*Memorandum*

**Behavioural and Societal Sciences**
Van Mourik Broekmanweg 6
2628 XE Delft
P.O. Box 49
2600 AA Delft
The Netherlands

www.tno.nl

T +31 88 866 30 00
F +31 88 866 30 10

**To**
Andre Rijnders

**From**
N.E. Ligterink

**Copy to**
Klaus Steininger
Rob Gardner
Alain Petit

**Subject**
COP test sequences

**Date**
10 March 2015

**Our reference**
<vnr-ext>

**Direct dialling**
+31 88 866 80 58



## Introduction

Manufacturers have their own independent quality control systems. However, for the testing of the conformity of production (COP) by the type-approval authorities a reference method is developed. This protocol must ensure a limited test burden and the average emission below the limit with a reasonable confidence.

## Protocol

The formal quality-control sequences have limitations in the large number of tests needed to arrive at a pass-fail decision in a critical case. Instead some prior requirements are set to arrive at a COP procedure which have a limited test burden, and focusses on the average emissions below the emission limit. The following requirements were agreed prior to the design of the protocol:

1. The average emissions should be below the emission limit
2. The measured emissions on the individual tests should be used in the approach (a distinction between small and large exceedances is to be made)
3. The minimal number of tests is three
4. The maximal number of tests is sixteen
5. The risk of unwarranted passes or failures should be more on the producer side than the consumer side

Based on these requirements are procedure was developed based on the progressing average of the tests and the variance in the test results as an estimate of the confidence.

The true average $X_{true}$ is not known. The test average $X_{tests}$ can be determined from the separate test values $x_i$:

$$X_{tests} = (x_1 + x_2 + x_3 + \ldots + x_N)/N$$

Moreover the variance VAR in the test results are an indication how how spread there is in the emission results, and the confidence in the $X_{test}$ average as an indication of the true average $X_{true}$:

$$VAR = ((x_1 - X_{tests})^2 + (x_2 - X_{tests})^2 + \dots + (x_N - X_{tests})^2)/(N-1)$$

The spread in the data is $(VAR)^{1/2}$, which corresponds to the standard deviation: $\sigma$, for a large sample of a normal distributed dataset. Hence, it is an indication of the accuracy with which $X_{tests}$ represents $X_{true}$.

Starting with three tests, the test average $X_{tests}$ is still a poor representative of $X_{true}$. If the margin is set by the spread in the data: $\sigma = (VAR)^{1/2}$ this will provide an measure for the accuracy with which the average is determined. It has however a problem for some distributions of test results. The square-root functional form may yield a margin above the limit, while all tests are below the limit, as pointed out by Alain Petit of Renault. Since the test results should typically be bounded in the domain from 0 to the limit value, the change from the square root to the normalized linear function solves such problems:

$$S = \sqrt{VAR} \to \frac{VAR}{limit}$$

It will ensure that in the case of all the test results are below the limit, the margin will not be above the limit:

$$X_{test} + \frac{VAR}{limit} \leq limit \quad \text{(if all the test values are below the limit)}$$

If the average of three tests (*N=3*) is above the limit, a failure decision is reached. If the average of all tests have a margin given by the spread below the limit, the pass decision is reached. For values in between additional tests are needed.

Average of all tests: $X = \frac{1}{N}\sum_{i=1}^{N} x_i$

Square of the spread in the test results: $VAR = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - X)^2$

Start with three tests (N=3) and continue till a decision is reached:

- Pass decision if:

$$X < 1.05 \cdot limit - \frac{VAR}{limit}$$

- Fail decision if:

$$X \geq 1.05 \cdot limit - \left(\frac{N-3}{13}\right) * \frac{VAR}{limit}$$

The factor “(N-3)/13” ensures that the pass and fail decisions are conclusive for N=16, while the strongest criterion is used for a fail decision at N=3, which slackens as a decision is not reached after a number of tests. The use of a linear interpolation between the criterion of “$X_{average}$ > limit” at N=3 and “pass-limit = fail-limit” at N=16 is for simplicity.

Without underlying assumptions about the probability distribution of test results, there is very little to say about the confidence with which the decision is reached. For example, if 1 in 100 tests will give a hundredfold emission value, the average is doubled thereby, with a small likelihood that the high emission is encountered in the COP testing. If, instead, a normal distribution is assumed, the Student t-test can be applied, however, in  the t-test approach the lower limit of 3 tests and the upper limit of 16 tests cannot be combined in a single set of producer risk and consumer risk with associated confidence levels.

This approach does not rely on the assumption of a underlying probability distribution. Furthermore, it depends only weakly on the underlying distribution of the data. At a later stage a factor of "1.05" of the limit value is introduced to allow for measurement uncertainties, etc..

## $CO_2$ values

For $CO_2$ emission there is no emission limit, but a declared or averaged measured value ($CO_2^{declared}$). This value may vary within a $CO_2$ family, with the optional weight and body. However, by using the normalized values for $CO_2$ the same COP procedure can be applied:

$$x_i = \frac{CO_2^{test-i}}{CO_2^{declared}}$$

## Simulation results

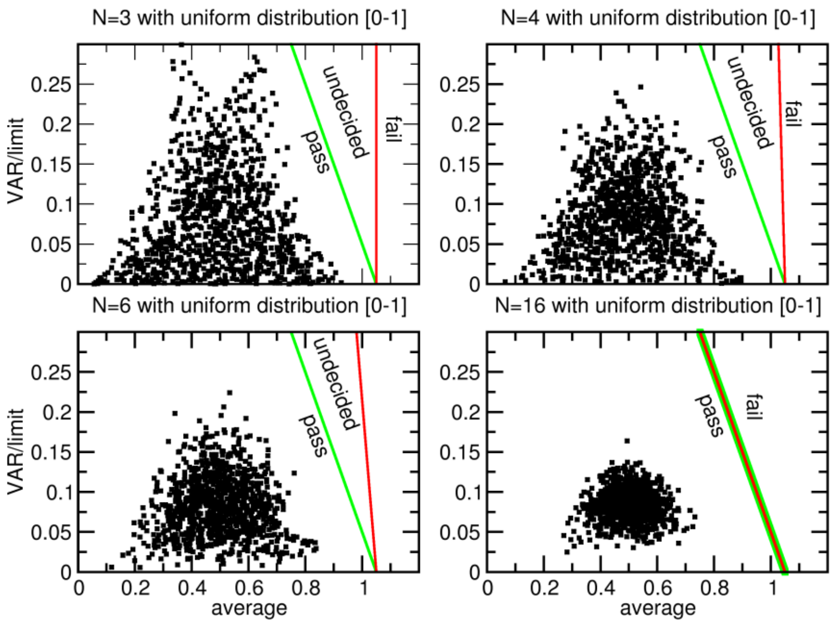


**Figure 1  Simulations with 1000 runs for a uniform distribution for  N=3, 4, 6 and 16 tests. The pass decision remains the same based on average and variance, the fail decision moves towards the pass decision as the spread decreases.**

In order to show how the procedure works in practice and that test values below the limit will not lead to fail decisions, simulations are carried out with an uniform distributed between 0 and 1. Basically the average and the margin (variance/limit) are determined for N tests, randomly thousand times repeated. In the case N=3 the spread is the largest over the 1000 cases, while for N=16 the spread is the smallest, as the result start to converge to the average of ½, and a spread of 0.08.

The use of an uniform distribution for the simulations have little restriction on the applicability of the outcome: The Central Limit Theorem ensures that for large number of tests the distribution of the average converges to a normal distribution. Hence, starting with a normal distribution will lead to similar results for this approach. It should be noted that the pass decision for different averages and variances have the same distance to the outer edge of the distribution of data points.

## Conclusions

The 5% margin of the average above the limit is set for pollutants in the pass and fail decisions. It is not likely such a margin is applicable for $CO_2$ values. Moreover, the procedure so far does not describe the vehicle for COP testing. For example, the maximal odometer mileage the selection of vehicles and the minimum checks by the authority should be included in the procedure.