Overview of Uncertainty from Sampling

Prof Michael H Ramsey

Chair of Eurachem WG on UfS
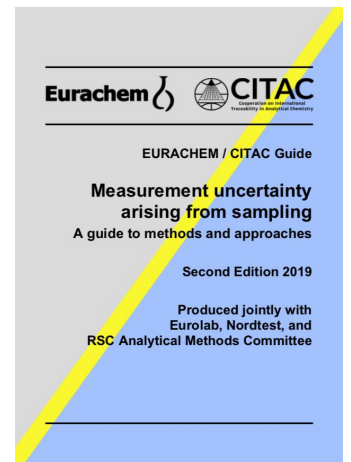
School of Life Sciences,

University of Sussex, Brighton, UK
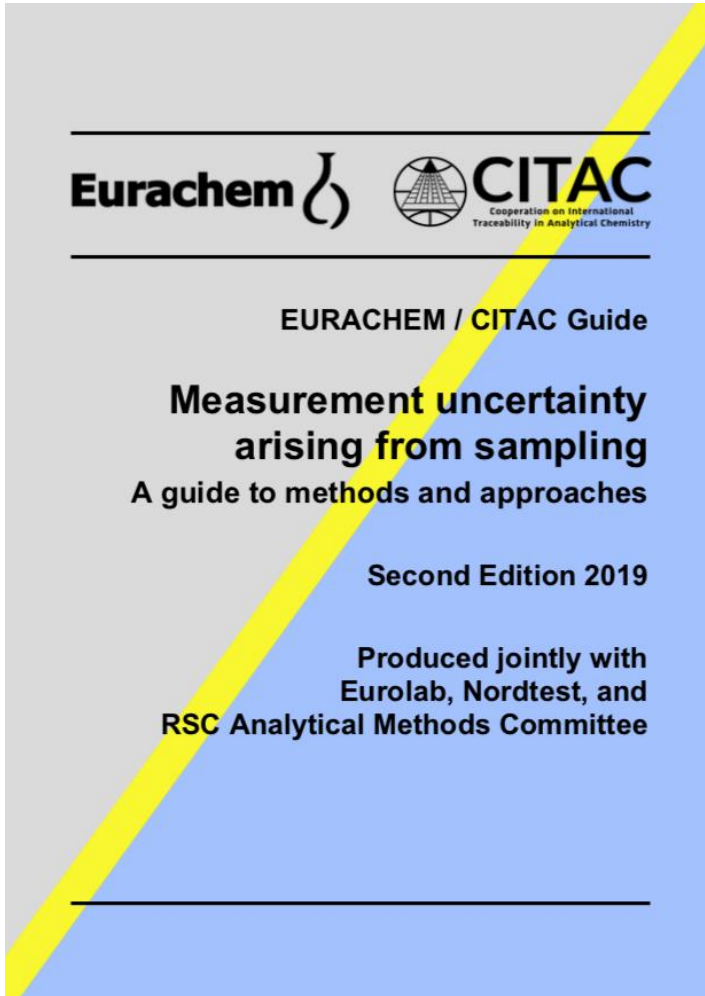
m.h.ramsey@sussex.ac.uk

Overview

- Sampling as part of the measurement process
  - So include Uncertainty from Sampling (UfS) within Measurement Uncertainty (MU)
  - UfS in Accreditation to ISO/IEC 17025:2017

- How to estimate UfS (and hence MU)

- Worked example for Lead in top soil
  - Also applicable to *in situ* measurements (- no physical sample taken) (e.g. PXRF)

- Advantages of knowing UfS within MU
  - More reliable compliance decisions,
  - Judge FFP of whole measurement process – for its Validation

- Worked example of Validation for Nitrate in Lettuce (*– Breakout Session)*

- Conclusions

US University of Sussex

Sampling within the Measurement Cycle
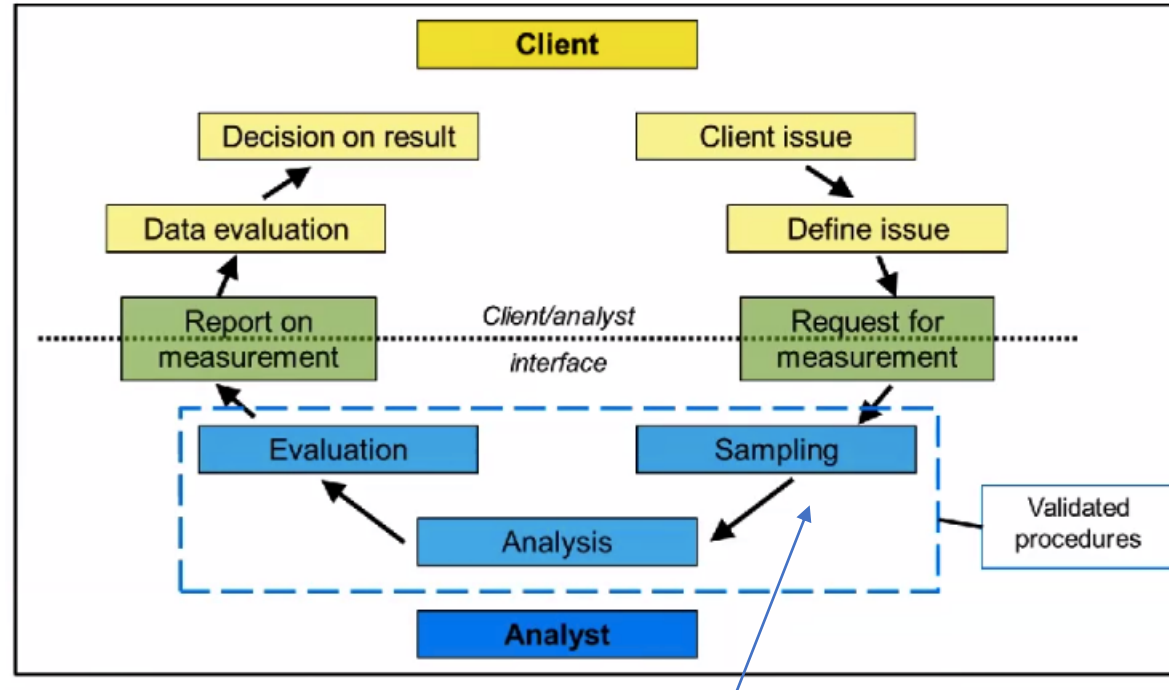
- Sampling and Analysis - both integral & essential parts of whole Measurement Cycle
  - which describes interaction between the analyst and client
    that drives the measurement and testing process

Estimation of UfS & MU – Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide*

Eurachem **CITAC**
Cooperation on International Traceability in Analytical Chemistry

EURACHEM / CITAC Guide

Measurement uncertainty arising from sampling
A guide to methods and approaches

Second Edition 2019

Produced jointly with Eurolab, Nordtest, and RSC Analytical Methods Committee

- Eurachem UfS Guide* describes several methods to estimate UfS
    - six worked examples for quantitative lab measurements made *ex situ* on wide range of analytes, in many different materials (e.g. food, feed, water and soil). Subsequently applied to gases, fuel etc.
- Most widely applicable approach for random components of MU (4 of 6 examples) - is 'Duplicate Method' based on a **balanced design** Only needs one sampler
- More sophisticated approach uses **multiple samplers**, e.g.:-
    - Sampling Proficiency Testing (SPT) results (briefly discussed here)
    - Collaborative Trial in Sampling (CTS) - mentioned

*Ramsey M.H., Ellison S. L. R., and Rostron P.(eds.) (2019) Eurachem/EUROLAB/CITAC/Nordtest/AMC Guide: *Measurement uncertainty arising from sampling: a guide to methods and approach,* Second Edition, Eurachem. http://www.eurachem.org/index.php/publications/guides/musamp

University of Sussex

Measurement Uncertainty (MU) -
including that arising from Sampling (UfS)

$x+U$ ↑ UCL

- MU (U) is *'an estimate attached to a test results (x)....*

  which characterises the range of values within which the true value is asserted to lie' [1]

  – 'True value' equivalent to '**Value of the Measurand**' in more recent definitions[2]
  – UfS mainly caused by small-scale heterogeneity of analyte within sampling target, so...

$x-U$ ↓ LCL

- Person resonsible for measurement needs to consider quality of primary sampling

  – as well as quality of instrumental analysis

- Primary metric for expressing quality of a measurement value is its uncertainty (MU)

  – To judge whether measurement value is Fit For its stated Purpose (FFP)

- **It is therefore essential to include UfS - to make a <u>realistic</u> estimate of MU**

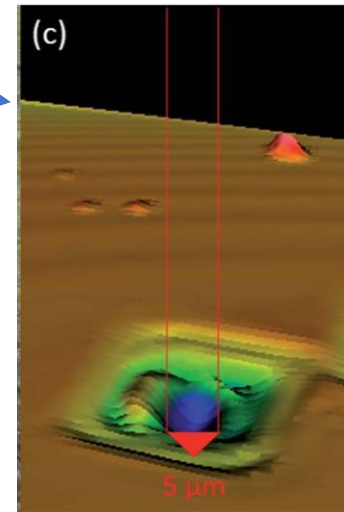 - Analytical MU alone may not include the true value
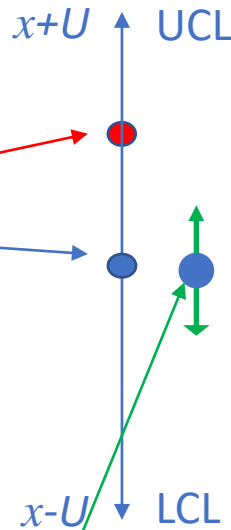
[1] Historic definition of MU from ISO 3534-1: 1993 Statistics – Vocabulary and Symbols, International Organization for Standardization, Geneva
[2] Parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand.
JCGM 100 (2008) / ISO/IEC Guide 98-3:2008

University of Sussex

Sampling as part of the measurement process

(a)

- Sampling is really the first step in the measurement process (traditional sampling at the macro scale, e.g. soil for *ex situ* analysis)

- *In situ* measurement techniques – sampling integral
  - Place the sensor→ make *in situ* measurement
    - taking a 'beam' sample at micro scale (e.g. mm or μm)
  - Uncertainty in sampling produces U in measurement value

- Physical sample preparation (in field or lab)
    - e.g. filter, acidify, dry, store, sieve, grind, split
  - is also part of the measurement process
  - and potentially important source of U

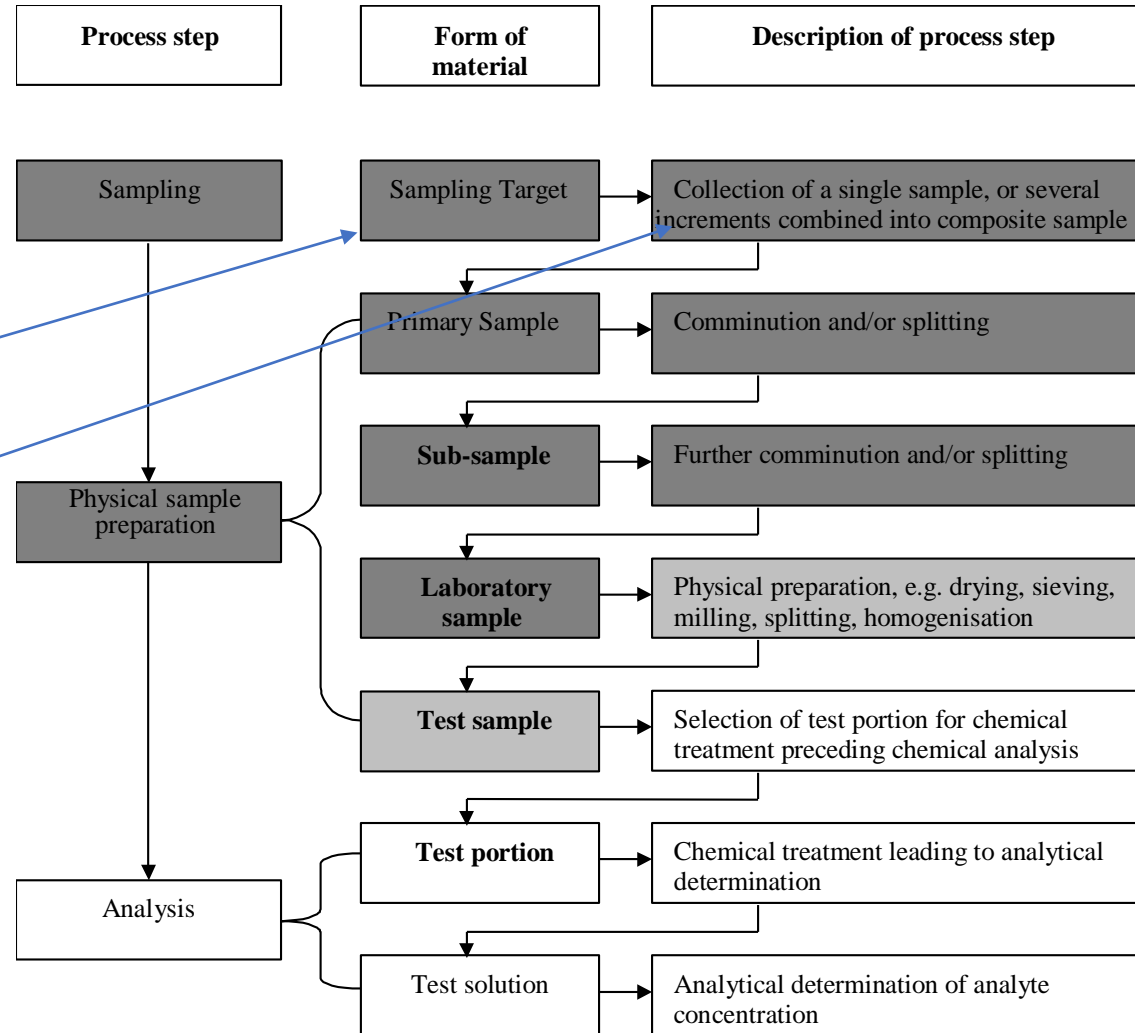- Include both in Validation and QC processes

(b)

hand-held portable X-ray Fluorescence (pXRF) on soil at 5 mm scale

(c)

Secondary Ion Mass Spectrometry (SIMS) on quartz, illustrating 5 μm beam scale

Analytical Methods Committee (2018) AMC Technical Brief No 84. Beam sampling: taking samples at the micro-scale, *Analytical Methods*, **10**, 1100-1102

US University of Sussex

Sampling as part of the measurement process

x10

Process step	Form of material	Description of process step
Sampling	Sampling Target	Collection of a single sample, or several increments combined into composite sample
Physical sample preparation	Primary Sample	Comminution and/or splitting
	Sub-sample	Further comminution and/or splitting
	Laboratory sample	Physical preparation, e.g. drying, sieving, milling, splitting, homogenisation
	Test sample	Selection of test portion for chemical treatment preceding chemical analysis
Analysis	**Test portion**	Chemical treatment leading to analytical determination
	Test solution	Analytical determination of analyte concentration

Sampling as part of the measurement process

- If objective is to measure true value of analyte concentration (or measurand)
  - in the **sampling target** (*i.e. portion of material, at a particular time, that a sample is intended to represent*) e.g. batch of food, area of soil, body of water, etc.
  - Then Sampling is included as part of the measurement process
  - UfS part of measurement uncertainty (& method validation and QC)
- If measurand (or true value) defined solely in terms of laboratory sample
  - Primary sampling is not included

- Most <u>users</u> of analytical measurements assume $x \pm U$ applies to target
  - not just to the lab sample

UfS in Accreditation to ISO/IEC 17025:2017

ISO/IEC 17025:2017 requires that UfS should be included in estimate of MU - unless it's explicitly excluded. i.e.

- Section 7.6.1 "Laboratories shall identify the contributions to measurement uncertainty. When evaluating measurement uncertainty all contributions that are of significance including those arising from sampling, shall be taken into account using appropriate methods of analysis".

- Section 7.6.3 "A laboratory performing testing shall evaluate measurement uncertainty. Where the test method precludes rigorous evaluation of measurement uncertainty, an estimation shall be made based on an understanding of the theoretical principles or practical experience of the performance of the method."

- ILAC Accreditation Committee recently concluded [4] that 7.6.3. is valid for sampling because sampling (in this context) is an activity associated with subsequent testing performed by a laboratory

Oehlenschlaeger E. (2019)

Example A2: Estimation of UfS in Soil - using Duplicate Method

Scenario:

- Former landfill, in West London
- 9 hectare = 90 000 m^2
- Potential housing development
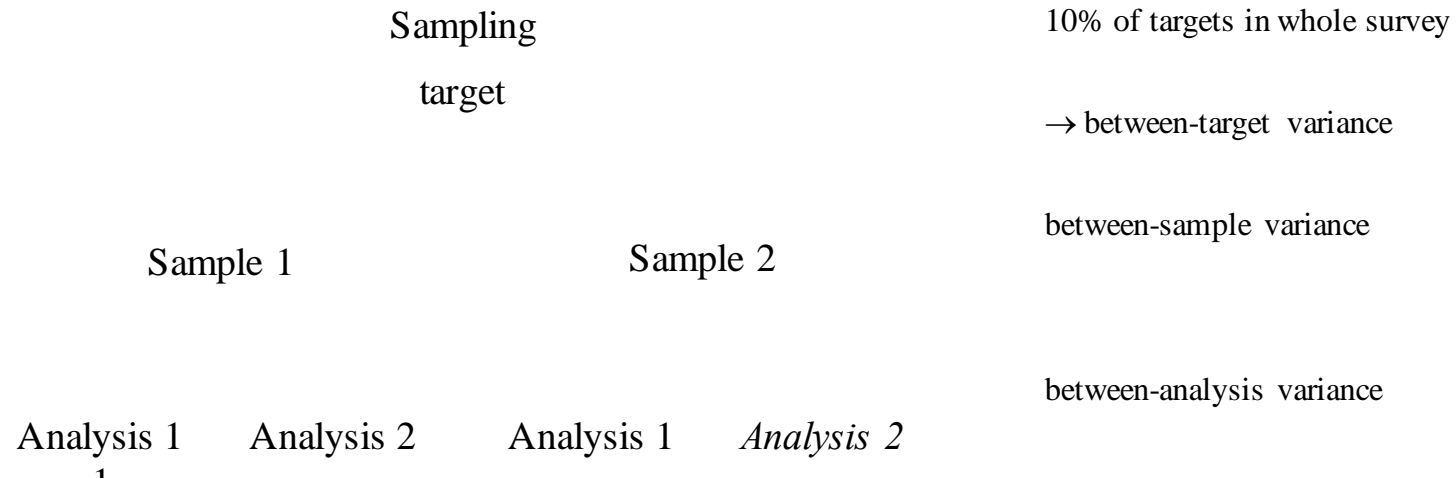- measurand → Pb conc. in each sampling target

Area of investigation:

- 300 m x 300 m area → depth of 0.15 m
- 100 sampling targets in a regular grid (10 x 10)
- 100 primary samples (taken with soil auger)
  - each intended to represent a 30 m x 30 m target

US University of Sussex

Application of Duplicate Method to estimate UfS

Figure 1: A balanced design

Sampling

target

10% of targets in whole survey

$\rightarrow$ between-target variance

between-sample variance

Sample 1   Sample 2

between-analysis variance

Analysis 1   Analysis 2   Analysis 1   *Analysis 2*

- Duplicate samples taken at 10/100 sampling targets (i.e. 10%)
  - randomly selected targets.
  - Duplicate sampling point 3 m from the original sampling point
    - within the sampling target, in a random direction, 3m = spatial uncertainty in survey
  - Reflects ambiguity in the sampling protocol
    - how differently could it be interpreted by a different samplers?

Sample prep and analysis in the lab

- Soil samples dried, sieved (<2 mm), ground (<100 µm)

- Test portions of 0.25g digested in nitric/perchloric acid

- Pb concentration measured with ICP-AES, under full AQC

- 6 soil CRMs for traceability and to estimate **analytical bias**
  - over range of concentration

- corrected for reagent blank concentrations
  - where statistically different to zero

- Raw measurements for use for estimation of uncertainty were:
  - **untruncated** – e.g. 0.0124 mg/kg, <u>not</u> < 0.1 or < detection limit
  - **unrounded**   – e.g. 2.64862 mg/kg, <u>not</u> 3 mg/kg

NIST – Wiley online

US University of Sussex

Spatial Map of Measured Pb concentration

Row	A	B	C	D	E	F	G	H	I	J
1	474	287	250	338	212	458	713	125	77	168
2	378	3590	260	152	197	711	165	69	206	126
3	327	197	240	159	327	264	105	137	131	102
4	787	207	197	87	254	1840	78	102	71	107
5	395	165	188	344	314	302	284	89	87	83
6	453	371	155	462	258	245	237	173	152	83
7	72	470	194	83	162	441	199	326	290	164
8	71	101	108	521	218	327	540	132	258	246
9	72	188	104	463	482	228	135	285	181	146
10	89	366	495	779	60	206	56	135	137	149

Argyraki (1997)

- Measured Pb concentration ranges from 56 to 3590 mg kg^{-1}

- Straddles then UK threshold of > 500 mg Pb kg^{-1} for action required (further risk assessment) – 8% of site

- Gives **Deterministic Map of the contamination (ignores MU) – 92% uncontaminated**

University of Sussex

Spatial Map of Measured Pb concentration

Row	A	B	C	D	E	F	G	H	I	J
1	474	287	250	338	212	458	713	125	77	168
2	378	3590	260	152	197	711	165	69	206	126
3	327	197	240	159	327	264	105	137	131	102
4	787	207	197	87	254	1840	78	102	71	107
5	395	165	188	344	314	302	284	89	87	83
6	453	371	155	462	258	245	237	173	152	83
7	72	470	194	83	162	441	199	326	290	164
8	71	101	108	521	218	327	540	132	258	246
9	72	188	104	463	482	228	135	285	181	146
10	89	366	495	779	60	206	56	135	137	149

Argyraki (1997)

- Uncertainty of measurements estimated by taking of Duplicate Samples at 10% of sampling targets

# Measurements from balanced design for UfS estimation

Sampling
target

- Large differences between some sample duplicates (e.g. D9) = high level of UfS

- Good agreement between analytical duplicates (< 10 % difference)

	Sample 1		Sample 2	
	Analysis 1	Analysis 2	Analysis 1	Analysis 2
Target #	S1A1	S1A2	S2A1	S2A2
A4	787	769	811	780
B7	338	327	651	563
C1	289	297	211	204
D9	662	702	238	246
E8	229	215	208	218
F7	346	374	525	520
G7	324	321	77	73
H5	56	61	116	120
I9	189	189	176	168
J5	61	61	91	119

mg kg^{-1}

- Needs inspection of frequency distribution to select the best approach to UfS estimation

Judge Frequency Distribution using Histograms

- Frequency distribution of Pb concentration <u>across the site</u> = long range heterogeneity
- Distribution of Pb measurements on 100 sampling targets is positively skewed = approximately log-normal
- Log-transformation necessary to remove skew
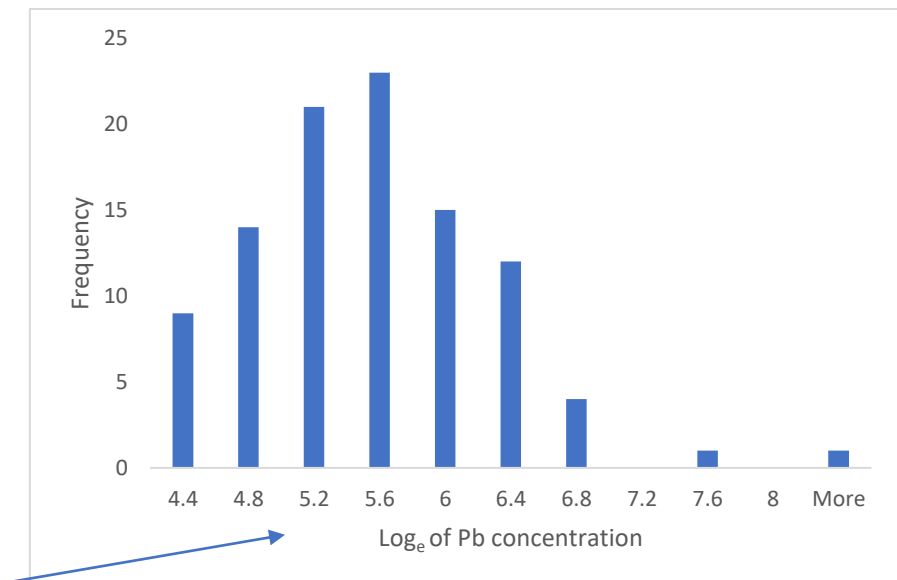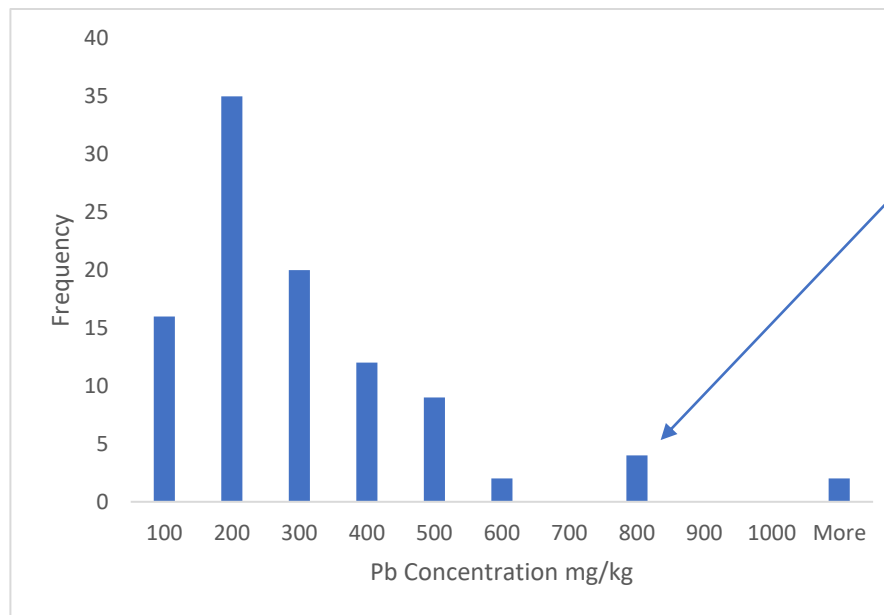
- Distribution closer to Normal after $\log_e$ transformation
  - Needed for use of ANOVA

University of Sussex

Need for log-transformation?

- Classical analysis of variance (ANOVA) assumes approximately normal distributions
- Robust ANOVA can accommodate up to 10% outlying values,
  - but not more, and not this heavy skew
- However, once transformed, measurement values (and ANOVA results) are no longer given in input units of concentration (e.g. mass fraction, mg kg $^{-1}$)

Measurement values of Pb concentration

In mg kg^{-1}

Target #	S1A1	S1A2	S2A1	S2A2
A4	787	769	811	780
B7	338	327	651	563
C1	289	297	211	204
D9	662	702	238	246
E8	229	215	208	218
F7	346	374	525	520
G7	324	321	77	73
H5	56	61	116	120
I9	189	189	176	168
J5	61	61	91	119

log$_e$-transformed

Target #	S1A1	S1A2	S2A1	S2A2
A4	6.67	6.65	6.70	6.66
B7	5.82	5.79	6.48	6.33
C1	5.67	5.69	5.35	5.32
D9	6.50	6.55	5.47	5.51
E8	5.43	5.37	5.34	5.38
F7	5.85	5.92	6.26	6.25
G7	5.78	5.77	4.34	4.29
H5	4.03	4.11	4.75	4.79
I9	5.24	5.24	5.17	5.12
J5	4.11	4.11	4.51	4.78

- Need a different way to express MU in this case = Uncertainty factor* $^{F}U = \exp(2s_G)$
  - s_G = standard deviation of the log$_e$-transformed values ($= s(\log_e(x))$)
  - $x ^{\times}/ {}^{F}U$ (rather than $x \pm U$)

What is the uncertainty factor? Eurachem-AMC Information Leaflet, https://www.eurachem.org/index.php/publications/leaflets/uncertainty-factor

University of Sussex

RANOVA3 output for Soil Example A2

Classical ANOVA

Mean	317.8	No. Targets		10
Total Sdev	240.19			
	Btn Target	Sampling	Analysis	Measure
Standard deviation	197.55	135.43	17.99	136.62
% of total variance	67.65	31.79	0.56	32.35
Expanded relative uncertainty (95%)		85.23	11.32	85.98
Uncertainty Factor (95%)		2.6032	1.12	2.6207

Robust ANOVA

Mean	297.31			
Total Sdev	218.49			
	Btn Target	Sampling	Analysis	Measure
Standard deviation	179.67	123.81	11.144	124.31
% of total variance	67.63	32.11	0.26	32.37
Expanded relative uncertainty (95%)		83.29	7.50	83.63

- Software RANOVA3* (in Excel) performs:-

- Classical ANOVA gives poor estimate of U' = 85.98%,

- but also estimate of ^FU as 2.62

  - after $\log_e$-transformation within RANOVA3

Robust U as 83.63% (for comparison)
Histogram suggests > 10% of outlying values,
so direct classical, and robust estimate are not very reliable
So log-transformation before classical ANOVA is likely to be a better option

- Analytical bias estimated with CRMs was - 3.41 % ± 1.34 %
  - has negligible effect when added into this MU

University of Sussex

Confidence Limits on Measurement Value

- *For $^F U$ = 2.62*, for a typical Pb measurement value of 300 mg kg^{-1}

  Upper confidence limit (UCL) = 784 mg kg^{-1} (300 x 2.62) $\longrightarrow$

  Measurement value of 300 mg kg^{-1} $\longrightarrow$

- Lower confidence limit (LCL) = 115 mg kg^{-1} (300 / 2.62)

- <u>Asymmetric confidence limits</u> around the measured value

- -185 and +484 mg kg^{-1} (away from 300)

- Reflects skew in frequency distribution of the uncertainty
    - as seen in histograms

- Not seen in <u>symmetrical confidence limits</u> from robust U' = 83.6% = 251 (300 * 0.836)

- = +/- 251 mg kg^{-1} UCL = 551 (300 + 251)

  (calculated without log-transformation).   LCL = 49   (300 - 251)
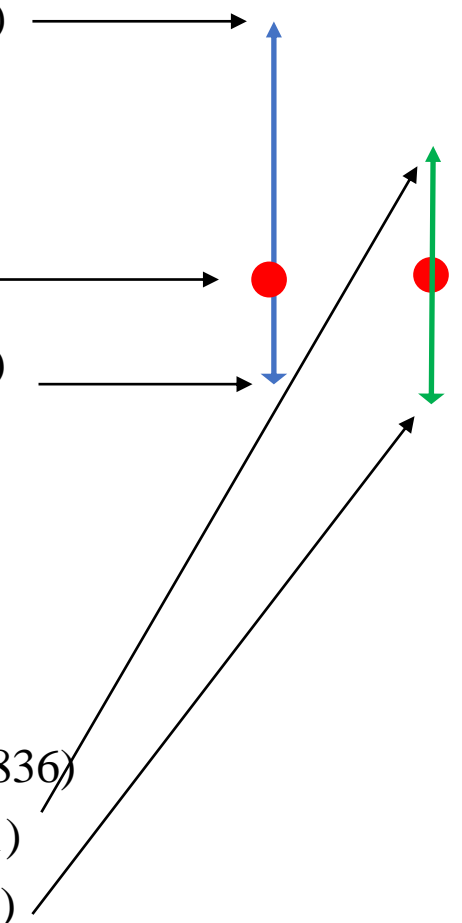
- $^F U$ **can be used to make probabilistic compliance map** of Pb contamination
- – only **46% definitely uncontaminated** (not 92% from deterministic) = <u>More reliable compliance decisions</u>

Example A1 from Eurachem UfS Guide:
Nitrate Concentration in Lettuce + Validation

- EU threshold 4500 mg kg^{-1} for nitrate concentration of Sampling Target[1]
  - i.e. ~ 12,000 – 20,000 heads in each bay/batch/target

- Current EU sampling protocol[2] specifies taking 10 heads (increments)
  - to make a single **composite sample** from each Sampling Target

- Analytical procedure/method (HPLC[3]) already validated using Collaborative Trial[4]
  - $U_{analysis}$ around 6% at that validation (RSDR = ~ 3%)

- <u>Need to validate the whole measurement procedure</u> (including sampling & sample prep)

- MU is key metric that affects compliance decisions
  - MU is affected by (and reflects) all of metrics for the measurement procedure
    - precision, bias, LOD, working range, selectivity, sensitivity, ruggedness
  - how much MU from the sampling (UfS)?

- Judge FFP of measurement procedure by the MU - is it close to Target MU?

1. Commission Regulation (EC) No 563/2002 of 2 April 2002 amending Regulation (EC) No 466/2001
2. European Directive 79/700/EEC. OJ L 207, 15.8.1979, p26.
3. BS EN 12014-2:1997, Foodstuffs. Determination of nitrate and/or nitrite content. General considerations
4. Farrington et al.,(2006), Journal of the Association of Public Analysts (Online), 34, 1-11

University of Sussex

UfS estimation for Lettuce using Duplicated 'W' Sampling Design
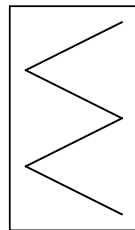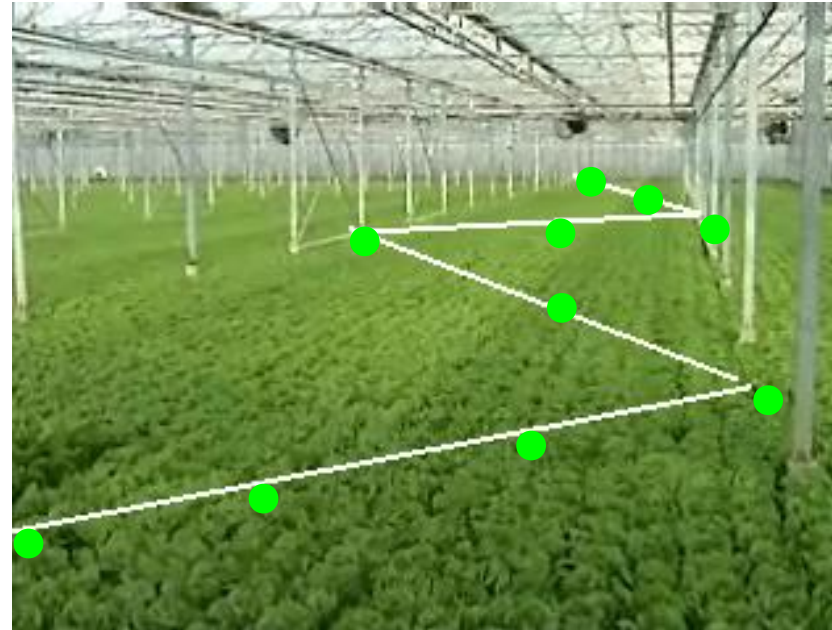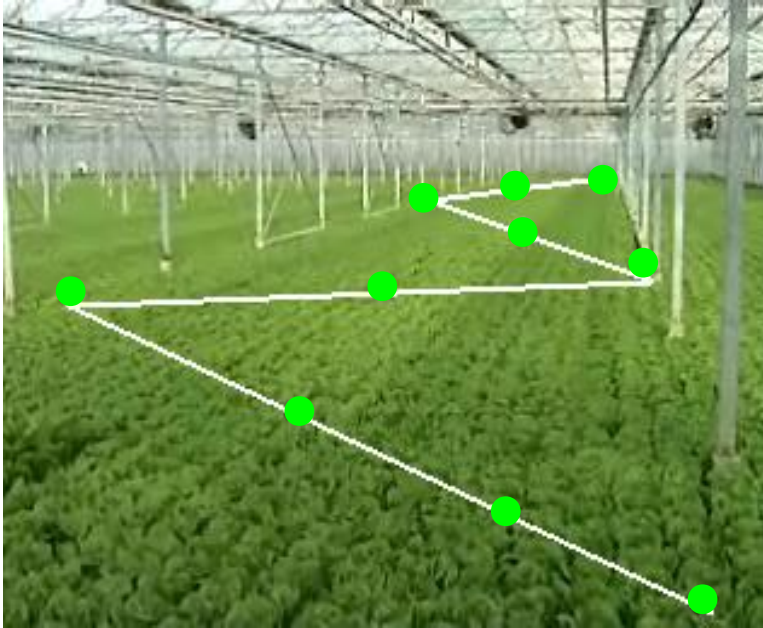
Sampling target
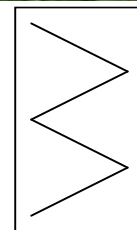
Duplicate sample is equally likely interpretation of 'W' design

Sample 1   Sample 2

Analysis 1   Analysis 2   Analysis 1   Analysis 2

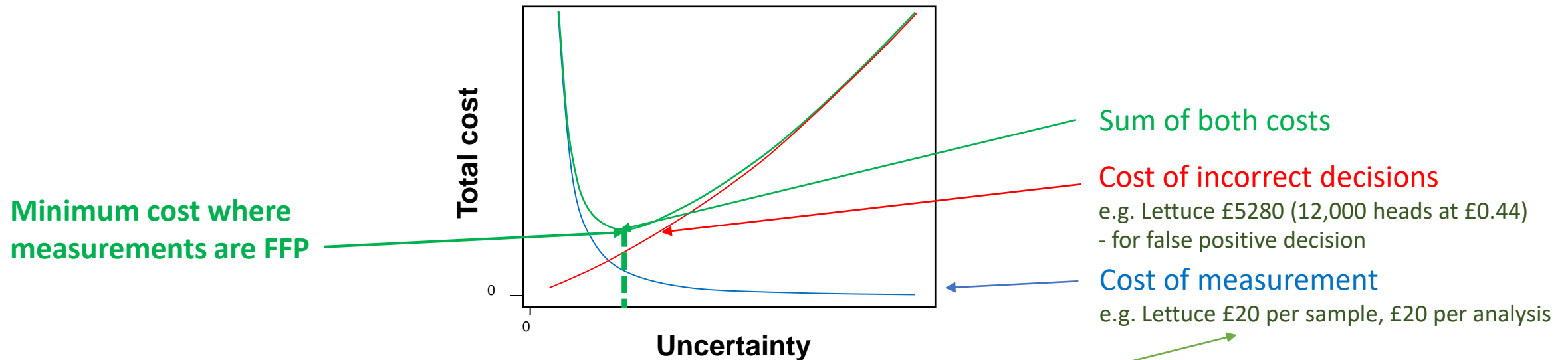# Estimating UfS (and MU) with Duplicate Method

- Use Duplicate Method – *as described earlier in talk*

- Selected 8 typical sampling targets (bays of ~20,000 lettuce) – sampled in duplicate

- Estimated UfS and MU using Analysis of Variance (ANOVA) – program RANOVA3*

- Selected Robust ANOVA - as it *accommodates < 10% outlying values*

- ***U' = 16.4%*** (**s_{meas} = 360 mg kg^{-1}**) – *as repeatability*

  - MU dominated by UfS (78% of MU)

  - UfS mainly caused by nitrate heterogeneity within the sampling targets

- *U'*$_{anal}$ = 7.6% – *as repeatability*

  - Very similar to MU = 6 % reported at separate validation of the analytical procedure[4]
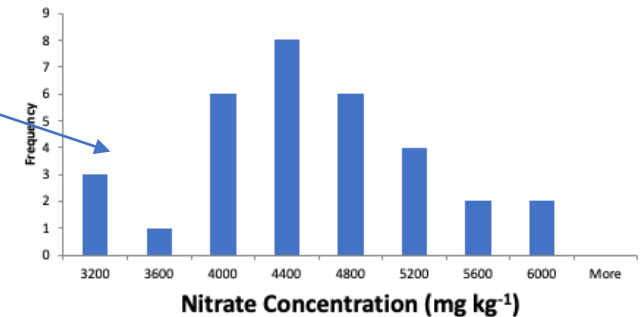
FFP at Target MU – e.g. set at minimum overall cost

- **Validation by judging Fitness for Purpose** (FFP)
- Target MU - can be Option (1) set externally (e.g. arbitrary 20%, 16% < 20% so FFP), or Option (2)…
- At MU that minimses the overall cost (including the consequences of incorrect decisions)

Minimum cost where measurements are FFP

Sum of both costs

Cost of incorrect decisions
e.g. Lettuce £5280 (12,000 heads at £0.44)
- for false positive decision

Cost of measurement
e.g. Lettuce £20 per sample, £20 per analysis

(y-axis) **Total cost**

(x-axis) **Uncertainty**

0

0

- By knowing UfS, can judge how <u>Target MU</u> (set by either Option) <u>achieved most cost-effectively</u> by:
  – Spending more (or less) on **chemical analysis** (e.g. more precise technique), or
  – Spending more (or less) on **sampling** (e.g. taking more increments)

Judge FFP - level of Uncertainty

- For lettuce example estimate MU (s_{meas}) using Duplicate Method
- Calculate Target MU using optimised uncertainty (OU) method
- **Measurement Procedure is judged as NOT FFP**

Cost ↑

Uncertainty→
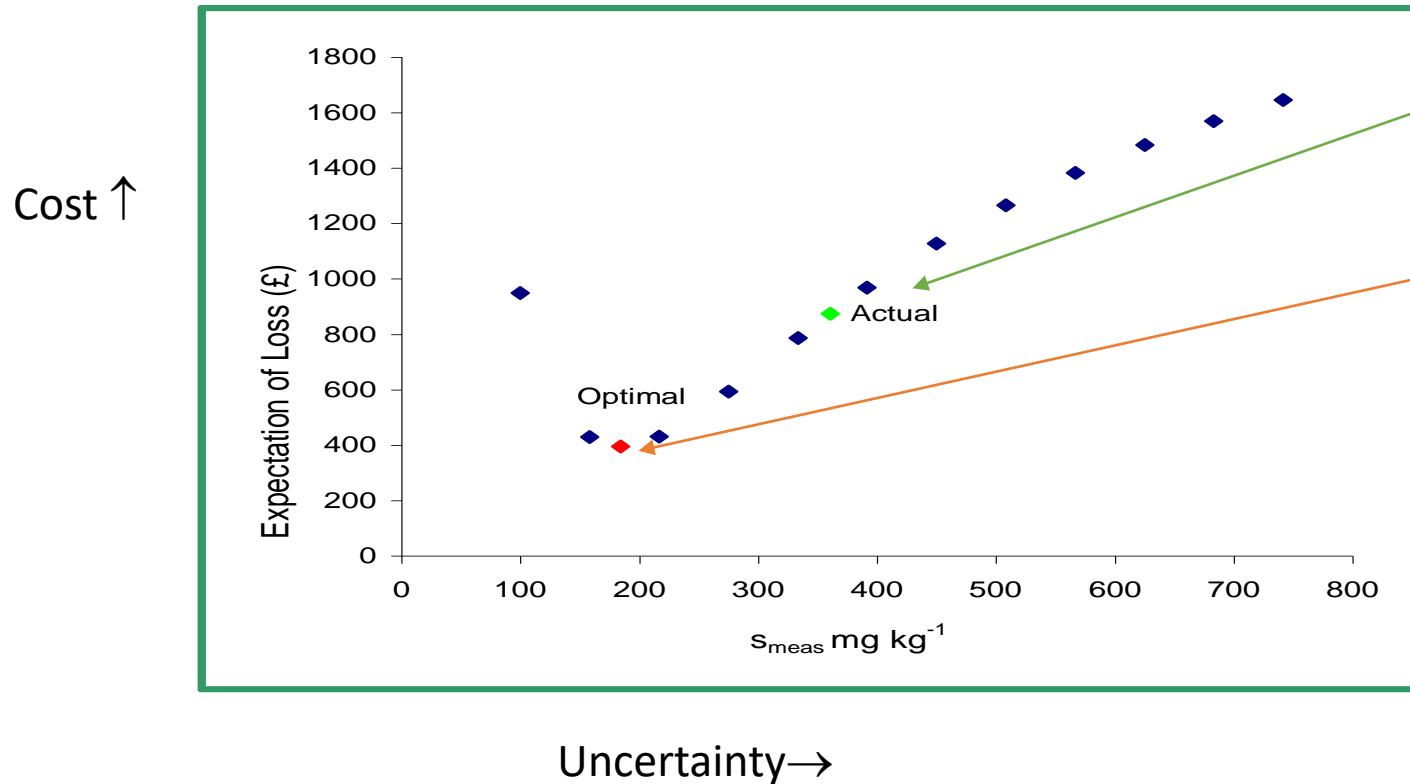
Actual MU (360 mg kg^{-1}) i.e. $U' = 16.4\%$
 - and consequent cost (£800 per target) is much higher than…

Optimal MU value (184 mg kg^{-1}) i.e. $U' = 8.3\%$
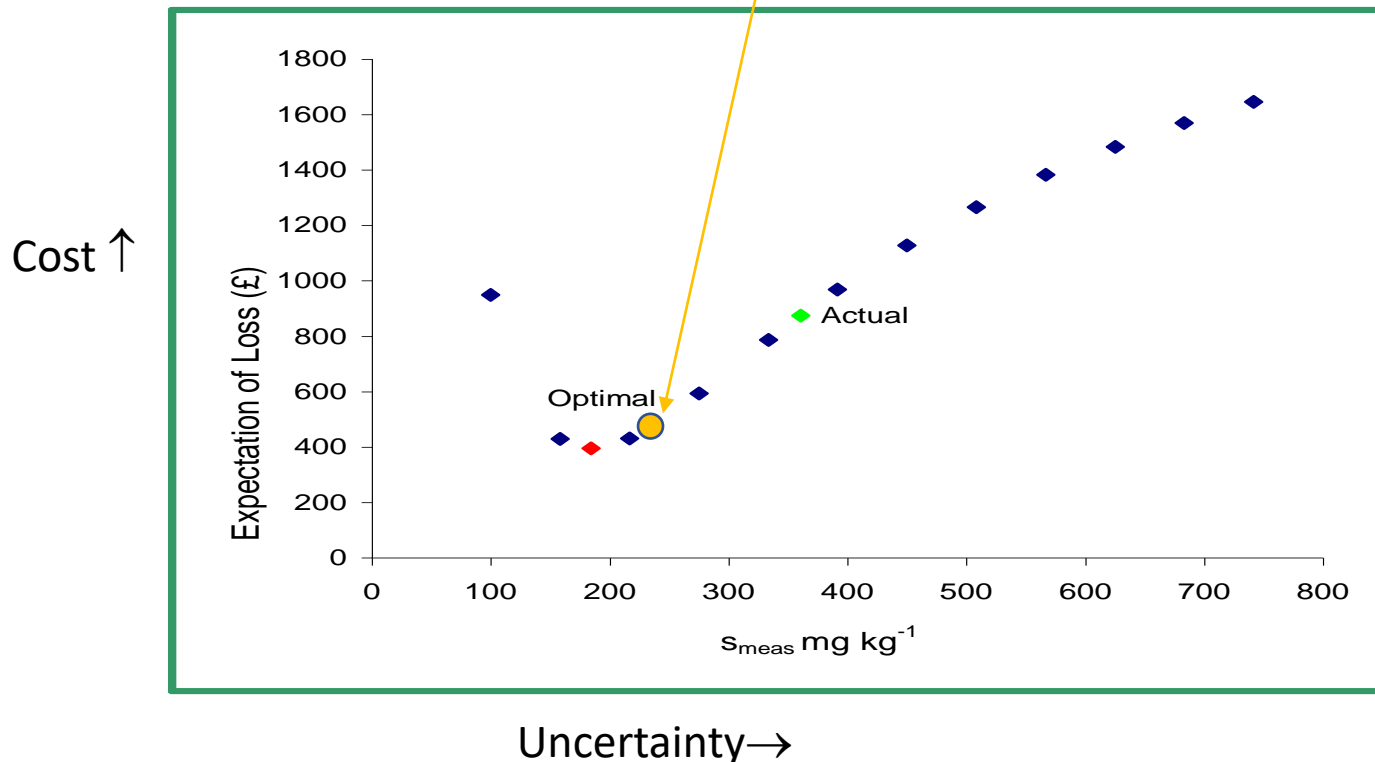At minimum cost (£400)

To achieve FFP - we need to reduce the MU by factor of 2

UfS accounts for 78% of MU (from ANOVA)
- So reducing UfS is most cost-effective

Sampling Theory predicts we can reduce UfS x2 by increasing sample mass by factor of 4 (= 2^2)

So take composite sample with 40 heads instead of 10 heads

Reducing the Uncertainty – to achieve FFP

- **Increasing number of increments from 10 to 40 heads**
- Reduces s_{samp} from 319 to 177 mg kg^{-1} - by a factor of x 1.8 (similar to model prediction of x2)
- Reduces MU (s_{meas}) from 360 to 244 mg kg$^{-1.}$ (U' from 16.4 % to 11.1%)
- Close to the optimal value (184 mg kg^{-1}) at similar Cost (~£500, down from £800 per target)
- **Achieves Fitness-for-Purpose (FFP)** = MU that minimises to overall financial loss

Cost ↑

Uncertainty→

Lyn, J.A., Palestra, I.M., Ramsey, M.H., Damant, A.P. and Wood, R. (2007) Modifying uncertainty from sampling to achieve fitness for purpose: a case study on nitrate in lettuce *Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement,* 12, 67-74

Estimate of Uncertainty using SPT - including Between-Sampler Bias
- Example using Sampling PT for moisture in butter*

ANOVA: U' as % on concentration of moisture in butter (200 tons)

$\approx$ Duplicate Method (single sampler) gives **$U' = 0.39$ %**

SPT (multiple samplers, n=9)   gives $U' = 0.87\%$

- U' larger x 2.2 - **includes bias between-samplers**

Remove two samplers with potentially non-proficient z-scores (RSz > 3)

SPT (n=7)   gives $U' = 0.69\%$

- **U' still larger x 1.8**

- **a more reliable estimate of Uncertainty**

- Ideally apply over multiple rounds of SPT, if targets comparable

- e.g. 16 rounds, stack-gas measurement SPT [Coleman et al ,2013, Accred Qual Assur 18:517–524]

- Multiple samplers (e.g. in CTS) better for Validation of Sampling

- More expensive than Duplicate Method, but sometimes justified

* Ramsey M.H. Geelhoed B, Damant, A.P., Wood, R. (2011) Improved evaluation of measurement uncertainty from sampling by inclusion of between-sampler bias using sampling proficiency testing. Analyst, 136 (7), 1313 – 1321. DOI:10.1039/C0AN00705F.

University of Sussex

Conclusions

- Including sampling within the measurement process:
  - Is essential for making reliable estimates of MU (including UfS)
  - Conforming to ISO/IEC 17025:2017
  - Being able to judge FFP, and hence validate the whole measurement process

- UfS (and hence MU) can be estimated with the Duplicate Method
  - Applicable to any sampling medium: soil, sediment, herbage, waters, gases etc.
  - Also applicable to *in situ* measurements (such as PXRF)*
  - Sampling PT (or CT) results can be used to include between-sampler bias within MU

- Including UfS within MU is crucial for:
  - Reliable compliance decisions
    - E.g. are concentration levels above from regulatory limits?
  - Rigorous validation of the whole measurement process *(later Breakout Session)*

* Worked example in Ramsey M.H. (2020) Measurement Uncertainty from Sampling: Implication for Testing, Diagnostics and Inspection. Presentation to 17th IMEKO TC 10 and EUROLAB Virtual Conference *"Global Trends in Testing, Diagnostics & Inspection for 2030"* October 20-22, 2020. https://www.imeko.org/publications/tc10-2020/IMEKO-TC10-2020-042.pdf

US University of Sussex

Effect of MU on geochemical interpretation

Concentration (C)

$C \times {}^{F}U$
$C \times 2.63$

$C / {}^{F}U$
$C / 2.63$

C+U

C

C-U

Threshold (T)   **500** – then UK limit

for risk assessment

Uncontaminated   Possibly Contaminated   Probably Contaminated   Contaminated   - Probabilistic Classification

Case Study   <190   190-500   **500-1318**   **>1318** mg/kg

Map Symbol

Uncertainty Factor ${}^{F}U$ = 2.63, so   190 = 500 / 2.63   1318 = 500 x 2.63

Probabilistic Geochemical Mapping using MU
Example for Pb at Hounslow Site

Deterministic Map
- *Ignores MU*
- **92% 'uncontaminated'**

Probabilistic Map
- *Allows for MU*
- **46% 'uncontaminated'**

Bettencourt da Silva, R., Argyraki, A., Borges, C., Ramsey, M.H. (2022) Spatial modelling of concentration in topsoil using random and systematic uncertainty components. Analytical Letters 210574656 https://www.tandfonline.com/doi/full/10.1080/00032719.2022.2050383