Uncertainty estimation for *in situ* measurement procedures, including the sampling component

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

ISS Workshop on UfS
9-10th March 2023, Rome
30 minutes + discussion

University of Sussex

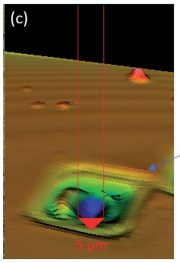
Overview

- What is an *in situ* measurement method?
- Why is Measurement Uncertainty (MU) important?
- Advantages/Disadvantages of *in situ* measurement methods[1]
- Estimation of MU for *in situ* measurements + Case Study
- Judging Fitness-for-Purpose (FFP) of *in situ* measurements
  - hence Validation of Measurement Procedures (Including Sampling)
  - For two different purposes
    1. Compliance with regulation
    2. Geochemical Mapping
- Conclusions

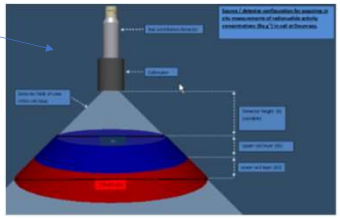
1. Ramsey M.H. (2020) Challenges for the estimation of uncertainty of measurements made *in situ*. Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement. 26(4), 183-192. 2020. https://doi.org/10.1007/s00769-020-01446-4

University of Sussex

What are *in situ* measurements?



- Taken *in situ* without disturbing or removing the test material
  - Different from 'on site' measurements, where physical sample is removed, measured locally
  - Sampling indivisible part of measurement process, e.g.
- Soils - handheld portable (P)XRF for some metals
- Minerals – SIMS at micron scale (e.g. Oxygen isotopes)
- Gases - many sensors – how most measurements are made!
  - e.g. Photolonisation detectors for VOCs, (e.g. benzene), IR for CH_4, CO_2
- Liquids - e.g. pH, UV-Vis for NO_3, TOC, H_2S
- Clinical - Transcutaneous Bilirubinometer (TcB) – jaundice ?
- Radioactivity – passive γ-ray detection of ^{137}Cs in very large sample mass (200 – 1000kg)



US University of Sussex

Advantages of *in situ* measurement methods

1. Measurement result virtually instantaneous (versus days/weeks for lab)
   - saving lives in clinical sector
   - saving money in commercial & environmental sectors
   - enables sensing networks to monitor analyte variability across time or space
2. Substantially lower cost, even if MU higher, enables:
   - taking of many more 'samples' - 1000 less exact measurements provide more information than 100 more exact = giving better coverage of target in space and/or time
   - Even 100% coverage – e.g. groundhog (γ–ray spec)
3. No sample preparation
   - avoid loss of analyte by removing, storing, & preparing sample
   - eliminating costs of taking, storing and disposing of samples
4. Ability to quantify heterogeneity of test material (U_{HET})
5. Can be made by less skilled personnel? – really a disadvantage –
   - *In situ* measurement scientists need to be more skilled
   - able to take most appropriate samples (of the 'undisturbed' kind) in real world
   - & take measurements of acceptable quality, both without local supervision

US University of Sussex

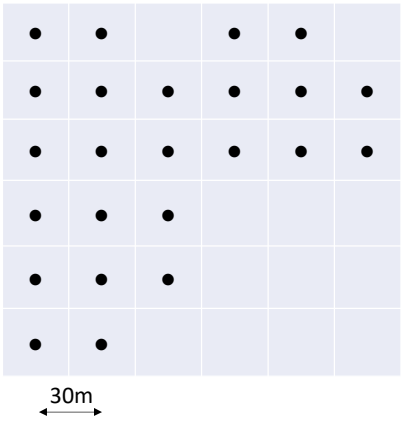
Disadvantages of *in situ* measurements

- *In situ m*easurements often have larger uncertainty (MU)
  - Due partially to heterogeneity of analyte concentration (not mixed)
  - Vertical heterogeneity in test portion + critical penetration depth (e.g. PXRF)
  - Detection limits often not as low as for *ex situ* measurements
- However, MU can be estimated by duplicate method (or SPT)
  - Can judge whether Measurements are fit-for-purpose (FFP) - if UfS quantified

US University of Sussex

Case Study: Estimation of UfS & MU for measurements made *in situ*

Site of a medieval Pb smelter at Wirksworth, Derbyshire, UK
Hand-held portable x-ray fluorescence spectrometer (PXRF) used to measure Pb concentration [Pb] in topsoil *in situ**



Grid of 24 sampling targets used to survey [Pb] across site

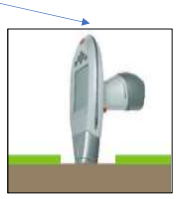
*Ramsey M.H. (2020) Measurement Uncertainty from Sampling: Implication for Testing, Diagnostics and Inspection. Presented to 17th IMEKO TC 10 and EUROLAB Virtual Conference *"Global Trends in Testing, Diagnostics & Inspection for 2030"* October 20-22, 2020. https://www.imeko.org/publications/tc10-2020/IMEKO-TC10-2020-042.pdf

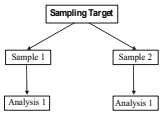
US University of Sussex

Case Study: Estimation of UfS & MU for measurements made *in situ* (Method)

- Duplicate Method used to estimate random components of MU of *in situ* measurements
  - as repeatability
- Equivalent of 'duplicate samples' are taken by placing the *in situ* measurement device twice, reflecting two independent interpretations of measurement protocol.
- In this study PXRF duplicates were 2m apart, in a randomly chosen direction, to reflect uncertainty in location







- These two sampling points are both equally likely interpretations of the protocol given that particular surveying technology
- **Simplified design** used for speed (no analytical duplicates)

University of Sussex

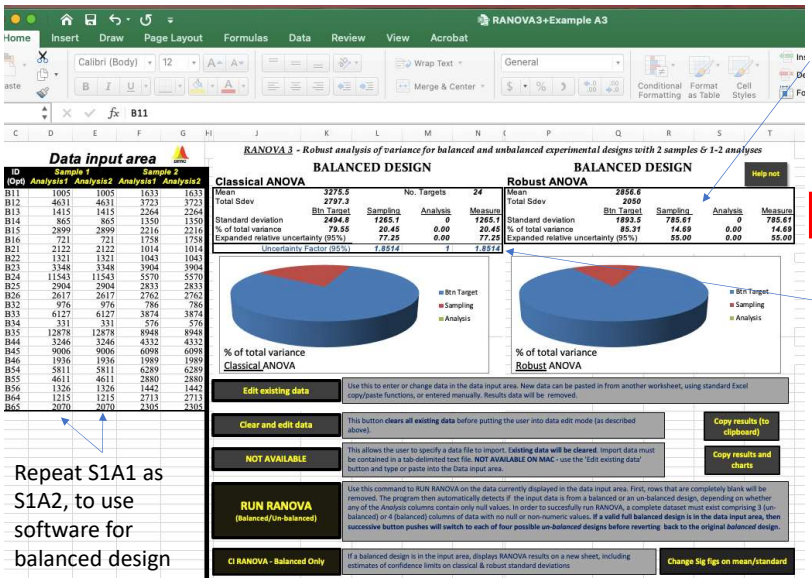
Duplicated PXRF measurements – for random component of UfS (Results)

Target Number	S1Pb mg/kg	S2Pb mg/kg
1	1005	1633
2	4631	3723
3	1415	2264
4	865	1350
5	2899	2216
6	721	1758
7	2122	1014
8	1321	1043
9	3348	3904
10	11543	5570
11	2904	2833
12	2617	2762
13	976	786
14	6127	3874
15	331	576
16	12878	8948
17	3246	4332
18	9006	6098
19	1936	1989
20	5811	6289
21	4611	2880
22	1326	1442
23	1215	2713
24	2070	2305

- Duplicated 'samples' show quite large variation (from small scale heterogeneity) RANOVA3 gave a robust estimate of MU (U'_{meas}) = 55%
  - Robust statistical methods much less susceptible to small proportion of outlying values (i.e. < 10%), which are evident in some duplicates (e.g. Targets 10 & 18).
- Need external estimate of $U'_{analysis}$ = 3%.
  - Made using additional *ex situ* PXRF measurements (made in lab on prepared versions of removed samples from same 24 targets), in fully balanced experimental design (i.e. with duplicated analyses)
  - Assumes instrumental performance of PXRF similar *in situ* to *ex situ*
- Gives $U'_{sampling,\ in\ situ} = 54.9\% = (\sqrt{55^2 - 3^2})$, using.... $s_{sampling} = \sqrt{s^2_{meas} - s^2_{analytical}}$
  - Sampling causes 99.7% of MU
- Any underestimate of $U'_{analysis}$ (as repeatability) has little effect on value of MU
- More generally, PXRF instrument reports uncertainty (U') of around 3%, but..
- **Actual MU is much higher at ~ 55 %** - when UfS included
- Better expressed as **Uncertainty Factor** ${}^{F}U = 1.85$
  - due log-normal distribution (explained later)

University of Sussex

UfS Estimation using RANOVA3



Robust ANOVA

Mean	2856.6			
Total Sdev	2050			
	Btn Target	Sampling	Analysis	Measure
Standard deviation	1893.5	785.61	0	785.61
% of total variance	85.31	14.69	0.00	14.69
Expanded relative uncertainty (95%)		55.00	0.00	55.00

Classical ANOVA

Mean	3275.5	No. Targets		24
Total Sdev	2797.3			
	Btn Target	Sampling	Analysis	Measure
Standard deviation	2494.8	1265.1	0	1265.1
% of total variance	79.55	20.45	0.00	20.45
Expanded relative uncertainty (95%)		77.25	0.00	[illegible]
Uncertainty Factor (95%)		1.8514	1	1.8514

Expressing MU as Uncertainty Factor (^{F}U)

- MU better expressed as an **Uncertainty Factor (^{F}U)** as MU is large ($U'_{meas} > 40\%$)
  (UfS Guide ([2]. p21,48)

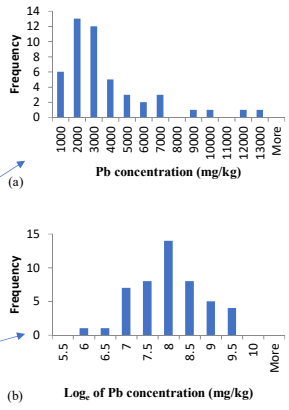
$$^{F}U = \exp(2s_{G,meas})$$

- Where, $\mathbf{s_{G,meas}}$ **is the the equivalent of** $\mathbf{s_{meas}}$
  - calculated within RANOVA3 as natural logarithms of measurement values
  - $s_{G,meas}$ is 0.308 for Case Study

$$^{F}U = \exp(2 * 0.308) = 1.85$$

- LCL and UCL then calculated as $x/^{F}U$ and $x*^{F}U$, respectively

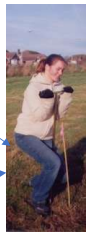
- Large MU often due to frequency distribution being log-normal
  - i.e. positively skewed - rather than normal (i.e. Gaussian)
- For Case Study distribution of PXRF measurements is log-normal
  - Made ~normal by taking logarithms of all measurement values



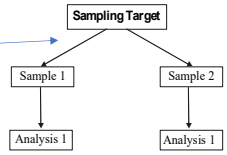
Estimation of UfS (and MU) for measurements made *in situ* (Include Analytical Bias)

- Systematic component of MU of *in situ* measurements from analytical bias (not sampling bias)
  - estimated by measurements made on matrix-matched CRMs (e.g. NIST 2710), but….
  - unlike most test materials in real world

CRM	Dried	Ground	Homogeneized	Compacted
Test material	Moist	Unground	Heterogeneous	Un-consolidated



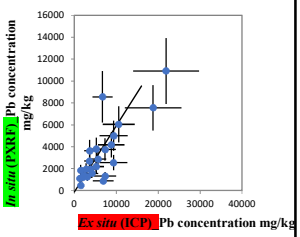
- To overcome this mis-match, compare *in situ* against *ex situ* measurements
  - made for same analyte on same sampling targets
- Need to also match value of the 'measurand', which is effectively the true value that is being estimated
  - i.e. total Pb concentration in dry soil
- In Case Study, also removed *ex situ* samples taken at same locations as PXRF measurement made
  - Also with simplified balanced design (i.e. without Analysis 2)
  - for all 24 sampling targets, but 8 targets would be OK for routine investigation
  - dry, disagregate, sieve (<2mm), grind and acid digestion in a remote laboratory (i.e. *ex situ*)
  - then analysed by ICP-AES (traceable to CRMs)



University of Sussex

'Bias' of *in situ* PXRF against *ex situ* ICP-AES measurements

Target ID	Ex situ ICP-AES	SD ICP (35.8%)	In situ P-XRF	SD PXRF (27.5%)
A6	**7340**	2628	**1319**	363
A5	**8815**	3156	**4177**	1149
A4	**1522**	545	**1840**	506
A3	**1290**	462	**1108**	305
A2	**9340**	3344	**2547**	700
A1	**3080**	1103	**1240**	341
B6	**4180**	1496	**1568**	431
B5	**1926**	690	**1183**	325
B4	**3670**	1314	**3626**	997
B3	**6718**	2405	**8555**	2353
B2	**5630**	2016	**2869**	789
B1	**3630**	1300	**2690**	740
C5	**6880**	2463	**881**	242
C4	**9370**	3354	**5002**	1376
C3	**1522**	545	**454**	125
C2	**21877**	7832	**10919**	3003
D3	**5230**	1872	**3788**	1042
D2	**18784**	6725	**7556**	2078
D1	**2800**	1002	**1963**	540
E3	**10584**	3789	**6050**	1664
E2	**7316**	2619	**3745**	1030
E1	**2235**	800	**1384**	381
F3	**3860**	1382	**1964**	540
F2	**5210**	1865	**2188**	602



- Systematic component of MU estimated as bias…
  - by comparing average value of both *in situ* PXRF measurements
  - against *ex situ* ICP-AES measurement
- Relationship modelled as a function of concentration -
  - Using FREML (Functional Relationship Estimation by Maximum likelihood [1, 2]
- In FREML uncertainty of both variables properly taken into account.
  - Also possible to use ordinary least-squares regression, but this can only allow for uncertainty in y-axis (e.g. PXRF) and ignores uncertainty for x-axis (e.g. ICP-AES)

Model → $[Pb]_{in\ situ} = b(1) \times [Pb]_{ex\ situ} + b(0)$

- Slope coefficient of linear model (b(1)) → rotational component of bias
- Intercept coefficient b(0) → translational component

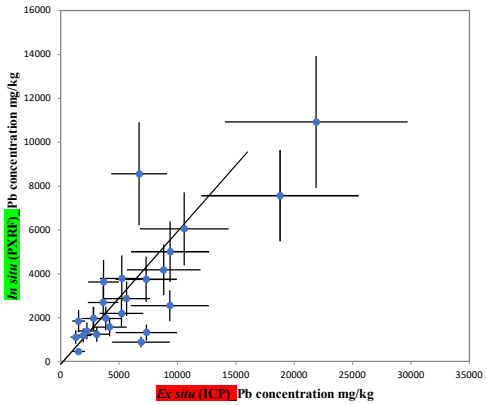
University of Sussex

[1] AMC Software, downloaded gratis from:https://www.rsc.org/Membership/Networking/InterestGroups/Analytical/AMC/Software/
[2] Analytical Methods Committee (2002) Fitting a linear functional relationship to data with error on both variables, Technical Brief No.10, Royal Society of Chemistry, London

'Bias' of *in situ* PXRF against *ex situ* ICP-AES measurements (2)

Equation describing relationship,
showing both coefficients and their standard errors (in parentheses):

$[Pb]_{in\ situ} = 0.60\ (\pm 0.09) \times [Pb]_{ex\ situ} - 120\ (\pm 288)$



- Estimated rotational bias of *in situ* PXRF measurements
  - compared against the *ex situ* ICP measurements
- calculated from slope coefficient, is **- 40%** (± 9%)
  - i.e. 100 x (1 - 0.60).
- No translational bias detected, as..
  - intercept coefficient = -120 mk/kg (± 288)
  - not statistically different from zero
- Possible causes of measurement bias identified as:
  - soil moisture
  - Material/particles > 2mm diameter (removed for *ex situ* sample)
  - surface roughness in the PXRF 'undisturbed sample'
  - Depth difference between undisturbed sample for *in situ* PXRF (~1mm)
  - and removed *ex situ* field sample for ICP-AES (150 mm) [1]

[1] Argyraki, A., Ramsey, M.H. and Potts P.J. (1997) Evaluation of portable XRF for in-situ measurements of lead on contaminated land. Analyst 122, 743-749

University of Sussex

Treatment of Systematic component of MU for *in situ* measurements - *Option 1*

Option 1 - 'correct' *in situ* measurements ($[Pb]_{PXRF,\ corr}$) to agree with *ex situ* values by applying a rearrangement of the bias model

- omitting the non-significant intercept for this Case Study

$$[Pb]_{PXRF,corr} = \frac{[Pb]_{PXRF,raw} - b(0)}{b(1)} = \frac{[Pb]_{PXRF,raw}}{0.60}$$

Uncertainty of correction (se' bias) = 0.09, as < 0.2, can be:-
Combined (as relative percentage 9%), into MU value *U'* =55% (*u'* = 27.5%)
([1], all expressed as relative standard uncertainty)

$$u'_{corr} = \sqrt{(u'^2 + (se^2_{b(1)})}$$

$$u'_{corr} = \sqrt{(27.5^2 + 9^2)} = 28.9\%$$

$U'_{corr} = 57.9\%$ - slightly up from 55%

[1] by approach described in UfS Guide, Section A2/6.4),

University of Sussex

Treatment of Systematic component of MU
for *in situ* measurements - *Option 1b*

Option 1b - 'correct' when MU is expressed as Uncertainty Factor:

${}^{F}U$,expressed as $s_{G,meas}$,

- using an approximation [1]

$$s_{G,meas,corr} = \sqrt{s_{G,meas}^2 + (s'_{bias})^2}$$

Expanded uncertainty factor ${}^{F}U$ **of 1.85** increases slightly to **1.88**

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

[1] Ramsey M H and Ellison S L R (2017) Combined uncertainty factor for sampling and analysis. Accreditation and Quality Assurance, 22(4), 187-189 DOI 10.1007/s00769-017-1271-y

University of Sussex

Treatment of Systematic component of MU
for *in situ* measurements - *Option 2*

Option 2 is not to correct, but to add the entire bias, and its uncertainty, to MU [2]

- again all terms expressed as relative percentage.

$$u'_{corr} = \sqrt{(u'^2 + ((b(1) - 1)^2) + (se^2_{b(1)})}$$

-
- For this case study:

$$u'_{corr} = \sqrt{(27.5^2 + 40^2 + 9^2} = 49\%$$

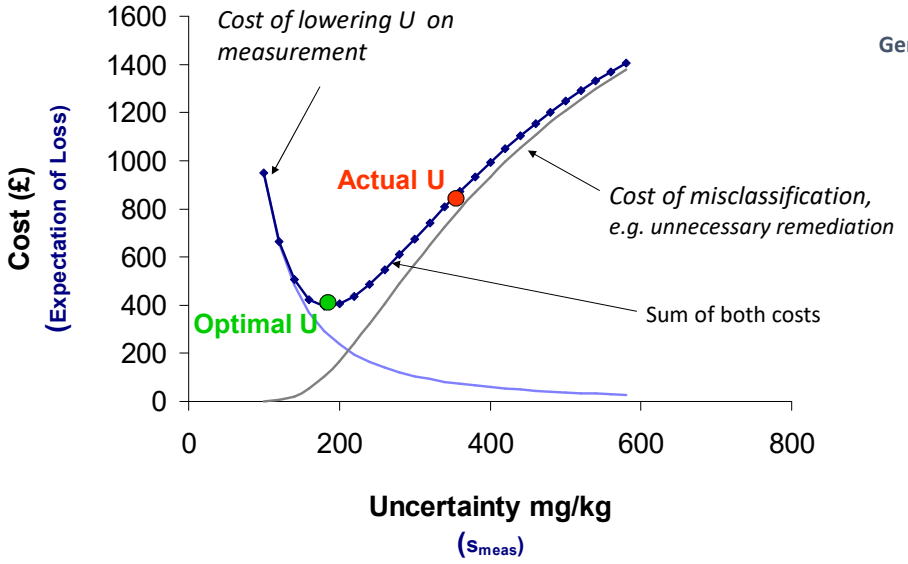
$$U'_{corr} = 98\%$$

[1] Ramsey M H and Ellison S L R (2017) Combined uncertainty factor for sampling and analysis. Accreditation and Quality Assurance, 22(4), 187-189 DOI 10.1007/s00769-017-1271-y
[2] UfS Guide, Section A2/6.4 page 56

University of Sussex

Judging FFP of *in situ* measurements – using Optimal U

hence validation of measurement methods (including sampling)



University of Sussex

Judging FFP using Optimized Uncertainty (OU) equation

$$E(L) = C\,[1 - \Phi\,(\varepsilon_1 / s_{meas})\,] + D/s^2_{meas}$$

E (L) – expectation of financial loss (= Total Cost)

s_{meas} – measurement uncertainty

Φ – standard normal cumulative distribution function

ε1 – error limit = $|\,T - c\,|$

(T = threshold value, c = contaminant concentration at which to optimise)

D – combined optimal cost for sampling and analysis

C – consequence costs (e.g. potential losses resulting from misclassification)

Thompson M, Fearn T (1996) What exactly is fitness for purpose in analytical measurement? Analyst, 121, 275–278.
Ramsey M.H., Taylor P. D. and Lee J.C. (2002) Optimized contaminated land investigation at minimum overall cost to achieve fitness-for-purpose, Journal of Environmental Monitoring, 4, 5, 809 – 814.

University of Sussex

Input data for Optimised MU

for *in situ* measurements

For Case Study

Item	Value	Units	*U'*
Sampling cost (each)	29	£	
Analytical cost (each)	12	£	
Consequenct cost per location, for false positive classification (i.e. unnecessary remediation)	10000	£	
$u_{sampling}$	784	mg kg^{-1}	54.9%
$u_{analytical}$	43	mg kg^{-1}	3%
$u_{measurement}$	786	mg kg^{-1}	55%
Threshold value of concentration	2000	mg kg^{-1}	
Concentration at which to optimise	2020	mg kg^{-1}	

Also use individual cost (& MU) of Sampling and Analysis
- to calculate which if more cost effective to reduce overall MU

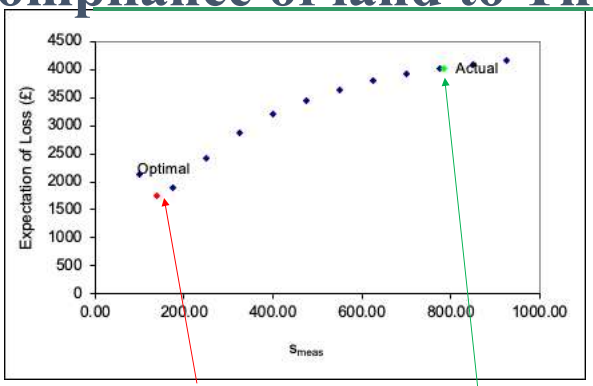
For 30m grid spacing

- From ANOVA + external u_{anal}

- UK limit at time of suvey

University of Sussex

FFP of *in situ* measurements? – Purpose 1: Compliance of land to Threshold (2000 mg kg^{-1})



Optimal uncertainty

138 mg kg^{-1}

£2286

Actual uncertainty

786 mg kg^{-1}

£4,026

In situ procedure currently **not Fitness-for-Purpose** (FFP)
- for classifying Pb against threshold (T) of 2000 mg kg^{-1}
- Because Actual U is >> than Optimal U

- To achieve FFP, need to:-
- Reduce MU x 5.7 (= 786/138) to achieve FFP
  - MU is 99.7% UfS (from ANOVA), so
- Reduce UfS x 5.8 to achieve FFP
  - Would require 34-fold composite measurement at each location (5.8^2) – impractical

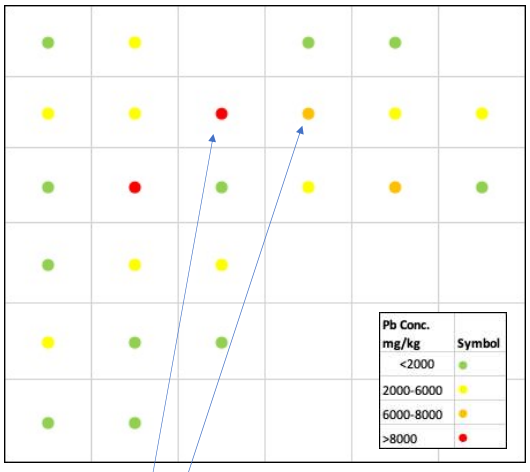
- To approach FFP:-
- Low cost of *in situ* allows closer grid spacing (e.g. 10m)
- Hence 9x lower remediation cost
  - gives 14x lower consequence costs (£278) at same MU [1]

[1] Taylor P D, Ramsey M H and Potts P.J. (2004) Balancing measurement uncertainty against financial benefits: a comparison of *in situ* and *ex situ* analysis of contaminated land. Environmental Science and Technology 38, 6824-6831.

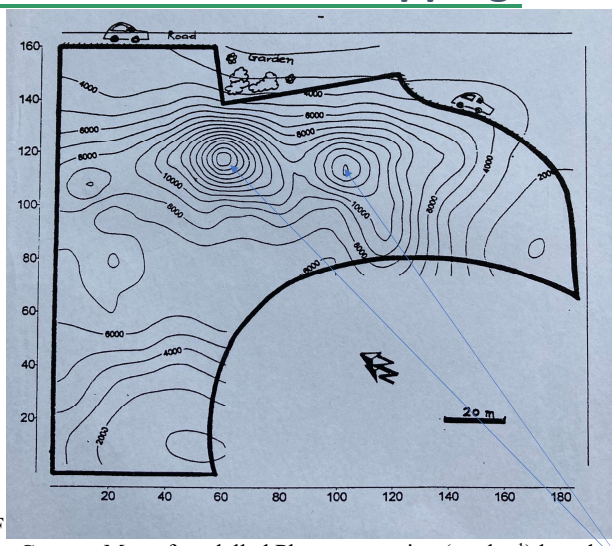
University of Sussex

FFP of *in situ* measurements? – Purpose 2:

In situ measurements for Geochemical Mapping - *FFP?*



Map of 24 Pb concentrations (mg kg^{-1}) measured by PRXF showing two 'hot spots' of high [Pb] [1]



Contour Map of modelled Pb concentration (mg kg^{-1}) based on 159 other measurements from several surveys, showing two clear peaks due to Pb smelters [1]

[1] Argyraki A (1997) Estimation of measurement uncertainty in the sampling of contaminated land. PhD Thesis. Imperial College. University of London

Alternative Purpose of *in situ* measurements?

- **Geochemical mapping** has a different FFP criterion
- Second FFP criterion that MU should not contribute > 20% of total variance [1]
- Target u_{meas} = 917 mg kg^{-1} (= 20% of total variance = SQRT(0.2*2050^2))
- Actual (robust estimate) u_{meas} = 786 mg kg^{-1} (ANOVA2 output)
- 786 < 917 mg kg^{-1} indicates that measurement results (& therefore measurement procedure) - are **fit for that purpose**

- Same conclusion FFP when Expressed In terms of relative uncertainty
- Actual robust MU of 55% < Target MU of 64%

- But 'Target' MU is more a preferred **maximum value** than a rigorous target
- i.e. Actual MU < Target MU is not a deficiency
- Even lower Actual MU beneficial as it would further improve reliability of geochemical map

- So *in situ* PXRF does give broadly reliable geochemical map of Pb for this site
- Can approximately locate location of Pb smelters

[1] UfS Guide, Section 16.2

University of Sussex

Conclusions

- *In situ* measurement method are useful, but have MU including UfS
- MU (inc UfS) can be estimated using the Duplicate Method
  - but systematic effects requires the use of an *ex situ* method for validation*
- Judging Fitness-for-Purpose (FFP) of *in situ* measurements
  - enabled Validation of measurement methods (including sampling)
  - Against whatever purpose is specified
  - Using a particular FFP criterion (e.g. either Optimal MU, or % of total variance)
- Case study *in situ* PXRF (for Pb at one particular site) proved to be:
  - FFP for Geochemical Mapping
  - But not FFP for Compliance with regulation (at relevant threshold)

*Ramsey M.H. (2020) Challenges for the estimation of uncertainty of measurements made *in situ*. Accreditation and Quality Assurance: Journal for Quality, Comparability and Reliability in Chemical Measurement. 26(4), 183-192. 2020. https://doi.org/10.1007/s00769-020-01446-4