Uncertainties related to analyte level – An R-package to use replicate measurements

S. Pospiech, K.G. van den Boogaart, A.D. Renno, R. Möckel, W. Fahlbusch

HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

HELMHOLTZ-INSTITUT FREIBERG FÜR RESSOURCENTECHNOLOGIE

Introduction

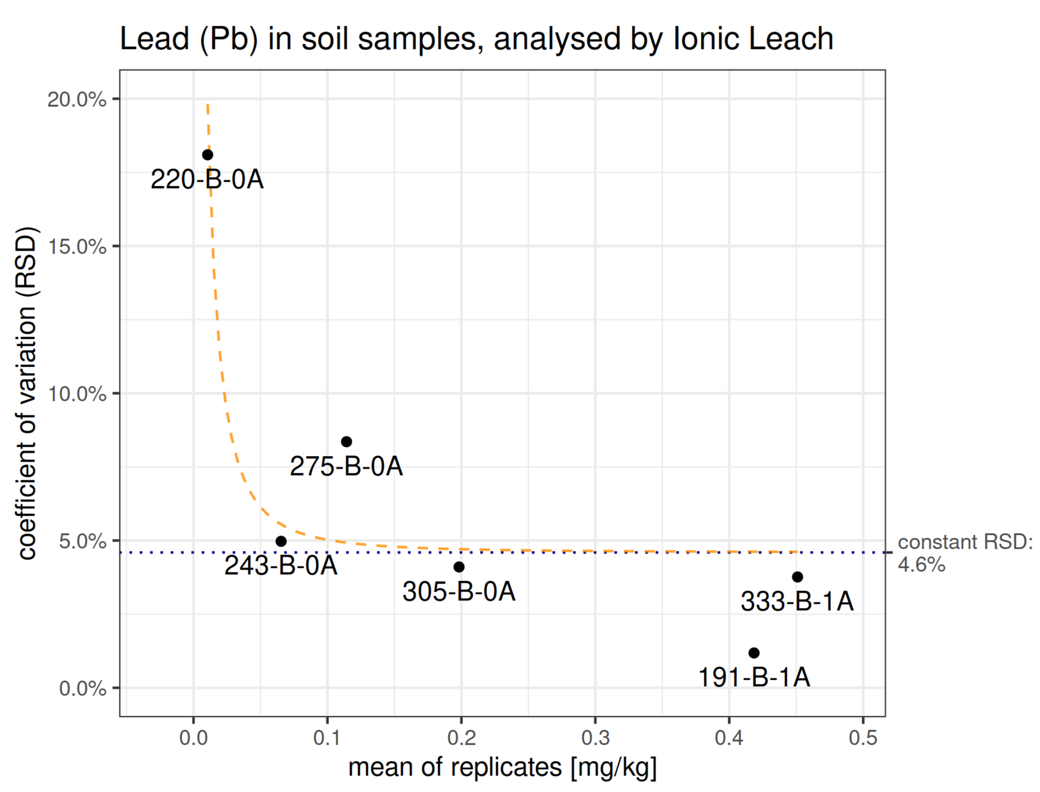
Measurement uncertainty (MU) often scales proportional to the analyte level (Ellison 2012). In this case, it can be expressed as a constant by using the relative standard deviation (RSD). This is called the multiplicative component of a MU model.

However, this underestimates the MU at very low analyte levels due to an additive component of MU. Estimating MU for a large range of analyte levels requires to use an uncertainty model with two components, a multiplicative *and* an additive.

Example case

A set of soil samples had been sent to a lab. Six randomly selected samples had been measured twice (7% of all samples), producing replicate measurements. Here the data for Pb (in mg/kg):

SampleID	Mean	SD	RSD [%]
220-B-0A	0.01055	0.0019	18.0
243-B-0A	0.06540	0.0033	5.0
275-B-0A	0.11425	0.0095	8.4
305-B-0A	0.19825	0.0081	4.1
191-B-1A	0.41850	0.0049	1.2
333-B-1A	0.45100	0.0170	3.8



Errors are additive if they are caused by random occurring effects of the measurement procedure.. Typical examples are the background 'noise' of a measurement device, or random occurring contamination. They must not be confused with errors introduced by systematic effects.

Multiplicative errors occur if errors in one step of the measurement scale the measurand value in the next step, e.g. if errors are introduced along the sample preparation and analysis procedure. In this case, the dispersion of value increases quadratically with the measurand value.

Theory

The uncertainty for a wide range of analyte level can be expressed by a combined uncertainty with two components (Ellison 2012, Hawkins 2014). The component describing the additive errors, σ_a^2, follows a normal (n) distribution. The second component describing the multiplicative errors, σ_m^2, follows a log-normal (ln) distribution.

The two-component model is thus

$$sd(\hat{x}|x_0) = \sqrt{\underbrace{\sigma_a^2}_{additive} + \underbrace{\sigma_m^2 \cdot x_0^2}_{multiplicative}}$$

where x_0 is the measurand value.

These model coefficients are estimated by fitting the model to the data, i.e. the means of replicate measurements and their respective variances.

For each sample the variance of the replicates, s_j^2, and the degree of freedom, $DF_j = M_j - 1$, can be calculated. Transforming the restricted likelihood function into a more general form, the log-likelihood $l(\sigma_a^2, \sigma_m^2)$ can be also given by the example of the two-component model with the additive and multiplicative error by (Hawkins 2014):

$$\frac{-1}{2}\sum_{j=1}^{g}\left[\frac{DF_j \cdot s_j^2}{\sigma_a^2 + \sigma_m^2 \cdot \overline{x}_j^2} + DF_j \cdot ln\left(\sigma_a^2 + \sigma_m^2 \cdot \overline{x}_j^2\right)\right]$$

DF_j, s_j^2 and $\overline{x}_j^2$ are data from the measurements, the σ^2 are the model coefficients, for which the values should be calculated, so that the sum reaches the maximal possible value.

Aim

Estimate individual uncertainties by analyte level:

- Parametrize the uncertainty model by data from replicate measurements
  - Reduce number of necessary replicates
  - Provide flexible choice of uncertainty model
  - Help avoiding common mistakes by providing warnings and error messages
- Use model parameter for further interpretation
  - Additive component for "background noise"
  - Multiplicative component as constant RSD

Usage

```
library(gmUncertainty) # load package
load(“my_data.RData”) # load your data
```

1. Calculate means, variances and degrees of freedom (DF) for all replicates:

```
data_summary = summarise_measurements(my_data, # the (tidy) data
                    col_value = “my_element”, # column name of variable
                    col_variable = “SampleID”) # column name of identifier column
```

Example output:

SampleID	Mean	RSD	Var	DF
191-B-1A	0.41850	0.01182735	0.000024500	1
220-B-0A	0.01055	0.18096572	0.000003645	1
243-B-0A	0.06540	0.04973534	0.000010580	1

2. Get model coefficients:

```
Myresult = get_mod_coef(data_summary, # input data
                    mod_par = c("n", "ln"), # design the model
                    col_mean = "Mean", # column name of the means
                    col_var = "Var", # column name of the variances
                    col_df = "DF") # column name of the degrees of freedom
```

Example output:

model_parameter	estimate
n	0.002035605
ln	0.045943332

3. Use results:

- Additive (normal, n): 0.002 mg/kg -> "noise" level
- Multiplicative (log-normal, ln) : 4.6% -> RSD

$$RSD(x) = \frac{\sqrt{\underbrace{0.002^2}_{additive} + \underbrace{0.046^2 \cdot x^2}_{multiplicative}}}{x}$$

References

Ellison, S. L. R., and A. Williams, eds. 2012. *Quantifying Uncertainty in Analytical Measurement*. Third edition. Eurachem/CITAC Guide.

Hawkins, Douglas M. 2014. "A Model for Assay Precision." **Statistics in Biopharmaceutical Research 6 (3):** 263–69. https://doi.org/10.1080/19466315.2014.899511.

JCGM 100:2008. 2010. *Evaluation of Measurement Data – Guide to the Expression of Uncertainty in Measurement (Jcgm 100:2008)*. online; BIPM; IEC; IFCC; ILAC; ISO; IUPAC; IUPAP; OIML. https://www.bipm.org/en/committees/jc/jcgm/publications

R Core Team. 2019. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Acknowledgement

This work is part of the NEXT project which receives funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 776804.

Helmholtz Institute Freiberg · Department of Modelling and Valuation
Dr. Solveig Pospiech · s.pospiech@hzdr.de · www.hzdr.de