Characterising inter-laboratory variability in environmental standard setting using weighted hierarchical bootstrapping

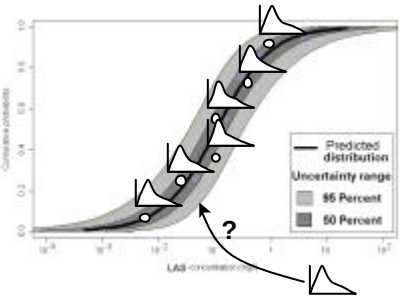
Frederik A.M. Verdonck, Olivier Thas and Peter A. Vanrolleghem

Ghent University - BIOMATH, Department of Applied Mathematics, Biometrics and Process Control, Coupure Links 653, B-9000 Gent, BELGIUM

Introduction (theory)

Complex data have several forms of **variations**:

- $X_1, X_2, \ldots, X_n$ are i.i.d. as X (——)
- sampling error : confidence bands around the cumulative distribution function (CDF) of X (using bootstrapping). ( )
- X_1 is a summary statistic, it basically is also a random variable and $X_2, \ldots, X_n$ are also random variables.
  E.g. X_i is the mean of $X_{i1}, X_{i2}, \ldots, X_{imi}$ (o)



Introduction (case study)

In the case study, X_i is the toxicity of a chemical towards a species.

Same forms of **variations**:

- variability between species (= inter-species sensitivity towards a chemical) (——)
- uncertainty: sampling error ( )
- X_i is the mean of several values found in literature (from inter-laboratory variations o)

Goal: How to include inter-laboratory variations?

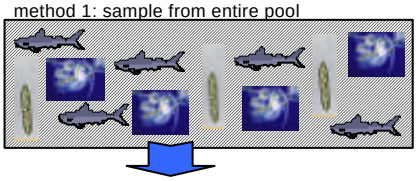
Proposed methodology

The **parametric bootstrap** method (assuming lognormal distribution) was selected as technique for characterising confidence intervals.

The answer on the question depends on the interpretation of the inter-laboratory variations: **variability** or **uncertainty**?

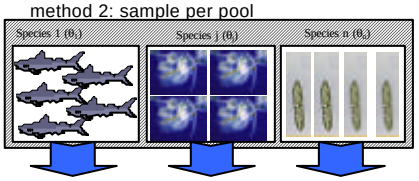
Variability: real variations, cannot be reduced through additional measurements
=> number of samples per shot = $m_1 + m_2 + \ldots + m_n$ samples
Two sampling strategies were investigated (same level or hierarchical level)

method 1: sample from entire pool



$m_1 + m_2 + \ldots + m_n$ samples

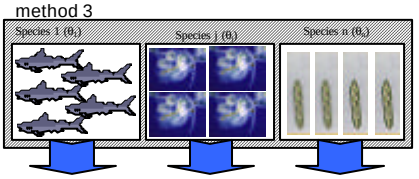
method 2: sample per pool



m_1 m_2 ... m_n

Uncertainty: error or ignorance, can partly be reduced through additional measurements
=> number of samples per shot = 1

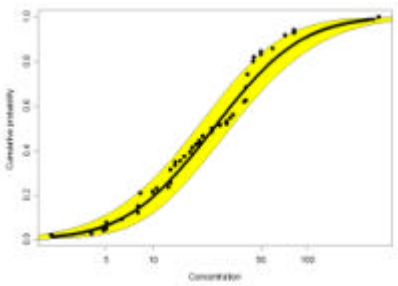
method 3



1 1 1

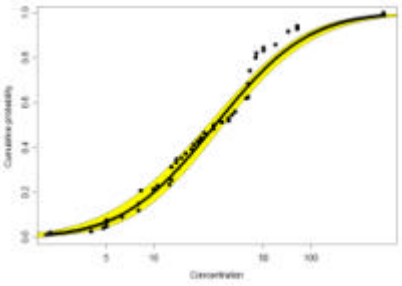
Results + Discussion

Depending on the method used, the interpretation of the black line and its uncertainty band is different:

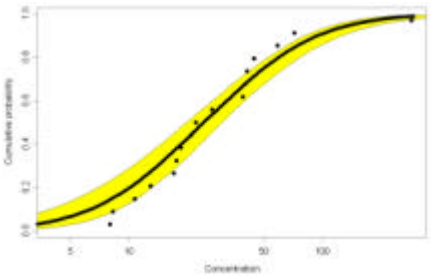


black line = inter-species + inter-laboratory variability
yellow band = sampling uncertainty

Interpretation: error due to sampling between all data i.e. from the entire pool



inter-species + inter-laboratory variability
sampling uncertainty



inter-species variability
sampling + inter-laboratory uncertainty

Interpretation: integrated sampling error of each species separately (i.e. between individuals per species conditioned on the species)

Based on expert knowledge, inter-laboratory variations should be interpreted as variability because the variations are not reducible (uncertainty can always partly be reduced).

TAKE HOME MESSAGE

- Treating all variations on the same level (method 1) was found to be the best method for environmental standard setting because:
  - inter-laboratory variations are interpreted as variability
  - the modelled uncertainty is sampling error for all data

Acknowledgement

This research has been funded by a scholarship from the Flemish Institute for the Improvement of Scientific-Technological Research in the Industry (IWT). The authors also like to thank Dr. Jaworska (Procter & Gamble) and Prof. Dr. Janssen (Ghent University - Laboratory for Environmental Toxicology and Aquatic Ecology) for their useful suggestions.