Law of Propagation of Standard Deviation with and without covariance

In any experimental work, the total uncertainty of the analytical process is combined by the uncertainty contributions of *all steps* of the procedure (e.g. sampling, sample preparation, dissolution, separation, extraction, measurement, etc). The combination of individual standard uncertainties in the form of standard deviations to give the combined standard uncertainty is determined partly by statistics and partly by functional relationships of the form:

$$y \quad = \quad f\,(x_1, x_2, x_3, \ldots, x_n) \qquad [1]$$

Since each step of the analytical procedure, x_i is independent to each other, the combined standard uncertainty $u(y)$ of these independent components is given by:

$$u(y)^2 = \sum_{i=1}^{n} \left[\frac{\partial f}{\partial x_i} \right]^2 u(x_i)^2 \qquad [2]$$

or, $u(y)^2 \quad = \sum\{[\partial f/\partial x_1]^2 u(x_1)^2 + [\partial f/\partial x_2]^2 u(x_2)^2 + [\partial f/\partial x_3]^2 u(x_3)^2 + \ldots.\}$

In other words, in order to get the total variance, we would only consider the addition of the variance of each independent component.

However, if the standard uncertainties or errors are NOT independent, there is an extra covariance factor to be considered:

$$u\ \ (y)^2 = \sum \left(\frac{\partial f}{\partial x_i} \right)^2 u(x_i)^2 + \sum \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} Cov(x_i, x_j) \qquad [3]$$

The $Cov(x_i,x_j)$ of component $(x_i,\ x_j)$ is estimated by the following equation:

$$Cov(x_i, x_j) = \frac{\sum (x_i - \overline{x}_i)(x_j - \overline{x}_j)}{n-1} = r_{xi,xj} u(x_i) u(x_j) \qquad [4]$$

where,

$r_{xi,xj}$ = correlation coefficient which is a normalized covariance, showing the

extent of linear relationship,

and,

$u(xi)$ and $u(xj)$ are the standard uncertainties of x_i and x_j in the form of standard deviation.

So, when do we need to consider the involvement of covariance in the above sample statistic?

In fact, there are certain types of laboratory analysis which finally give rise to multiple measurands, i.e. to a set of results of individual related analyte components in a single analysis, of which a total of its components is to be presented in the final report, such as the analysis of total aflatoxins in feeds in the form of B1, B2, G1 and G2 by HPLC, the analysis of total PCBs (Polychlorinated biphenyls) and total dioxins/furans in a chemically contaminated soil sample and the analysis of gaseous hydrocarbons in a natural gas mixture by gas chromatography.

In these examples, the combined standard uncertainty of its total is not just a simple sum of the individual variance of its component analytes but also involves interactions of variances between the components. This point can be illustrated in the following example.

Let's take a look at a set of aflatoxin results in μg/kg presented by 5 laboratories in a collaborative study as summarized in the table below:

Lab #	B1	B2	G1	G2	Total Aflatoxins
1	6.5	4.3	2.5	2.4	15.7
2	4.4	2.5	2.7	1.8	11.4
3	6.1	3.6	3.1	2.5	15.3
4	5.9	3.4	2.3	2.2	13.8
5	4.6	2.9	1.9	2.3	11.7
Mean	5.500	3.340	2.500	2.240	
Std Dev, u	0.941	0.688	0.447	0.270	1.987
Variance	0.885	0.473	0.200	0.073	3.947

It is noted that the sum of variances of B1, B2, G1 and G2 = (0.885+0.473+0.200+0.073) = 1.631 was very much smaller than the estimated variance of the total aflatoxins (3.947), which is a square of standard deviation of the total aflatoxins reported by the 5 laboratories. This observation highlighted a point that there were inter-relationship amongst the four aflatoxin components during the HPLC analysis.

The covariances between the pairs of aflatoxin components are calculated and summarized in the following table:

	B1	B2	G1	G2
B1	0.885	0.615	0.150	0.183
B2	0.615	0.473	0.060	0.141
G1	0.150	0.060	0.200	0.010
G2	0.183	0.141	0.010	0.073

Notice that instead of using the statistical equation [4] to calculate these covariance, we can also make use of the Microsoft Excel ® function "=COVARIANCE.S(array1,array2)" where arrays 1 and 2 are the two laboratory results of the aflatoxin components reported.

From the equation [4], we can also calculate their respective correlation coefficients $r_{xi,xj}$ as summarized below:

	B1	B2	G1	G2
B1	1.000	**0.951**	0.357	**0.718**
B2	**0.951**	1.000	0.195	**0.756**
G1	0.357	0.195	1.000	0.083
G2	**0.718**	**0.756**	0.083	1.000

When there is no correlation we assume that the analysis data are randomly related to one another。If we arbitrarily set $r_{xi,xj} > 0.5$ to be significant in substantially affecting the combined standard deviation, it is obvious that B1, B2 and G2 results had some form of correlation between any pair of them. This is confirmed when we apply the sample statistic equation [3] to re-estimate the total standard uncertainty expressed as standard deviation with covariance factors considered. The final combined standard uncertainty = 1.987 is exactly the same as the standard deviation calculated from the total alfatoxins reported by the 5 laboratories.

Furthermore, it is noted that:

if the total standard deviation of each component calculated *without* considering the covariance fictitiously is equal to 1.80, and the standard deviation of the covariance factor fictitiously is equal to 0.60, the combined standard uncertainty with covariance considered = $\sqrt{(1.80^2 + 0.60^2)}$= 1.90.

Now, the error by omitting the covariance is equal to (1.90 – 1.80) or 0.10 and the relative error % committed by ignoring this covariance factor of 0.60 is actually equal to (0.10x100/1.90) or 5.1%. Such level of error is statistically acceptable with 95% confidence and hence we can 'safely' ignore the contribution of that covariance which is about 1/3 of the other combined variance.