Measurement uncertainty from sampling and its roll in validation of measurement procedures

Prof. Michael H Ramsey

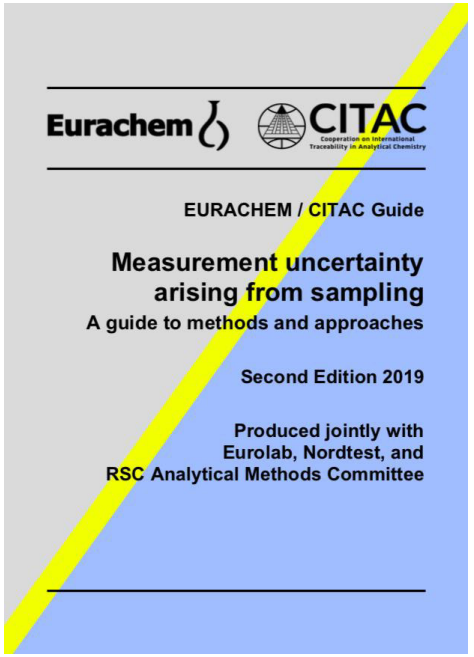
Chair of Eurachem UfS Working Group

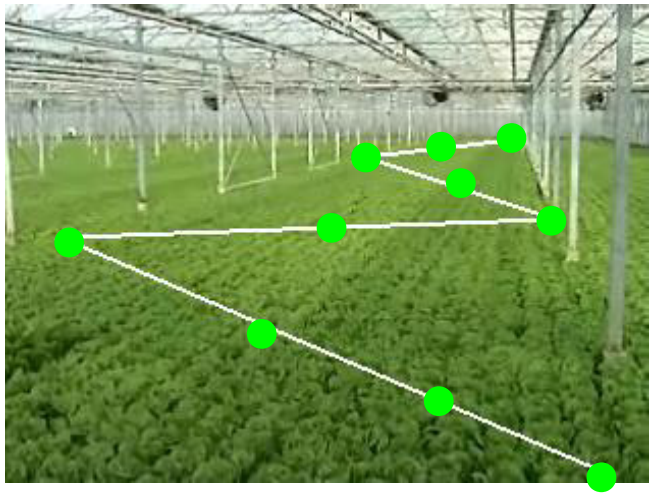
School of Life Sciences,

University of Sussex, Brighton, UK

m.h.ramsey@sussex.ac.uk

Eurachem Workshop 2023
Bern-Wabern, 22nd to 23rd June 2023





University of Sussex

Overview of Talk

- Sampling as part of the measurement process
- Uncertainty (U) of measurement values – *the key metric*
    - why to include the contribution from sampling (UfS)
- Estimation of UfS (& MU) - Mainly using Duplicate Method
    - but also using Sampling Proficiency Testing
- Validation of measurement procedures including sampling (VaMPIS)
  - by judging of fitness for purpose (FFP) of measurement values & procedures
  - example for an *ex situ* measurement procedure
  - gives improved reliability of compliance decisions
- Conclusions

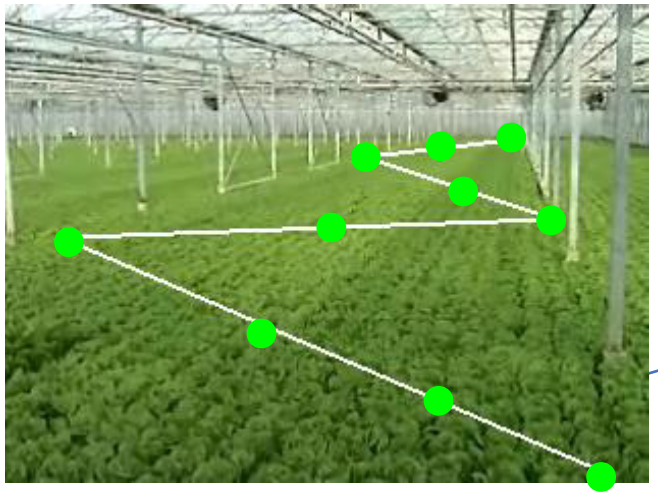
Sampling as part of the measurement process

- Sampling really the first step in the measurement process
- *In situ* measurement techniques reveal this
  - Place the sensor→ make measurement = taking a sample
  - Uncertainty from sampling produces MU in measurement
- 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 MU
  - include in the validation and QC processes (often omitted by labs)

siebtechnik-tema.com

University of Sussex

Sampling as part of the measurement process

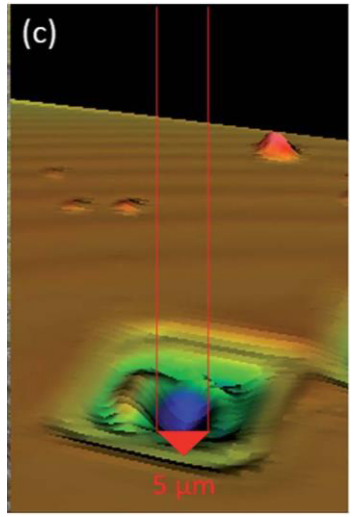


Sampling Target (at macro scale)
Portion of material, at a particular time, that a sample is intended to represent

x10

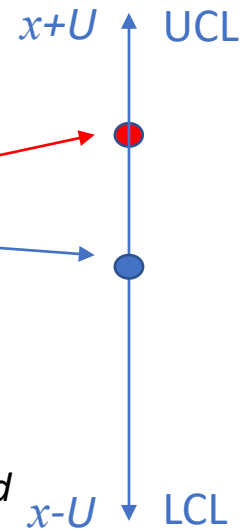
Process step	Form of material	Description of process step
Sampling	Sampling Target	Collection of a single sample, or several increments combined into composite sample
	Primary Sample	Comminution and/or splitting
Physical sample preparation	**Sub-sample**	Further comminution and/or splitting
	Laboratory 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

More careful use of the word ‘sample’



Sampling Target at micron scale

Measurement Uncertainty (MU) – *the key metric*



- Historially: 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
  - *Parameter, associated with the result of a measurement, that characterises the dispersion of the values that could reasonably be attributed to the measurand.* [2]
  - UCL = Upper Confidence Limit, LCL = Lower Confidence Limit.
  - **Confidence Interval** (CI) is between LCL and UCL
- Includes both Random effects (e.g. precision) and Systematic effects (e.g. bias)
- MU arises from all steps in measurement (e.g. sampling & physical sample prep.) - *in ISO/IEC 17025*
- Key parameter of measurement (and sampling) quality
- Doesn't assume measurements (or sampling) are 'correct' – hence 'representative'
  - traditional approach to Sampling Quality

[1] Historic definition of MU from ISO 3534-1: 1993 Statistics – Vocabulary and Symbols, International Organization for Standardization, Geneva
[2] JCGM 100 (2008) / ISO/IEC Guide 98-3:2008

University of Sussex

Statistical model

for ***Empirical*** estimation of uncertainty - One Sampling Target

$$x = X_{true} + \varepsilon_{sampling} + \varepsilon_{analytical}$$

x = ***measured*** **value** of the analyte concentration in one sampling target

X_{true}= ***true*** **value** of the analyte concentration in the sampling target

$\varepsilon_{sampling} + \varepsilon_{analytical}$ = effects on measured concentration from sampling and analysis

Variance (standard deviation squared) **of measurement value** = σ^2_{meas}

$$\sigma^2_{meas} = \sigma^2_{sampling} + \sigma^2_{analytical}$$

$\sigma^2_{sampling}$ is the between-sample variance on one target (largely due to analyte heterogeneity)
$\sigma^2_{analytical}$ is the between-analysis variance on one sample (as Repeatability)

For ***estimates*** of variance, we have:

$$s^2_{meas} = s^2_{sampling} + s^2_{analytical}$$

Statistical model

for ***Empirical*** estimation of uncertainty - **Multiple Sampling Targets**

Multiple sampling targets (n > 8) are needed for more realistic estimate of MU & UfS – using SPT

$$x = X_{true} + \varepsilon_{target} + \varepsilon_{sampling} + \varepsilon_{analytical}$$

ε_{target} represents the variation of concentration between the targets

and has variance $\sigma^2_{between-target}$.

Variance of measurement value = $\sigma^2_{meas} = \sigma^2_{sampling} + \sigma^2_{analytical}$

$$\sigma^2_{total} = \sigma^2_{between-target} + \sigma^2_{sampling} + \sigma^2_{analytical}$$

For our estimates of variances, we have:-

$$s^2_{total} = s^2_{between-target} + s^2_{sampling} + s^2_{analytical}$$

s^2_{meas}

University of Sussex

How MU is expressed & reported

- MU usually expressed using standard deviation (*s*), e.g.:-

1. Standard uncertainty (*u*)

$$u = s_{meas} \text{ (often } = s_{analytical})$$

2. Expanded uncertainty (*U*)

$$U = ks_{meas} = 2s_{meas}$$

with coverage factor (*k*) of 2 for 95% confidence

- may need k > 2 for *U* based upon small number of samples*

3. Expanded relative uncertainty (*U′*)

$$U' = 100\frac{2s_{meas}}{x}\%$$

– for measurement value (x)

MU can also be expressed as a Confidence Interval, e.g. $= x \pm U$

* Rostron P.D., Fearn T., and Ramsey M.H. Improved coverage factors for Expanded Measurement Uncertainty calculated from two estimated variance components. ACQUAL - submitted 2023

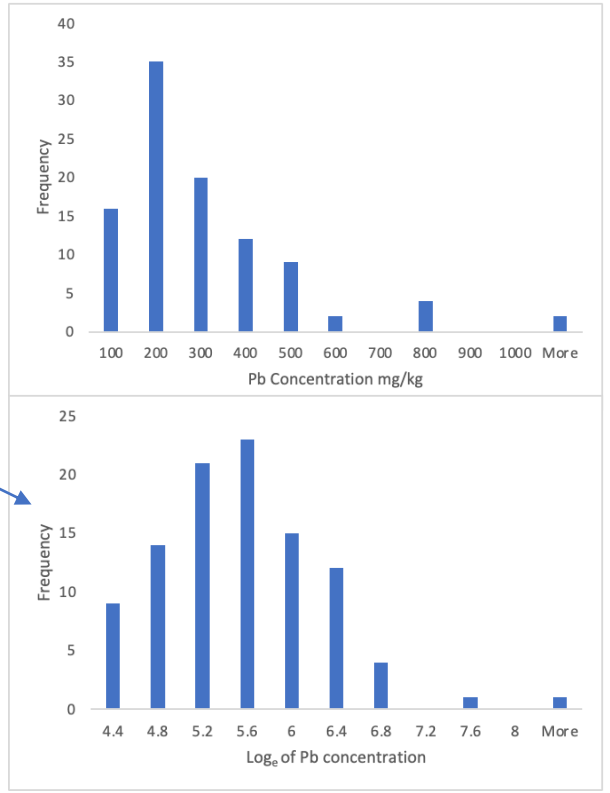
University of Sussex

MU expressed as Uncertainty Factor

4. Uncertainty Factor (^{F}U)

$$^{F}U = \exp\left(2s_{G,meas}\right)$$

– $s_{G.meas}$ is SD of $\log_e$–transformed measurement values [1]



Confidence Interval = x ˣ/ ^{F}U

ˣ/ called *'times over'*

[1] What is the uncertainty factor? Eurachem-AMC Information Leaflet, May 2021
https://www.eurachem.org/index.php/publications/leaflets/uncertainty-factor

University of Sussex

Four empirical methods for estimating uncertainty
including that from sampling

Method #	Method description	Samplers (People)	Protocols/ Procedures	Component estimated			
				Sampling Precision	Sampling Bias	Anal. Precision	Anal. Bias
1	Duplicates	single	single	Yes	No	Yes[3]	No[1]
2	Multiple protocols	single	multiple	between protocols		Yes[3]	No [1]
3	CTS	multiple	single	between samplers		Yes	Yes [2]
4	SPT	multiple	multiple	between protocols +between samplers		Yes	Yes [2]

Example of SPT later

CTS = Collaborative Trial in Sampling (use in Validation), and SPT = Sampling Proficiency Test.

Simplest Empirical method is 'Duplicate Method' (#1) – *applied in 4 Examples in UfS Guide*

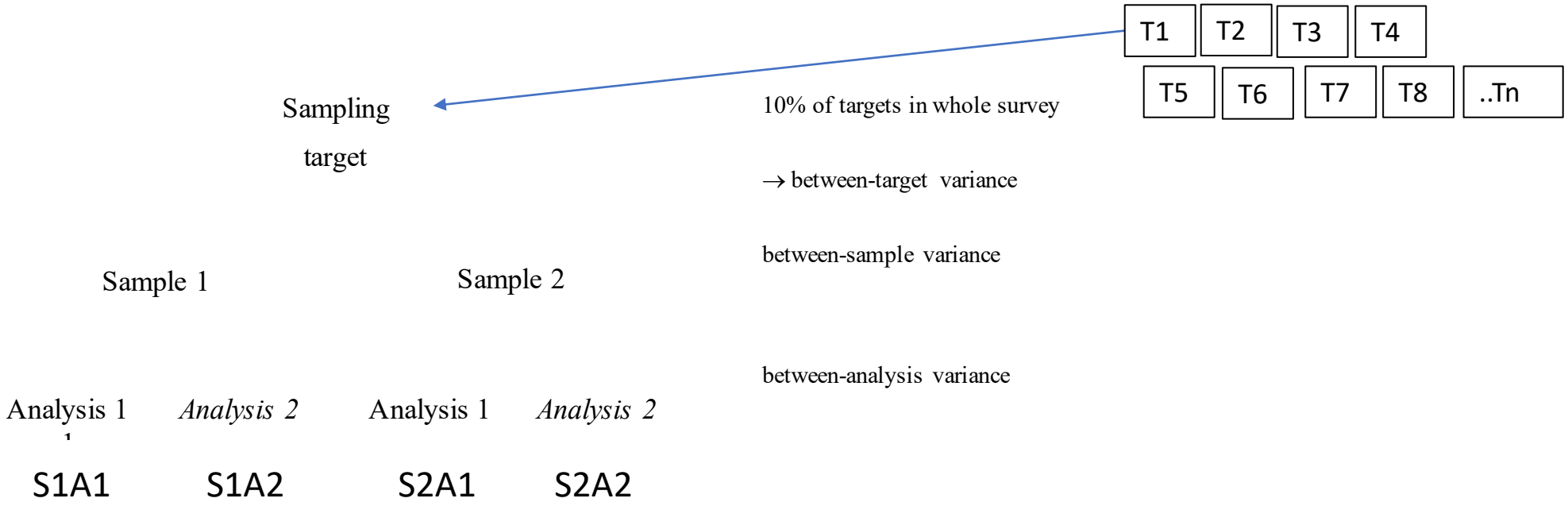
1 estimate analytical bias using CRM, 2 Analytical bias partially or completely included where multiple labs involved 3 Repeatability conditions

University of Sussex

Duplicate Method of UfS Estimation – General Principles


- Duplication is most cost-effective form of replication
  - Apply to both duplicate samples and duplicate chemical analyses
    - using two-stage nested experimental design (balanced or unbalanced)
  - But can have large confidence interval of resulting estimates of MU
  - Unless it is applied to at least 8 sampling targets (ideally more, e.g. 20)
- Realistic taking of duplicate samples is crucial
  - Not just the splitting of a single sample
- Take duplicate samples independently by fresh interpretation of the sampling procedure
  - How far away (in space or time) might duplicate sample be taken? Reflects..
    - ambiguity in sampling procedure
    - spatial uncertainty in the surveying device in use
    - Example below for *ex situ* measurement of Nitrate in lettuce (UfS-A1, VaMPIS-B1)

Estimation of MU (including UfS)
Using Duplicate Method – **Full Balanced Design**

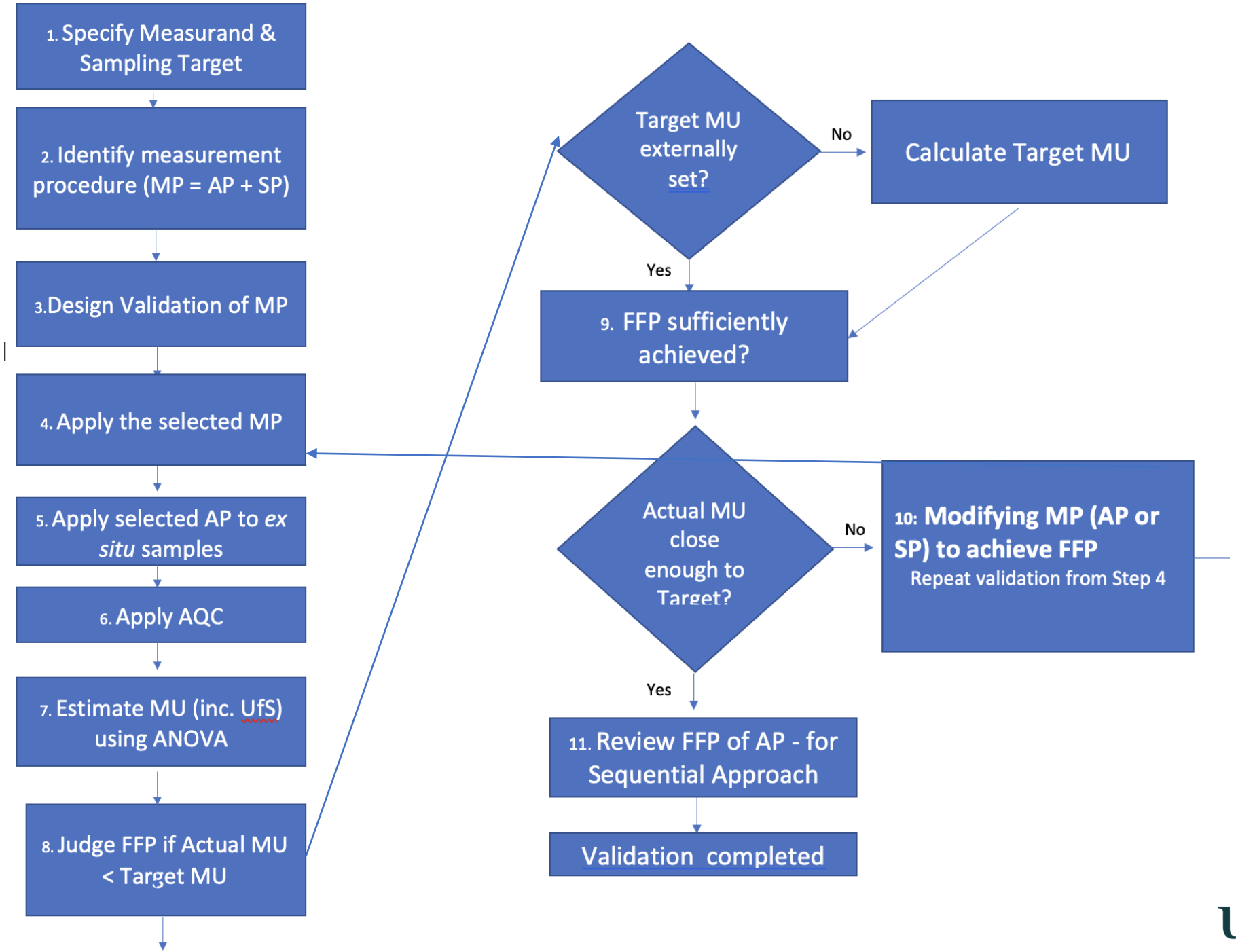


- Usually uses this full balanced experimental design *(unbalanced - no S2A2 - reduces cost)*
- 8 typical Sampling Targets chosen
- Only requires one 'sampler' (or measurement scientist)
  - Can be improved using multiple 'samplers' - using SPT results *(see later slide, and UfS Guide)*
- Explain Duplicate Method for Case Studies – followed by ANOVA
  - Applicable to Validate both *ex situ* and *in situ* measurement methods – *flow chart*

University of Sussex

Validation of Measurement Procedures Including Sampling (VaMPIS)

- Flow Chart



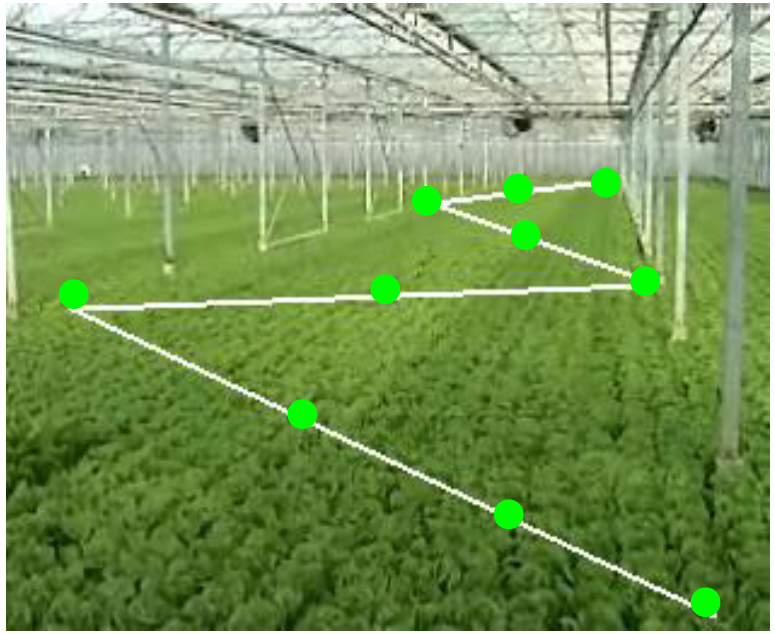
University of Sussex

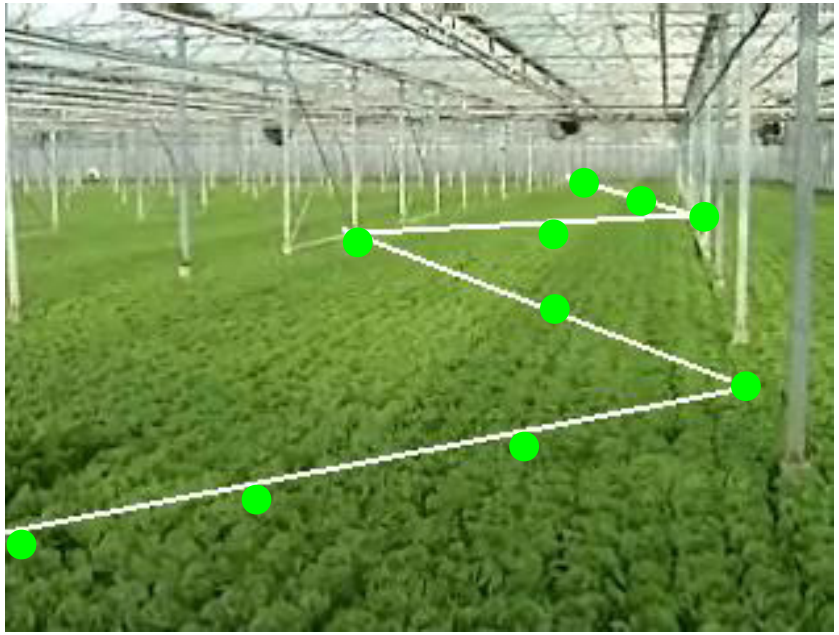
Validation using MU/UfS - Nitrate Concentration in Lettuce

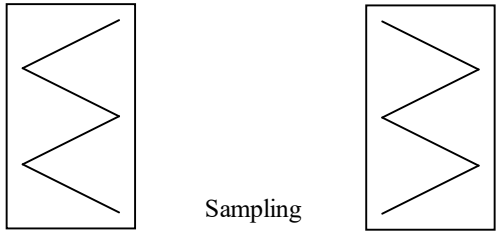
- EU threshold 4500 mg kg^{-1} for nitrate concentration of Sampling Target[1]
  - i.e. ~ 12,000 – 20,000 heads in each bay/batch/target
- Current EU sampling procedure[2] specifies taking 10 heads (increments)
  - to make a single **composite sample** from each Sampling Target
- Analytical procedure/method (HPLC[3]) already validated using Collaborative Trial[4]
  - $U_{analysis}$ around 6% at that validation ($RSD_{Reproducibility}$ = ~ 3%)
- Need to validate the whole measurement procedure
  - including sampling & sample preparation
- 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

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*

University of Sussex

Estimating UfS (and MU) for Nitrate in Lettuce

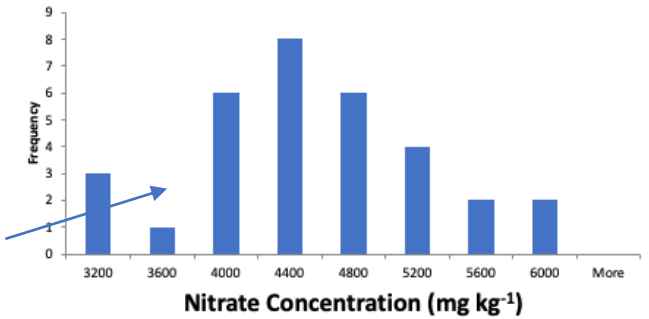
Sampling target	Sample 1		Sample 2	
	Analysis 1	Analysis 2	Analysis 1	*Analysis 2*
Sample target	**S1A1**	**S1A2**	**S2A1**	**S2A2**
	Nitrate concentration (mg kg^{-1})			
A	3898	4139	4466	4693
B	3910	3993	4201	4126
C	5708	5903	4061	3782
D	5028	4754	5450	5416
E	4640	4401	4248	4191
F	5182	5023	4662	4839
G	3028	3224	3023	2901
H	3966	4283	4131	3788

Analytical duplicates generally show ~10% precision

Sampling duplicates generally differ by <20%

Target C has greater difference (~50%) – outlying values?

Distribution approx. Normal + <10% outliers



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

University of Sussex

RANOVA3 output for Nitrate in Lettuce (Example A1)

Classical ANOVA

Mean	*4345.6*	No. Targets		*8*
Total Sdev	*774.53*			
	Btn Target	Sampling	Analysis	Measure
Standard deviation	*556.28*	*518.16*	*148.18*	*538.93*
% of total variance	*51.58*	*44.76*	*3.66*	*48.42*
Expanded relative uncertainty (95%)		*23.85*	*6.82*	*24.80*
Uncertainty Factor (95%)		*1.2432*	*1.0738*	*1.2574*

Robust ANOVA

Mean	*4408.3*			
Total Sdev	*670.58*			
	Btn Target	Sampling	Analysis	Measure
Standard deviation	*565.4*	*319.05*	*167.94*	*360.55*
% of total variance	*71.09*	*22.64*	*6.27*	*28.91*
Expanded relative uncertainty (95%)		*14.47*	*7.62*	*16.36*

- Software RANOVA3* (in Excel) performs:-
- Classical ANOVA gives poor estimate of *U'* = 24.8 %
  - Due to presence of outlying values
- also gives estimate of ^{F}U as 1.26 (~ 26% - similar)
  - after $\log_e$-transformation within RANOVA3
  - but distribution NOT log-normal in this case

Robust U' as 16.4%. (u = 360 mg kg^{-1})
Most reliable estimate of MU -
As approximately Normal distribution-
- but with < 10% outliers

U' anal = 7.6% – as repeatability
Very similar to MU = 6 % reported at separate validation of analytical procedure

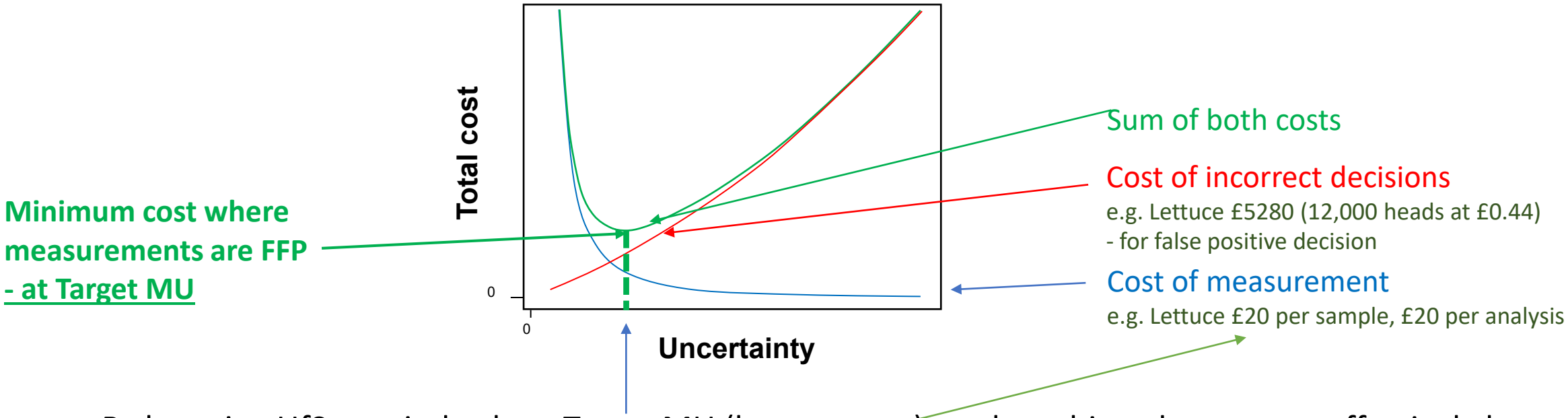
- Analytical recovery not statistically different from 100%
- Therefore no analytical bias detected

* http://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/

University of Sussex

Validation of Measurement Procedure - Judge FFP against Target MU

- **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 **Optimal MU*** that minimises the overall cost (including the consequences of incorrect decisions)



- By knowing UfS, can judge how Target MU (however set) can be achieved most cost-effectively 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)

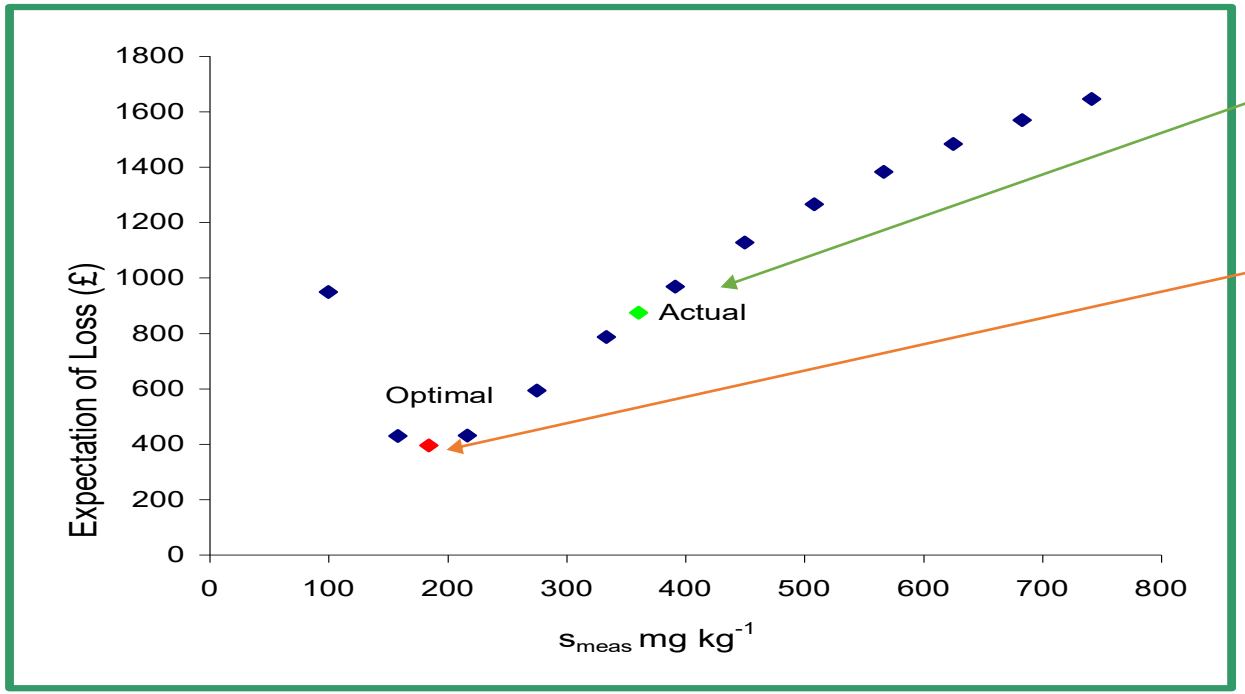
Optimal MU* method explained more in VaMPIS-SG

University of Sussex

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 – to make FFP

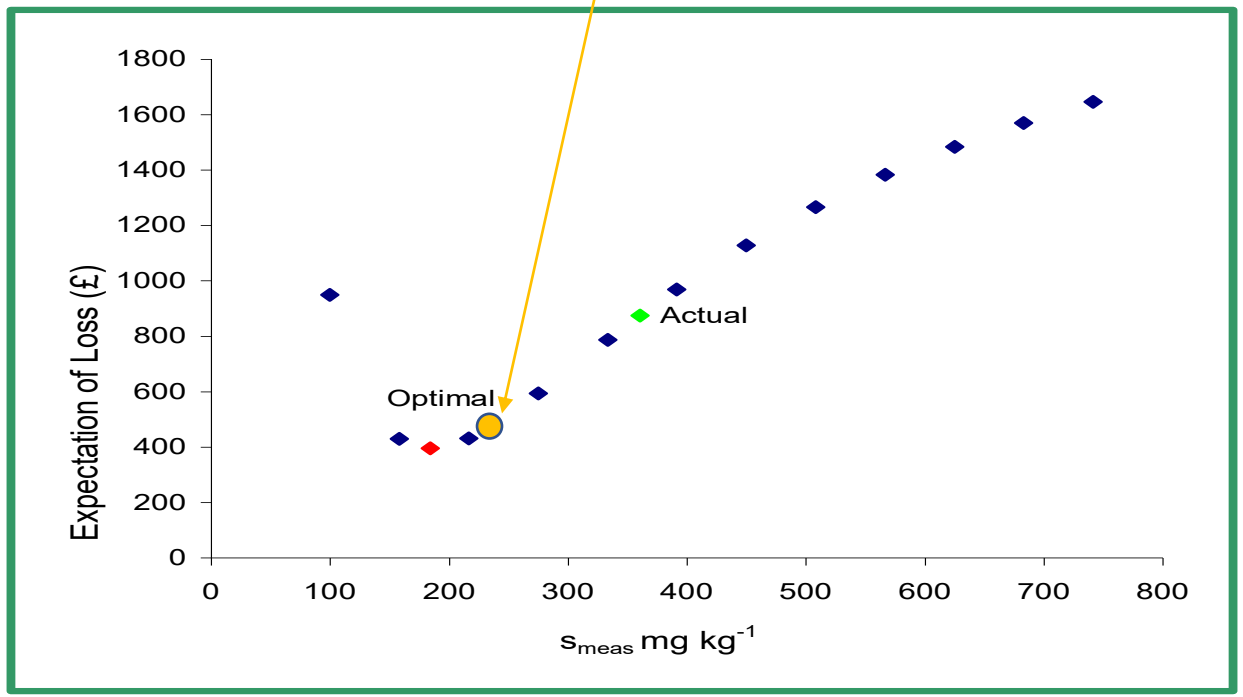
* In upcoming SG-VaMPIS – not in UfS Guide

University of Sussex

Reducing the Uncertainty – to achieve FFP

- **Increasing number of increments from 10 to 40 heads**
- Reduced s_{samp} from 319 to 177 mg kg^{-1} - by a factor of x 1.8 (similar to model prediction of x2)
- Reduced 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)
- A**chieves Fitness-for-Purpose (FFP)** = MU that minimises to overall financial loss

Cost ↑



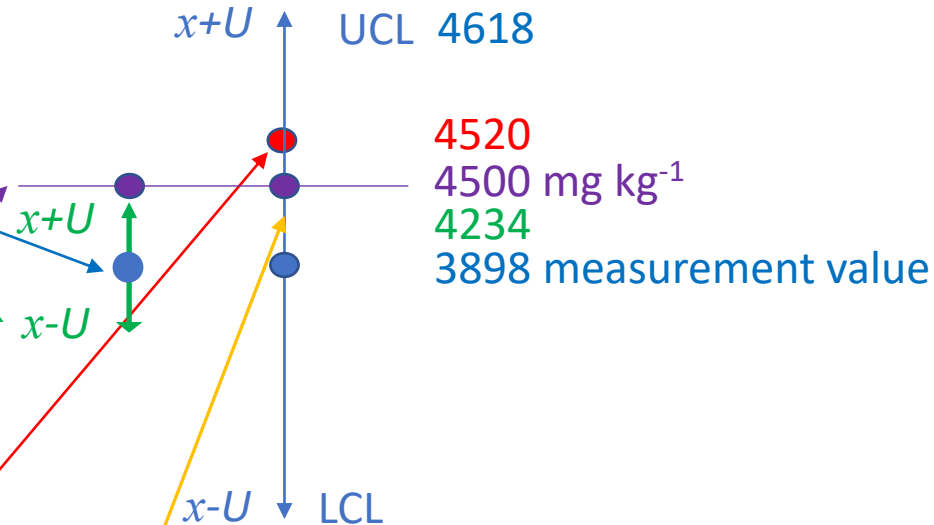
Uncertainty→

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

University of Sussex

Compliance decision - More reliable using UfS

Nitrate in Lettuce (Target A)

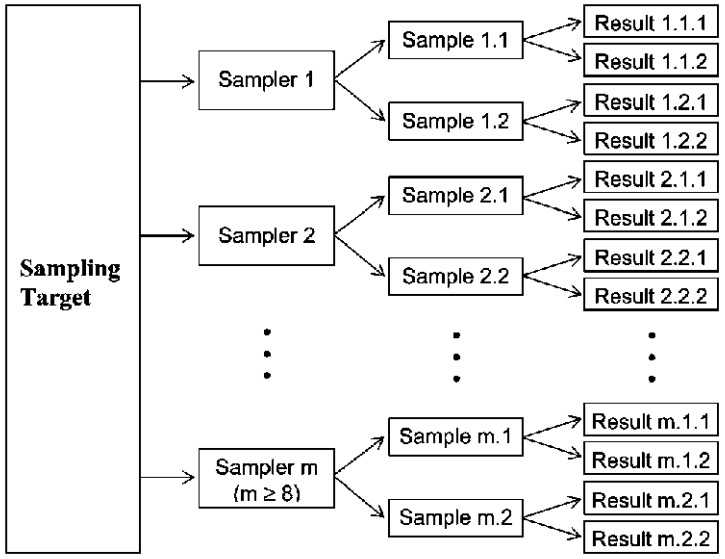


- Measurement value *(x)* 3898 for Target A (S1A1)
- and its MU ($U_{analytical}$) u_{anal} = 168, U = 336 mg kg^{-1}
  - = range within which true value lies
- Threshold value for compliance 4500 mg kg^{-1}
- Appears that true value cannot exceed the Threshold
- = measurement indicates Target A is **Compliant**
- Use MU that includes UfS u_{meas} = 360, U = 720 mg kg^{-1}
- True value possible for Target A 4520 mg kg^{-1}
- Over Threshold, therefore measurement indicates Target A is **Non-Compliant**
- Non-Compliance of Target A (False positive) appears impossible with MU based only upon $U_{analytical}$
- **Only by including UfS within MU can non-compliant batch (Target A) be rejected reliably**
- Reduced U with 40 heads (488 mg kg^{-1}) gives UCL 4386 which is < Theshold of 4500 mg kg^{-1}
  - Target A might have been shown to be **Compliant** (if that FFP measurement procedure had been used)

University of Sussex

Estimate of Uncertainty using SPT - including Between-Sampler Bias

- Example using Sampling PT for moisture in butter*



* 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.

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

≈ Duplicate Method (single sampler) gives $\boldsymbol{U' = 0.39\ \%}$

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 $\boldsymbol{U' = 0.69\%}$

- **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 using one procedure (CTS) better for VaMPIS
- More expensive than Duplicate Method, but sometimes justified

University of Sussex

Conclusions

- Eurachem UfS Guide explains importance of UfS (& MU), and how to estimate it
- Including sampling within the measurement process:
  - Is essential for making reliable estimates of MU (including UfS)
    - e.g. for Compliance Decisions: e.g. are concentration levels above from regulatory limits?
    - Conforms to ISO/IEC 17025:2017
  - Being able to judge FFP, and hence validate the whole measurement process
  - Hence rigorous Validation of the whole Measurement Process (Including Sampling)
    - Upcoming Supplementary Guidance on VaMPIS
- UfS (and hence MU) can be estimated with Duplicate Method *(most practical)*
  - Applicable to any sampling medium: soil, sediment, herbage, waters, gases etc.
  - Also applicable to *in situ* measurements (such as PXRF – Example B2 in SG-VaMPIS)
  - Sampling PT (or CT) results can be used to also include between-sampler bias within MU
- *Questions?*

University of Sussex

Compliance decision - More reliable using UfS

Nitrate in Lettuce (ALL 8 Targets)

Ignores MU – 4 Batches rejected

Sample Target	Nitrate Conc (x) in S1A1 mg/kg	Deterministic Classification x <4500
A	3898	Y
B	3910	Y
C	5708	N
D	5028	N
E	4640	N
F	5182	N
G	3028	Y
H	3966	Y
Batches Accepted		**4**

Classification of 8 batches of lettuce based upon deterministic compliance decision (i.e. ignoring MU),

University of Sussex

Compliance decision - More reliable using UfS

Nitrate in Lettuce (ALL 8 Targets)

Ignores MU – 4 Batches rejected

Allows for MU, (using 10-fold composites) – 7 Batches rejected

Sample Target	Nitrate Conc (x) in S1A1 mg/kg	Deterministic Classification x <4500	10-head U'_{10} = 16.4%	$x + U_{10}$	Probabilistic Classification $x+U_{10}$ < 4500
A	3898	Y	639.3	4537	**N**
B	3910	Y	641.2	4551	**N**
C	5708	N	936.1	6644	N
D	5028	N	824.6	5853	N
E	4640	N	761	5401	N
F	5182	N	849.8	6032	N
G	3028	Y	496.6	3525	Y
H	3966	Y	650.4	4616	**N**
Batches Accepted		**4**			**1**

Classification of 8 batches of lettuce based upon <u>probabilistic</u> decision using MU <u>for 10-head composite</u> U'_{10} =16.4%
- caused further three batches (A, B & H) to be rejected as potential false positives

University of Sussex

Compliance decision - More reliable using UfS

Nitrate in Lettuce (ALL 8 Targets)

Allows for MU, (using 40-fold composites) – 4 Batches rejected

Sample Target	Nitrate Conc (x) in S1A1 mg/kg	Deterministic Classification x <4500	10-head U'_{10} = 16.4%	$x + U_{10}$	Probabilistic Classification $x+U_{10}$ < 4500	40-head U'_{40}= 11.1%	$x + U_{40}$	Probabilistic Classification $x + U_{40}$ < 4500
A	3898	Y	639.3	4537	**N**	432.7	4331	**Y**
B	3910	Y	641.2	4551	**N**	434.0	4344	**Y**
C	5708	N	936.1	6644	N	633.6	6342	N
D	5028	N	824.6	5853	N	558.1	5586	N
E	4640	N	761	5401	N	515.0	5155	N
F	5182	N	849.8	6032	N	575.2	5757	N
G	3028	Y	496.6	3525	Y	336.1	3364	Y
H	3966	Y	650.4	4616	**N**	440.2	4406	Y
Batches Accepted		**4**			**1**			**4**

Classification of 8 batches of lettuce based upon <u>probabilistic</u> decision using <u>MU for 40-head</u> (U'_{40}) composite samples
Validated (FFP) procedure gave lower MU of 11.1%
- three marginal batches (A, B & H) now accepted

University of Sussex