HOW TO TREAT CORRELATION IN THE UNCERTAINTY BUDGET,
WHEN COMBINING RESULTS FROM DIFFERENT MEASUREMENTS

RÜDIGER KESSEL, MICHAEL BERGLUND, PHILIP TAYLOR AND
ROGER WELLM

European Commission, Joint Research Centre
Institute for Reference Materials and Measurements
Retieseweg, B-2440 Geel, Belgium
e-mail: ruediger.kessel@irmm.jrc.be

1 Abstract

The ISO/BIPM Guide to the Expression of Uncertainty in Measurement (GUM) describes a
method to evaluate the associated uncertainty of a measurement result. It is still an ongoing
challenge to adapt the Guide to the different fields of metrology. In chemical analysis results
from different measurements must often be combined. This paper will discuss cases where
correlation can have an import influence on the uncertainty of the result.
A scheme will be presented for the calculation of the correlation between results using the
uncertainty budgets. Implementing a model, which includes the correlation, can significantly
change the importance of some parameters. It also gives the analyst a better understanding of
the major sources of uncertainties in the measurement process.

2 Introduction

The ISO/BIPM Guide to the Expression of Uncertainty in Measurement (GUM)
provides an international accepted method to quantify the quality of a measurement. Many
metrology institutes have adopted it. Without a full uncertainty budget for the
measurements, it would be difficult to confirm that results are properly anchored
and traceable to standard units (SI).

A lot of effort is invested to achieve a certain quality of the results expressed as
a specific uncertainty. It is important to understand all aspects of the uncertainty
budgeting if the results are used for international standardization and comparisons.

In the certification work of chemical reference materials, measurements are
often carried out several times with partly independent methods, to get a better
understanding of the measurand. An arithmetic or weighted mean is used to
combine these results. The calculation of the uncertainty is problematic when the
standard error propagation is used. The paradoxical situation can often occur that
the uncertainty of the result is smaller than the uncertainty of the reference material
used, especially if more than 6 values are combined.
The nature of metrology does not allow the uncertainty of a result to be smaller than the uncertainty contribution of the reference values used. The measurement result cannot be better than the value anchoring it to the unit system (SI).

In the past the uncertainty of the result in our certificates was not directly derived from the calculation of the value. Instead the uncertainty of the result was estimated by an expert judgement based on the uncertainty of the combined values. The GUM, however, only allows expert judgement for input quantities not for results. So this method should be replaced by a better founded procedure.

The effect of the averaging of the contributions of the reference values on the final uncertainty can be avoided if the correlation between the combined values is included.

An approach to understand the source of correlations between results will be presented in this paper. The concept of the uncertainty contribution coefficient will be introduced as well as a scheme to calculate the correlation. The correlation between result and input quantities will be investigated and the problem of combining results from different sources will be treated. With a short discussion about the correlation in a test of equivalence a pragmatic solution for the multiple-replicate problem will be presented. In a short example it will be shown what influence correlation may have on regression. The influence on chemical inter-comparisons will be discussed and finally some implications on future certificates will be discussed.

3 Correlation introduced by the model of evaluation

One of the biggest improvements of the GUM method compared to the former concepts is that the estimation of the measurement uncertainty is derived from the model for evaluating the result. All the knowledge of a specific measurement can be implemented in the measurement evaluation model. With the freedom to use an appropriate model, the GUM method can be tailored to nearly every measurement. Very often, however, when the GUM method is implemented in practice, the biggest problem is to define the complete measurement evaluation model.

One very useful method to determine the evaluation model is to divide the problem into smaller sub-problems by introducing interim results (substitution). These sub-equations normally describe well-defined steps of the measurement procedure. The advantage is that the sub-equations can be derived directly and they are understandable and easy to handle. It is usually an advantage to split up a complicated problem into smaller well-defined parts. Figure 1 and Figure 2 illustrate two of the different possible evaluation models for a measurement.

The interim results can then be calculated individually and later combined to get the final result. It makes no difference for the value of the result how the calculation is done. A step by step calculation gives the same result as constructing one big equation. However there can be a difference when the uncertainty is calculated if correlation is not taken into account. It is often the case that the same
quantities with the same estimations are used in several sub-equations. There can be a significant difference in the final uncertainty if the uncertainties of the interim results are calculated separately and then combined or if everything is calculated in one single, big model.

The reason for this is that interim results can be correlated if they share common input quantities. Theory and practical calculations show that the uncertainty for a stepwise calculation is the same as with one big model if the correlation is taken into account.

The same situation arises if previous measured results are combined to a new result. Former end results now become interim results for the calculation of a new end result. If values with a combined uncertainty are used then correlations can arise.

Figure 1: Example of a measurement evaluation in one step.

Figure 2: Example for a step by step evaluation. Possible correlation of interim result must be taken into account.
4 Calculating the correlation between results

Suppose a set of $m$ results $Y_j$ ($j = 1, \ldots, m$) depends on the set of $n$ input quantities $X_i$ ($i = 1, \ldots, n$), although some of these quantities $X_i$ may not necessarily appear in all functions.

$$Y_j = f_j(X_1, \ldots, X_n) \quad (j = 1, \ldots, m)$$ (0)

The correlation between the results $Y_j$ can be calculated if the functions can be linearized and the full uncertainty budgets for these results are known. We define the uncertainty contribution coefficient of $n$ input quantities to $m$ results as:

$$h_i(y_j) = \frac{\partial Y_j}{\partial X_i} \cdot \frac{u(x_j)}{u(y_j)} \quad (i = 1, \ldots, n) \quad (j = 1, \ldots, m)$$ (1)

The uncertainty contribution coefficient is the uncertainty of an input quantity multiplied with the sensitivity of the input to the result (partial derivative) divided by the uncertainty of the result. The values for the sensitivities and uncertainties must be obtained from the individual budgets of the results.

The uncertainty contribution coefficient has no unit and is smaller than or equal to unity if the input quantities are not correlated. Note that percent figures cannot be used for these parameters, because they do not sum up to one.

The correlation can be calculated with equation 2a, if the input quantities are not correlated. It is derived from an equation given in [2]:

$$r(y_k, y_l) = \sum_{i=1}^{n} h_i(y_k) \cdot h_i(y_l) \quad (k, l = 1 \ldots m)$$ (2a)

If the input quantities are correlated equation 2b can be used.

$$r(y_k, y_l) = \sum_{i=1}^{n} \sum_{j=1}^{n} h_i(y_k) \cdot h_j(y_l) \cdot r(x_i, x_j) \quad (k, l = 1 \ldots m)$$ (2b)

The following calculation scheme is useful for calculations if the input quantities are not correlated. It can be easily implemented in a spreadsheet program. In one table the uncertainty contribution coefficients for all results and all common input quantities are listed. Only input quantities, which are used for more than one result should be taken into account. If a quantity is not used for the calculation of a result then the uncertainty contribution coefficient is zero.

Table 1 shows an example of this table for 4 results and seven common inputs.
Table 1: Example of a table of uncertainty contribution coefficients for four results.

<table>
<thead>
<tr>
<th>$h_i(y)$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>0.1</td>
<td>0</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0.8</td>
<td>0.3</td>
<td>0.1</td>
<td>0</td>
</tr>
<tr>
<td>$x_3$</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>$x_4$</td>
<td>0.2</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$x_5$</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>$x_6$</td>
<td>0.3</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$x_7$</td>
<td>0.4</td>
<td>0.2</td>
<td>0.1</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2 shows the correlation matrix between the results. The diagonal values of any correlation matrix must be one, so they do not need to be calculated. The correlation matrix is symmetric to the diagonal, so only one half have to be calculated. The correlation coefficients are calculated by summing up the product between two columns of Table 1.

\[
\begin{align*}
\text{r}(y_1, y_2) &= 0.1 \cdot 0 + 0.8 \cdot 0.3 + 0.1 \cdot 0.1 + 0.2 \cdot 0.8 + \ldots \\
\text{r}(y_1, y_3) &= 0.1 \cdot 0.5 + 0.8 \cdot 0.1 + 0.1 \cdot 0.1 + 0.2 \cdot 0 + \ldots \\
\vdots
\end{align*}
\]

Table 2: Correlation matrix derived from Table 1.

<table>
<thead>
<tr>
<th>$r(y_i, y_j)$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>$y_3$</th>
<th>$y_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>1</td>
<td>0.55</td>
<td>0.18</td>
<td>0.03</td>
</tr>
<tr>
<td>$y_2$</td>
<td>0.55</td>
<td>1</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>$y_3$</td>
<td>0.18</td>
<td>0.06</td>
<td>1</td>
<td>0.27</td>
</tr>
<tr>
<td>$y_4$</td>
<td>0.03</td>
<td>0.01</td>
<td>0.27</td>
<td>1</td>
</tr>
</tbody>
</table>

If we want to investigate the correlation between different results then, in addition to the uncertainty of the result, a list of uncertainty contribution coefficients must be available.

The correlation needs to be investigated only if the uncertainty contribution coefficient is greater than about 0.1. The maximum effect to the final uncertainty is then always smaller than 10%, but usually much smaller (see equation 2).

5 Correlation between result and input quantities

The discussion on correlation of results with common input quantities leads to the question of how a result is correlated with the input quantities used in calculating the result. The answer can be found if we assume that a second result $Y_j^*$ exists which is equal to each input quantity.
\[ Y = f(X_1, \ldots, X_n) \]
\[ Y^*_i = X_i \]  
Equation 2 above gives the following solution in the case the input quantities are not correlated.

\[ r(y, x_i) = h_i(y) \]  
(4a)

The correlation between the result and a contributing input quantity is equal to the uncertainty contribution coefficient if all input quantities are non-correlated. If the input quantities are correlated the equation becomes a bit more complicated:

\[ r(y, x_i) = \sum_{j=1}^{N} r(x_j, x_i) \cdot h_j(y) \]  
(4b)

This gives again equation 4a for non-correlated input quantities.

6 Applications where correlation between results can be important

In this paragraph some cases are discussed where correlation between results should be investigated.

6.1 Combining results from different sources

For the certification of reference material, results from different measurements are often combined. If the same input quantity is used for different measurement results possible correlations must be considered. This is the case if the evaluation of the results contain common calibration data or tabulated values.

As an example, tabulated reference values from IUPAC or certified reference materials are often used in the calculation of analytical results in chemical measurements. For isotopic measurements the Isotopic Composition of the Elements table from IUPAC is widely used. These references sometimes contribute considerably to the uncertainty of the final result. For instance, an isotopic spike is calibrated against a reference material, which is assumed to have natural isotopic composition (as published by IUPAC). When the spike is applied to the measurement of a sample, the same constants (e.g. atomic weights, isotope composition) that were applied in the spike calibration are used again in calculating the result of the sample.

All results, which are calculated with the use of common values, are possibly correlated. The uncertainty contribution coefficient (defined above in equation 1) is a useful way to judge if results are correlated or not. If the uncertainty contribution coefficient to any result of a common quantity is greater than 0.1 then the correlation should be investigated.

The presented method is based on the decision of which input quantities are common in more than one calculation. Common quantities must be identified. There
are two possible methods to organise the calculations. The first is that a unique quantity name can be used for an input quantity in all calculations. If the same estimation of a quantity is used in different calculations then it shares a common unique name. This can lead to complicated naming rules and a high degree of discipline is needed if results from different authors are to be combined. The second solution is to use a clear reference scheme. All quantities, which are estimated from an external source and could be shared between calculations must be defined and must have a unique source (reference). If a quantity is used in a calculation the reference to the source must be stated. Input quantities are the same if they have a reference to the same source, even if they are named differently.

The second approach is preferred, because the naming of all quantities is hereby independent from the calculation of possible correlation and only quantities with a stated reference need to be investigated.

### 6.2 Test of Equivalence

The uncertainty of measurement is an important concept because it allows us to compare results in a reliable way. Only if the uncertainty or the reasonable spread of values is known is it possible to judge if the difference of two values is significant or not. It is very important to investigate any possible correlation if a difference between results is calculated. Any positive correlation will reduce the uncertainty of the difference. A non-significant difference can become very significant, if the correlation is taken into account.

The question of whether or not two results are equal is similar. Two values are often assumed equal if they fulfil the following condition.

\[ |Y_1 - Y_2| \leq k \cdot u(Y_1 - Y_2) \quad \text{(usually } k = 2 \text{) (6)} \]

It is important for this test that the correlation between the two results is investigated and included if known. If for example the uncertainty of the two results is dominated by a common reference value then the criterion for equivalence must be much smaller than if the values were independent. There are other ways to judge if results are equal or not, but if the uncertainty is used in the judgement then correlation should be investigated.

### 6.3 Multiple replicates

Chemical measurements are often done multiple times to check for possible instabilities. In contrast to a multiple reading of an instrument the whole measurement procedure is redone. A full uncertainty budget is calculated for every replicate. For combining the results of the replicates the arithmetic mean is used.

Due to the fact that the same reference and calibration materials are used for all replicates the results of the replicates are correlated. This correlation was not taken into account in the past. Instead a typical uncertainty for one replicate was used for
the mean value of all replicates. The disadvantage of this method is that it was up to the analyst performing the measurement to decide which replicate was representative for the measurement and which uncertainty should be used for the mean value.

If the correlation between the replicates is taken into account for calculating the uncertainty of the result, then all replicates can contribute to the uncertainty. No selection of a typical replicate is necessary any more.

Sometimes the complicated nature of chemical measurement gives rise to an unexpected spread of the replicates. Therefore an additional equivalence check should be performed on the results of the replicates:

\[ |Y_i - \bar{Y}| \leq k \cdot u(Y_i - \bar{Y}) \] (7)

All correlation between the replicates should be included in the calculation of \( u(Y_i - \bar{Y}) \). This shows if the variation between the results of the replicates is smaller than or equal to the uncertainty introduced by all statistical independent input quantities.

If this test fails there is a strong hint of an unknown effect between the replicates. The measurement should be investigated in order to find and correct this effect. There are cases where the resources are not available to study this effect in detail. In these cases additional quantities should be added to every replicate. It should be indicated that the quantities are included because of an unknown between replicate effect. The value of all these quantities is zero. The analyst should estimate the standard uncertainty of these added quantities (type B evaluation according to the GUM method). As long as there is no knowledge about replicates that are more reliable than others, the influence of the between replicate effect should be considered as the same for all replicates.

A possible criterion for the estimation of the unknown between replicate effect can be derived from the equivalence check above.

The uncertainty of the unknown effect can be chosen so that the equivalence check will be just passed. This value will be dependent on how the results were combined. If a non-weighted arithmetic mean is used the minimal additional uncertainty that must be added to the \( n \) replicates can be calculated from the following equation:

\[
\begin{aligned}
\delta_{\text{between}}(\syntext{\text{\textbf{u}}}) \geq & \sqrt{n - 1} \left[ \left( \frac{Y_{i,\text{max}} - \bar{Y}}{k} \right)^2 - u^2 \left( Y_{i,\text{max}} - \bar{Y} \right) \right] \\
Y_{i,\text{max}} & : \max \left\{ |Y_i - \bar{Y}| - k \cdot u(Y_i - \bar{Y}) \right\}
\end{aligned}
\] (8)
The equation doesn’t give any information about the between replicate effect at all. It only gives a minimum value for the uncertainty, which need to be added so that all replicates will pass this particular check. It is the responsibility of the analyst to decide if the value is reasonable or if a larger uncertainty is necessary to cover the possible between replicate effect.

The discussed procedure can be generalized. Multiple replicate measurements must be checked for a possible between replicate effect. Possible correlation between replicates should be investigated and must be included in the check. Any check can be used as long as it is fit for purpose and clearly defined.

If the test fails then additional quantities must be added to the replicates in order to model the (unknown) between replicate effect. The effect should of course be investigated and corrected if possible. If the effect cannot be corrected then an uncertainty must be estimated using scientific judgement. The minimal uncertainty that must be added can be derived from the check criterion as described above.

6.4 Regression

Another application where correlation can have a significant influence is the calculation of the parameter of a fitted curve by linear regression. It is clear that the uncertainty of the result of regression is bigger if the points are not statistically independent. Again common calibration factors are a typical source of correlation. The uncertainty contribution coefficients of common input quantities to every point must be known to calculate the correlation.

An example data set for a linear least square regression on $y$ for a simple model is shown in Table 3.

$$Y = V_m \cdot X$$

(9a)

Table 3: Example Data set for regression

<table>
<thead>
<tr>
<th>Point</th>
<th>$x_i$</th>
<th>$u(x_i)$</th>
<th>$y_i$</th>
<th>$u(y_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>2329.0456</td>
<td>0.0017</td>
<td>0.02808552</td>
<td>0.11$\times 10^{-9}$</td>
</tr>
<tr>
<td>$P_2$</td>
<td>2329.0455</td>
<td>0.0017</td>
<td>0.028085529</td>
<td>0.11$\times 10^{-9}$</td>
</tr>
<tr>
<td>$P_3$</td>
<td>2329.0254</td>
<td>0.0016</td>
<td>0.028085283</td>
<td>0.11$\times 10^{-9}$</td>
</tr>
<tr>
<td>$P_4$</td>
<td>2329.0435</td>
<td>0.0016</td>
<td>0.028085506</td>
<td>0.11$\times 10^{-9}$</td>
</tr>
<tr>
<td>$P_5$</td>
<td>2329.0272</td>
<td>0.0017</td>
<td>0.028085312</td>
<td>0.11$\times 10^{-9}$</td>
</tr>
<tr>
<td>$P_6$</td>
<td>2329.0277</td>
<td>0.0017</td>
<td>0.028085334</td>
<td>0.11$\times 10^{-9}$</td>
</tr>
</tbody>
</table>

$V_m$ represents a material property and is of interest. The slope was calculated with equation 9b. The standard uncertainty was calculated by error propagation.
\[ V_m = \frac{\sum_{i=1}^{n} x_i \cdot y_i}{\sum_{i=1}^{n} x_i^2} \quad n : \text{number of points} \]  

(9b)

The value for \( V_m \) was calculated to 12.05881584\( \times 10^{-6} \) with an associated standard uncertainty \( u(V_m) \) of 3.5\( \times 10^{-12} \) if no correlation is taken into account. The uncertainty contributions of the X and Y values are both dominated by common calibration factors. The correlation between the values was investigated. A correlation of \( r(X_i, X_j) = 0.99 \) and \( r(Y_i, Y_j) = 0.95 \) for \( i \neq j \) was found. If the correlation was included then the standard uncertainty \( u(V_m) \) is 8.6\( \times 10^{-12} \), more than doubled.

The data set in this example may be not very typical for regression, but it shows clearly that the uncertainty can be underestimated if an existing correlation is not taken into account.

6.5 Inter-comparison

In chemical measurements the same standards and reference materials are often used for different measurements. So the results share common input quantities. The repeatability of chemical analysis has increased a lot with modern instrumentation. As a consequence the contribution of the reference material to the uncertainty of the result can become more significant. If results are compared in an inter-comparison, possible correlation should be investigated. An indication of a possible correlation is the usage of the same reference material (or reference materials from the same source). Also values derived from IUPAC or other common sources should be investigated if the uncertainty contribution coefficient is greater than 0.1, as suggested in paragraph 4.

7 Conclusion

The fact that correlation can have an effect if results with associated uncertainties are combined is well known. In this paper we have shown, that the uncertainty contribution coefficient is a useful parameter to investigate possible correlation between results and calculate them. The importance of including correlation in uncertainty calculations has been shown for a number of cases. However, this leads to some requirements for the quantities used. For instance, if certified values are combined all information must be available to investigate if the values are correlated or not. The user of the certified values must perform this check, because
only they know how they combine the certified values. As a consequence the user must have all this information.

The standard certificate with a value and an associated uncertainty doesn’t provide this information. A clear reference to all standards and reference values used is necessary. The user must have a complete picture of the traceability chain so that they can investigate all possible correlation. One way to solve this problem is to include the uncertainty contribution coefficient of all relevant references and if any correlation between these references was taken into account (usually not) into the certificate. In practice, this can be limited only to values from external sources which have an uncertainty contribution coefficient greater than 0.1.

Only if all this information is provided can the result of a measurement be used in combination with other results and a complete uncertainty budget can be given for a final result.

References