A procedure for estimating fundamental variability

Steffen Uhlig[1], Bertrand Colson[1] and Petra Gowik[2]

[1] QuoData GmbH, Prellerstr. 14, 01309 Dresden, Germany
uhlig@quodata.de
tel: +49 (0)351 40 28 86 711
fax: +49 (0)351 40 28 86 719

[2] Federal Office of Consumer Protection and Food Safety, Mauerstr. 3942, 10117 Berlin, Germany
petra.gowik@bvl.bund.de
tel: +49 (0)30 18444 00 0
fax: +49 (0)30 18444 89 99 9

Abstract

The question whether a given set of test items can be considered "identical" is often addressed in terms of the homogeneity of the test material from which said items were taken. However, for some types of matrices – in particular, for matrices consisting of minute separate particles, only some of which carry the analyte under consideration – even in the case of homogenous test material, an irreducible source of variability between test items may remain: the fundamental variability. In this paper, the concept of fundamental variability is explained, and procedures for reducing and characterizing it are described.

Keywords: Fundamental variability, homogeneity, variance, method validation, proficiency testing, measurement uncertainty

Introduction

In the approach to characterizing the precision of an analytical method described in ISO 5725 [1], total variability is partitioned into two different components: laboratory bias and analytical error. Sample bias (i.e. differences between samples) is not considered a component of precision. For this reason, it is required in ISO 5725 that, for a given test material and level, differences between the test items sent to the participating laboratories should be negligible. In Section 6.4.2 of ISO 5725-1, this requirement is stated as follows: "If different items are to be used in different laboratories, then they shall be selected in such a way as to ensure that they can be presumed to be identical for practical purposes." In Section 6.4.4 of ISO 5725-1, it is further explained that if the requirement regarding "identical" test items cannot be met, then "…inhomogeneity of the test material will form an essential component of the precision of the measurement…" The question of sample inhomogeneity is thus of considerable importance.

The question whether a given set of test items can be considered "identical" is often addressed in terms of the homogeneity of the test material from which said items were taken. The tacit assumption is that if the test material is homogenous, then the test items will be "identical." There are cases, however, where test items taken from perfectly homogenous

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2022 doi:10.20944/preprints202211.0460.v1

test material will not meet the “identical” requirement, leading to an additional component of variability.

In this paper, this additional component of variability is identified as the fundamental variability, defined as the irreducible variation between test items or test portions which remains even under the highest achievable degree of homogeneity. The concept of fundamental variability as presented here is not new. Pierre Gy published an elaborate discussion of the origins of fundamental variability in 1979 [2]. Unfortunately, his insights did not reach the scientific community responsible for method validation and measurement uncertainty, so that the fundamental variability is not discussed in the majority of the relevant standards and guides (ISO 5725, ISO 21748 [3], JCGM GUM series [4] and VIM [5]).

In the following, the concept of fundamental variability will be explained, and procedures for reducing and characterizing it will be described.

The concept of fundamental variability

In this paper, the term *sample* is used to denote the result of obtaining an individual sampling unit from a larger mass of material. Thus, the term sample can be used to denote a test item taken from bulk material prepared for a method validation study or a proficiency test, and subsequently packaged and sent to a participating laboratory. However, the term sample can also be used to denote a *test portion* taken from a laboratory sample inside a given laboratory. In a chain of sampling steps in which a sample from a given step becomes the basis for obtaining samples at the following step, it is convenient to use the term *subsample* to denote the individual sampling unit and the term *sample* to denote the larger mass of material from which said subsample was taken.

The term *homogenous* is often used to mean that the concentration of target analyte is *uniform* across the material from which samples are obtained. It is often assumed that samples taken from homogenous material will result in homogenous samples, in the sense that the concentration of target analyte is uniform across the samples. This assumption does not always hold, due to the fundamental variability.

Fundamental variability reflects heterogeneity at the level of the material’s constituent particles. Fundamental variability is considerable when the target analyte is distributed among sparsely distributed *carrier particles*, each of which has a relatively high concentration in the sense that the presence of even a single such carrier particle in a given sample will affect the measured concentration.

Consider the following example. Say that in a *perfectly homogenous* sample, the target analyte is distributed among carrier particles, with 10 ng of target analyte per carrier particle, and an average of 5 carrier particles per g. This does not mean that there will be exactly 5 carrier particles in every 1 g subsample. Rather, the number of carrier particles per subsample will vary randomly, following a Poisson distribution. The following table provides an overview of possible outcomes along with the corresponding probabilities.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2022 doi:10.20944/preprints202211.0460.v1

Table 1: Probability that a subsample contains a given number of carrier particles if, on average, there are 5 carrier particles per 1 g subsample (the mean target analyte concentration in the sample is 50 ng/g)

Number of particles in 1 g subsample	Concentration of target analyte	Probability
0-3	0-30 ng/g	26.5 %
4-6	40-60 ng/g	49.7 %
≥ 7	≥ 70 ng/g	23.8 %

As can be seen from this simple example, the fundamental variability can have a considerable impact on analyte concentration: there is > 50 % probability that the target analyte concentration in a subsample will lie outside the concentration range 40-60 ng/g, which corresponds to ± 20 % of the mean concentration in the sample from which the subsamples were obtained.

Two questions thus arise. First: how can we reduce the fundamental variability, and second, how can we measure or characterize it.

In practice, the fundamental variability can be reduced by increasing subsample size, and by finer grinding. Both of these will in effect increase the mean number of carrier particles per subsample. The impact of increasing subsample size will be illustrated on the basis of the previous example. Since the sample was perfectly homogenous, doubling the size of subsamples will double the mean number of carrier particles per subsample. The following table provides the corresponding probabilities.

Table 2: Effect of doubling subsample size. Probability that a subsample contains a given number of carrier particles if, on average, there are 10 carrier particles per 2 g subsample (the mean target analyte concentration is 50 ng/g)

Number of particles in 2 g subsample	Concentration of target analyte	Probability
0-7	0-35 ng/g	22.0 %
8-12	40-60 ng/g	57.1 %
≥ 13	≥ 65 ng/g	20.8 %

As can be seen, the probability that the concentration of a given subsample will lie within ± 20 % of the mean concentration has increased from roughly 50 % to 57 %.

Turning to the effect of finer grinding, consider the simple case that the carrier particles are ground in such a way as to double their number and halve each carrier particle's target analyte content. We will then obtain the same probabilities as in Table *2*. But what happens if finder grinding is combined with increasing subsample size? The following table provides the probabilities corresponding to this combined impact.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2022 doi:10.20944/preprints202211.0460.v1

Table 3: Combined effect of doubling subsample size and finer grinding. Probability that a subsample contains a given number of carrier particles if, on average, there are 20 carrier particles per 2 g subsample, each with a target analyte concentration of 5 ng/g (the mean concentration is 50 ng/g)

Number of particles in 2 g subsample	Concentration of target analyte	Probability
0-15	0-37.5 ng/g	15.7 %
16-24	40-60 ng/g	68.7 %
≥ 25	≥ 62.5 ng/g	15.7 %

As can be seen, the probability that the concentration of a given subsample will lie within ± 20 % of the mean concentration has now increased to roughly 69 %.

Alternative procedures for reducing fundamental variability

In many cases, due to constraints arising from the analytical procedure, it is not possible to increase the size of the test portions themselves. In such cases, two procedures can ensure the fundamental variability is under control: gradual sampling and sieving.

Procedure 1: Gradual sampling

Typically, fundamental variability appears twice: first, during sampling (i.e. taking test material from a container or batch of products to form a laboratory sample), and second, during subsampling within the laboratory, e.g. taking a test portion from the homogenized laboratory sample.

The proposed procedure consists in adding intermediate sampling steps, and in controlling the fundamental variability at each step via sample size and grinding. The main point is that fundamental variability can be reduced even though a previously specified test portion size must be respected.

The procedure will be illustrated on the basis of an example.

Take the case that there are 5000 carrier particles in 1 tonne of perfectly homogenous material in a container, with 1 mg of target analyte per carrier particle. The mean concentration is thus 5 mg/L. Further assume that the size of laboratory sample and test portion must be 1 L and 10 g, respectively.

In a two-step sampling process, the 1 L laboratory sample is taken directly from the container. The fundamental variability expressed as a Poisson RSD for this first step is quite considerable (ca. 45 %). Prior to obtaining the test portion, it is assumed that grinding can be applied to ensure the fundamental variability is negligible in the second step.

In the gradual sampling procedure, the first sampling step consists in taking a 100 L sample from the container. The fundamental variability expressed as a Poisson RSD in this first step is 4.5 %. The second sampling step consists in taking a 10 L subsample from the 100 L sampe. The Poisson RSD for this second step depends on the number of carrier particles in the 100 L sample. We assume here that there are 500 particles in the 100 L sample. The

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2022 doi:10.20944/preprints202211.0460.v1

Poisson RSD for the second step is thus calculated as 14.5 %. The remaining steps consist in taking a 1 L laboratory sample from the 10 L sample, and then to take a 10 g test portion. It is assumed that grinding can be applied to the 10 L sample and to the 1 L laboratory sample, so that the Poisson RSD is negligible in these 2 remaining steps. Combining the two Poisson RSD values from the first two sampling steps, a total relative fundamental variability of < 15 % is obtained.

The gradual sampling procedure is illustrated in the following table.

Table 4: Gradual sampling procedure. Comparison of fundamental variability values (expressed as Poisson RSD) with and without gradual sampling.

<table>
<tr><th colspan="3">No gradual sampling</th><th colspan="3">Gradual sampling</th></tr>
<tr><th>Sampling step</th><th>Description of sample</th><th>Fundamental variability</th><th>Sampling step</th><th>Description of sample</th><th>Fundamental variability</th></tr>
<tr><td></td><td>Container
1 tonne</td><td></td><td></td><td>Container
1 tonne</td><td></td></tr>
<tr><td colspan="3" rowspan="3"></td><td>1</td><td>Intermediate sample 1
100 l</td><td>Poisson RSD
≈ 4.5 %</td></tr>
<tr><td>2</td><td>Intermediate sample 2
10 l</td><td>Poisson RSD
≈ 14.5 %</td></tr>
<tr><td colspan="3">+ grinding</td></tr>
<tr><td>1</td><td>Laboratory sample
1 l</td><td>Poisson RSD
≈ 45 %</td><td>3</td><td>Laboratory sample
1 l</td><td>Poisson RSD
negligible due to grinding</td></tr>
<tr><td colspan="3">+ grinding</td><td colspan="3">+ grinding</td></tr>
<tr><td>2</td><td>Test portion
10 g</td><td>Poisson RSD
negligible due to grinding</td><td>4</td><td>Test portion
10 g</td><td>Poisson RSD
negligible due to grinding</td></tr>
</table>

It should be noted that the Poisson RSD depends on the number of particles per subsample. Thus, if there are 450 instead of 500 carrier particles in the 100 L sample[1] (first sampling step in the gradual sampling procedure), the Poisson RSD increases to 4.7 %. In the above example, this has no noticeable impact on the total fundamental variability, which remains < 15 %. However, there are cases where this needs to be taken into account. The combined Poisson RSD must then be computed via computer simulation. A more general point is that a

[1] The probability (Poisson distribution) that the number of carrier particles is less than or equal to 450 is 0.0124.

 doi:10.20944/preprints202211.0460.v1

correct evaluation of fundamental variability requires an appropriate consideration of all the sampling steps as one single process.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2022 doi:10.20944/preprints202211.0460.v1

Procedure 2: Sieving

The sieving procedure can be applied to reduce fundamental variability in cases where there is a range of different particles sizes, and the target analyte is concentrated in sparsely distributed in larger particles. Just as in the case of gradual sampling, the main point is that fundamental variability can be reduced even though a previously specified test portion size must be respected.

The starting point is a small enough sample, that sieving is practicable. It is important that the fundamental variability arising from previous sampling steps is acceptable. (See for instance the first two sampling steps in the gradual sampling example, resulting in a 10 L sample.) If the target analyte is concentrated in larger carrier particle which make up p % of the sample, then these larger carrier particles can be separated, yielding a $p \cdot 10$ L sieving product. For instance, if $p = 0.1$, the sieving product is a 1 L sample. The proportion of carrier particles in the sieving product is very high, so that the fundamental variability associated with taking a test portion from the sieving product is negligible. Once the concentration in the sieving product is determined, it is multiplied with p in order to obtain the actual concentration in the starting sample.

Procedure for characterizing fundamental variability

We now turn to the second question: How can we characterize or measure fundamental variability in order to determine whether it is acceptable?

One possible approach involves the use of handheld *in situ* sensors or spectrometers, sufficiently sensitive to identify differences in target analyte concentrations at the molecular level.

Another approach is the application of specially designed experiments. Fundamental variability cannot be characterized by means of classical homogeneity studies such as the standard designs described in ISO 13528 [6] and Guide 35 [7]. Indeed, in these designs, it is not possible to distinguish fundamental variability from sample heterogeneity *per se*, so that the former may be mistaken for the latter. We will return to this aspect in the next section.

The following procedure allows a characterization of fundamental variability.

Step 1
Check whether one of the following criteria is met:

Criterion 1: The in-house repeatability standard deviation is larger than 3 times the expected value.
Criterion 2: The in-house repeatability standard deviation is larger than the Horwitz SD value.
Criterion 3: Conspicuous “upper” outliers are present in QC data. (The presence of such outliers constitutes a further indication that the unexpectedly large observed variability may be due to fundamental variability.)

If at least one of these criteria is met, proceed to Step 2.

Step 2
Conduct the following experiment:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2022 doi:10.20944/preprints202211.0460.v1

1. Obtain 20 test results under repeatability conditions. Calculate the corresponding variance s_1^2.
2. Increase test portion size by a factor k (e.g. triple test portion size, $k = 3$). If it is not possible or practical to increase test portion size, grinding and homogenizing a volume corresponding to a k-fold increase in test portion size prior to taking a test portion with the original size is another option.
3. Obtain 20 test results under repeatability conditions on the basis of the increased test portion size (or finely ground test material). Calculate the corresponding variance s_2^2.
4. If the ratio $\frac{s_1^2}{s_2^2}$ is greater than 2.17 (quantile of the corresponding F-distribution), then calculate the SD characterizing fundamental variability as follows:

$$s_F = \sqrt{\frac{k}{(k-1)} \cdot (s_1^2 - s_2^2)}$$

The procedure just described assumes that the data are normally distributed. However, in certain cases it may be sensible to apply a procedure based on the Poisson distribution.

It should be noted that 20 test portions will not always be sufficient. Consider, for example, the case that there are only 5 carrier particles in the laboratory sample, from which – on the basis of the specified test portion size – as many as 1000 test portions could be obtained. In this scenario it is unlikely that any of the 20 test portions will contain at least one carrier particle. In such a situation, it would be advisable to incorporate *a priori* information – such as expert knowledge based on similar test materials – in the evaluation of the fundamental variability.

The procedure will now be illustrated on the basis of an example. The following table provides data for the calculation of the fundamental variability following the procedure just described.

Table 5: Target analyte concentrations [ng/g] from an experiment for the calculation of fundamental variability

	Experiment 1: Original test portion size	Experiment 2: Test portion size is tripled
Sample 1	14.0	15.1
Sample 2	11.9	13.8
Sample 3	10.5	11.8
Sample 4	14.9	14.0
Sample 5	13.1	11.4
Sample 6	9.5	15.7
Sample 7	15.6	12.4
Sample 8	18.3	11.5
Sample 9	12.5	12.1
Sample 10	16.4	13.7
Sample 11	18.0	15.8
Sample 12	14.0	12.5
Sample 13	13.0	12.8
Sample 14	20.8	15.1
Sample 15	10.2	11.8

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2022 doi:10.20944/preprints202211.0460.v1

Sample 16	21.5	10.6
Sample 17	13.9	11.1
Sample 18	17.8	12.9
Sample 19	7.7	11.4
Sample 20	12.2	16.3

Note that, in Experiment 1, several conspicuously large values are obtained – an indication that fundamental variability is non-negligible.

The following variances and corresponding ratio are obtained:

Table 6: Variances and their ratio

s_1^2	s_2^2	s_1^2/s_2^2
13.54	3.05	4.44

As can be seen, the ratio s_1^2/s_2^2 is greater than the value 2.17 (see point 4 in Step 2, above). Accordingly, the fundamental variability is calculated as

$$s_F = \sqrt{\frac{3}{2} \cdot (s_1^2 - s_2^2)} = 3.97.$$

Distinction between fundamental variability and heterogeneity

The following examples clarify the distinction between fundamental variability and heterogeneity according to ISO 13528 [6].

Table 7 provides fundamental variability SD values for 5 experiments in which 30 % between-sample heterogeneity is added to the test results provided in Table 5.

Table 7: Fundamental variability SD values for five experiments in which ± 30 % between-sample heterogeneity was added to the data from Table 1

Experiment	**1**	**2**	**3**	**4**	**5**
Fundamental variability SD [ng/g]	3.68	4.35	4.79	2.76	3.99

As can be seen, the standard variation values vary randomly, sometimes less than, sometimes greater than the 3.97 value obtained in the previous section. In other words, the above procedure for the calculation of the fundamental variability is not affected by the presence of between-sample heterogeneity.

Conversely, fundamental variability will have practically no effect on a standard homogeneity experiment following Annex B.1.1 in ISO 13528 [6].

Table 8 provides test results from such a homogeneity experiment.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 November 2022 doi:10.20944/preprints202211.0460.v1

Table 8: Data from homogeneity experiment

	Test result 1 [ng/g]	Test result 2 [ng/g]
Test item 1	13.61	12.38
Test item 2	10.36	11.12
Test item 3	8.87	8.15
Test item 4	10.47	9.03
Test item 5	8.64	6.09
Test item 6	6.09	5.56
Test item 7	10.92	9.20
Test item 8	11.08	10.64
Test item 9	11.03	8.96
Test item 10	8.18	7.74

The following analytical (within test item) and heterogeneity (between test item) standard deviation values are obtained:

Analytical standard deviation:	0.98 [ng/g]
Heterogeneity standard deviation:	1.92 [ng/g]

If fundamental variability is added to the data (e.g. 20 %, that is to say, the same relative magnitude as the heterogeneity standard deviation), then the new heterogeneity standard deviation hardly changes, though the analytical standard deviation has doubled (which may admittedly lead to a violation of the IUPAC 0.5 SDPA criterion [8], thus disqualifying the homogeneity experiment).

Table 9: Data from homogeneity experiment, with added 20 % fundamental variability

	Test result 1 [ng/g]	Test result 2 [ng/g]
Test item 1	13.61	14.38
Test item 2	10.36	13.12
Test item 3	8.87	10.15
Test item 4	12.47	13.03
Test item 5	10.64	8.09
Test item 6	10.09	5.56
Test item 7	10.92	9.2
Test item 8	13.08	10.64
Test item 9	13.03	12.96
Test item 10	8.18	15.74

The following analytical and heterogeneity standard deviation values are obtained for the modified homogeneity experiment:

 doi:10.20944/preprints202211.0460.v1

Analytical standard deviation:	2.27 [ng/g]
Heterogeneity standard deviation:	2.18 [ng/g]

The reason why fundamental variability hardly affects the heterogeneity standard deviation is that fundamental variability manifests itself as random variation between all test portions, i.e. it manifests itself in the same manner both within and between test items. Thus it can be expected to have practically no impact on differences between test item-specific mean values. This is the reason why the experimental designs proposed in ISO 13528 [6] and ISO Guide 35 [7] for homogeneity testing are not suitable to characterize fundamental variability.

Closing remarks

Depending on the type of sample, fundamental variability can have a considerable effect on test results in areas such as food or soil analysis. This is particularly so in the case of small test portion sizes and low concentrations, where the target analyte is concentrated among sparsely distributed carrier particles.

The fundamental variability can be brought under control by increasing sample size and by finer grinding of samples. Both of these measures are tantamount to increasing the average number of carrier particles per subsample. In cases where, due to analytical constraints, it is impossible to increase the test portion size, then the fundamental variability can be brought under control by introducing intermediate sampling steps, each with carefully determined sample size and grinding procedures. Another possibility is to separate carrier particles via a sieving procedure, and to take the test portion from such a sieving product.

It is important to note that, in any investigation of fundamental variability, the central quantity is the average number of carrier particles per subsample at each sampling step. This quantity depends on the distribution of particle sizes, the volume of the sample and the volume of the subsample. Another relevant question is whether the target analyte concentration is uniform across carrier particles.

The procedure for characterizing fundamental variability is a top-down approach and thus yields a measure of fundamental variability without requiring any particle-level understanding of the test material. However, as pointed out in the text, it is assumed that the data are normally distributed. Accordingly, the validity of this assumption should be checked before the procedure is applied.

The procedure proposed in this paper for characterizing the fundamental variability requires a high number of individual analyses, which in most fields of application can be realized only exceptionally. It therefore makes practical sense to link the proposed procedure with suitable model assumptions in order to then arrive at an estimate of fundamental variability in routine applications with much less effort. For example, the approach can be applied to different particle size fractions as examples, since the effect of the fundamental variability is very strongly dependent on the particle size distribution. A routine estimation of the fundamental variability could then be achieved by determining the mass fractions as well as the associated contents in the two or more particle size fractions.

In the introduction, it was briefly mentioned that the case of heterogenous samples lies beyond the scope of ISO 5725. The question whether irreducible sample heterogeneity

 doi:10.20944/preprints202211.0460.v1

should be considered an actual component of an analytical method's precision has perhaps not yet received sufficient attention. One possible approach would be to distinguish method precision *per se* (laboratory bias and repeatability precision) from matrix heterogeneity, and to provide information regarding both. Certainly, if the fundamental variability is non-negligible, then it must be accounted for in any evaluation of measurement uncertainty. Accordingly, if test portion size or grinding procedures differ in routine measurement from the procedures followed in the method validation study (say, to ensure homogenous samples were sent to the participants), the missing component of measurement uncertainty must be evaluated separately and duly accounted for. The procedure described in this paper can then be applied.

References

[1] ISO 5725-2:2019. Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method.

[2] Gy P. Sampling of particulate materials theory and practice (1979), Elsevier scientific publishing company.

[3] ISO 21748:2017. Guidance for the use of repeatability, reproducibility and trueness estimates in measurement uncertainty evaluation.

[4] JCGM 100:2008. Evaluation of measurement data — Guide to the expression of uncertainty in measurement.

[5] JCGM 200:2012. International vocabulary of metrology — Basic and general concepts and associated terms (VIM).

[6] ISO 13528:2015. Statistical methods for use in proficiency testing by interlaboratory comparison.

[7] ISO Guide 35:2017. Reference materials — Guidance for characterization and assessment of homogeneity and stability.

[8] Thompson M., Ellison S.L.R., Wood R. The International Harmonized Protocol for the proficiency testing of analytical chemistry laboratories (IUPAC Technical Report). Pure Appl. Chem. 2006, 78 (1) pp. 145–196.