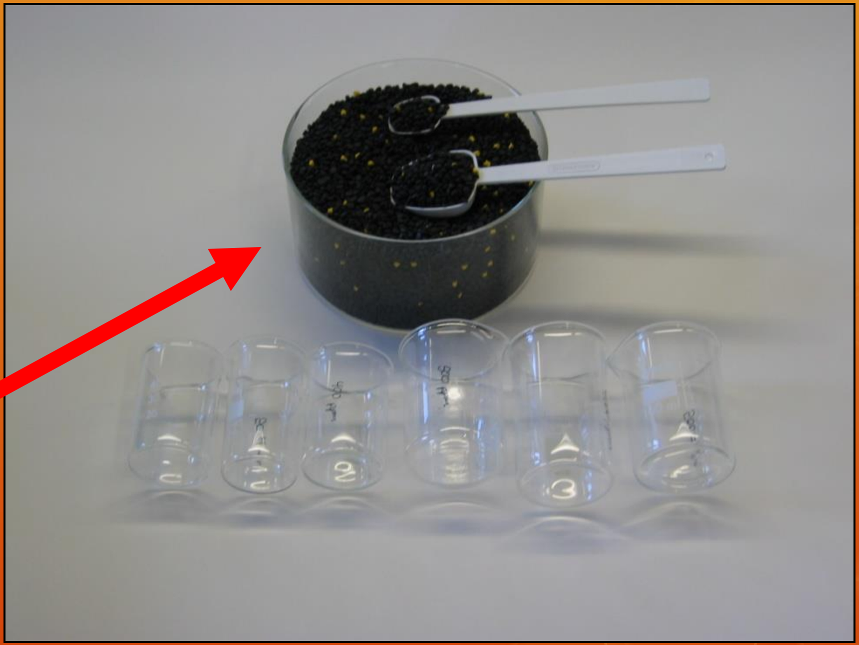


WHAT did we just learn about grab sampling?
For all heterogeneous materials – at all stages!

- **Heterogeneity as a function of:**

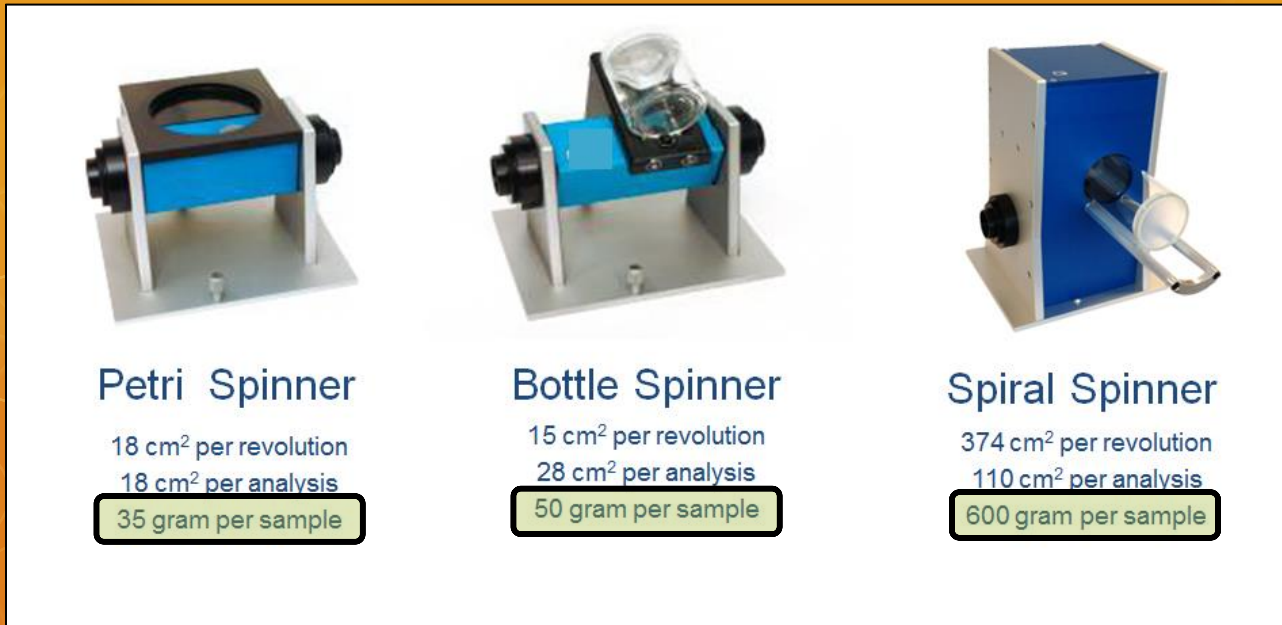
- *Concentration*
- *Spatial distribution*

- *Sampling tool size*

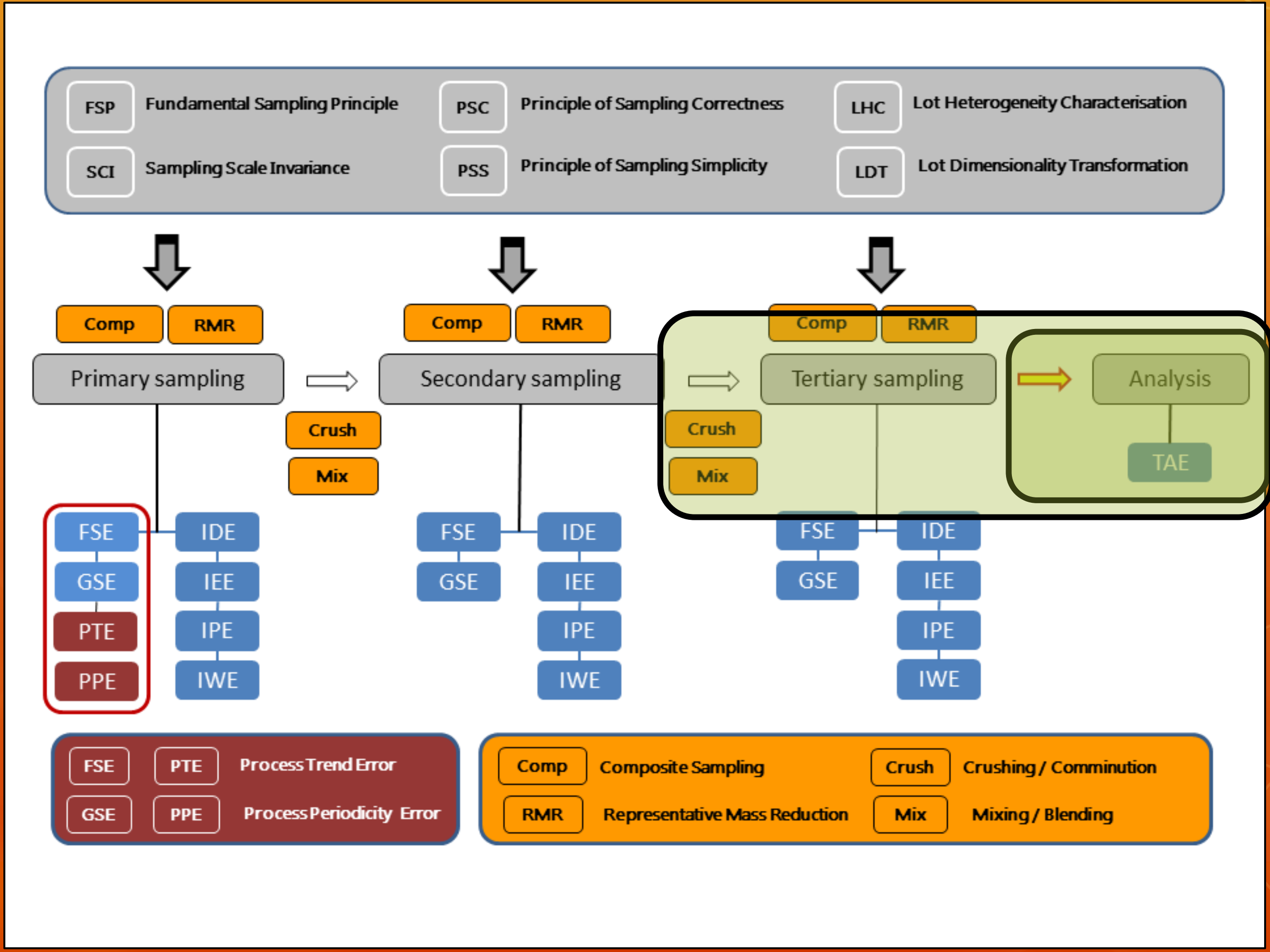


Sampling rate - 1:50

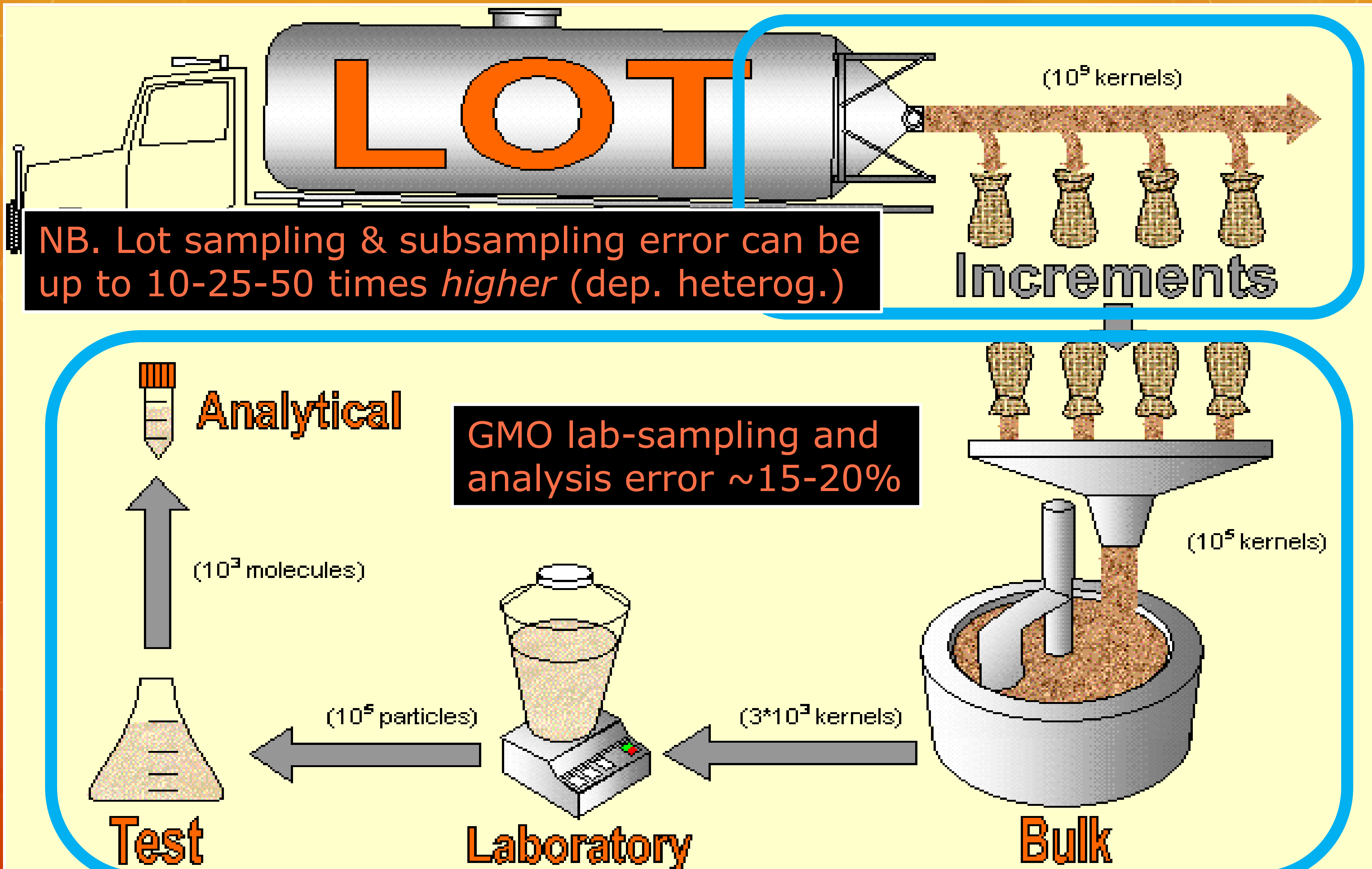
These sampling tool sizes are very unrealistic w.r.t. real-life situations ...


Petri Spinner
18 cm2 per revolution
18 cm2 per analysis
35 gram per sample
Bottle Spinner
15 cm2 per revolution
28 cm2 per analysis
50 gram per sample
Spiral Spinner
374 cm2 per revolution
110 cm2 per analysis
600 gram per sample

Theory of Sampling (TOS) – everything in a glance

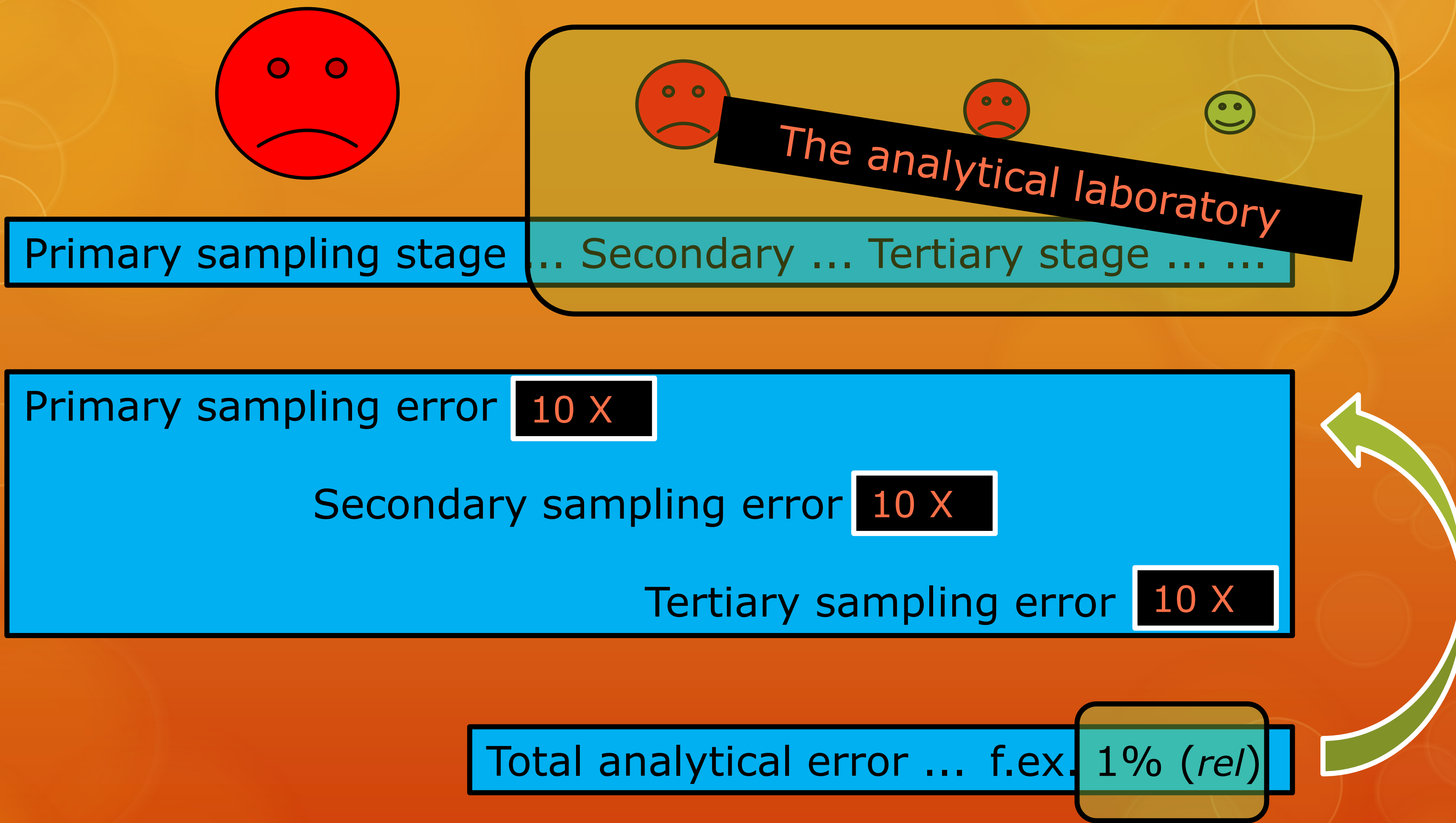


- the analytical result <u>depends</u> on the sampling procedures used ....



NB. Lot sampling & subsampling error can be up to 10-25-50 times *higher* (dep. heterog.)

GMO lab-sampling and analysis error ~15-20%


The analytical laboratory
Primary sampling stage ... Secondary ... Tertiary stage ... ...
Primary sampling error
10 X
Secondary sampling error
10 X
Tertiary sampling error
10 X
Total analytical error ... f.ex. 1% (*rel*)

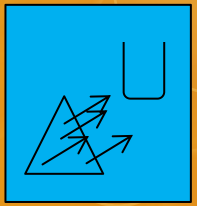
Representative Sampling: Theory of Sampling (TOS)

TOS - Axiomatic exposé

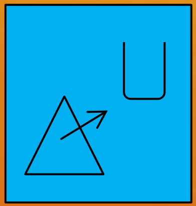
Governing principles (GP) – Sampling unit operations (SUO)

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. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

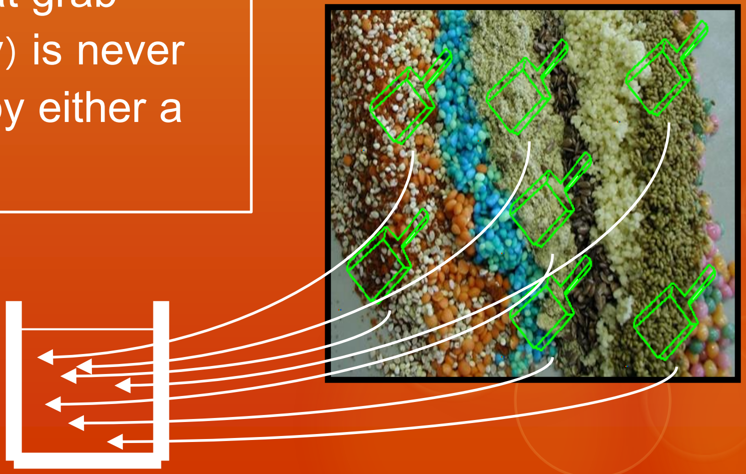
Composite sampling employs Q increment extractions with the aim to 'cover' the lot volume (only Q = 4 increments shown in this principal illustration). Proportional to the heterogeneity encountered, a higher number of increments will be required. Comp samp. must always respect FSP !!!

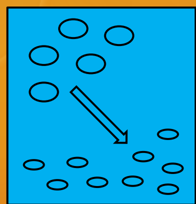


GP (6) Lot Heterogeneity Characterisation (LHC) guarantees that no sampling plan, sampling procedure nor sampling equipment is employed without a mandatory heterogeneity characterisation of the lot material.



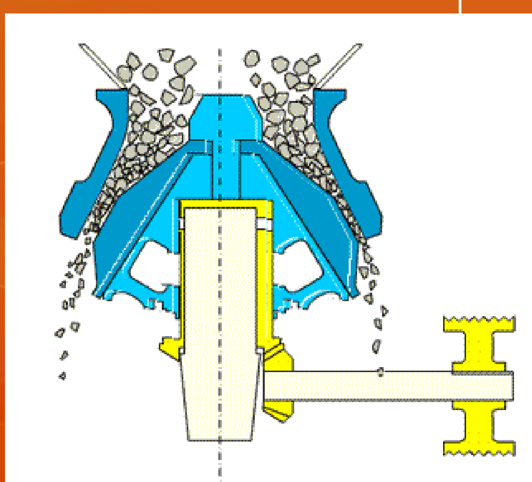
Composite sampling is specifically demanding that grab sampling (extraction of one single increment only) is never invoked, unless thoroughly tested and accepted by either a Replication Experiment (RE) or by variographics.



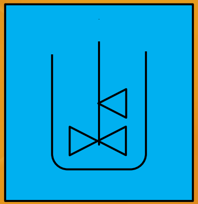


Crushing (**comminution**) is a sampling unit operation which is only brought to bear when necessary, i.e. when the top particle size is contrasting too much with respect to smaller size ranges in order for sampling to be effective and representative. Comminution is the technical process in which the top particle sizes is preferentially crushed first.

A consequence of crushing/comminution is that the majority of particle sizes tend to become more similar, with the further advantage that mixing becomes more effective.

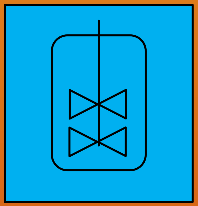


ceration, crushing or shredding in the presense of a
ilitating liquid (often used for selective extraction), as
plied to biological materials also lead to reduced general
rticle sizes.



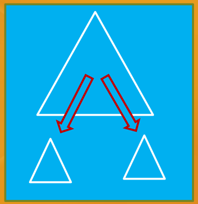
Mixing is a forced mechanical process designed to reduce the distributional heterogeneity (DH) of a material system.

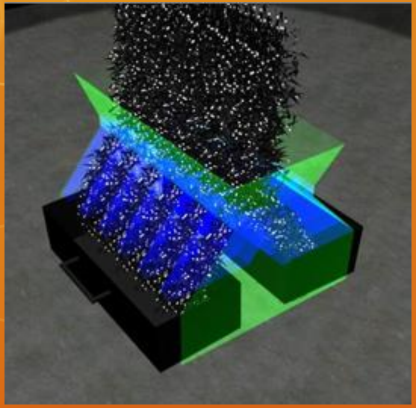
It is always advantageous to mix the results of a sampling or a sub-sampling process *before* further processing (sub-sampling or a next stage mass reduction).

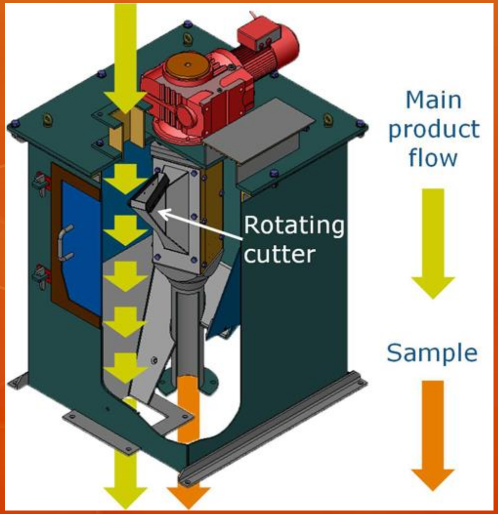


Blending is mixing under stoichiometric constraints, i.e. the final mixing product, a blend, must satisfy compositional constraints e.g. tea, tobacco, cement, pharmaceutical drugs.

Mixing / blending can be applied to both polyphase dry systems (aggregates) and to slurries (solid – liquid systems).







Representative Mass Reduction (RMR) is the key sampling unit operation connecting all sampling stages. Often the terms mass reduction and sub-sampling are used inter alia. There are very many sub-sampling procedures and types of equipment offered on the market, but far from all deliver representative solutions.

For stationary lots, the benchmark study by Petersen et al. (2004) showed conslusively that only the *riffle-splitting principle* lead to Representative Mass Reduction (RMR). Riffle splitters have different physical manifestations; both stationary and roraty solutions exist.

For dynamic lots, lots in movement, the *Vezin sampler* is by far the most effective, fully representative RMR equipment in existence. The Vezin sampler is also superior regarding slurries a.o.

Representative Sampling: Theory of Sampling (TOS)

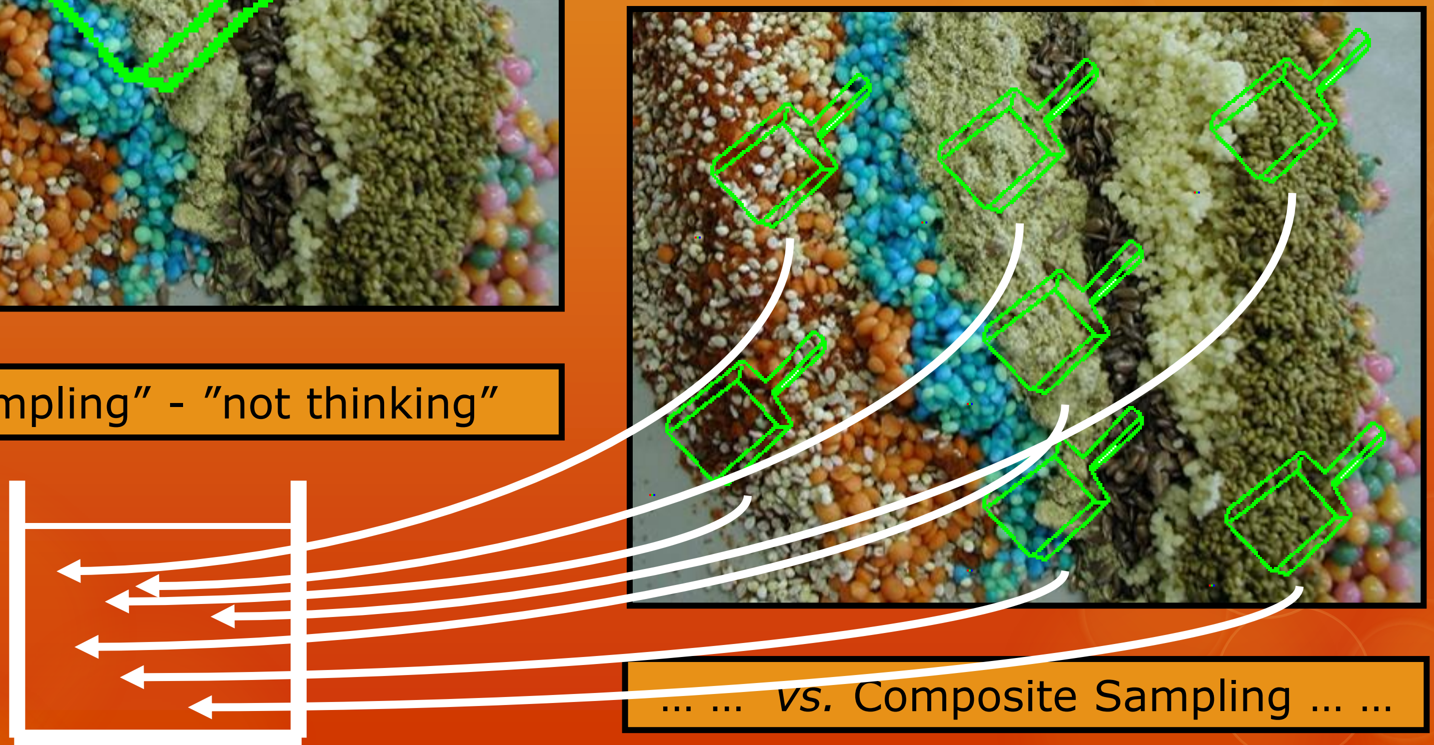
TOS - Axiomatic exposé

Governing principles (GP) – Sampling unit operations (SUO)

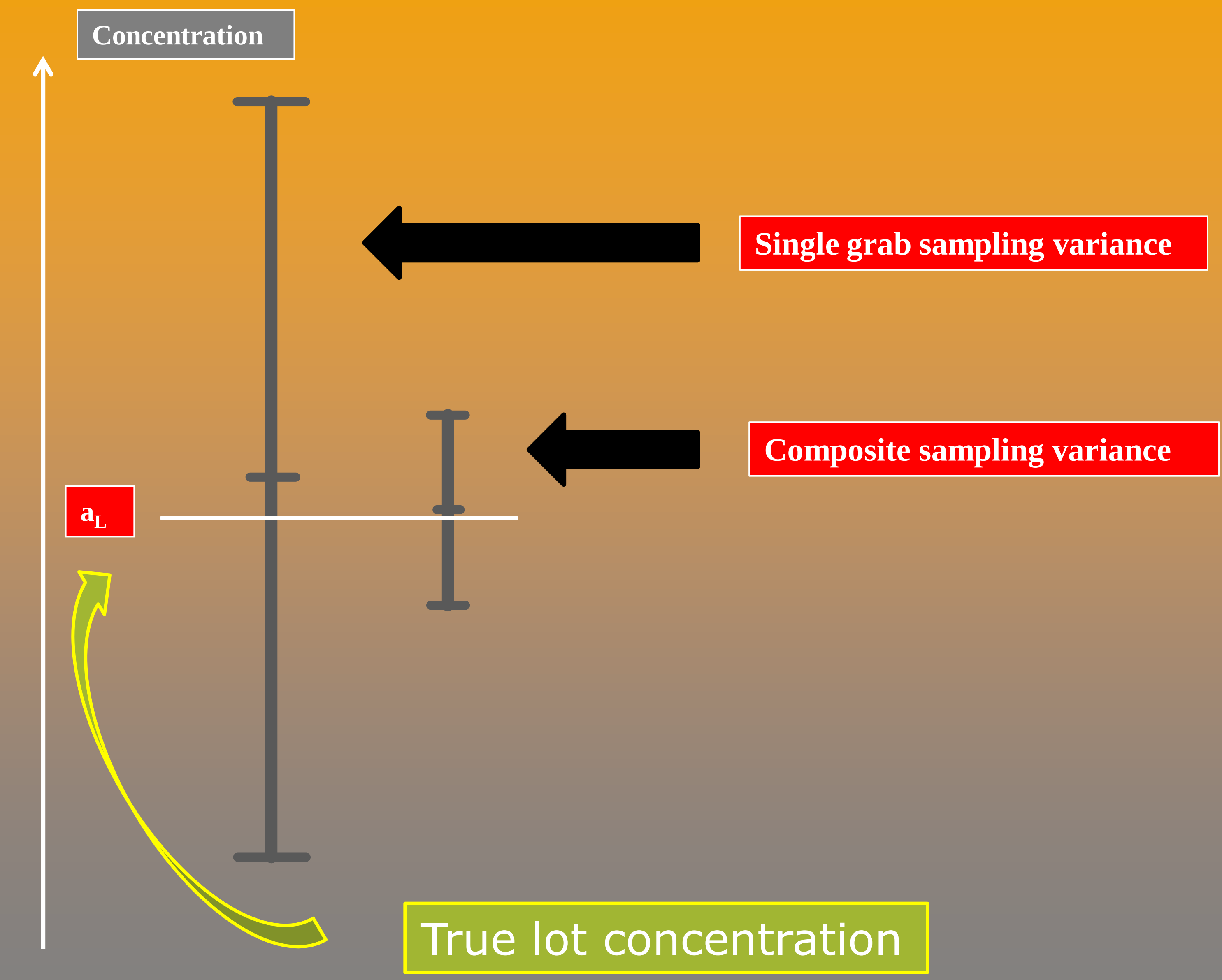
1. FSP: Fundamental Sampling Principle
2. SSI: Sampling Scale Invariance
3. PSC: Sampling Correctness (bias-free sampling)
4. PSS: Sampling Simplicity (primary sampling + mass-reduction)
5. LDT: Lot Dimensionality Transformation
6. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

Sampling Unit Operations: Composite Sampling

"Grab Sampling" - "not thinking"

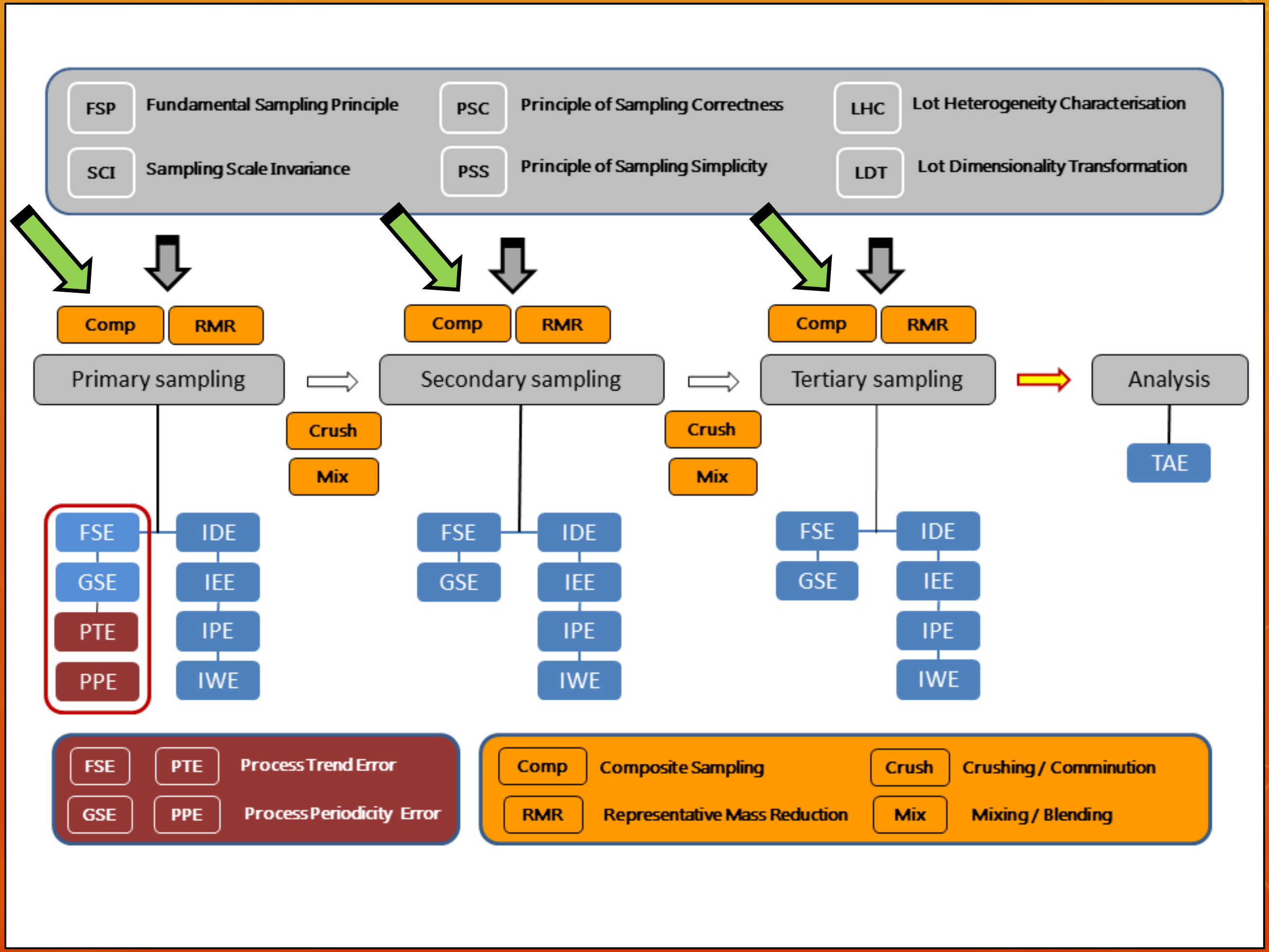


… … *vs.* Composite Sampling … …



Concentration
Single grab sampling variance
Composite sampling variance
a_L
True lot concentration

Theory of Sampling (TOS) – everything in a glance



Representative Sampling: Theory of Sampling (TOS)

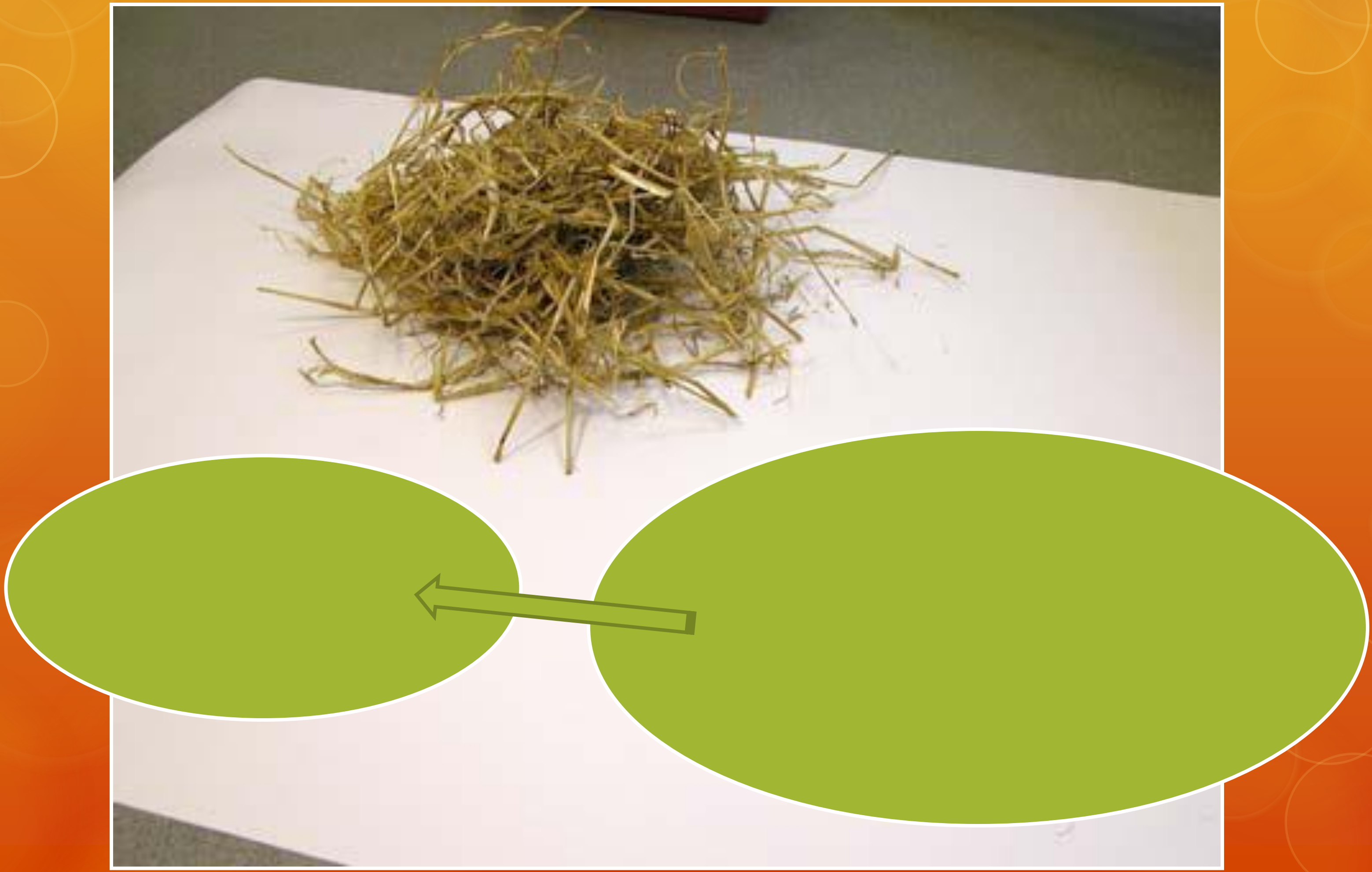
TOS - Axiomatic exposé

Governing principles (GP) – Sampling unit operations (SUO)

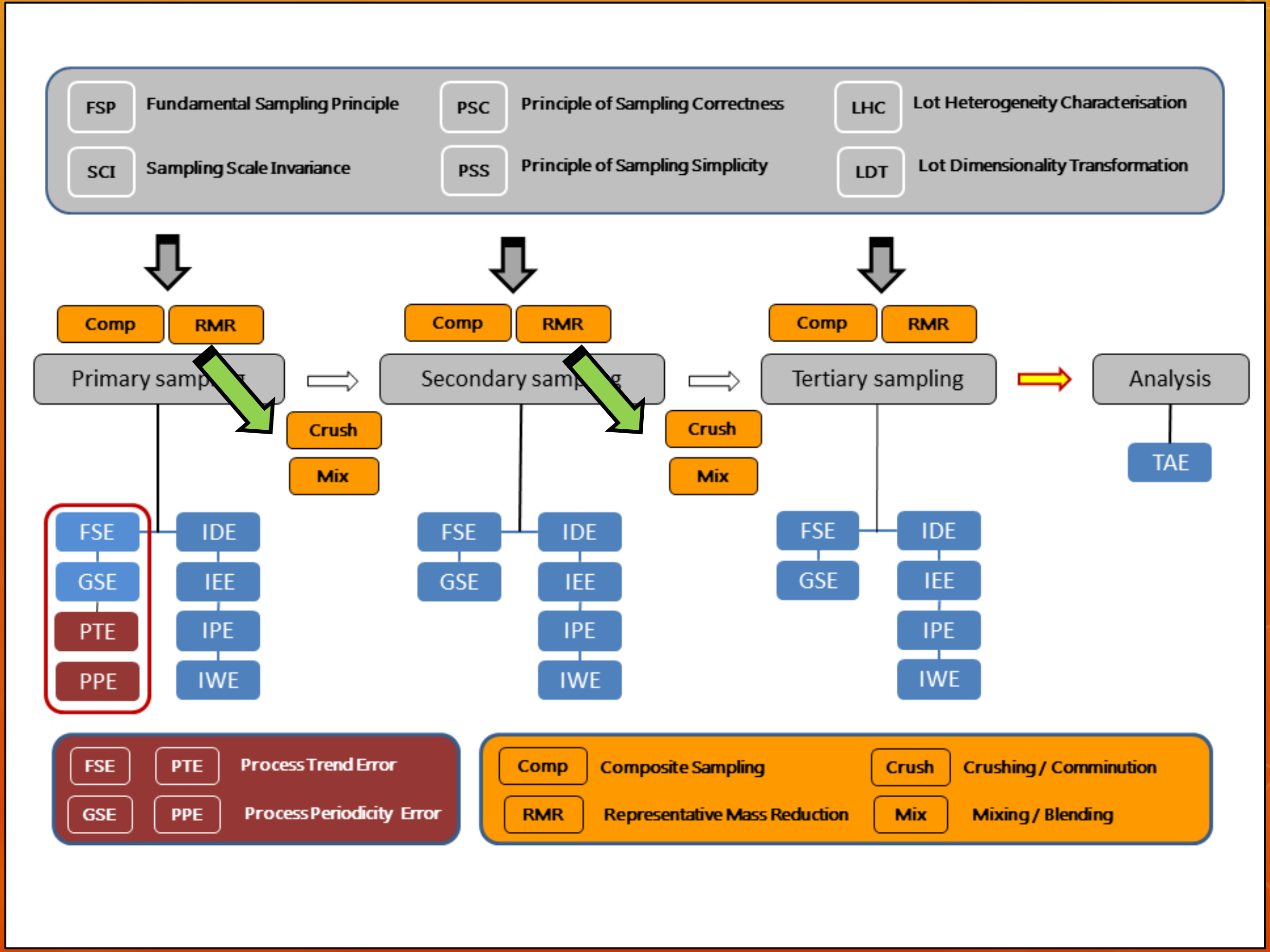
1. FSP: Fundamental Sampling Principle
2. SSI: Sampling Scale Invariance
3. PSC: Sampling Correctness (bias-free sampling)
4. PSS: Sampling Simplicity (primary sampling + mass-reduction)
5. LDT: Lot Dimensionality Transformation
6. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

Crushing / Maceration

REALLY difficult materials ... Que faire?

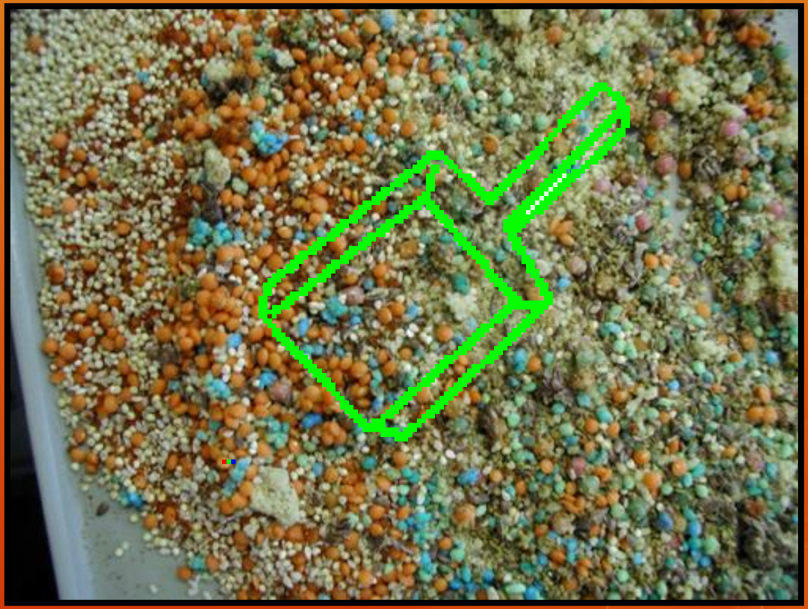


Theory of Sampling (TOS) – everything in a glance



Sampling Unit Operations: Mixing / Blending

A pro pos CRUSHING + MIXING before further sub-sampling:




KE

In general one cannot cannot crush the entire primary lot!

But *sometimes*, you do this anyway. BIG advantages ahead!

Sampling Unit ration: Comminution

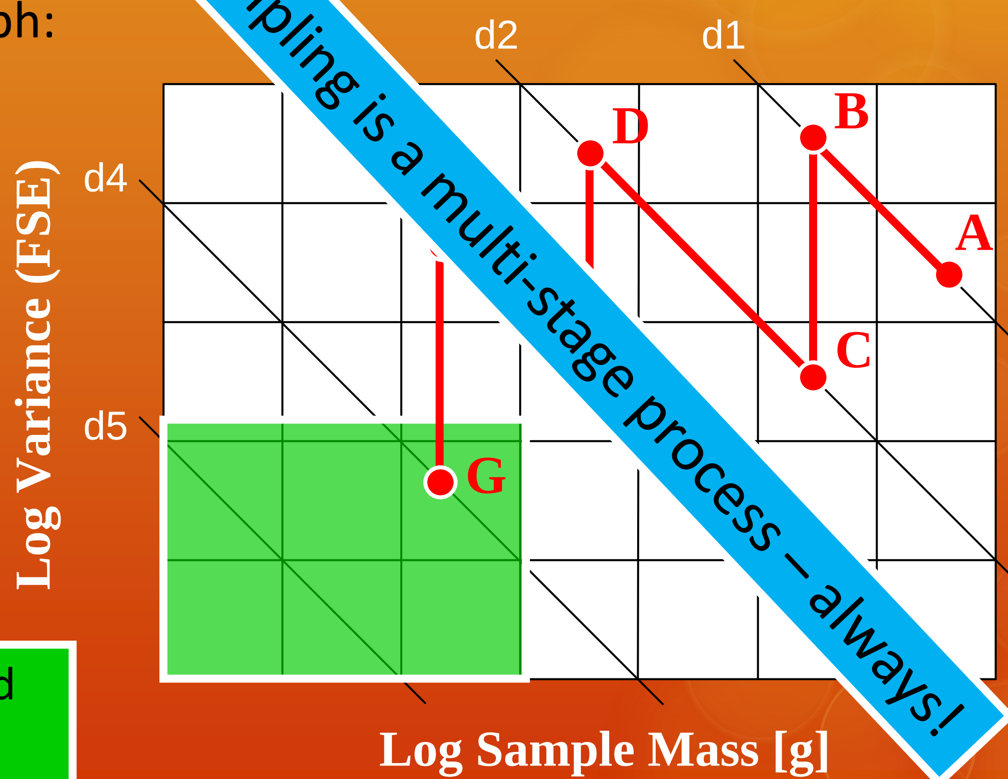
Representative sampling is a multi-stage process – always!

Particle size reduction (crushing)

Sampling nomograph:

$$\mathrm{var(FSE)} = C \times d^3 \left(\frac{1}{M_S} - \frac{1}{M_L} \right)$$

size d1 > d2 > d3 > d4 > d5



Acceptable treshold region of sampling error vs sample mass

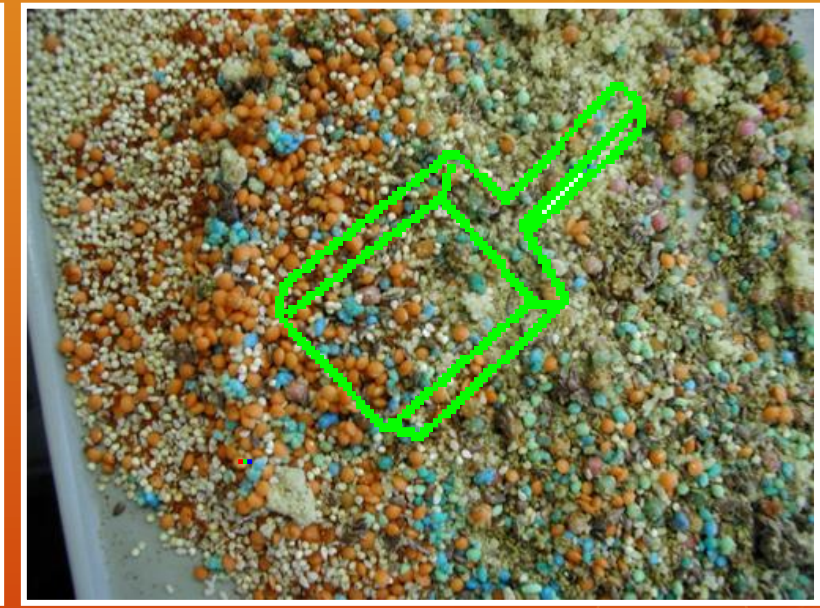
Representative Sampling: Theory of Sampling (TOS)

TOS - Axiomatic exposé

Governing principles (GP) – Sampling unit operations (SUO)

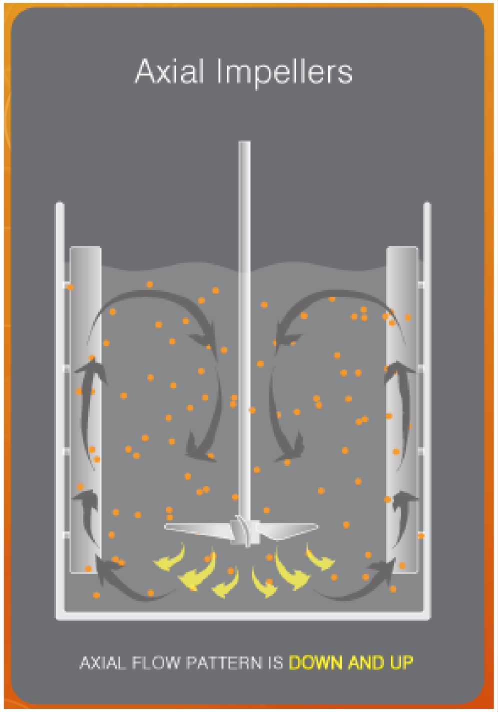
1. FSP: Fundamental Sampling Principle
2. SSI: Sampling Scale Invariance
3. PSC: Sampling Correctness (bias-free sampling)
4. PSS: Sampling Simplicity (primary sampling + mass-reduction)
5. LDT: Lot Dimensionality Transformation
6. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

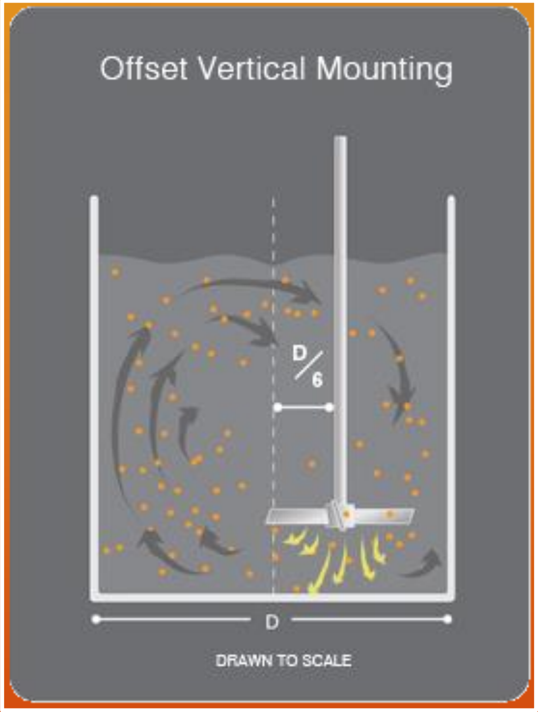
Sampling Unit Operations: Mixing / Blending



Mixing – intuitively clear. However, forceful mixing is a much less effective process than commonly known ...

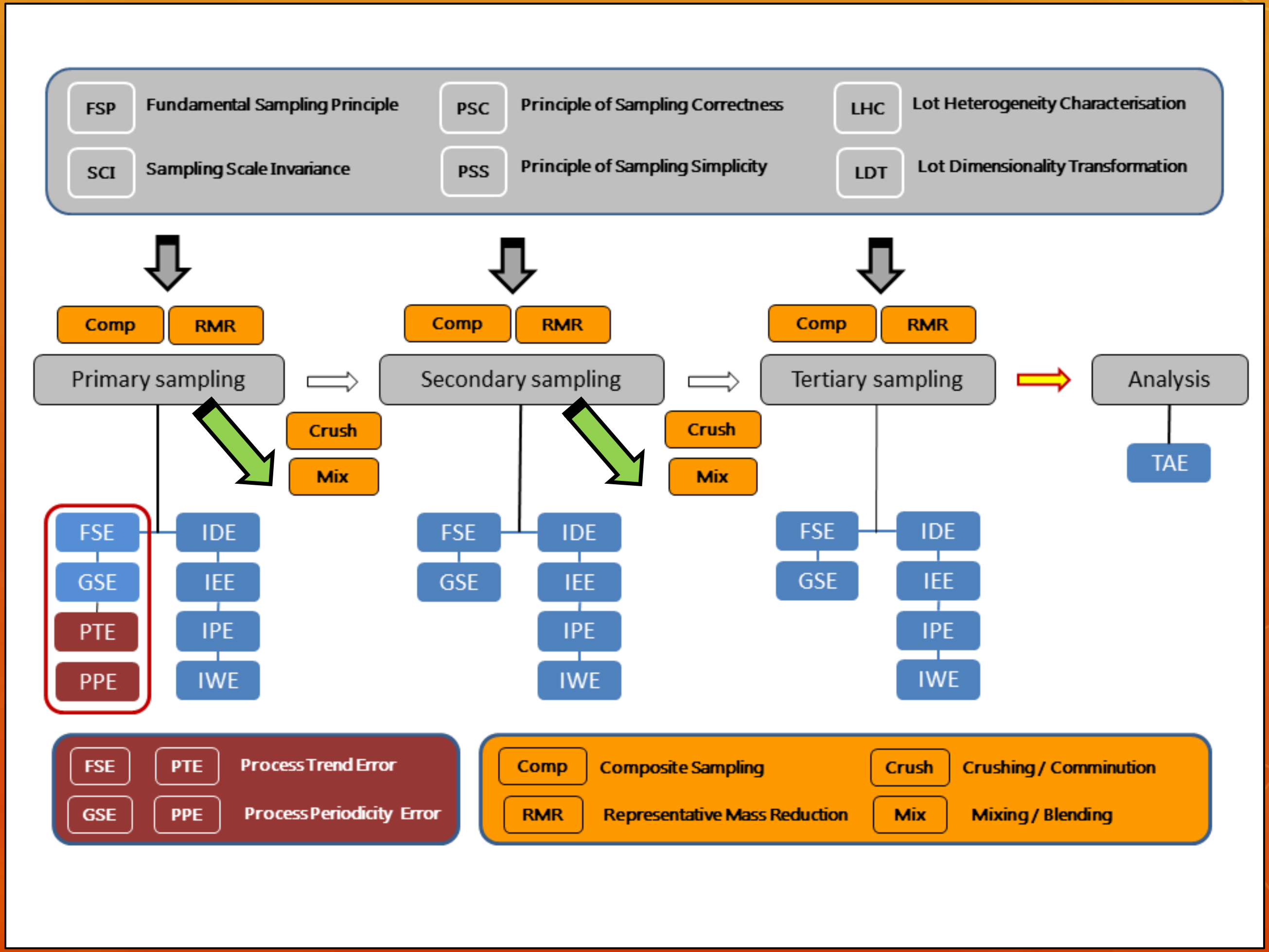
Mixing – It is manifestly not enough just ”to mix” and then hope all is OK





Mixing – It is always necessary to conduct a baseline *validation* of the mixing/blending operation in use, e.g. A Replication Experiment (RE).

Theory of Sampling (TOS) – everything in a glance



Sampling Unit Operation: Mixing / Blending

Reduces contributions to sampling variation from the Grouping and Segregation Error (GSE)

$$s^2(GSE) = \zeta \cdot \gamma \cdot s^2(FSE)$$

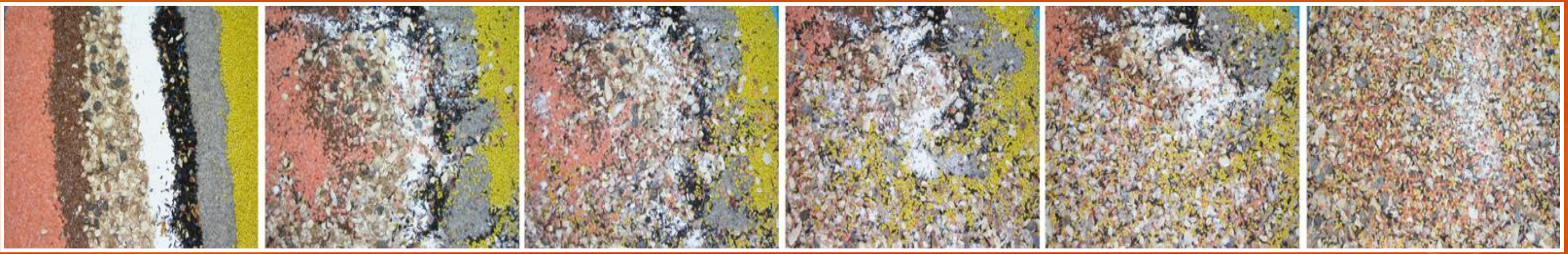
$$\gamma = \frac{N_F - N_G}{N_G - 1} \approx \boxed{\frac{N_F}{N_G}}$$

Segregation factor

Grouping factor (unaffected by mixing - reduced only by selecting smaller increments)

$\zeta \approx 1$

$\zeta \approx 0$



Four practical Sampling Unit Operations (SUO)

1. Composite Sampling
2. Particle Size Reduction (comminution)
3. Mixing / blending
4. Representative Mass Reduction (- sample preparation)

Theory & Practise of Sampling (TOS)

Practical procedures … …

Four Sampling Unit Operations (SUO)

Used as active steps in the sampling process (often used several times, in combination)

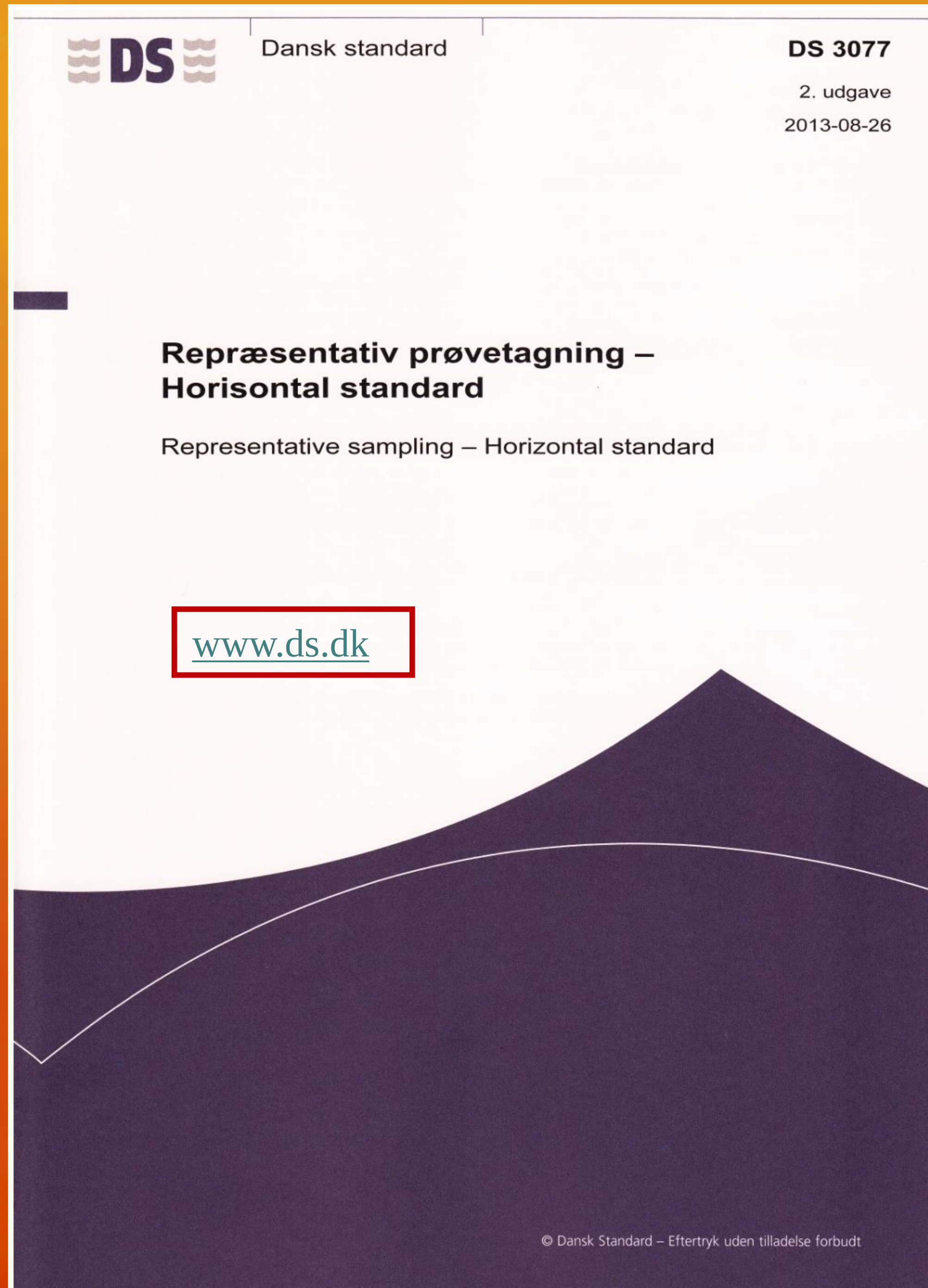
Fundamental insights - I

- The empowering role of universal principles & SUO’s

Governing principles (GP) & Sampling Unit Operations (SUO)

1. FSP: Fundamental Se Invariance
2. PSC: Sampling Correctness (bias-free sampling)
3. PSS: Sampling Simplicityampling Principle
4. SSI: Sampling Scal (primary sampling + mass-reduction)
5. LDT: Lot Dimensionality Transformation
6. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (sub-sampling/splitting)

All GP's & SUO's are <u>not</u> involved in <u>all</u> sampling tasks. The analysis & the sampling objective determines which GP's and SUO's to use. The Theory of Sampling (TOS) to the fore ... DS 3077: First horisontal standard (2013)



Dansk standard

DS 3077

2. udgave

2013-08-26

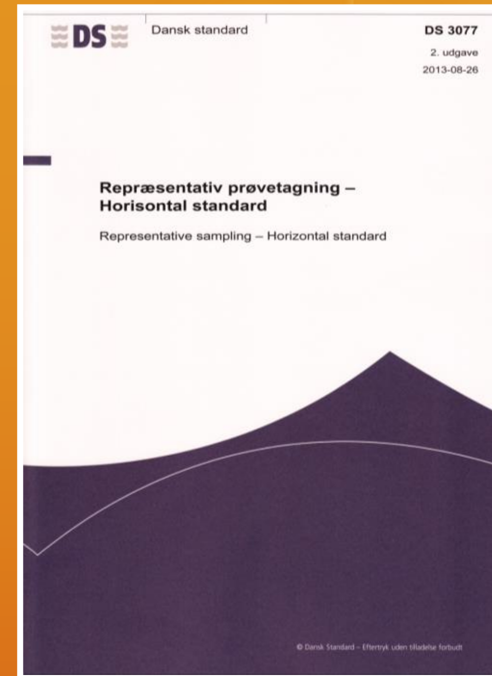
Repræsentativ prøvetagning – Horisontal standard

Representative sampling – Horizontal standard

www.ds.dk

Horizontal – Representative Sampling

DS-3077



DS Dansk standard

DS 3077
2. udgave
2013-08-26

**Repræsentativ prøvetagning –
Horisontal standard**

Representative sampling – Horizontal standard

© Dansk Standard – Eftertryk uden tilladelse forbudt

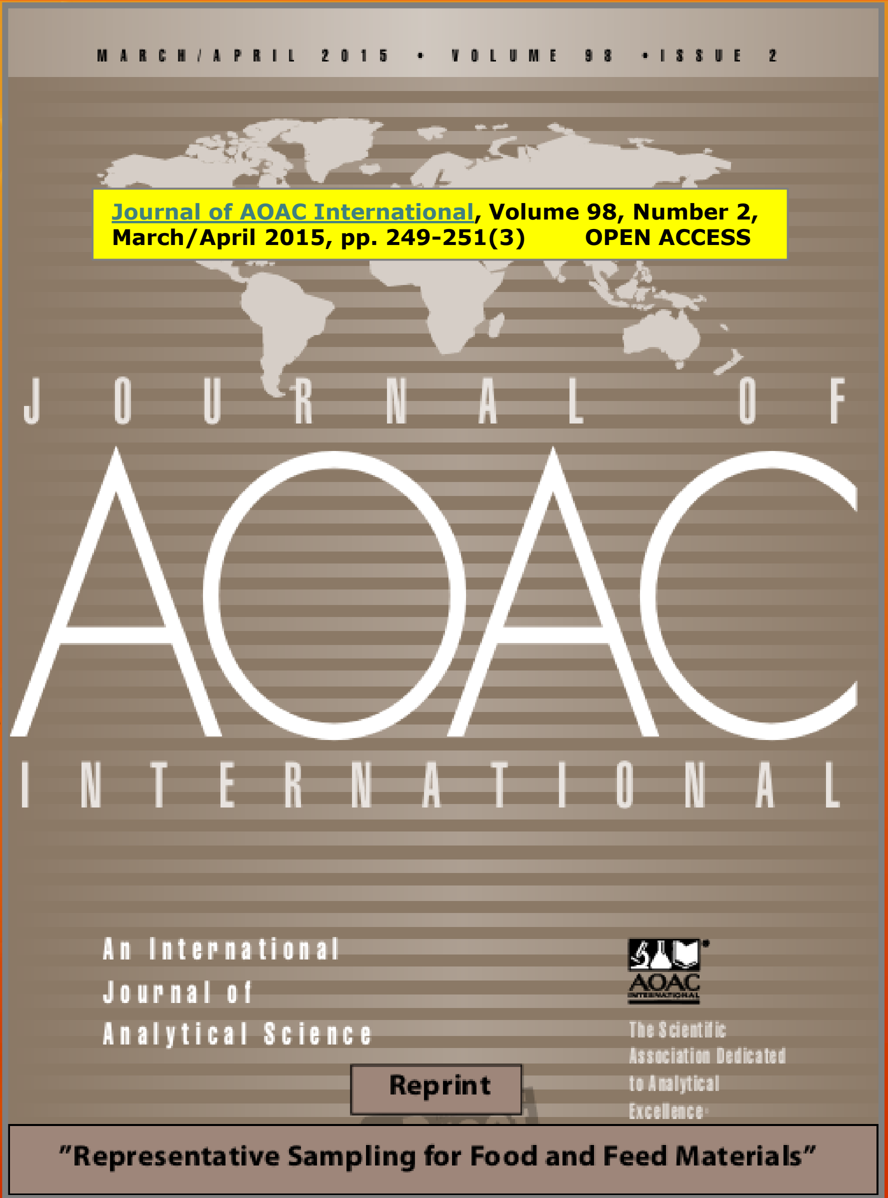
This standard outlines a practical, iterative, self-controlling approach with minimal complexity, based on the Theory of Sampling (TOS). The generic sampling process described and all elements involved are sufficient and necessary for the stated objective, with the consequence that no exceptions can 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.

An important analogy (for *some* ..)

1. $\nabla \cdot \mathbf{D} = \rho_V$
2. $\nabla \cdot \mathbf{B} = 0$
3. $\nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t}$
4. $\nabla \times \mathbf{H} = \dfrac{\partial \mathbf{D}}{\partial t} + \mathbf{J}$

Maxwell's Equations describe the world of electromagnetics. The four equations describe how electric and magnetic fields *propagate*, *interact*, and how they are *influenced* by objects

2015: SPECIAL SESSION ON J. AOAC



World's first attempt to bring the Theory of Sampling (TOS) and foods/feed safety assessment together!

Transatlantic Special Section taskforce, Oct. 2014, Windsor, CO:
Nancy Thiex, Kim H. Esbensen, Charles A. Ramsey, Claas Wagner, Claudia Paoletti.

Fundamental insights - II

- The role of *statistics* – in sampling and analysis

A (most) surprising and critical distinction

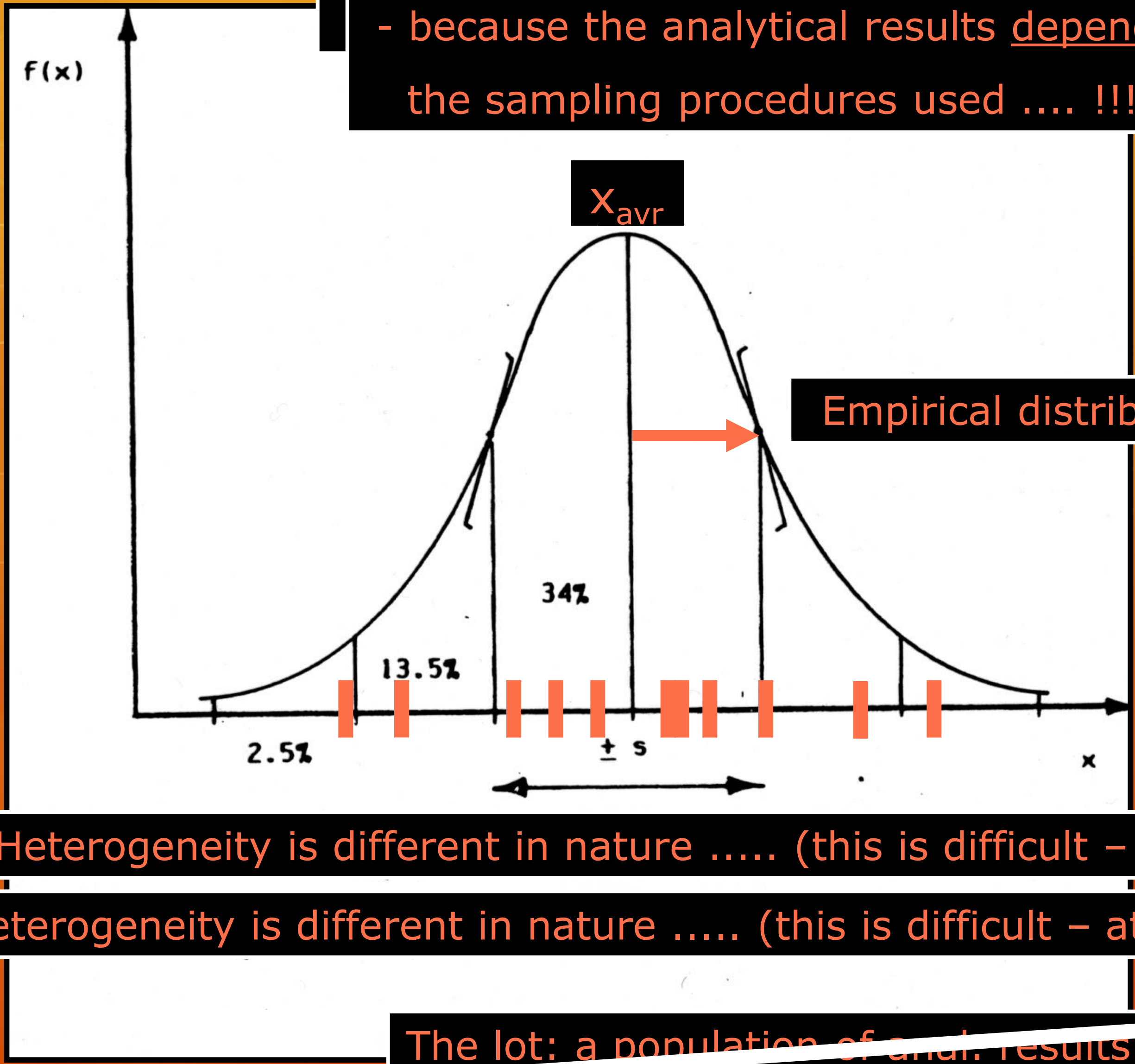
Because the sampler is free to choose between alternative sampling procedures, - equipment and – conditions ....

- based, or <u>not</u> based, on a competent understading and accept of the adverse role of heterogeneity when ***<u>interacting</u>*** with a specific sampling process ....

- the analytical result will <u>depend</u> on these choices!

All analytical result will depend on the specific sampling procedure(s) used to deliver the analytical aliquot!

- very difficult to avoid conventional
- because the analytical results <u>depend</u> on the sampling procedures used .... !!! !!!



Empirical distribution of anal. results

Heterogeneity is different in nature ..... (this is difficult – at first)

Heterogeneity is different in nature ..... (this is difficult – at first)

~~The lot: a population of anal. results~~

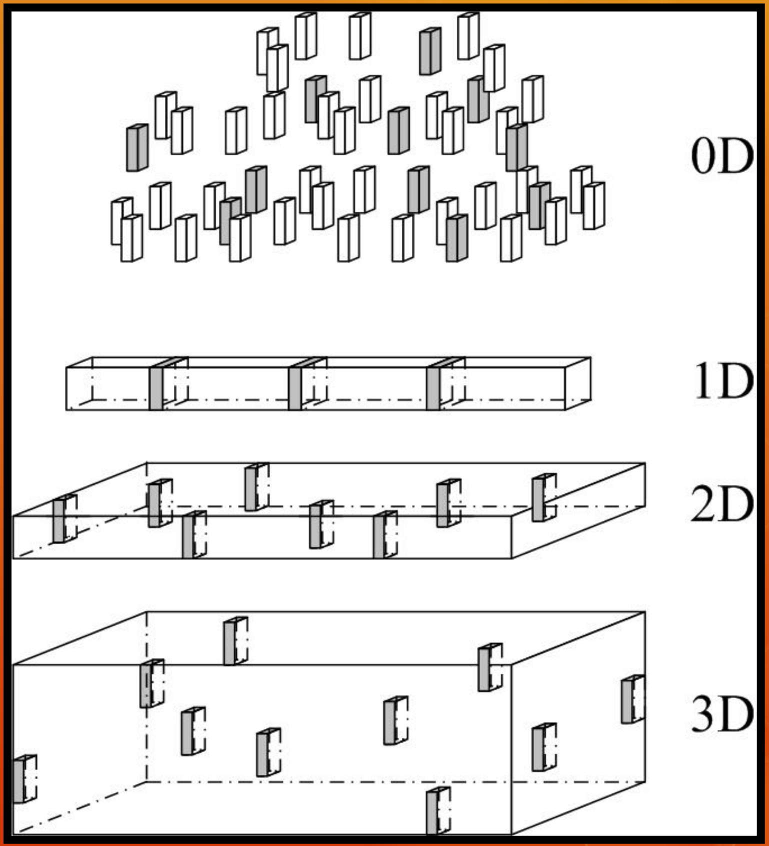
Representative Sampling: Theory of Sampling (TOS)

TOS - Axiomatic exposé

Governing principles (GP) – Sampling unit operations (SUO)

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. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

Lot Dimensionality (definition via practical *increments*)



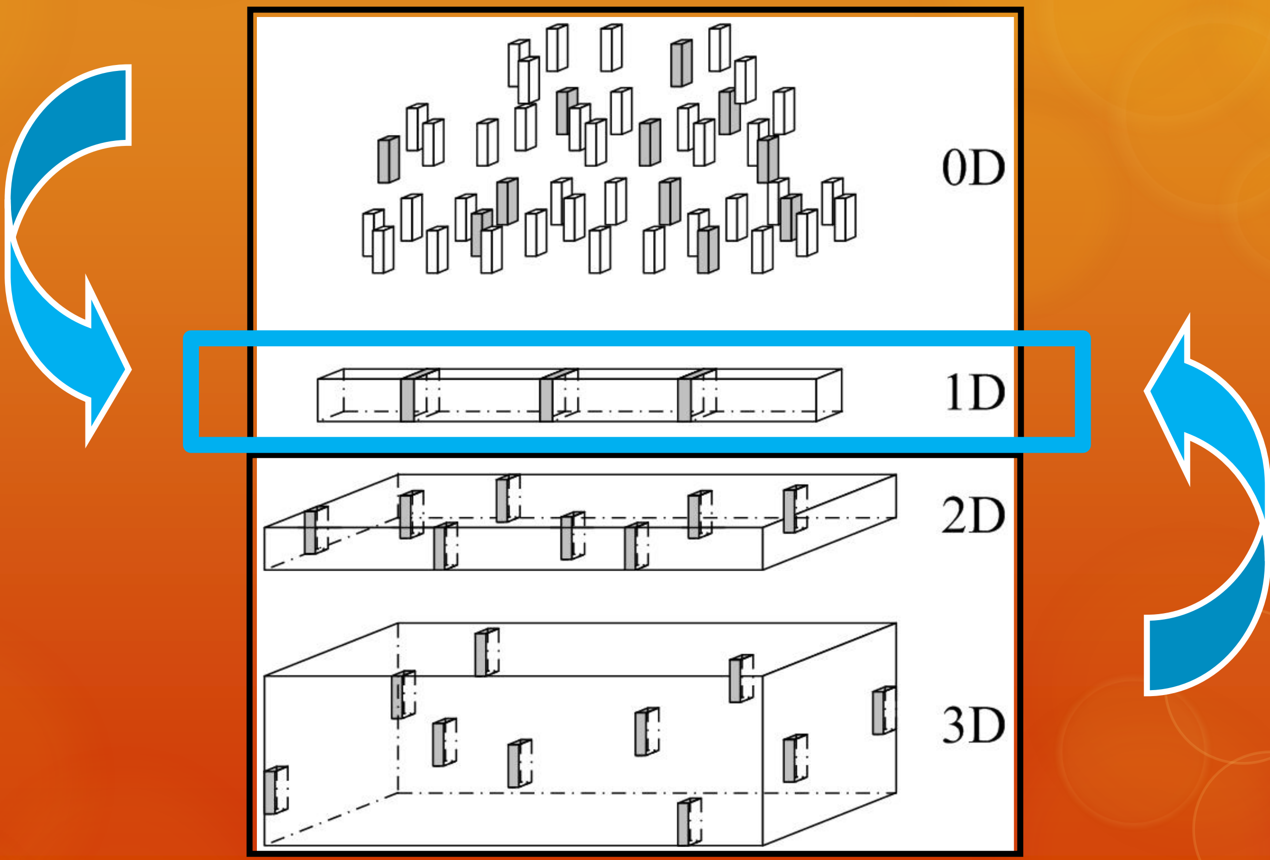
No correlation exists between increments. Total access to the complete lot volume. Increments can be picked at will without untoward effort or difficulty (practical def)

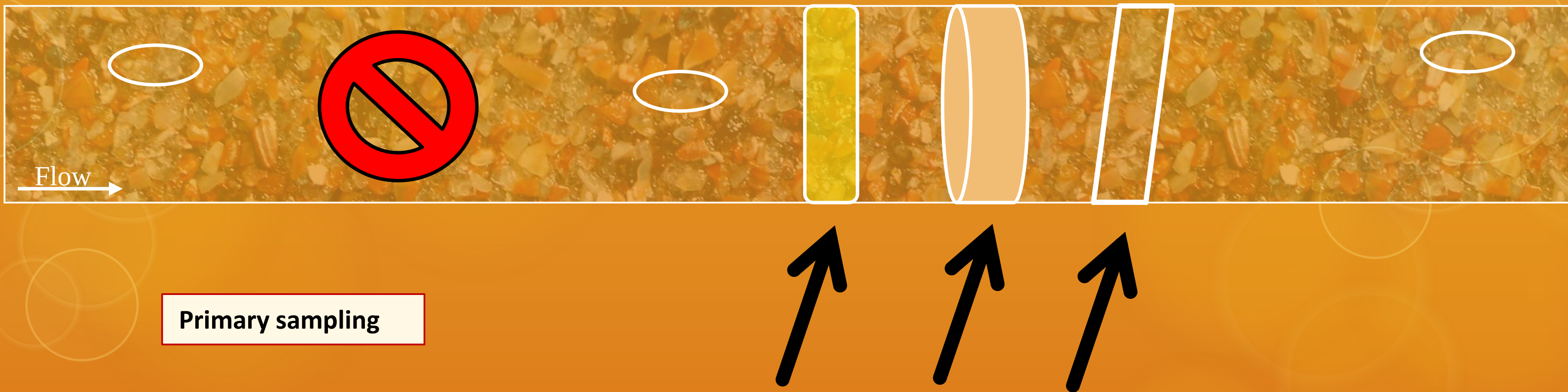
Increments only cover *two* lot dimensions (e.g. a cross-wise planar-parallel “slice”)

Increments only cover *one* lot dimension (e.g. a drill-core; a profile)

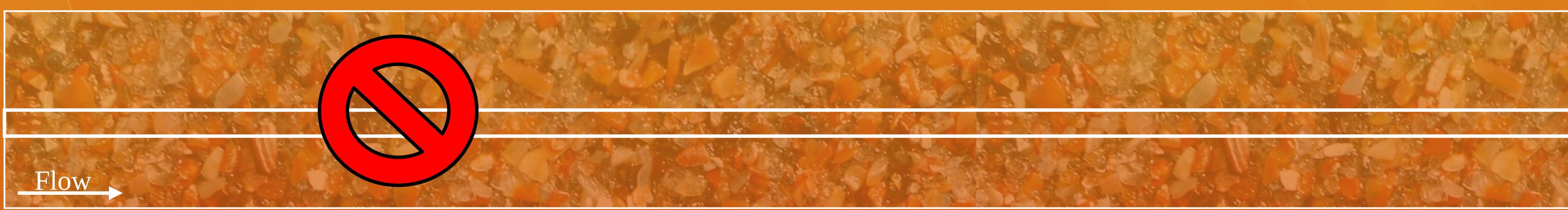
Increments do <u>not</u> fully cover <u>any</u> of the lot dimensions (e.g. a “depth sample”)

LDT: Lot Dimensionality Transformation





Primary sampling



Primary sampling

Fundamental Sampling Principle (FSP): 1-D lots

Fundamental Sampling Principle (FSP):

All increments must have the *same* (non-zero) probability of ending up in the sample

EXAMPLE: lot dim. transformation (LDT)

Optimal sampling localisation: 1-D transportation "bottle-necks"

Off-loading - or

transportation …

from 3-D lots

TOS-correct

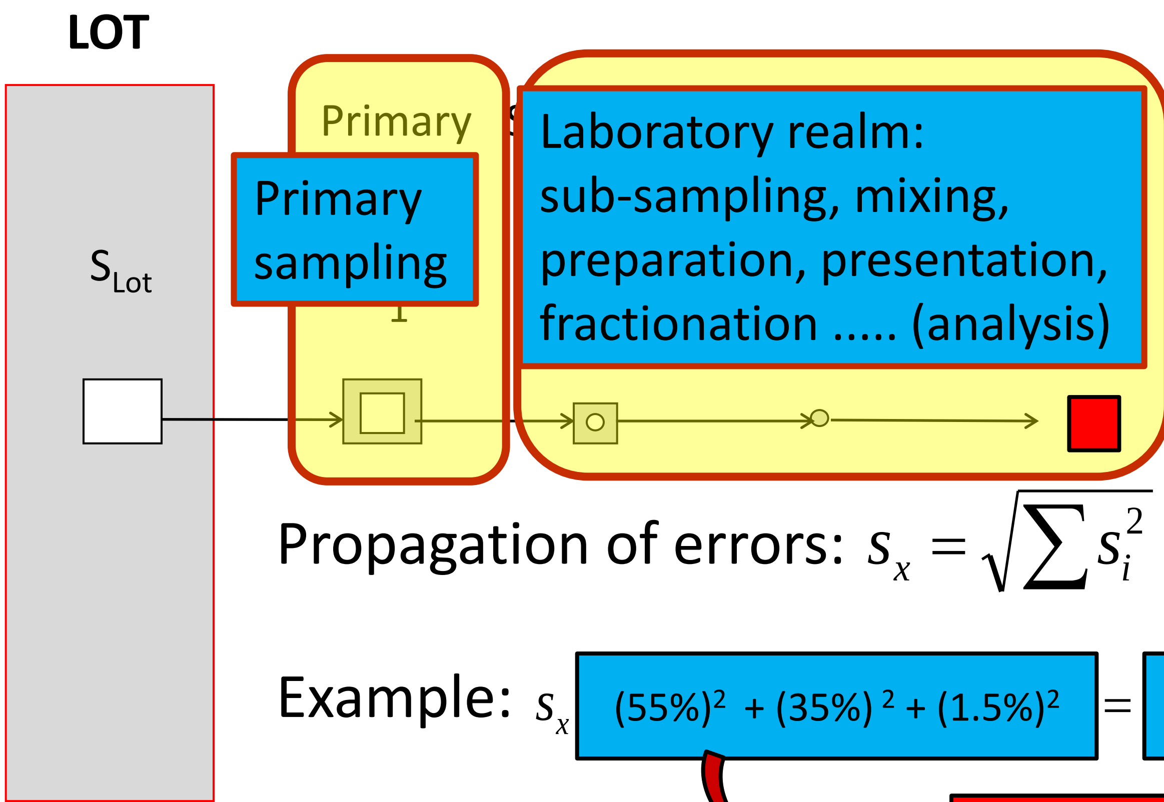
sampling site: 1-D

Representative Sampling: Theory of Sampling (TOS)

TOS - Axiomatic exposé

Governing principles (GP) – Sampling unit operations (SUO)

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. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)



Propagation of errors: $s_x = \sqrt{\sum s_i^2}$

Example: s_x $\boxed{(55\%)^2 + (35\%)^2 + (1.5\%)^2} = \boxed{(65\%)^2}$

Goal: $S_X \curvearrowright S_{Lot}$

Who's responsible .... ?

The analytical process always contains several sampling and preparation steps, but <u>usually</u> primary sampling <u>dominates</u>

Representative Sampling: Theory of Sampling (TOS)

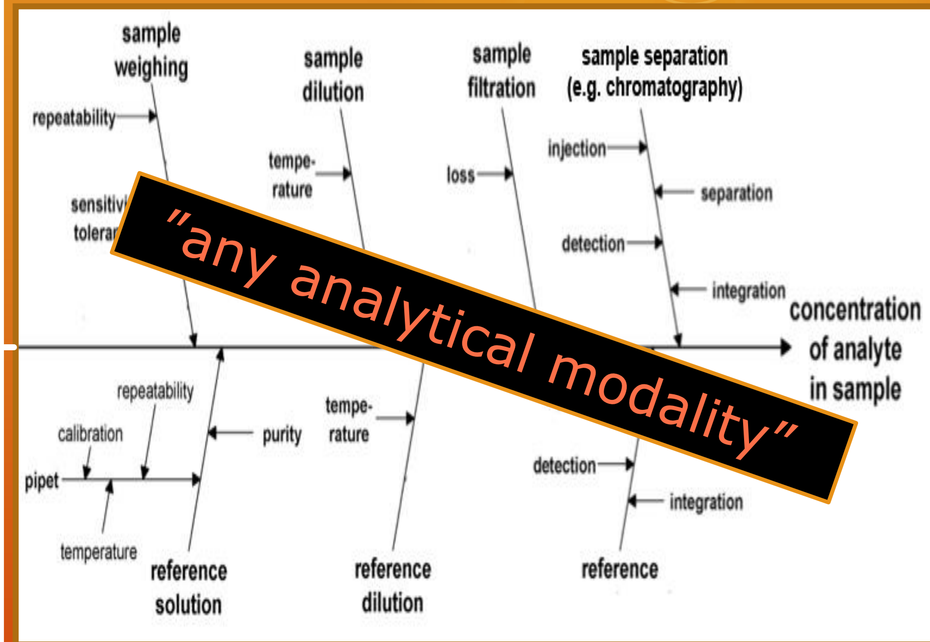
TOS - Axiomatic exposé

Governing principles (GP) – Sampling unit operations (SUO)

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. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

Theory of Sampling (TOS)

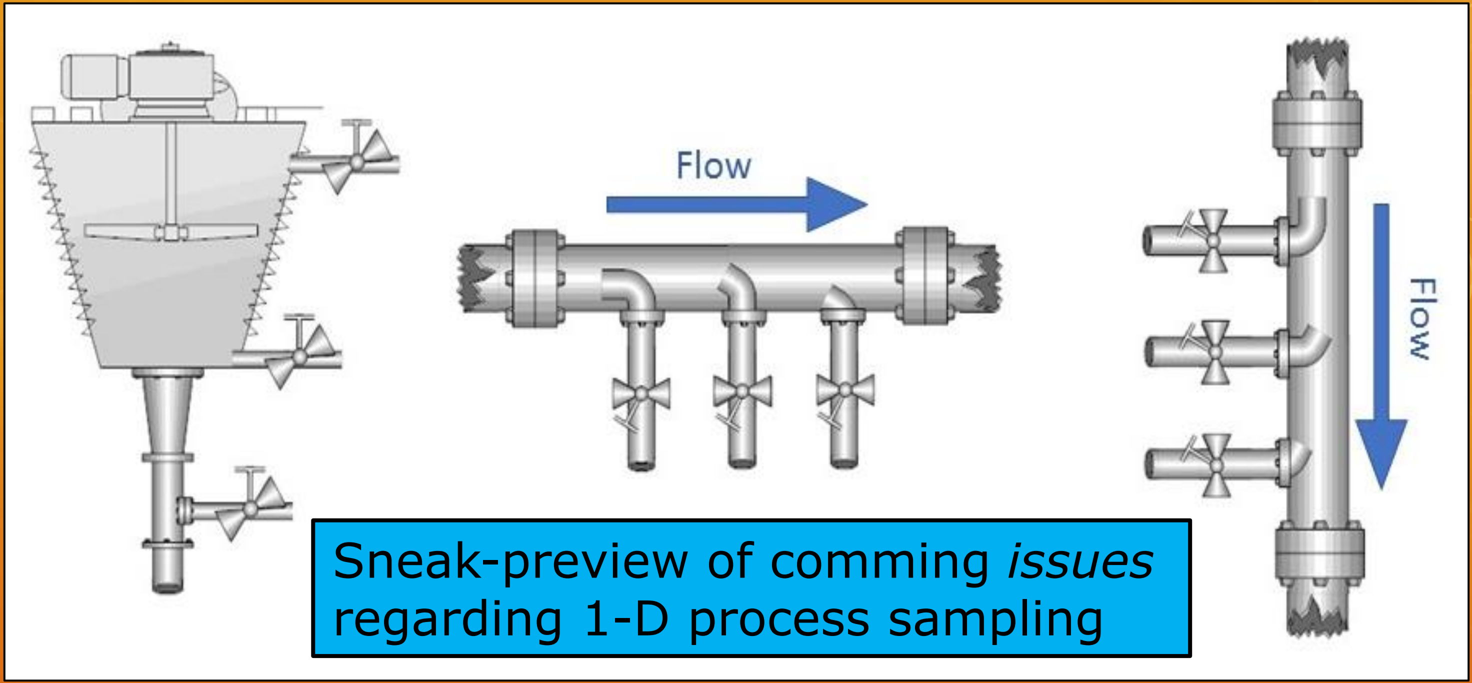
Measurement Uncertainty (MU)



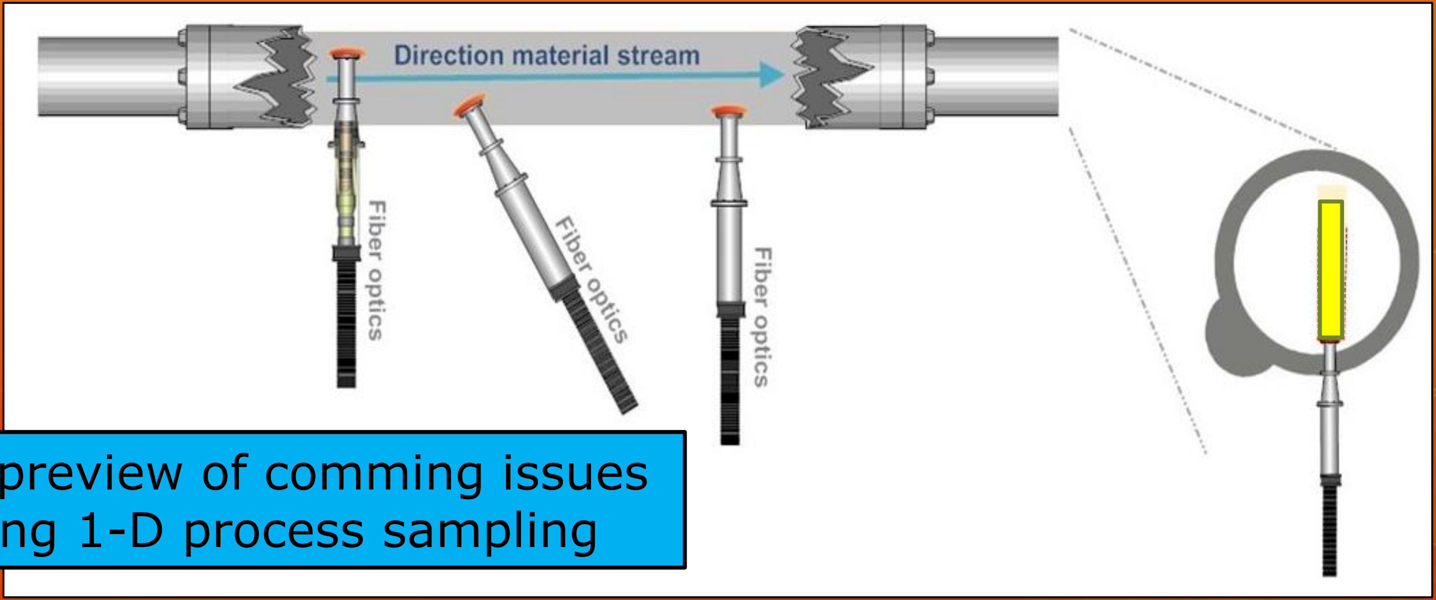
$$MU_{total} = MU_{sampling} + MU_{analysis}$$

Fundamental insights - III

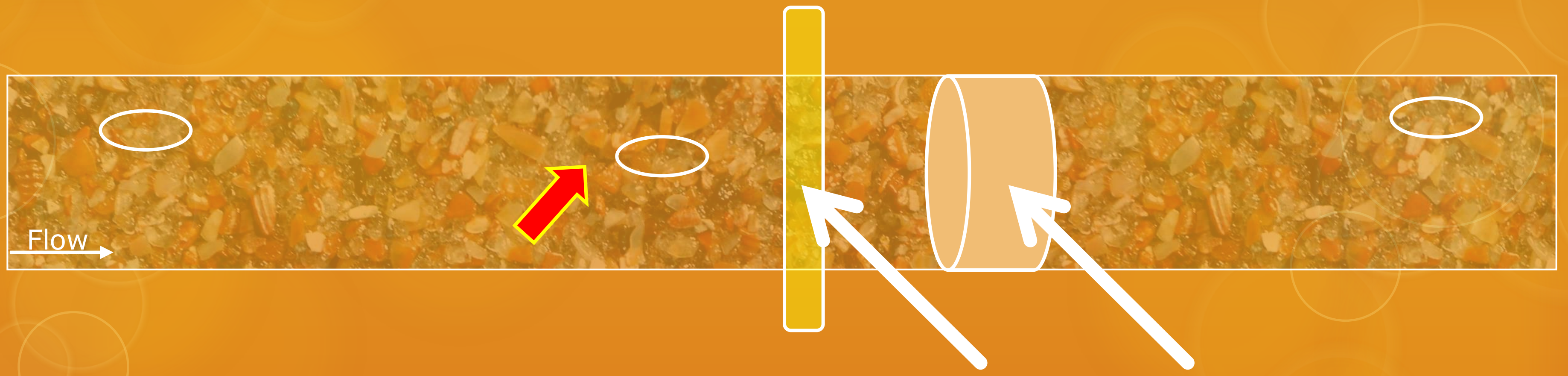
- The empowering role of universal principles & SUO's –
- Sampling of <u>stationary lots</u> vs. <u>moving lots</u> (*process sampling*)
- Illustration of the origin of Incorrect Sampling Errors (ICS) - and their effects ..... (unnecessary, but fatal bias)



Sneak-preview of comming *issues* regarding 1-D process sampling

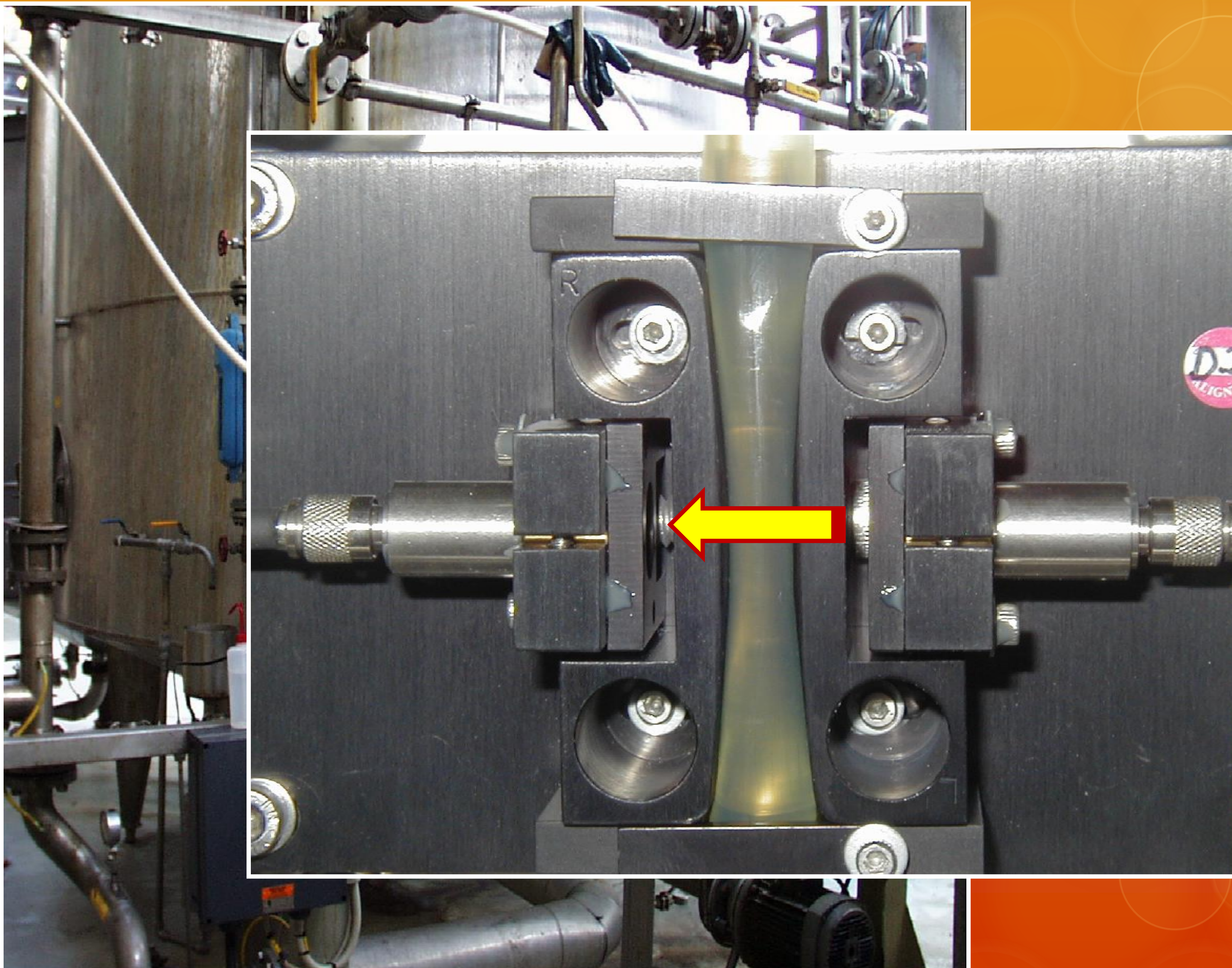


Sneak-preview of comming issues regarding 1-D process sampling


Flow


Flow

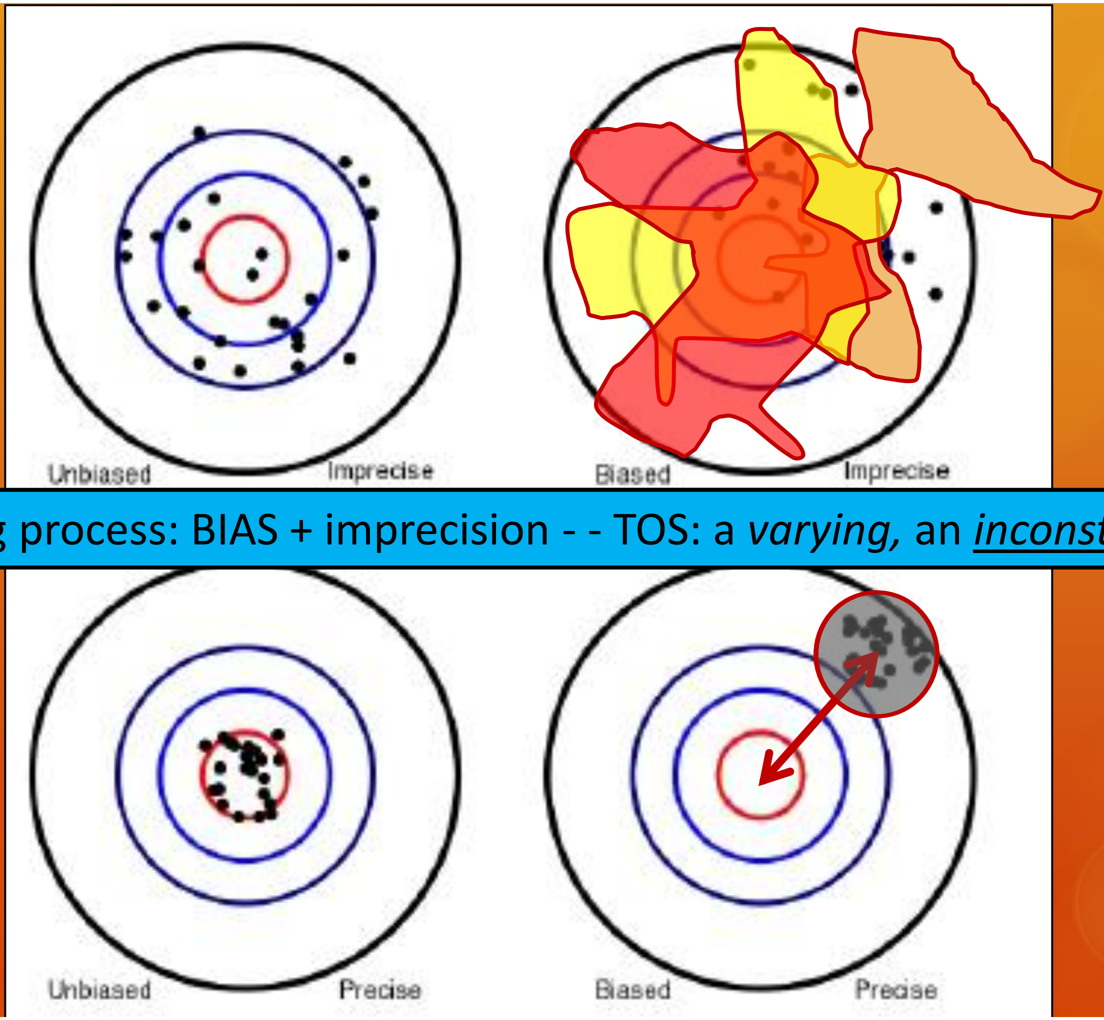

Flow


R

Fundamental insights - IV

- The most important distinction: analysis vs. sampling bias

Critical distinction between analysis (*sensu strictu*) – and sampling_plus_analysis! MUST be understood!



Sampling process: BIAS + imprecision - - TOS: a *varying,* an *inconstant* bias!

Analytical process: bias + imprecision (statistical concept: a constant bias)

Representative Sampling: Theory of Sampling (TOS)

TOS - Axiomatic exposé

Governing principles (GP) – Sampling unit operations (SUO)

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. The Replication Experiment, RE(**r**)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

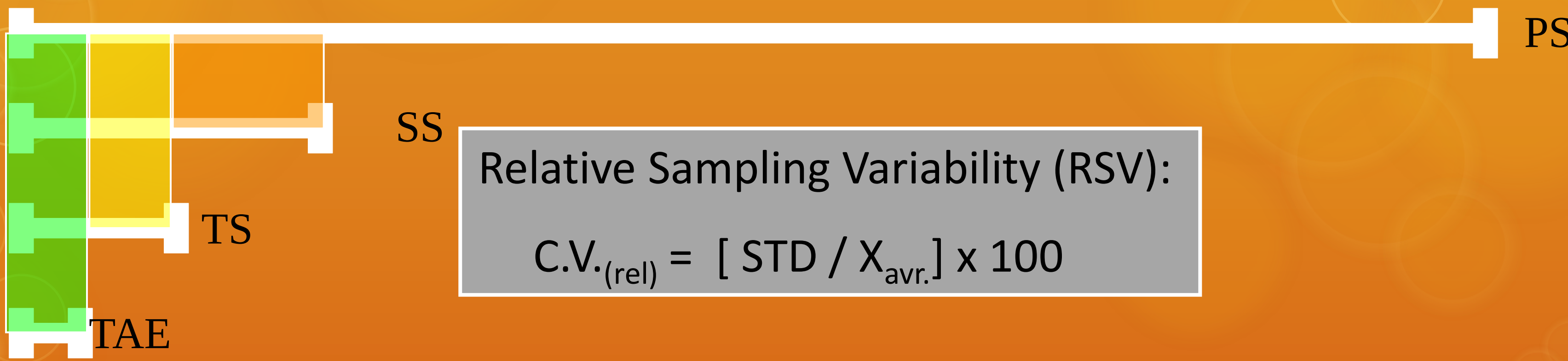
The Replication Experiment, RE(r)

HOW TO? - What you can do one time:

-you can <u>replicate</u>, f.eks. 10 times

Replication Experiments need <u>only</u> to be carried out for *each* ***principal*** type of material as sampled by a ***specific*** sampling procedure ... NB NB NB NB

Relative Sampling Variance (RSV) – C.V.$_{(rel)}$



Relative Sampling Variability (RSV):

C.V.$_{(rel)}$ = [STD / X$_{avr.}$] x 100

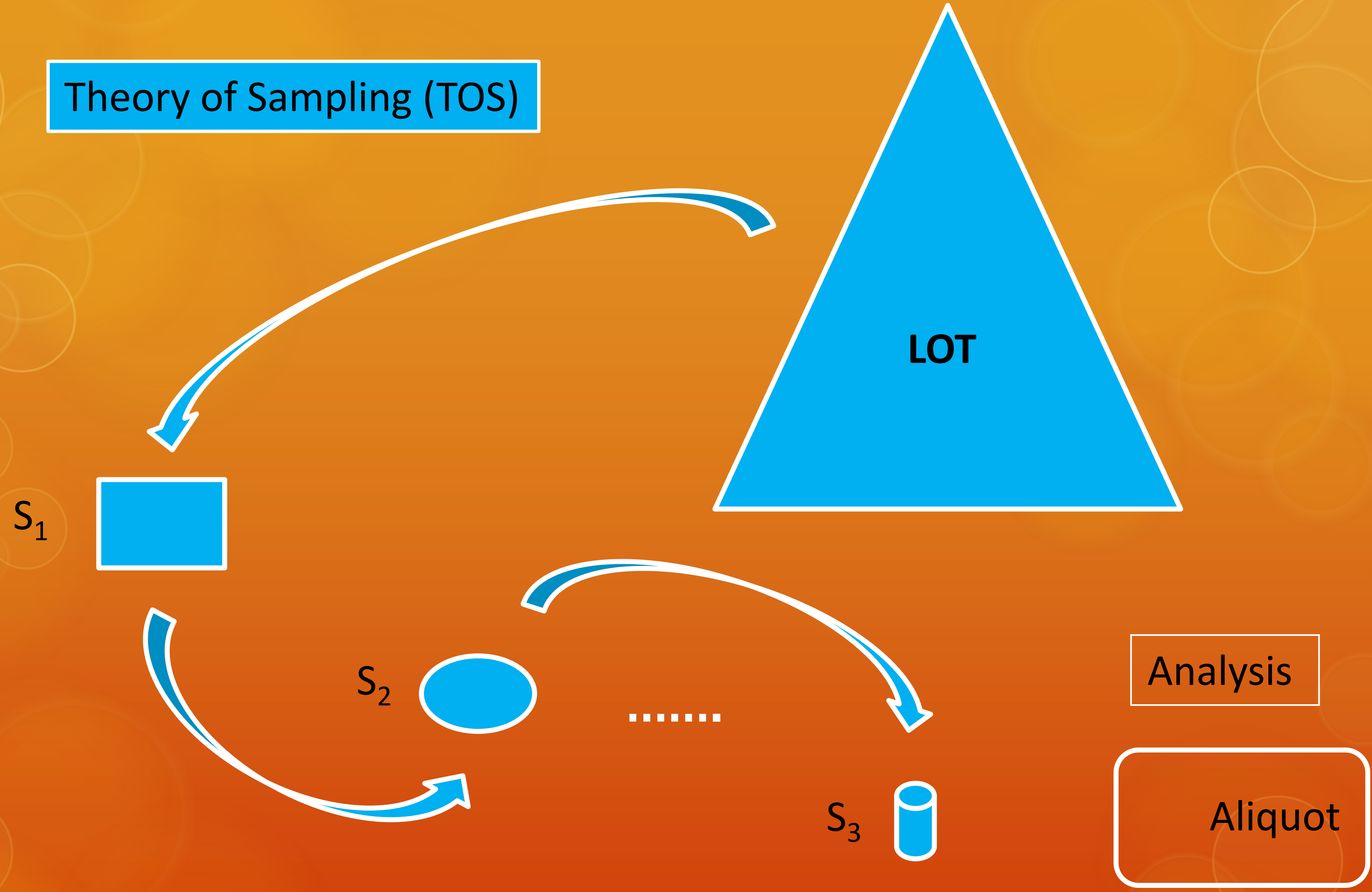
Relative Sampling Variability (RSV), aka

relative Coefficient of Variation, C.V.$_{(rel)}$

at all scales

Replication Experiment: RSV[%] characterises all sampling + analysis operations – for all types of equipment and all types of materials !!!

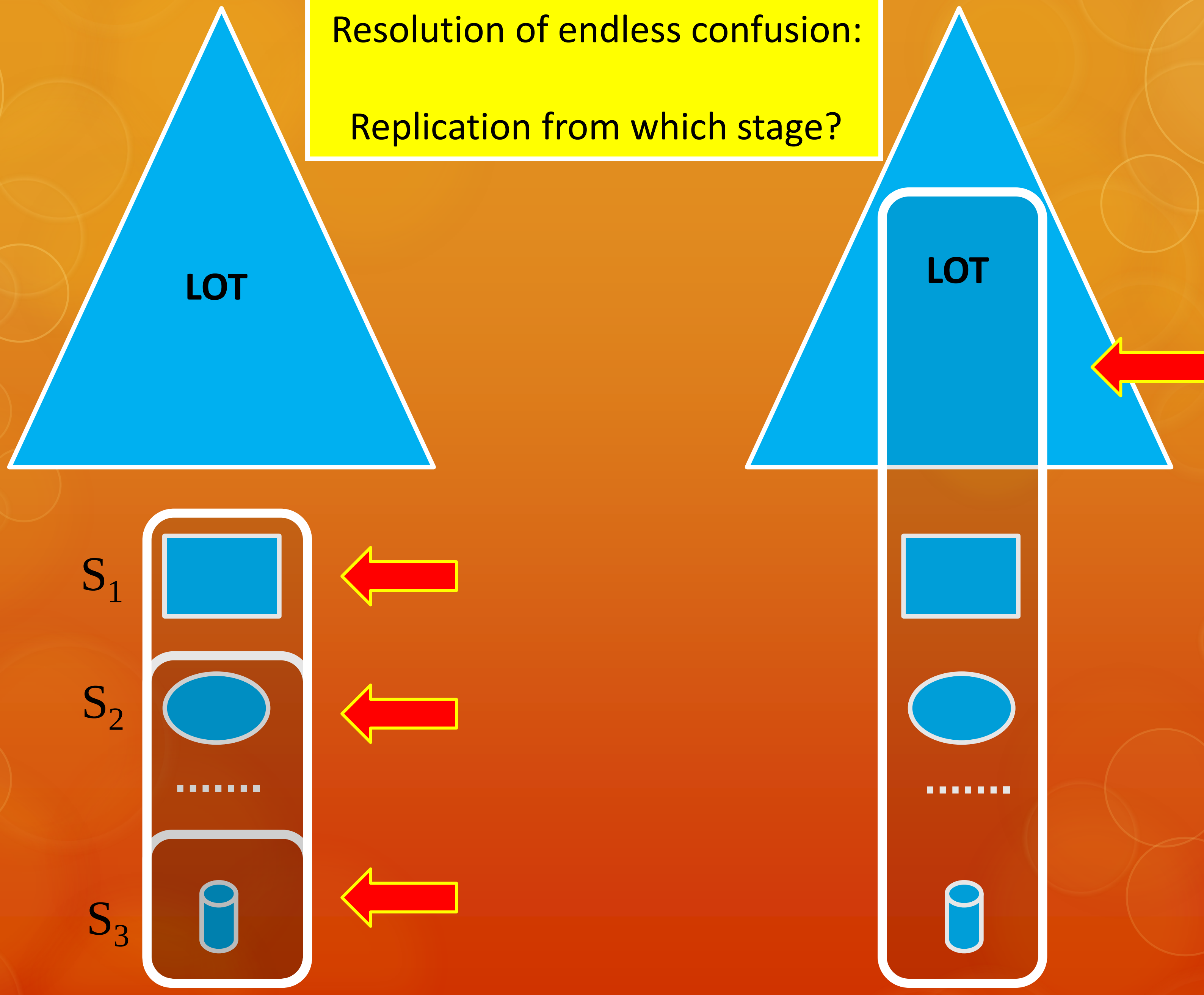
Theory of Sampling (TOS)

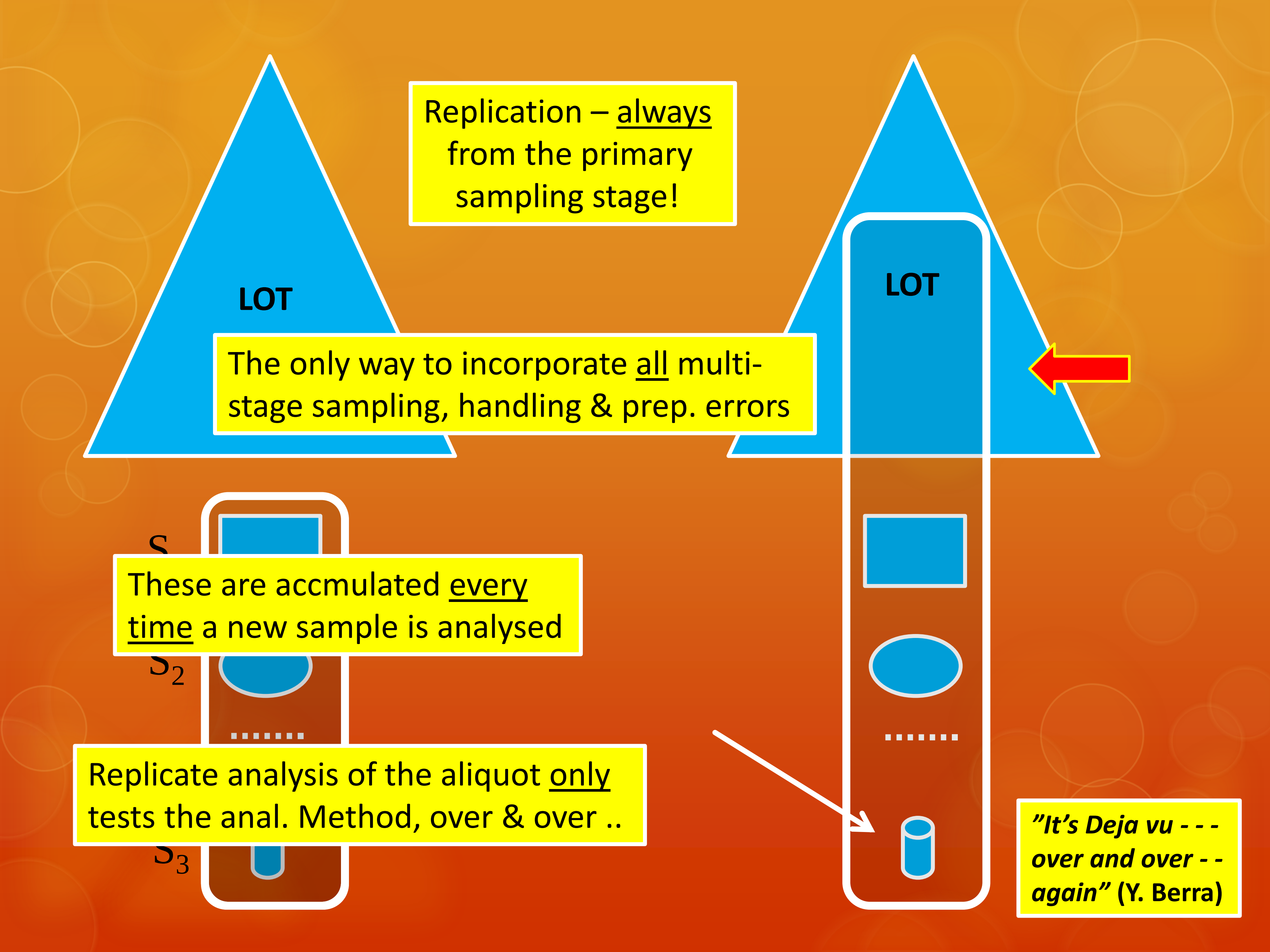


Framework for understanding sampling + analysis variability ... TOS

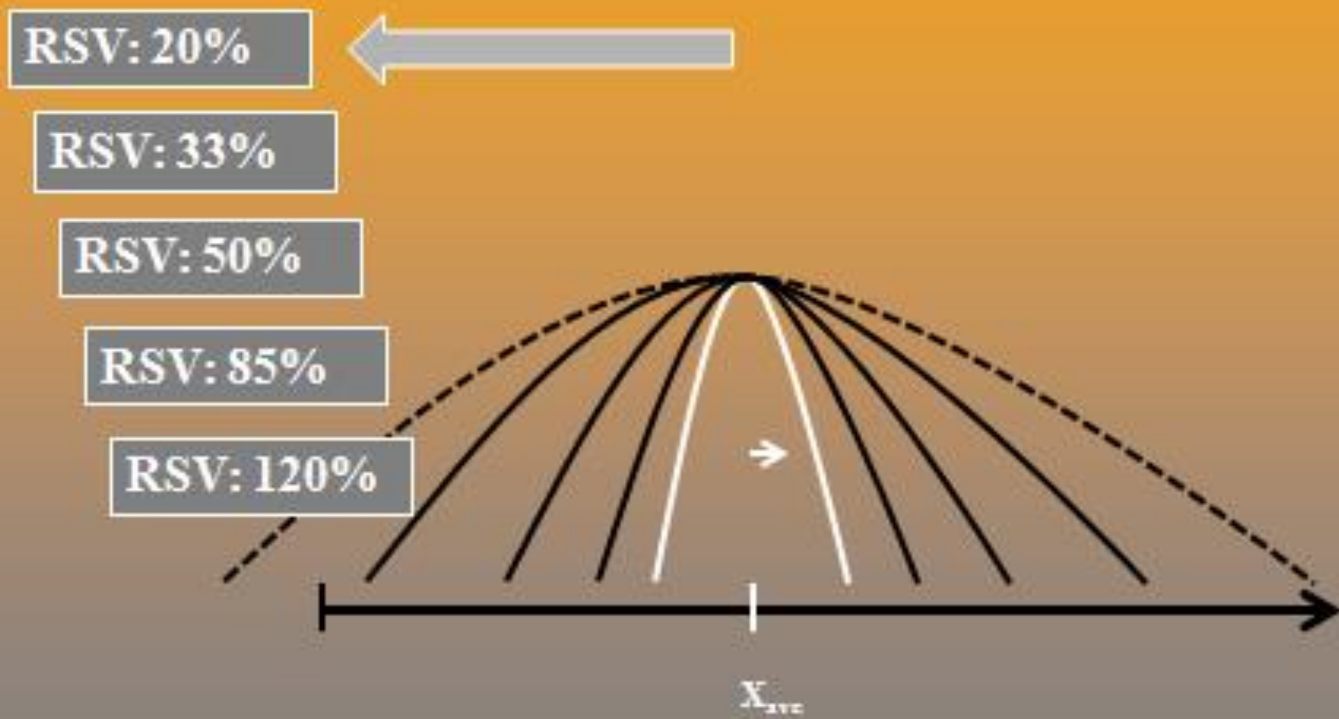
Resolution of endless confusion:

Replication from which stage?




Replication – always from the primary sampling stage!
LOT
LOT
The only way to incorporate all multi-stage sampling, handling & prep. errors
These are accmulated every time a new sample is analysed
S_2
Replicate analysis of the aliquot only tests the anal. Method, over & over ..
S_3
"It's Deja vu - - - over and over - - again" (Y. Berra)

RSV threshold(s)



CV_1 - Variability among analytical replicates = analytical uncertainty

CV_2 - Variability among increments = analytical + sampling uncertainty

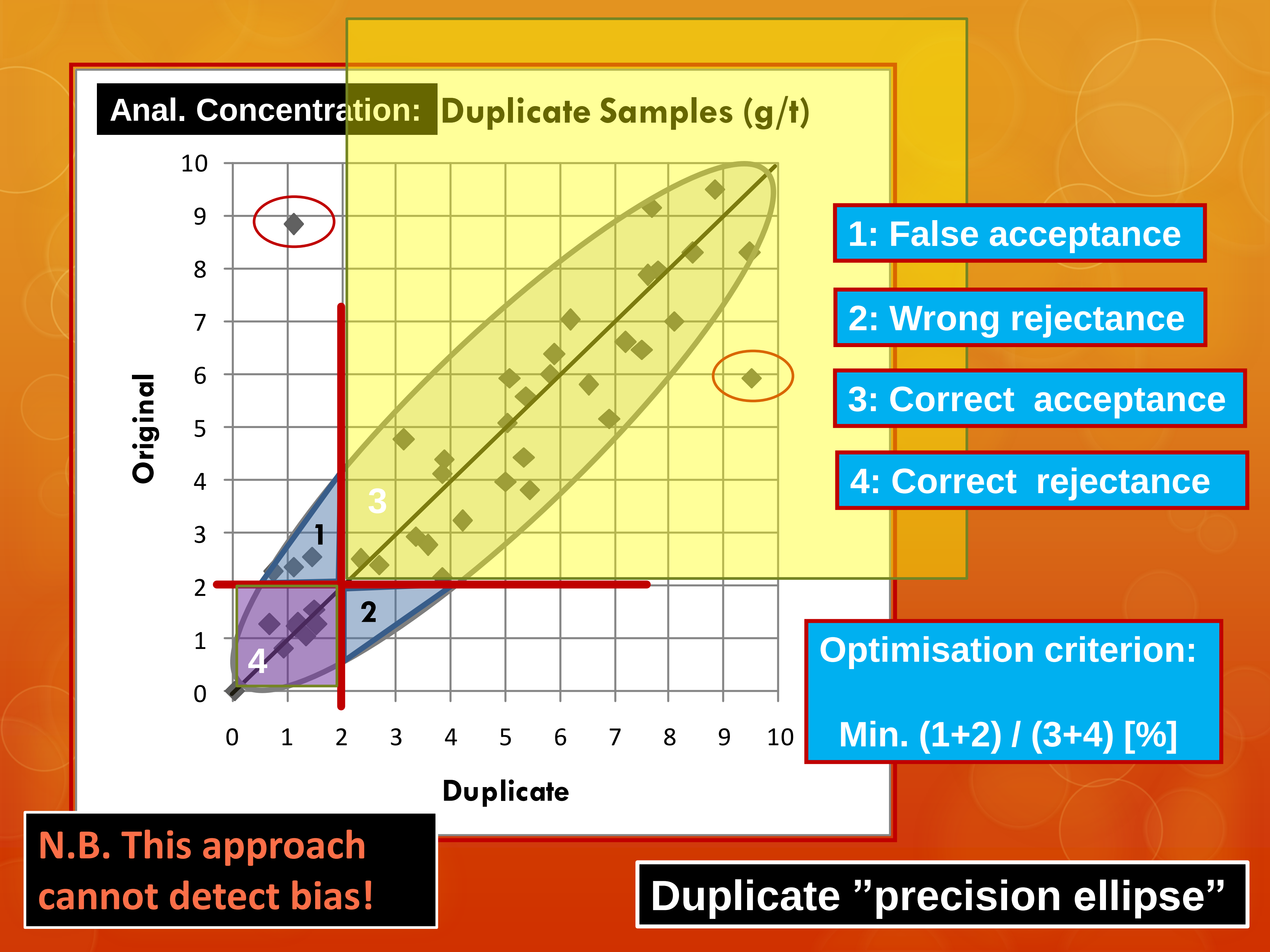
Lot	CV_1 ⟷	CV_2
1	10,18	116,85
2	7,77	44,5
3	6,85	40,8
4	6,52	51,59
5	22,34	249,88
6	30,96	133,64
7	20,12	60,63
8	19,65	57,54
9	5,06	39,74
10	6,52	62,13
11	5,68	52,88
12	6,25	38,04
13	4,69	35,03
14	15,48	218,12
15	34,09	171,81

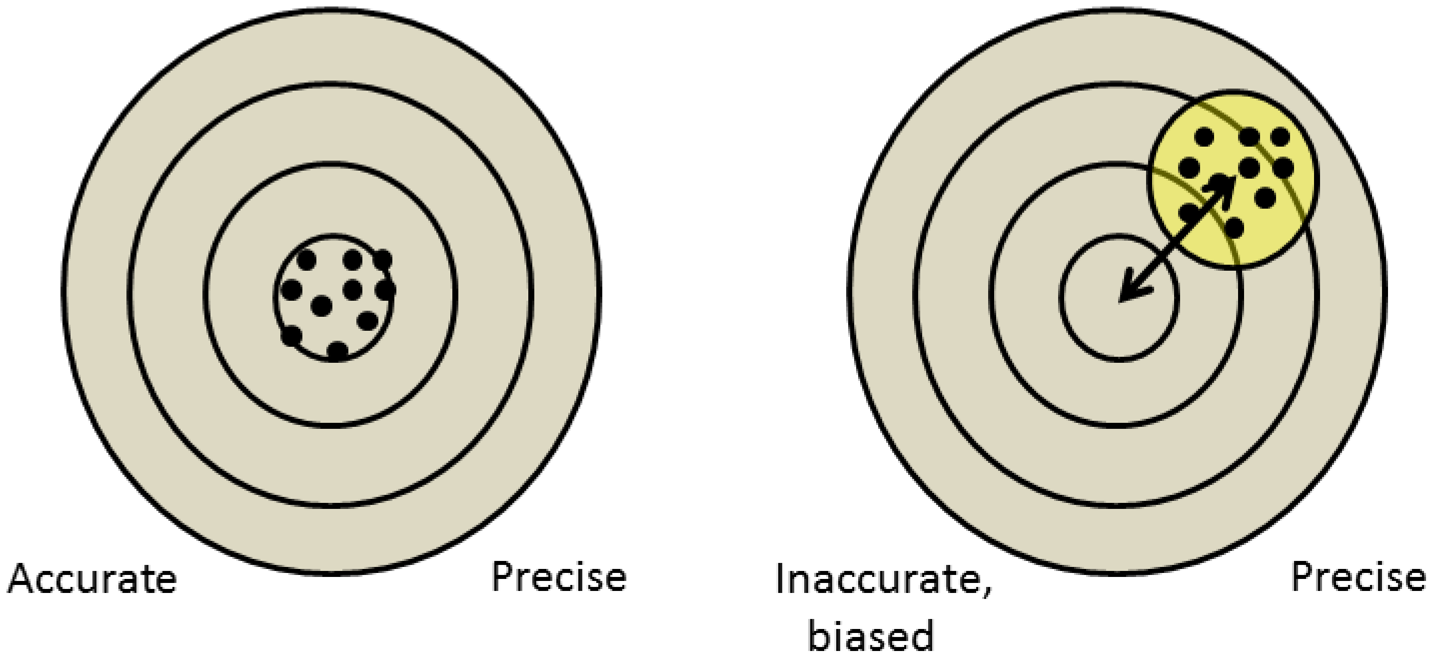
There never was an analytical result – <u>without</u> preceding sampling

The specific analytical mesurement uncertainty CV_1 .... should not be used as a measure of the <u>total</u> measurement uncertainty CV_2.

What would be the consequence(s)?

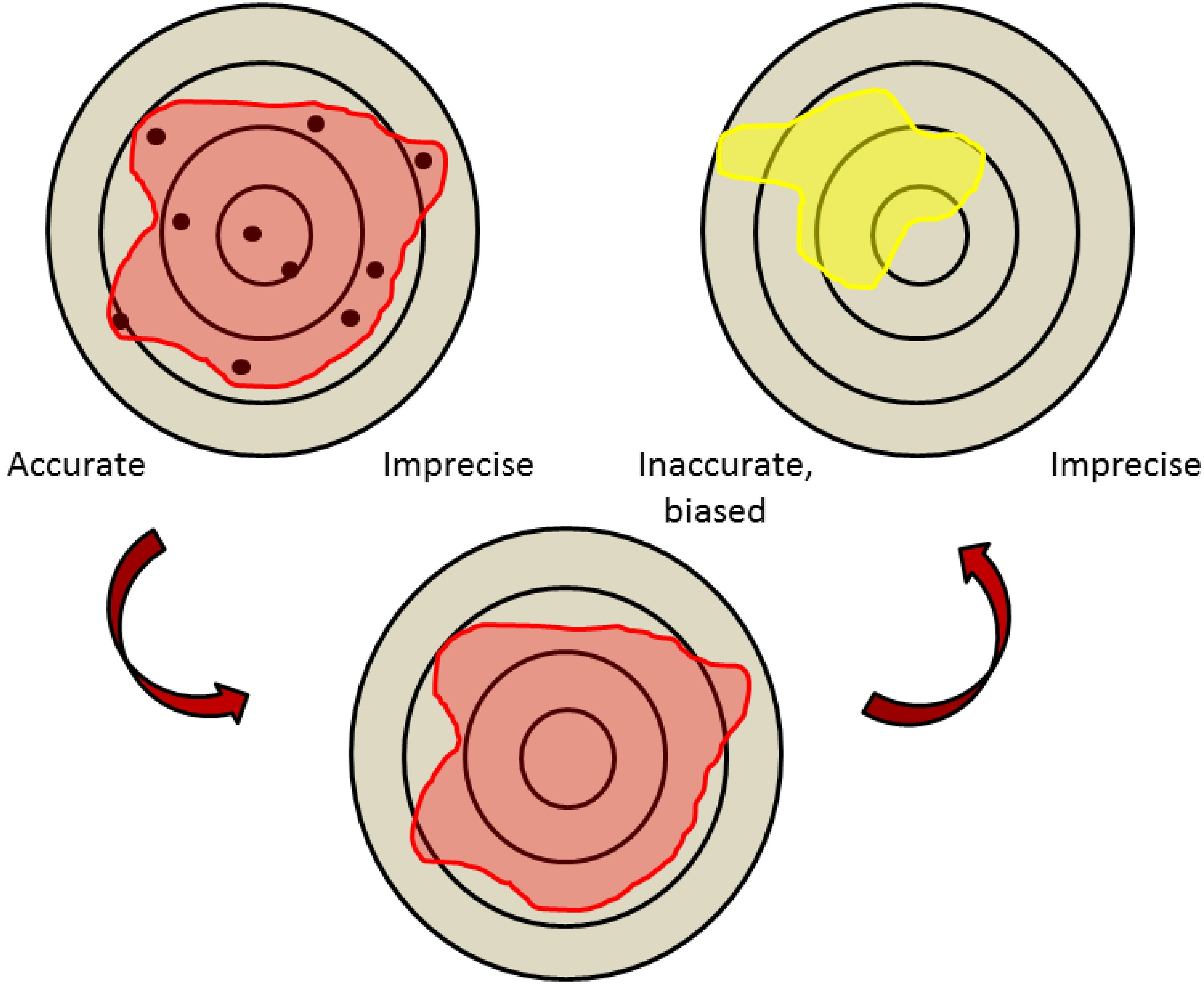
Who is responsible?

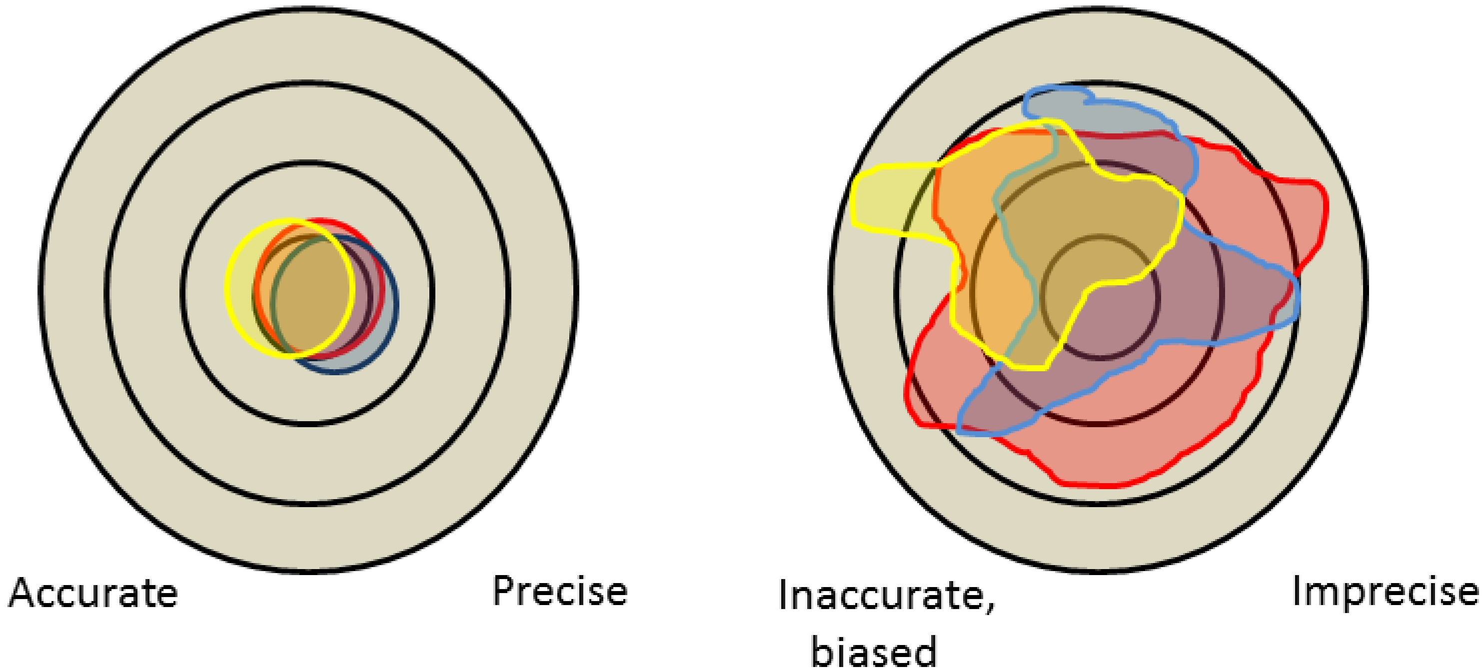

Anal. Concentration: Duplicate Samples (g/t)
Original
Duplicate
0
1
2
3
4
5
6
7
8
9
10
1
2
3
4
1: False acceptance
2: Wrong rejectance
3: Correct acceptance
4: Correct rejectance
Optimisation criterion:
Min. (1+2) / (3+4) [%]
N.B. This approach cannot detect bias!
Duplicate "precision ellipse"



Replication Experiment, RE(10) delineating *ideal* characteristics: ***unbiased*** analytical process

Replication Experiment, RE(10) delineating *realistic* characteristics: ***constant*** analytical bias


Accurate
Imprecise
Inaccurate,
biased
Imprecise



3 X Replication Experiment, RE(10)
delineating *unrealistic* expectations:
constant, vanishing sampling bias

3 X Replication Experiment, RE(10)
delineating *realistic* characteristics:
inconstant sampling bias

Representative Sampling: Theory of Sampling (TOS)

TOS - Axiomatic exposé

Governing principles (GP) – Sampling unit operations (SUO)

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. LHC: Lot Heterogeneity Characterization (0-D, 1-D)
7. SUO: Composite Sampling
8. SUO: Comminution
9. SUO: Mixing / Blending
10. SUO: Representative Mass Reduction (Sub-sampling)

The analytical laboratory

TOS in the laboratory – TOS in the laboratory

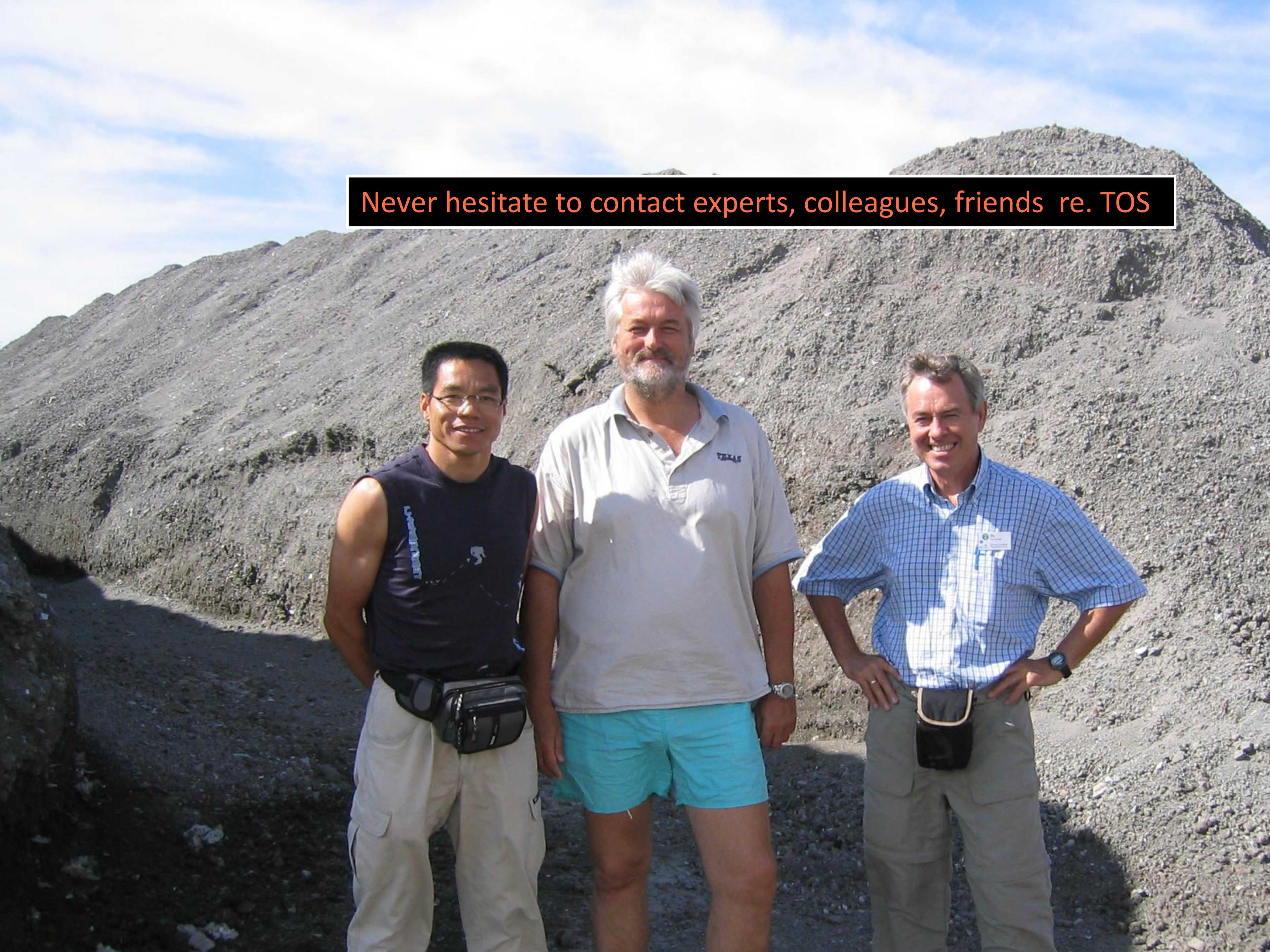
References

Esbensen, K.H. & Wagner, C. (2014)
Why we need the Theory of Sampling.
The Analytical Scientist, vol 21 (Oct. 2014) pp. 30-38*

DS 3077 (9 page free preview)*
http://webshop.ds.dk/product/M278012/standard.aspx

http://www.impublications.com/tos-forum

http://www.spectroscopyeurope.com/articles/sampling


Never hesitate to contact experts, colleagues, friends re. TOS

khe@gmail.com

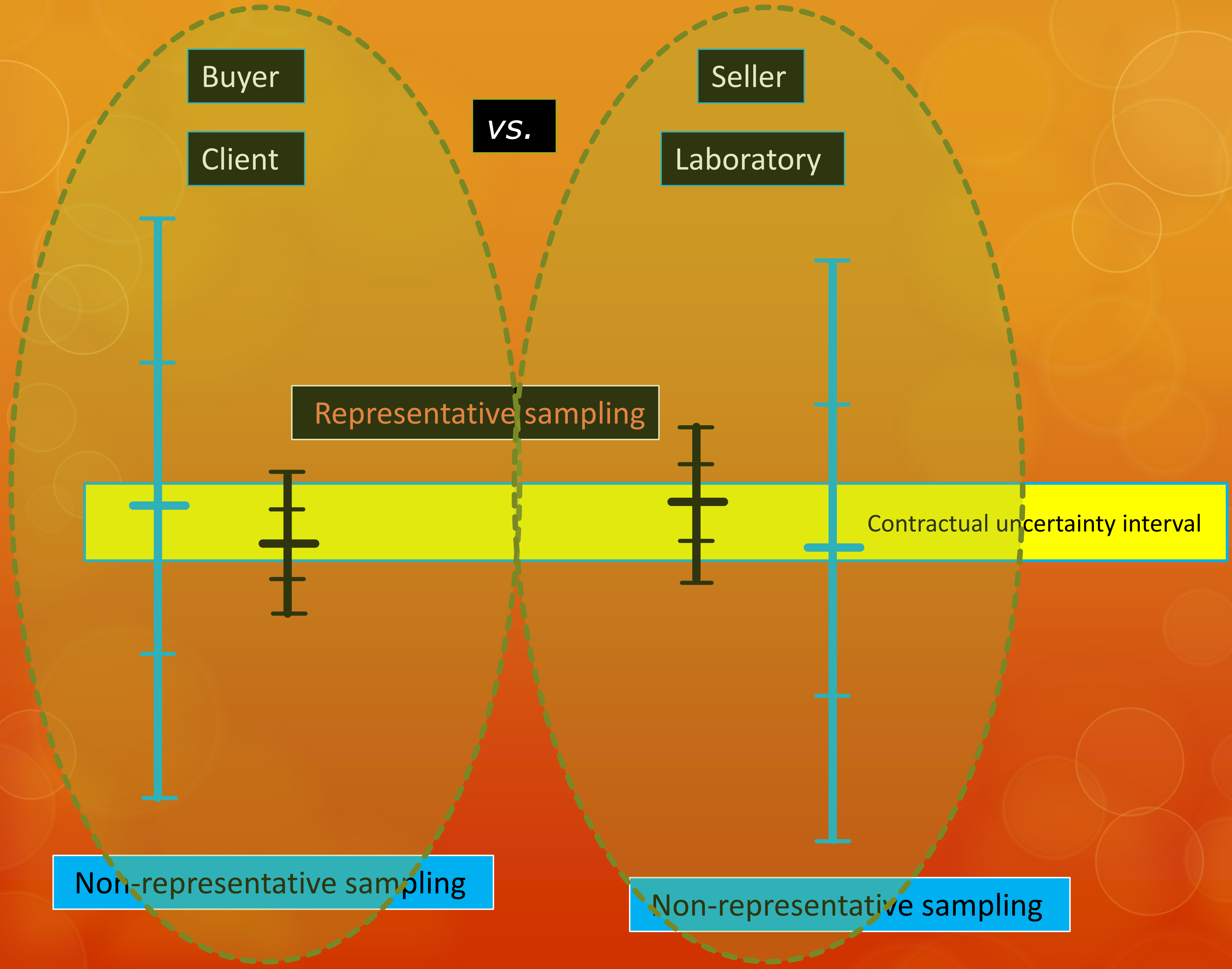
www.khe.consult.com

Thank you for your attention!

Fundamental insights -

- ...a HUGE business opportunity ...

Can be discussed later


Buyer
Client
vs.
Seller
Laboratory
Representative sampling
Contractual uncertainty interval
Non-representative sampling
Non-representative sampling

Advanced stuff – only if there is time

Effect of
problem

Causes: sampling variation FSE GSE IEE IPE IWE TAE	IDE *(cyclicity)*
(trends)	Bias Test(s) Round Robin test(s) “Correction factors” **“Effects estimation”**

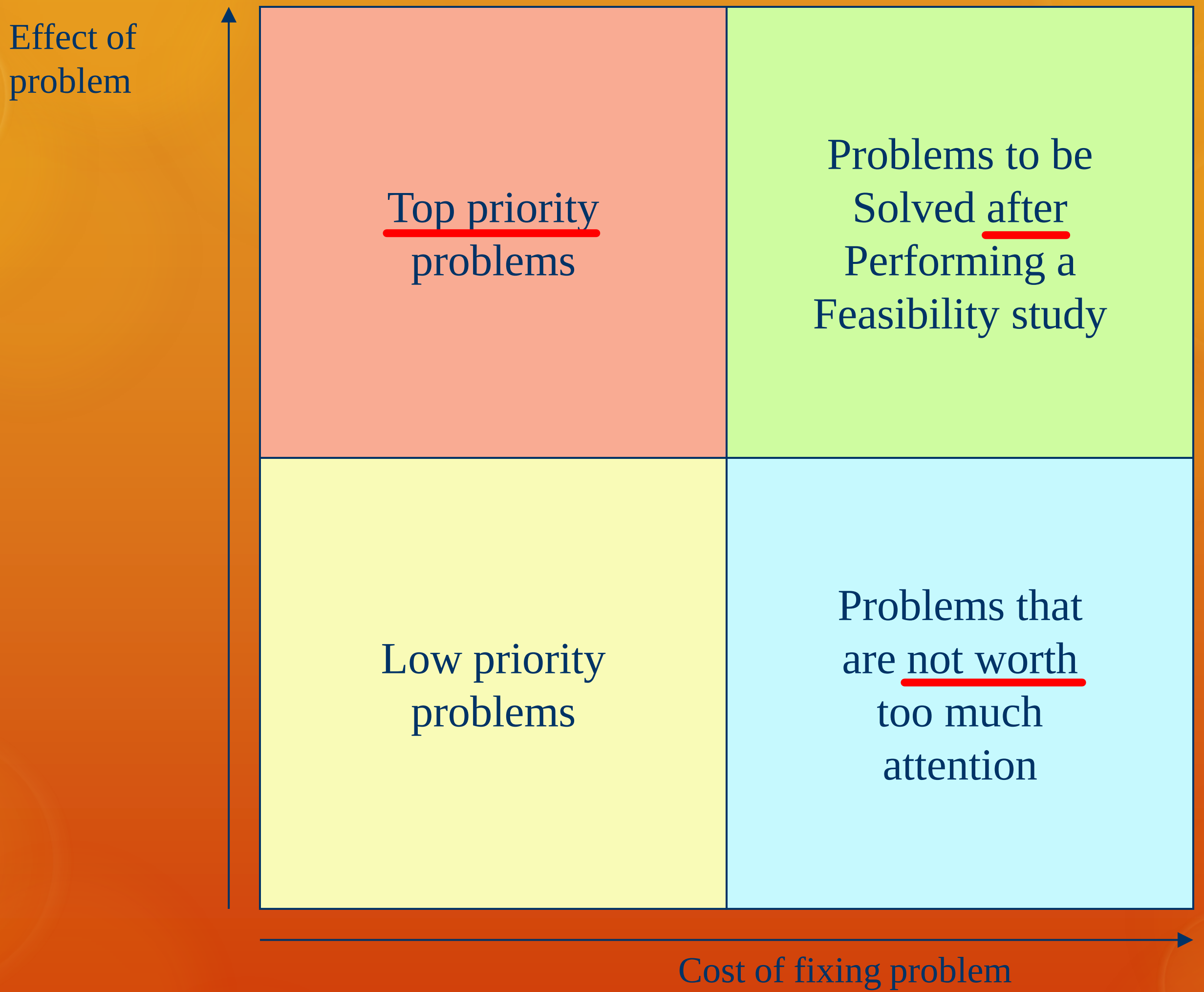
Cost of fixing problem

Note:
Elimination of (all) Incorrect Sampling Errors (ISE)

TOS’ Preventive Paradigm (TPP)

Principle of Sampling Correctness (PSC)

Courtesy: Francis Pitard Sampling Consultants



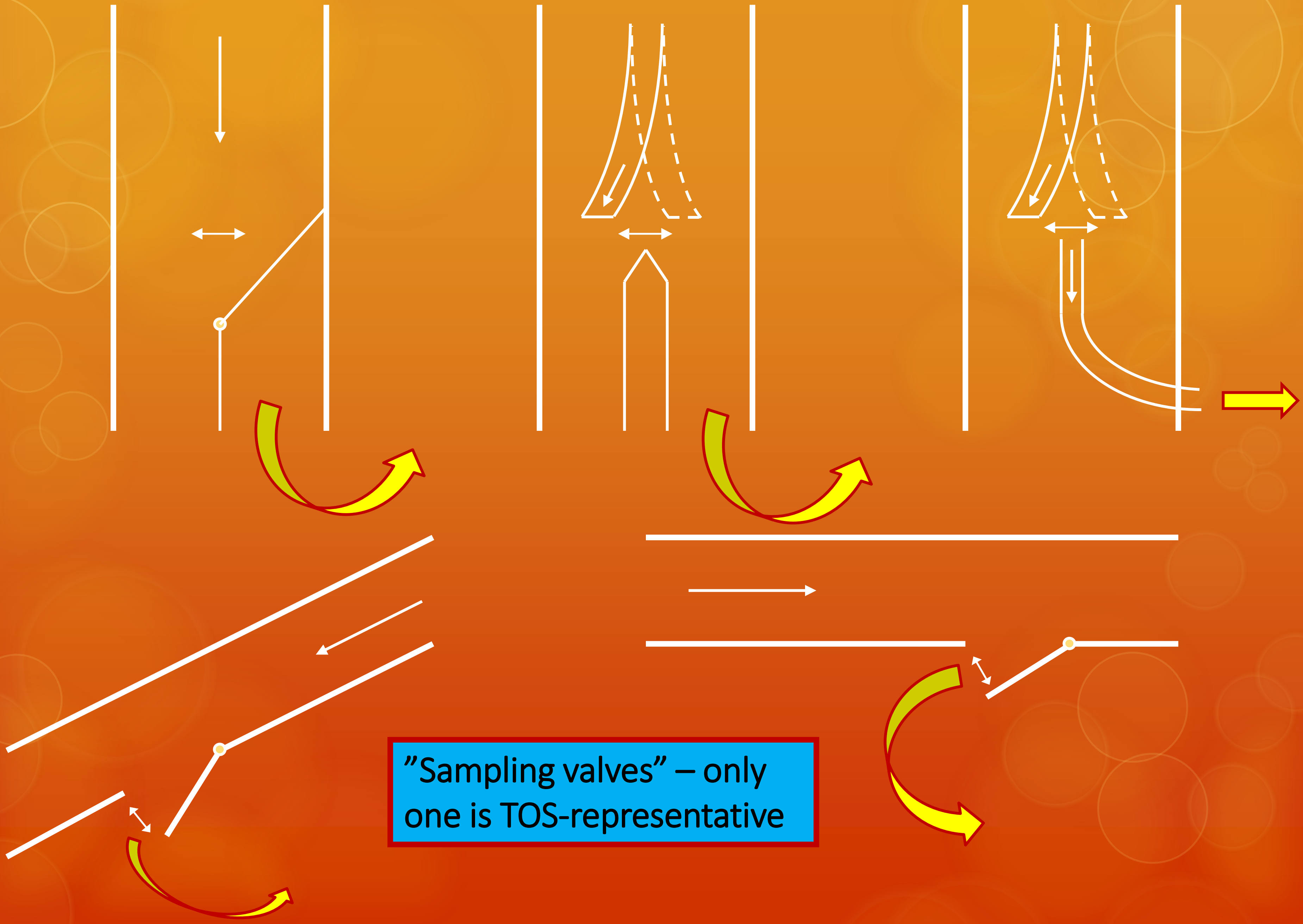
Courtesy: Francis Pitard Sampling Consultants

A hidden elephant in the room …

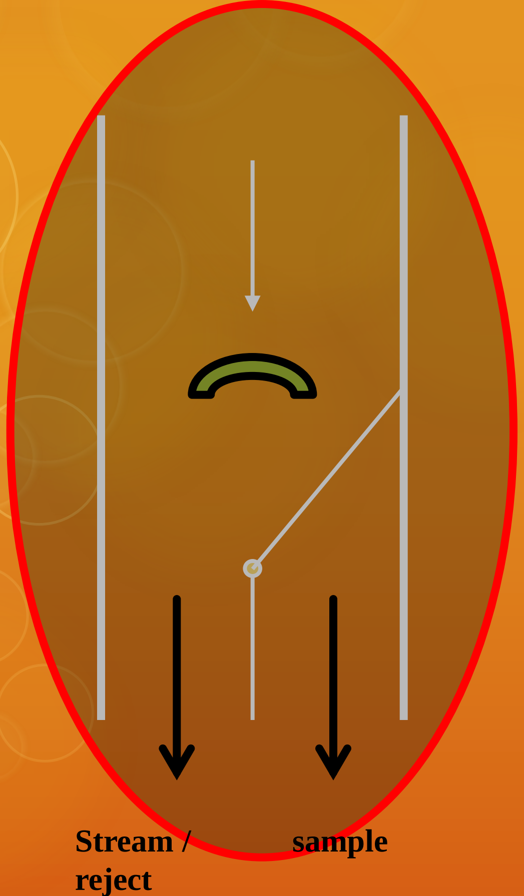
- \- a fundamental difference of critical importance:
- a sample cell … … is containing the sample to do analysis

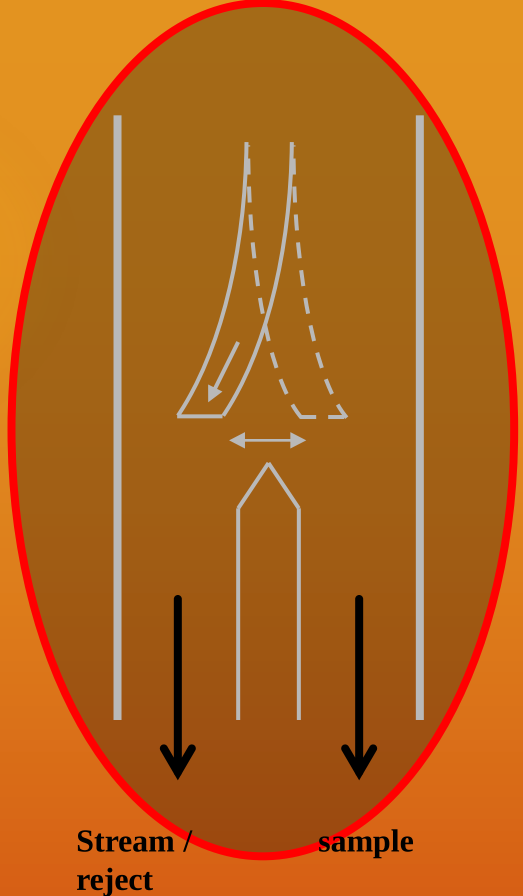
A hidden elephant in the room ...

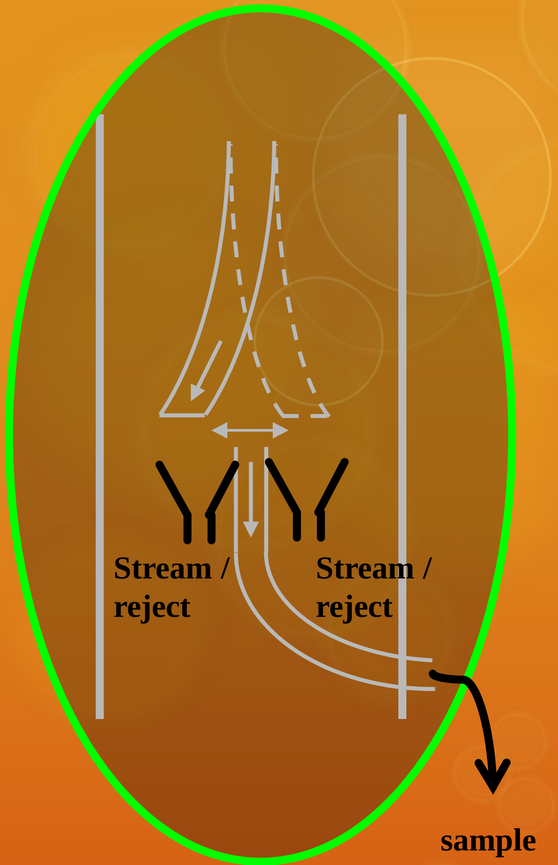
- - a fundamental difference of critical importance:
- a sampling cell ... ... is sampling and doing the analysis



"Sampling valves" – only one is TOS-representative







Flap valve design

"Inversed" flap valve

In-line outtake valve

1.

2.

3.


33
11 22 '00

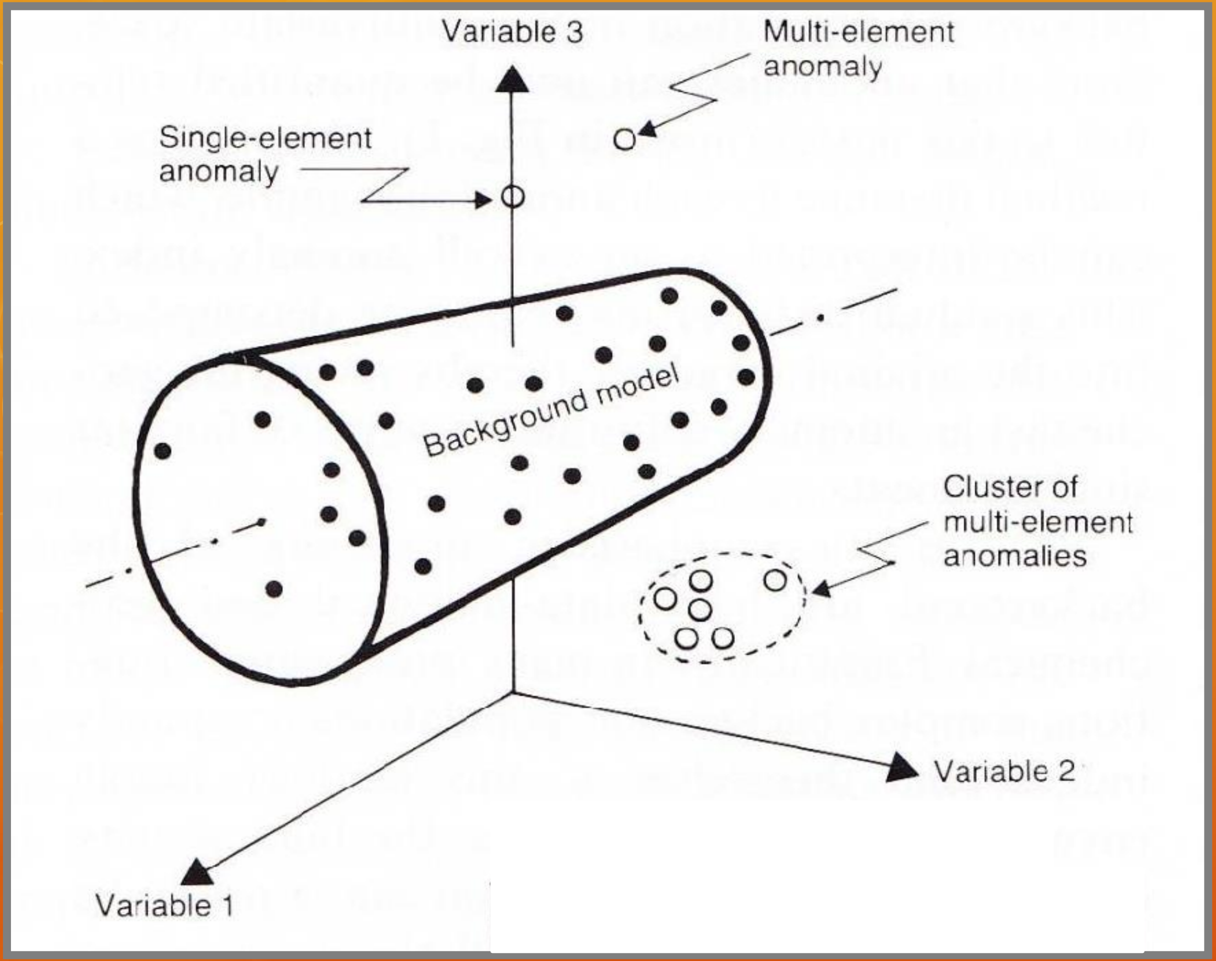
The multivariate data matrix: embedded errors in all cells

199.1	205.6	210.7	213.2	216.0	200.6	208.5	212.6	216.5	218.5	202.4	209.9	214.8	217.9	220.4	4.1	4.5	4.8
198.5	204.1	210.0	212.0	214.6	199.4	206.3	211.9	215.1	217.5	202.0	208.8	214.5	217.7	218.9	4.1	4.4	4.7
194.6	200.7	205.4	207.2	209.0	198.1	204.9	209.3	211.5	214.3	201.3	207.6	213.4	215.9	218.0	3.9	4.2	4.5
194.8	200.1	206.8	210.3	212.6	196.4	204.1	210.0	213.4	216.6	198.6	206.6	213.0	216.1	219.2	1.7	1.9	2.0
194.1	201.3	207.1	211.3	213.7	196.6	[illegible]	210.9	215.1	217.5	199.8	208.1	213.7	218.7	220.2	1.9	2.2	2.4
193.2	201.2	206.9	210.1	213.2	197.0	204.9	[illegible]	214.2	217.8	199.5	207.1	214.3	218.6	219.9	1.7	2.0	2.3
193.8	200.4	205.6	208.4	209.5	196.8	203.4	208.8	[illegible]	214.3	198.5	206.2	212.8	215.7	216.9	2.2	2.4	2.5
193.5	201.5	205.3	208.0	211.1	[illegible]	[illegible]	[illegible]	[illegible]	214.4	199.2	205.6	212.4	215.1	217.9	[illegible]	[illegible]	[illegible]
194.6	202.0	207.1	209.8	212.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	200.1	206.5	212.9	216.7	218.8	[illegible]	[illegible]	[illegible]
179.5	187.2	192.8	194.7	197.4	[illegible]	[illegible]	[illegible]	[illegible]	203.9	[illegible]	196.3	202.2	205.3	207.3	[illegible]	[illegible]	[illegible]
178.4	185.9	189.8	193.8	195.7	[illegible]	[illegible]	[illegible]	[illegible]	203.3	[illegible]	197.9	202.1	206.0	208.6	[illegible]	[illegible]	[illegible]
181.4	188.8	192.7	195.9	197.8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
194.0	199.7	205.5	208.3	210.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
195.2	201.3	207.4	209.6	212.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
193.8	200.8	206.2	210.1	211.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
189.9	197.9	202.7	207.4	209.6	[illegible]	[illegible]	[illegible]	[illegible]	213.4	195.9	204.8	211.3	214.6	218.0	[illegible]	[illegible]	[illegible]
193.4	201.6	206.9	210.4	213.7	[illegible]	[illegible]	[illegible]	[illegible]	217.3	199.3	206.9	214.2	217.1	219.8	[illegible]	[illegible]	[illegible]
193.9	200.9	207.7	210.9	213.4	197.5	205.6	211.5	215.4	218.4	200.9	209.0	216.0	219.0	223.0	1.6	1.9	2.1
200.5	206.8	212.1	214.2	217.3	202.7	209.8	214.7	217.8	220.8	204.0	212.2	216.5	220.3	224.0	3.2	3.3	3.5
193.3	201.2	206.5	209.3	211.9	196.7	203.6	210.4	213.8	216.5	198.5	206.1	214.0	216.3	218.8	1.9	2.2	2.4
189.9	196.3	202.3	205.4	208.3	192.9	201.3	207.1	210.7	214.1	194.9	204.0	210.4	213.6	216.3	1.3	1.5	1.7
180.7	186.5	192.1	194.4	197.5	186.6	194.6	199.5	202.8	204.8	191.1	198.8	205.1	207.5	210.6	0.9	1.2	1.4
193.3	200.0	206.5	210.1	212.0	195.9	203.2	209.9	212.3	215.7	197.2	205.2	212.0	215.8	218.2	1.9	2.1	2.2
194.4	201.5	208.3	211.8	215.1	197.4	205.5	212.4	216.0	218.1	200.4	208.3	215.9	218.5	221.8	1.4	1.9	2.2
200.6	209.7	218.6	222.8	226.8	202.0	212.3	220.2	224.8	227.6	203.5	213.4	221.1	226.2	229.6	0.7	1.1	1.4

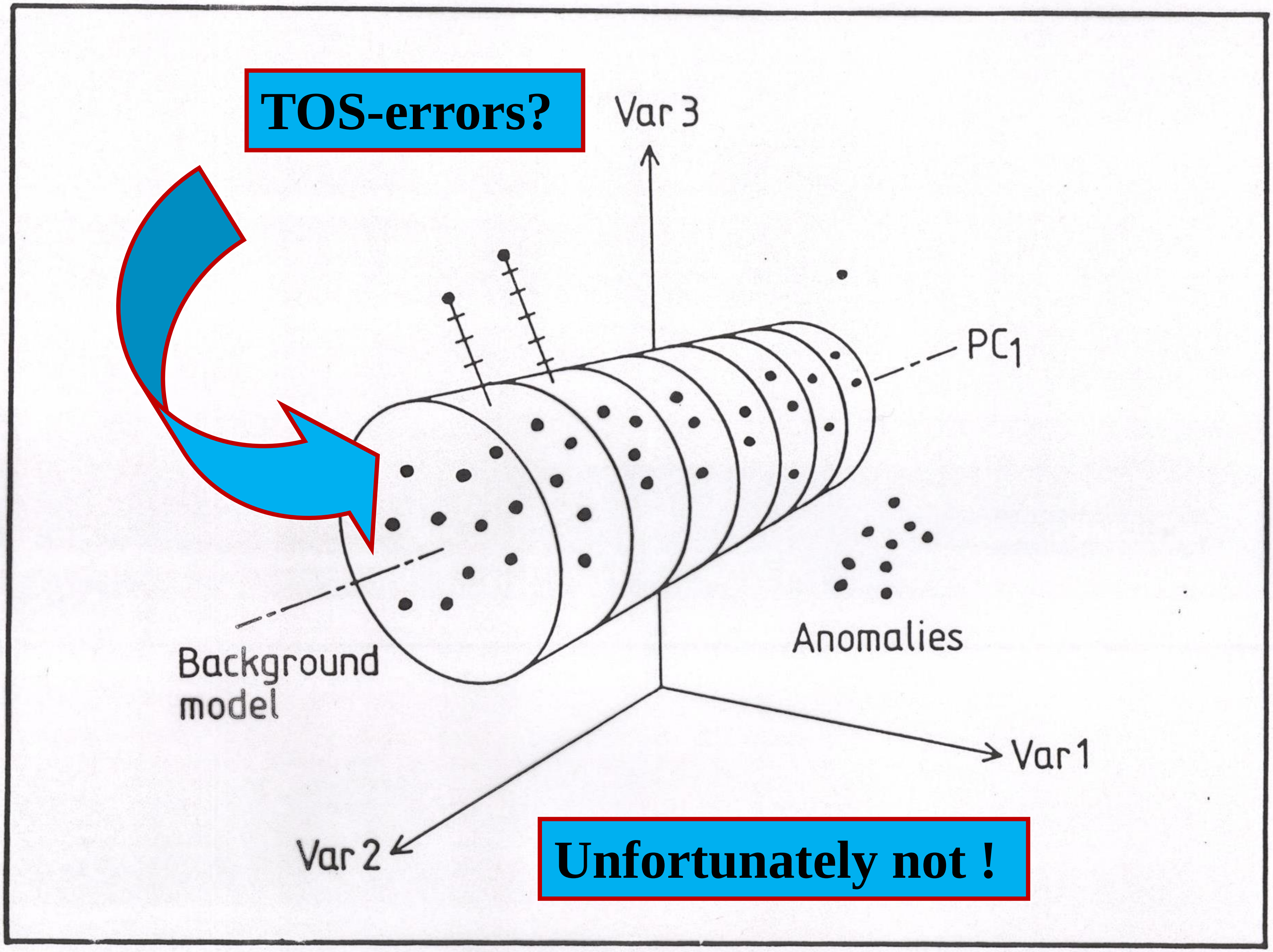
X

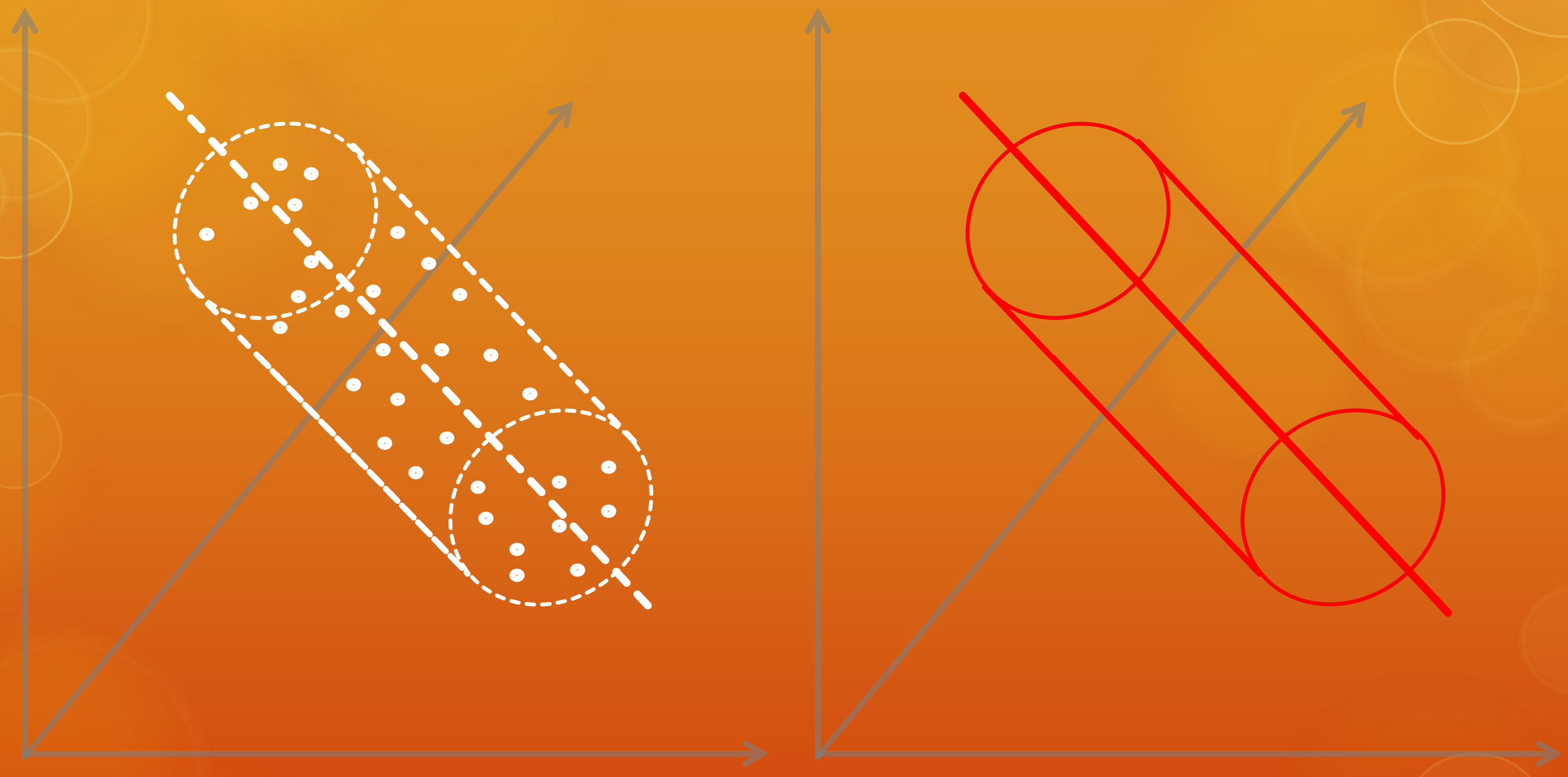
Y

How to work with such errors? (TSE + TAE + CME)

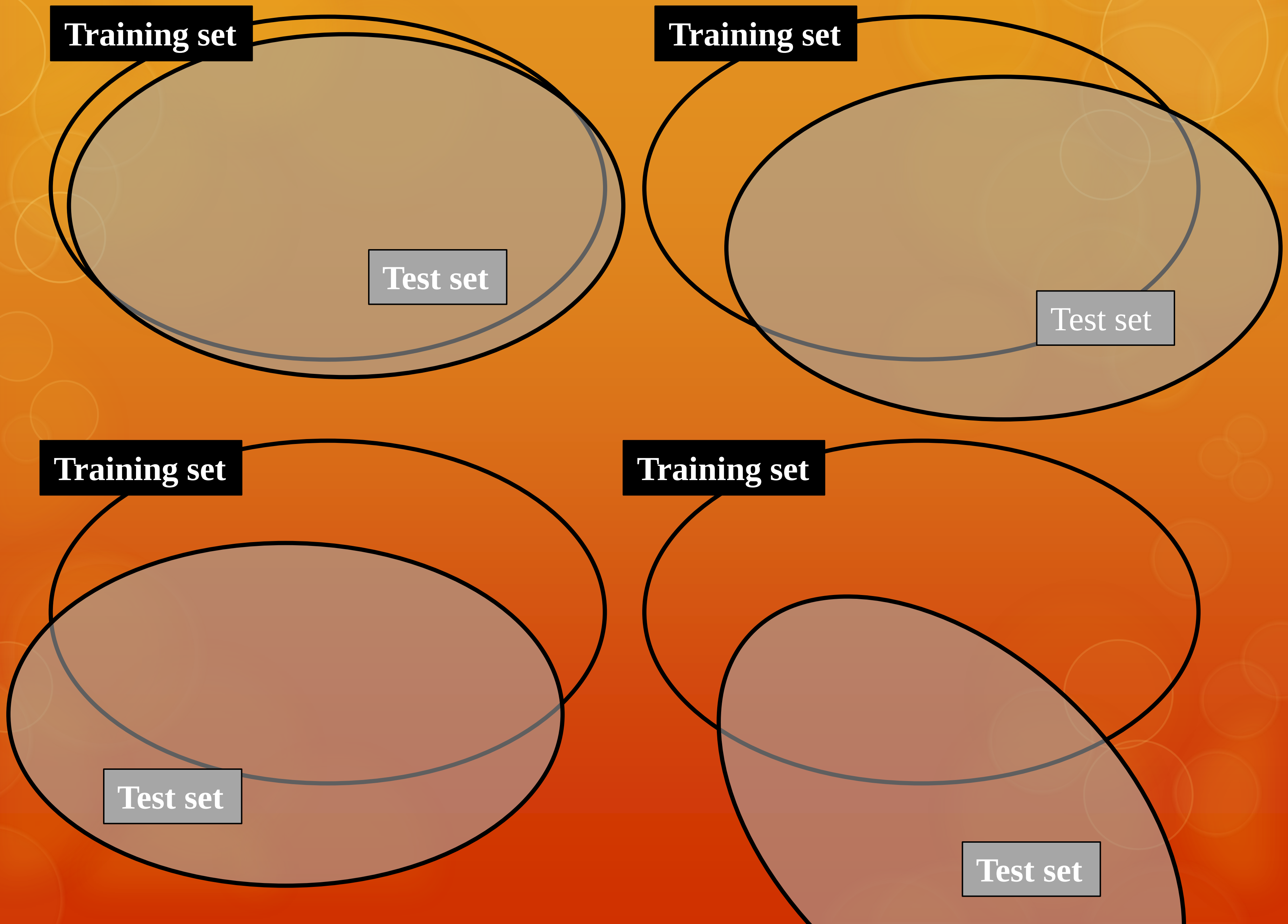


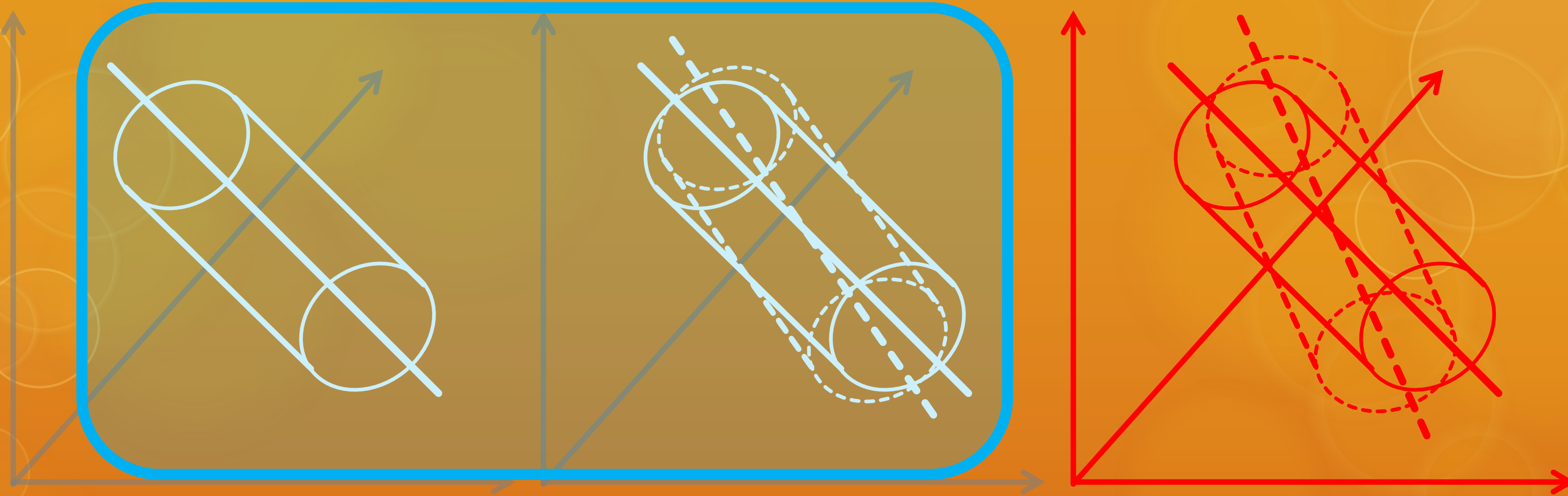
Multivariate Data Models
aka
”Bilinear Data Models”
(PCA, PLS. SIMCA)

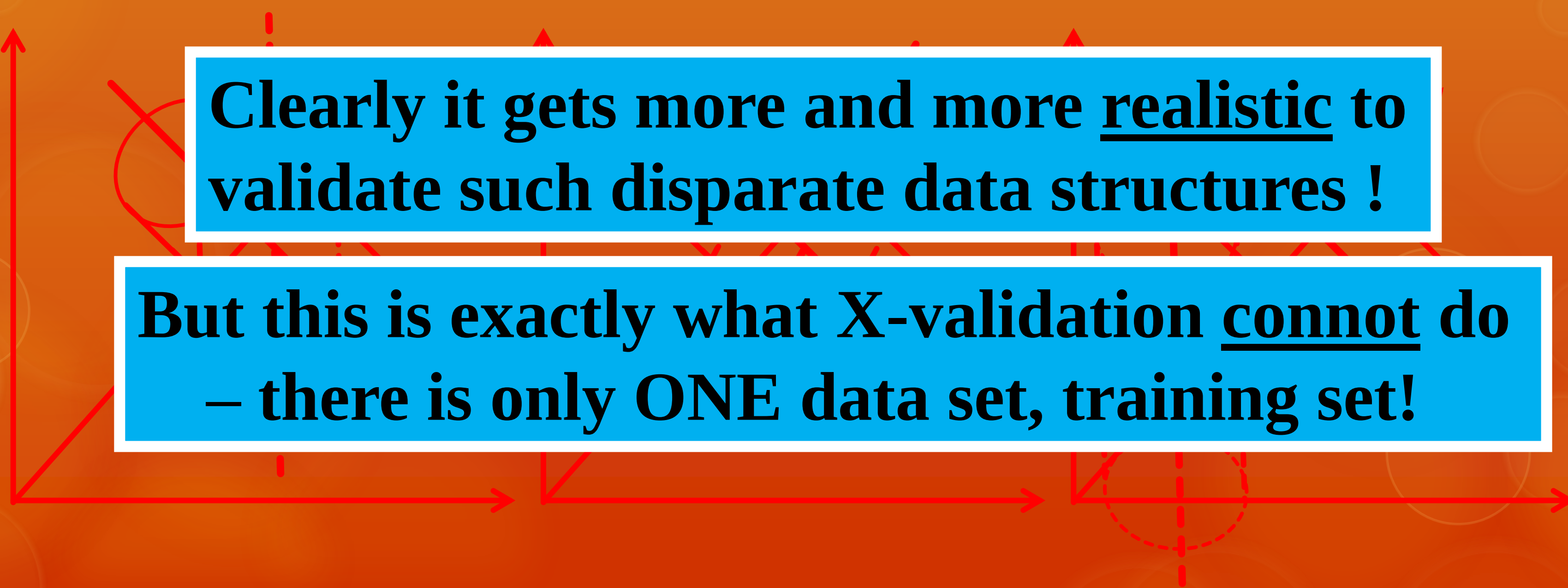

TOS-errors?
Var 3
PC1
Anomalies
Background model
Var 1
Var 2
Unfortunately not !



A bilnear data model (PCA, PLS, SIMCA)

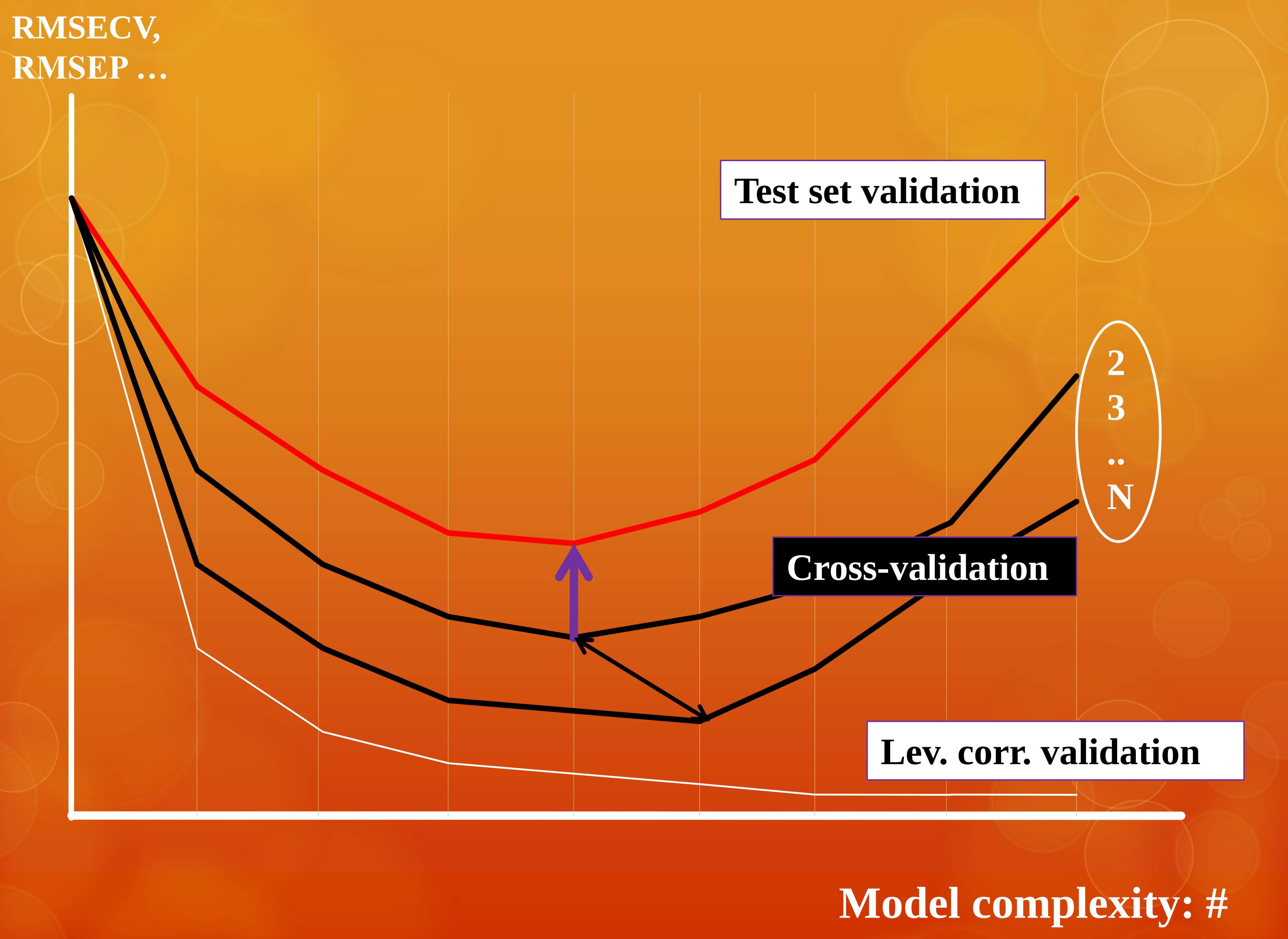

Training set
Test set
Training set
Test set
Training set
Test set
Training set
Test set

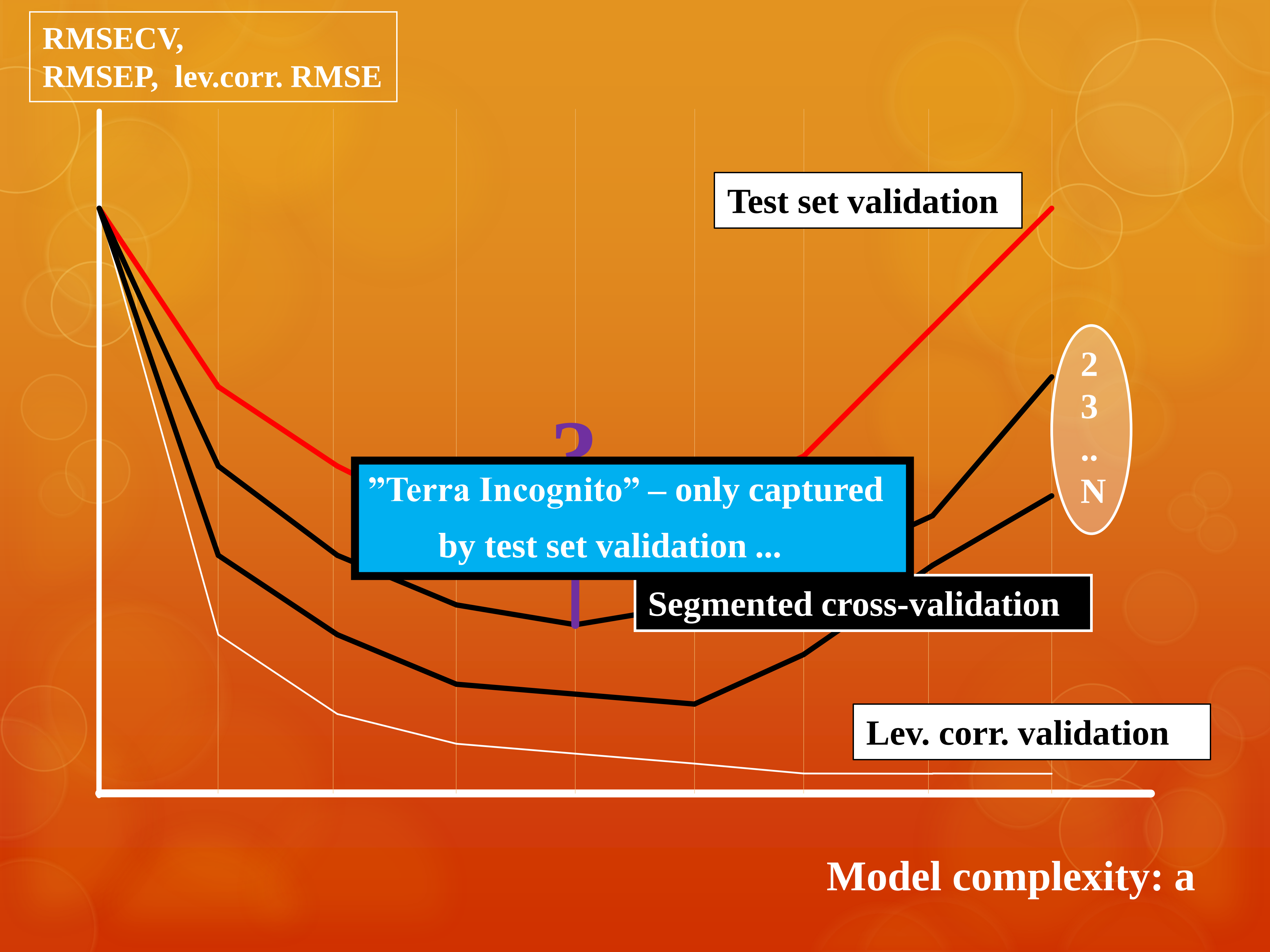


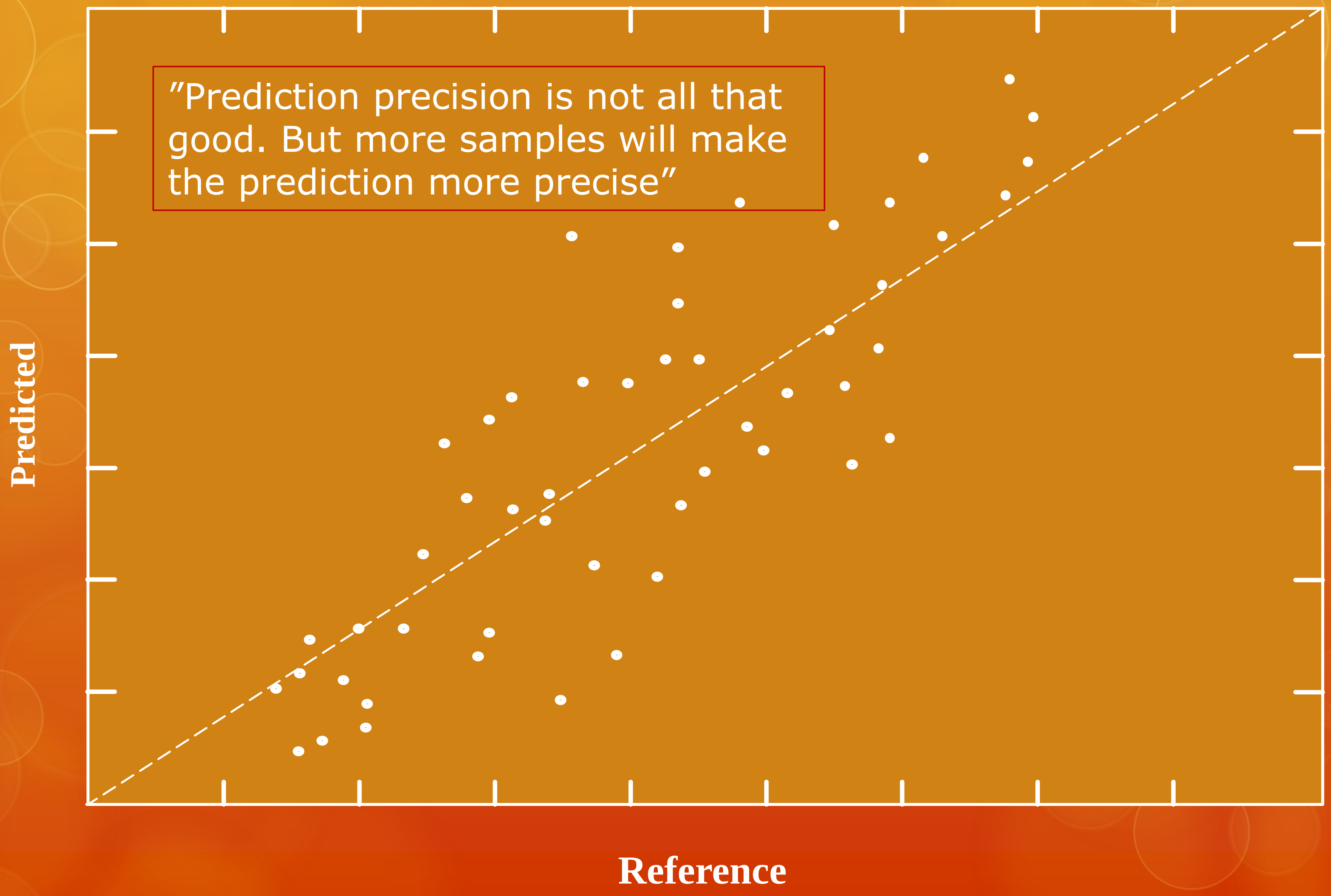


Clearly it gets more and more realistic to validate such disparate data structures !

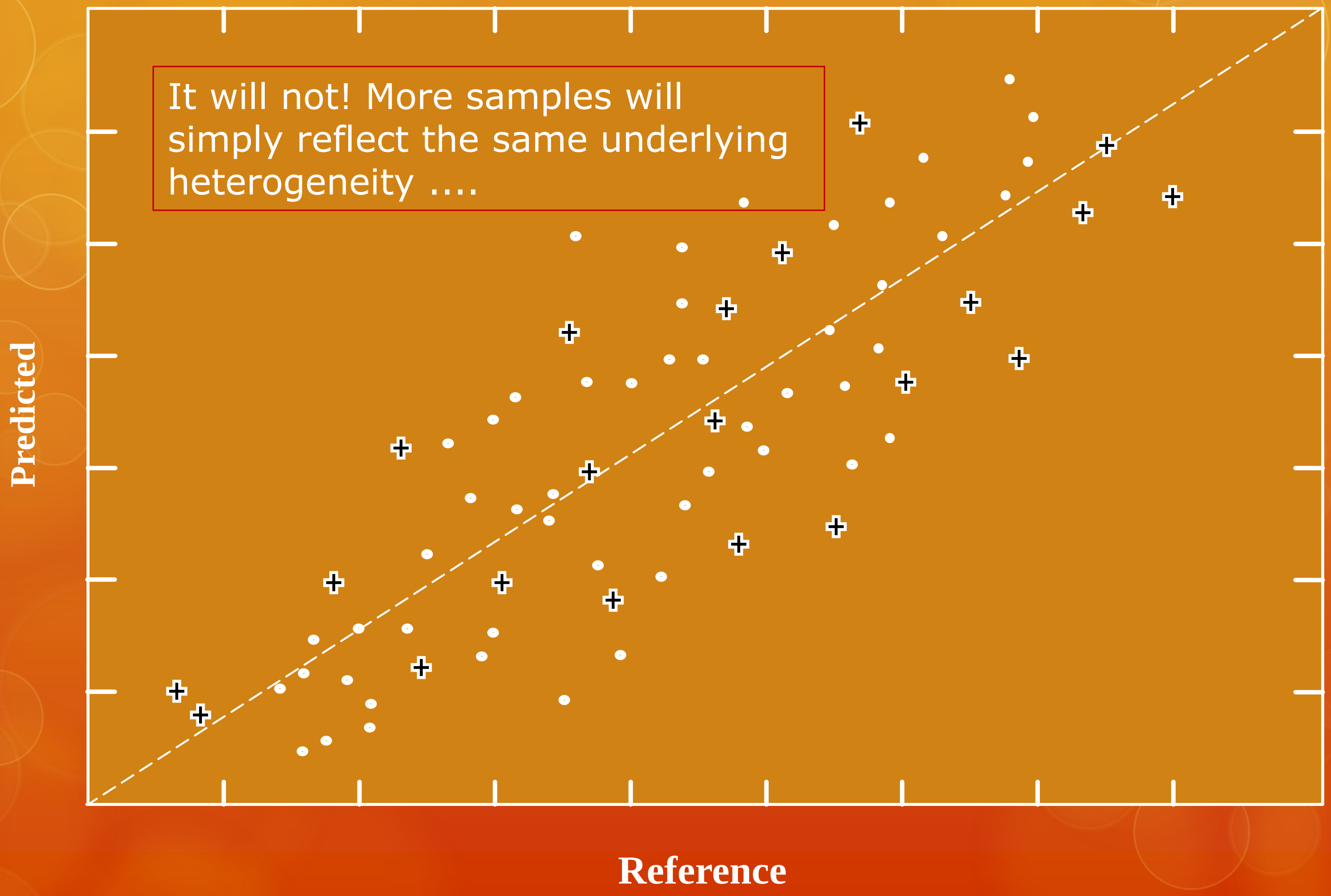
But this is exactly what X-validation connot do – there is only ONE data set, training set!


RMSECV,
RMSEP …
Test set validation
2
3
..
N
Cross-validation
Lev. corr. validation
Model complexity: #


RMSECV,
RMSEP, lev.corr. RMSE
Test set validation
?
2
3
..
N
”Terra Incognito” – only captured
by test set validation ...
Segmented cross-validation
Lev. corr. validation
Model complexity: a



”Prediction precision is not all that good. But more samples will make the prediction more precise”
Predicted
Reference


It will not! More samples will simply reflect the same underlying heterogeneity ....
Predicted
Reference

Special Issue Article

Journal of CHEMOMETRICS

Received: 24 January 2010, Revised: 21 February 2010, Accepted: 23 February 2010, Published online in Wi[illegible]ence: 2010

(www.interscience.wiley.com) DOI: 10.1002/cem.1310

See literature ;-)

Principles of Proper Validation: use and abuse of re-sampling for validation

Kim H. Esbensen[a*] and Paul Geladi[b]

The final word !!!

Validation in chemometrics is presented using the *exemplar* context of multivariate calibration/prediction. A phenomenological analysis of common validation practices in data analysis and chemometrics leads to formulation of a set of generic Principles of Proper Validation (PPV), which is based on a set of characterizing distinctions: (i) Validation cannot be understood by focusing on the methods of validation only; validation must be based on full knowledge of the underlying definitions, objectives, methods, effects and consequences—which are all outlined and discussed here. (ii) Analysis of proper validation objectives implies that there is one valid paradigm only: test set validation. (iii) Contrary to much contemporary chemometric practices (and validation *myths*), cross-validation is shown to be unjustified in the form of monolithic application of a one-for-all procedure (segmented cross-validation) on all data sets. Within its own design and scope, cross-validation is in reality a sub-optimal *simulation* of test set validation, crippled by a critical sampling variance omission, as it manifestly is based on one data set only (training data set). Other re-sampling validation methods are shown to suffer from the same deficiencies. The PPV are universal and can be applied to all situations in which the assessment of performance is desired: prediction-, classification-, time series forecasting-, modeling validation. The key element of PPV is the Theory of Sampling (TOS), which allow insight into all variance generating factors, especially the so-called incorrect sampling errors, which, if not properly eliminated, are responsible for a fatal inconstant sampling bias, for which no statistical correction is possible. In the light of TOS it is shown how a second data set (test set, validation set) is critically necessary for the inclusion of the sampling errors incurred in all 'future' situations in which the validated model must perform. Logically, therefore, all one data set re-sampling approaches for validation, especially cross-validation and leverage-corrected validation, should be terminated, or at the very least used only with full scientific understanding and disclosure of their detrimental variance omissions and consequences. Regarding PLS-regression, an emphatic call is made for stringent commitment to test set validation based on graphical inspection of pertinent *t*–*u* plots for optimal understanding of the *X*–*Y* interrelationships and for validation guidance. QSAR/QSAP forms a partial exemption from the present test set imperative with no generalization potential.

Keywords: Principles of Proper Validation (PPV); future performance assessment; test set validation; cross-validation; re-sampling; predictive regression; Theory of Sampling (TOS)


Never be afraid to contact experts, colleagues, friends re. TOS

Sampling Principles: LHC (3-D vs. 1-D cases)

Lot Heterogeneity Characterization (LHC)

- » Why? It makes *no sense* to design a sampling procedure without knowing the lot heterogeneity quantitatively
- » A ***replication experiment*** will reveal TSE as well as the sampling steps generating the largest variation
- » 10 (40,60) samples provides all the necessary information:
- » Complete empirical sampling error estimation –

VARIOGRAPHIC ANALYSIS (1-D case)

Sampling Principles: LHC (3-D vs. 1-D cases)

SUO: Variography (variographics)

1-dimensional heterogeneity characterization

1-D lots: processes, long stationary piles, ordered series etc.

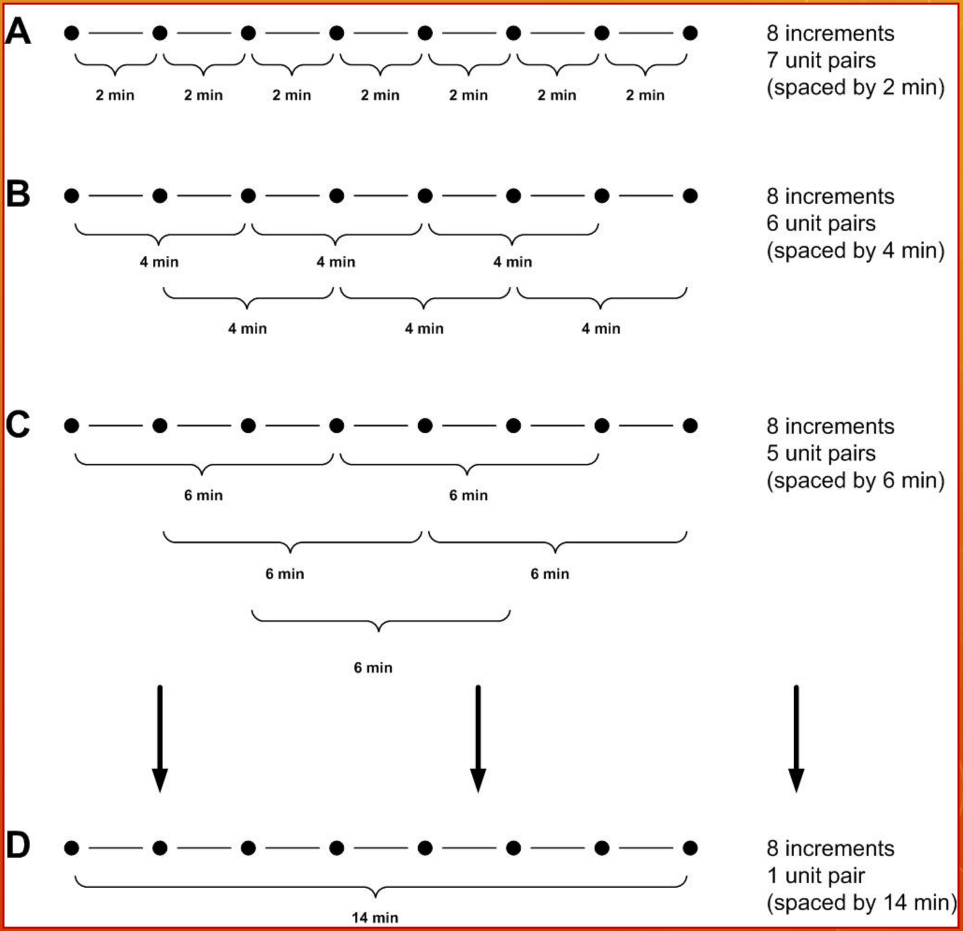
Valuable information about lot *variation* (trends, upsets, periodic phenomena …)

Sampling Principles: LHC (3-D vs. 1-D cases)

SUO: Variography

"Lag" is the distance between increments (samples) along the time or spatial dimension

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



Lag = 1

Lag = 2

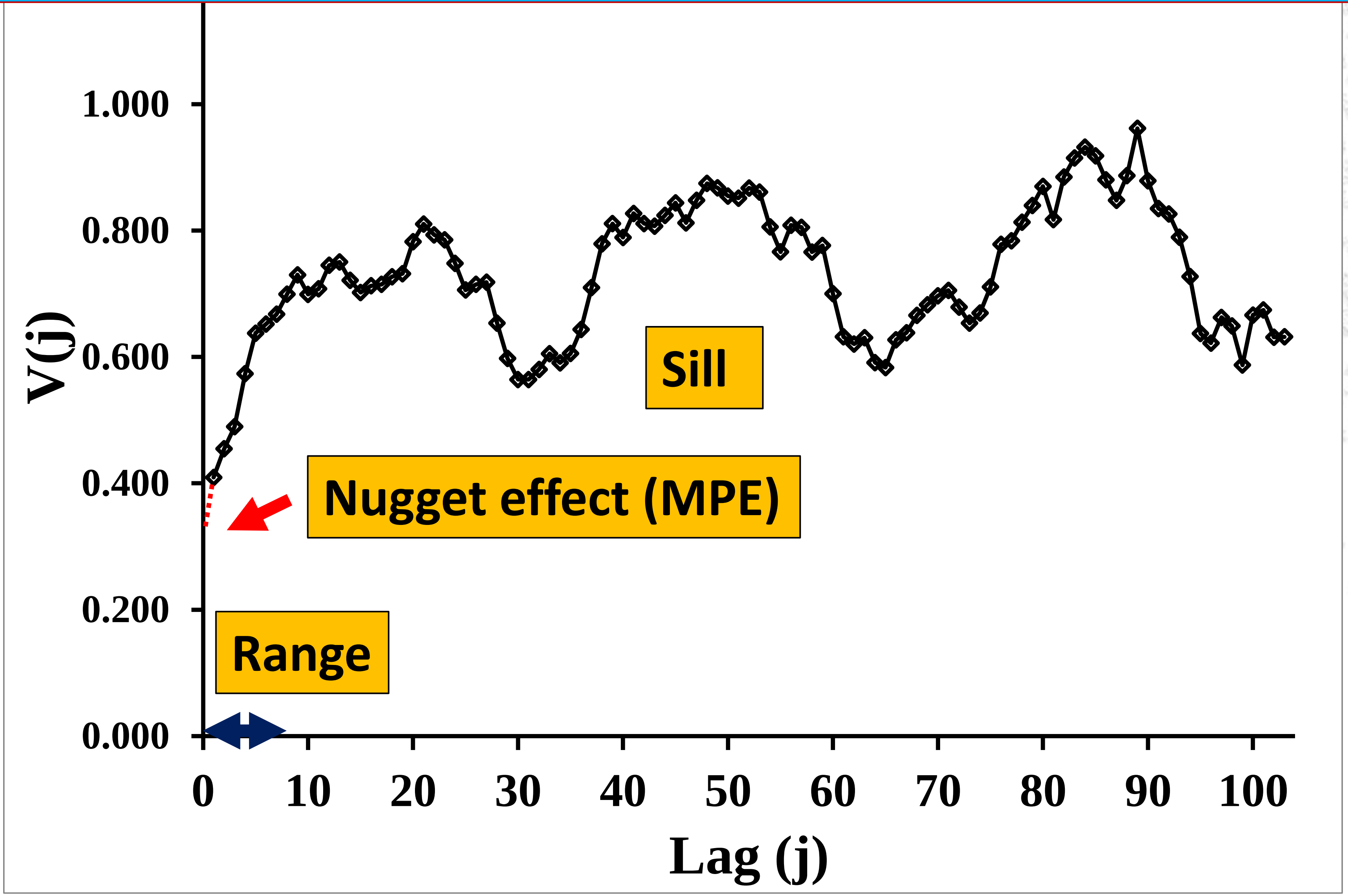
Lag = 3

Lag = 7

	B	C	D	E	F	G	H	I	J	K	L	M
1	Hq (Pred Values)	Lag (j)	V(j)	Q-j	1/(2(Q-j))	sum(hq+j-hq)^2	Lag1	Lag2	Lag3	Lag4	Lag5	Lag6
2												
3	**15.04**	1	0.245	207	0.002	102	0.00214	0.21641	0.15936	0.04268	0.64609	2.13949
4	**14.99**	2	0.409	206	0.002	169	0.17548	0.12454	0.06396	0.57381	2.00619	1.34351
5	**14.57**	3	0.455	205	0.002	186	0.00436	0.45132	0.11465	0.99501	0.5479	1.09161
6	**14.64**	4	0.490	204	0.002	200	0.36699	0.1637	1.13103	0.64996	0.95805	3.47114
7	**15.24**	5	0.573	203	0.002	233	1.02091	2.78656	1.99374	0.13913	1.5808	0.07054
8	**14.23**	6	0.638	202	0.002	258	0.43415	0.16128	1.9138	5.14246	1.62818	0.74909
9	**13.58**	7	0.652	201	0.002	262	0.0662	4.17099	8.56499	3.74384	2.3238	3.2436
10	**13.83**	8	0.668	200	0.003	267	3.18623	7.12516	2.81434	1.60554	2.38301	5.35876
11	**15.62**	9	0.699	199	0.003	278	0.78199	0.01153	0.26822	0.05823	0.28079	1.30599
12	**16.50**	10	0.730	198	0.003	289	0.98347	1.96616	1.26698	0.1256	4.10913	6.36351
13	**15.51**	11	0.699	197	0.003	275	0.16851	0.01793	0.40615	1.07205	2.34365	2.22189
14	**15.10**	12	0.708	196	0.003	277	0.07651	1.09788	0.3905	1.2553	1.16662	1.67314
15	**15.38**	13	0.745	195	0.003	291	0.59475	0.8127	1.95161	1.84063	2.46521	3.23137
16	**16.15**	14	0.750	194	0.003	291	2.79793	4.70109	4.52796	5.48169	6.59873	3.62027

$$V(j) = \frac{1}{2(Q_{total} - j)} \sum_{q=1}^{Q_{total}-j} \left(h_{q+j} - h_q \right)^2$$

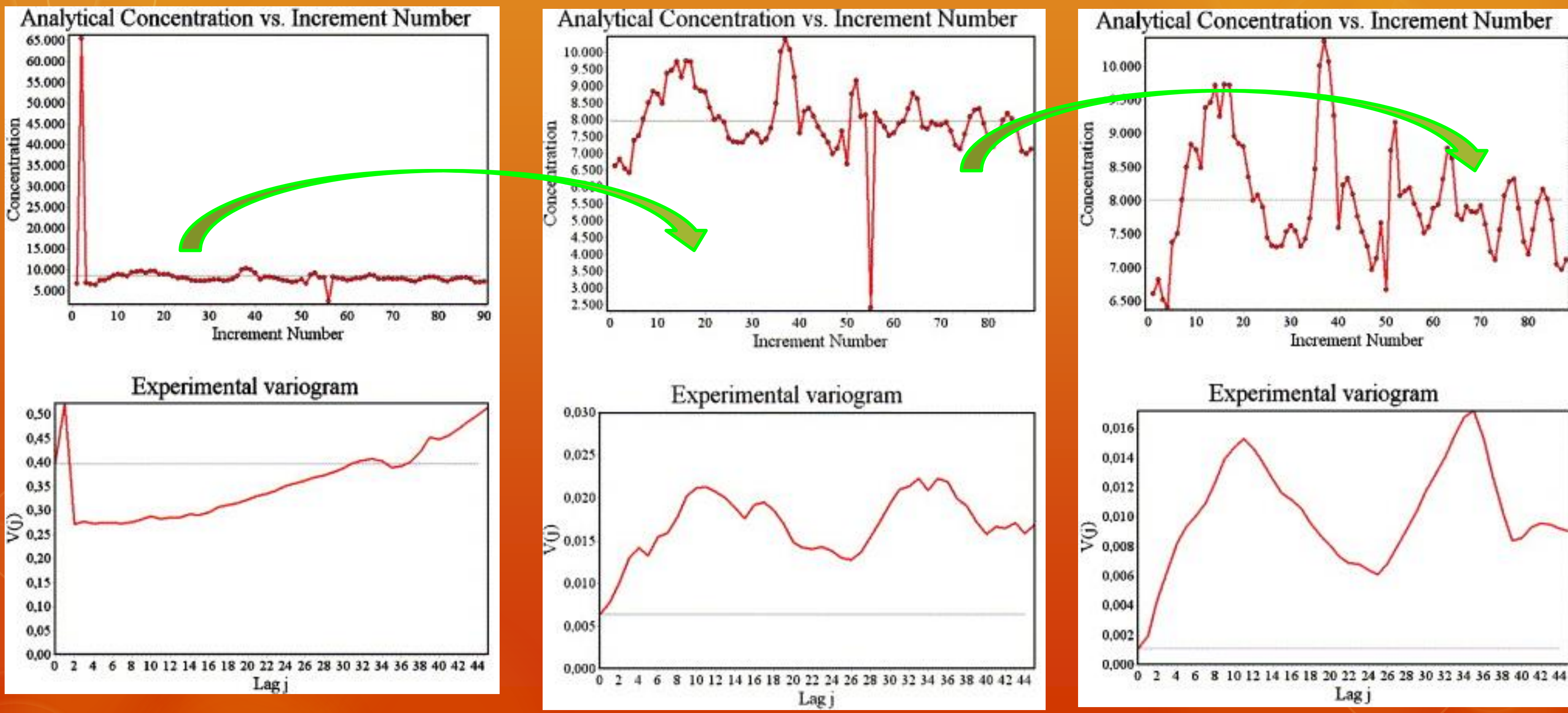
Extracting TSE+TAE information from real time measurements through variographic analysis



From raw data to variogram

Outliers significantly afect the variogram.

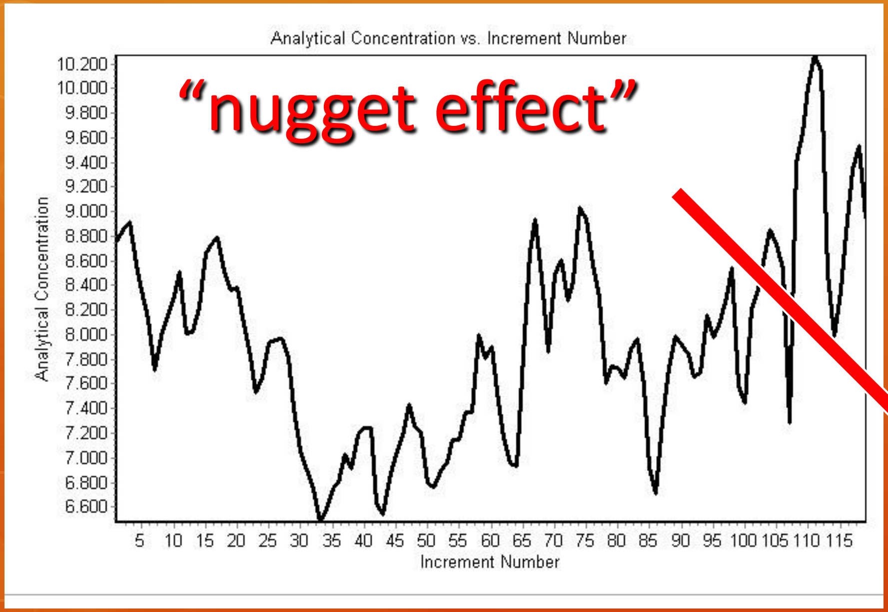
- ALL outliers should be removed - *sequentially*

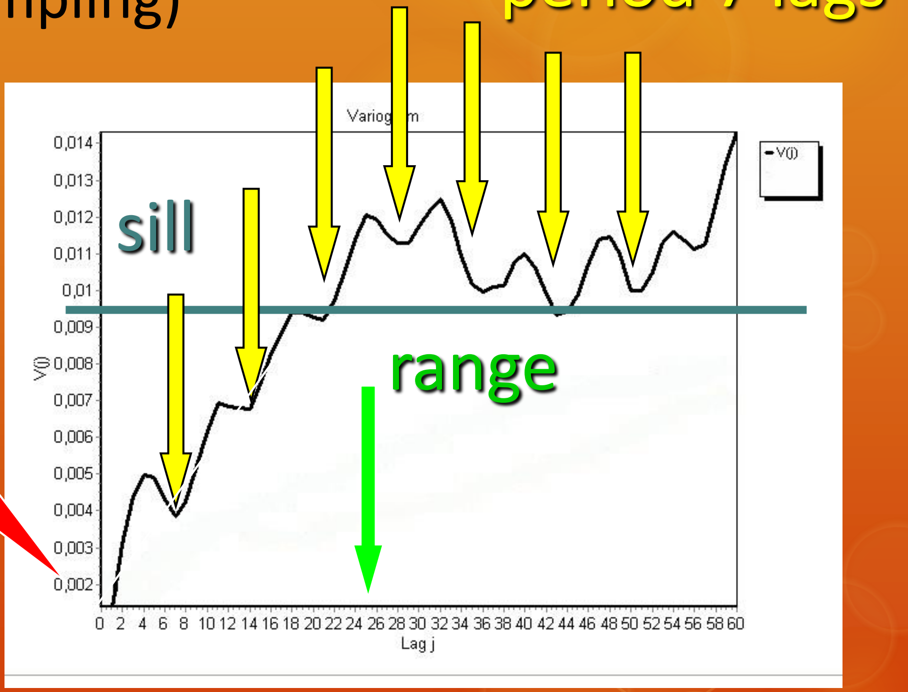


Sampling Principles: LHT (1-D case)

Variographics:

Ribe Biogas powerplant (daily sampling)





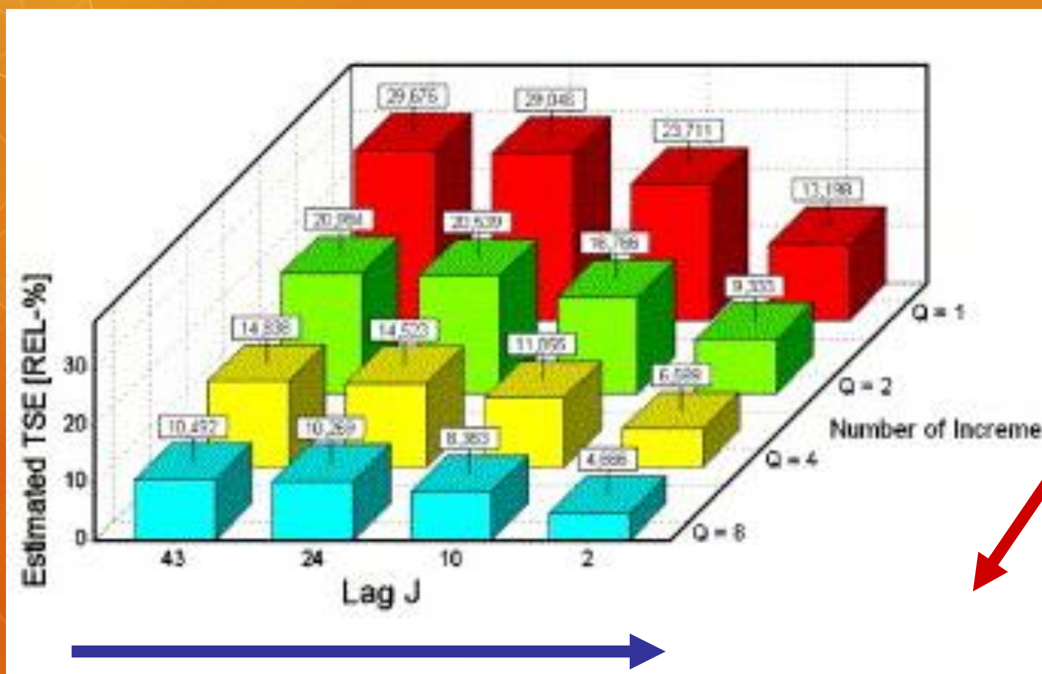
Analytical grade (CH_4 yield m^3)

Variogram and characteristic features

Decreasing Total Sampling Errors (1-D lots)

Sampling error plots

Total Sampling Error - f(V(j), Q, j)



Q ↑
TSE error ↓

Lag ↓, sampling rate ↑
TSE ↓

Direction of steepest error descent ...
Conditioned on a preselected error level