

Precision of Olfactometry and Odor Testing Results

Authored by:

Michael A. McGinley, P.E.
St. Croix Sensory, Inc.

And

Charles M. McGinley, P.E.
St. Croix Sensory, Inc.

Presented at
Water Environment Federation / Air & Waste Management Association
Specialty Conference: Odors and Air Emissions 2006
Hartford, CT: 9-12 April 2006

Copyright © 2006



St. Croix Sensory Inc.
P.O. Box 313, 3549 Lake Elmo Ave. N.
Lake Elmo, MN 55042 U.S.A.
800-879-9231
stcroix@fivesenses.com

PRECISION OF OLFACTOMETRY AND ODOR TESTING RESULTS

Michael A. McGinley, P.E. and Charles M. McGinley, P.E.

St. Croix Sensory, Inc.
3549 Lake Elmo Ave. N.
P.O. Box 313
Lake Elmo, MN 55042

ABSTRACT

Olfactometry is a precise testing method. How precise depends on the quality assurance / quality control (QA/QC) statistics associated with the laboratory's results. The European Olfactometry testing standard, EN13725:2003, provides a formalized method for monitoring the performance of panel members (assessors) and test results. This is accomplished through testing with the standard odorant 1-butanol (n-butanol).

Olfactometry precision can be understood through four different variance values: 1. specific panel variance, 2. within panel variance, 3. inter-panel variance, and 4. inter-laboratory variance.

Inter-laboratory studies have shown standard deviation of log threshold values as high as 0.30 and individual laboratories have shown inter-panel standard deviations of 0.10 and within panel standard deviations as low as 0.05. This paper defines the four types of variances and shows how the precision of odor panel results can be expressed as analogous to a group of assessors determining sound level within ± 1 dB in 95% of the cases.

This paper will present these different measures of precision with test data and examples of applications of how this information can be utilized by various stakeholders.

KEYWORDS

Olfactometry, odor testing, odor, olfactometer, EN13725

INTRODUCTION

Olfactometry standard EN13725 provides a framework for monitoring both the performance of the individual assessors and the odor panel as a whole. Measurement and documentation of performance allows a laboratory to clearly define the precision of their results.

The perception of odors is a logarithmic phenomenon [Stevens 1960]. An understanding of the logarithm scale of odors provides a basis for understanding the true precision of olfactometry. Olfactometry testing involves presenting the odorous air sample to assessors following an ascending series of concentrations in a geometric progression, i.e. each dilution level is twice the concentration (one half the dilution ratio) as the previous level.

This testing scale does not have an equal spread between values, and thus introduces the need for the use of logarithm base 10 transformations when dealing with any statistical review of odor evaluation data [Dravnieks 1968; Stevens 1962]. More specifically, the transformations are performed in order to stabilize (make uniform) the variance. With the uniform variance, the linear transformed data will show symmetry around the group average (panel average result in log base 10). However, this data will be

asymmetrical around the reported odor concentration value (odor unit value) of detection and recognition thresholds. All statistical calculations, which are based on a normal distribution, therefore, must be conducted with the transformed values, in this case, the logarithm base 10 values.

DEFINING ACCURACY AND PRECISION IN EN13725

EN13725 defines accuracy and precision and how they relate to olfactometry. Accuracy is the closeness of a value to the true or correct result. Accuracy is a function of both bias and precision. Bias is a measure of the systemic errors in a result. Precision measures the scatter in data.

When olfactometry is used to quantify environmental odors, the true result is not known. Thus the accuracy of a specific sample cannot be directly defined. Without a known true value, bias also cannot be estimated (Berthouex et. al., 1994). Precision, on the other hand, can be measured and defined by a laboratory based on the variance (standard deviation) of repeated measurements of the same sample.

Olfactometry Variance

When analyzing olfactometry data there are four different variances to understand and consider:

1. Specific panel variance: This is the variance between the results of the assessors in one test session. This is usually expressed as the standard deviation of the logarithms of the individual threshold estimate (ITE) values. When conducting odor panels following EN13725, this standard deviation usually ranges between 0.00 and 0.30. However, this is not an appropriate value to define the variance of the threshold result. The final threshold value result (detection or recognition threshold) can only be defined by the average of the group's individual thresholds, thus any measure of precision of olfactometry must compare the variance of multiple determinations of this group average and not the variance among the assessors.
2. Within panel variance: This is the variance that defines the repeatability of one panel of assessors evaluating the same sample multiple times on the same test day. This value is usually the smallest test variance since the variability associated with different assessors and different test periods are removed.
3. Inter-panel variance: This variance defines the reproducibility of one sample tested in one laboratory during different lab sessions with potentially different panels of assessors. This value is determined for a laboratory by using n-butanol as a standard test odorant. A test sample of n-butanol is evaluated on a regular basis. The laboratory's inter-panel average n-butanol threshold and variance of this result can be determined (see EN13725, Section 5.3.2). This variance for n-butanol is used to estimate the inter-panel variance for all other environmental odor testing results. EN13725 requires the standard deviation between tests (inter-panel variance) to be less than 0.172 (CEN, 2003).
4. Inter-laboratory variance: This is the variance of one sample (usually a standard odorant) evaluated by multiple laboratories. This variance may be much higher since it will take into account all components of variability between laboratories, e.g. sample of assessors, olfactometer design, quality performance of laboratory, etc. It is not uncommon for the inter-lab standard deviation to be as high as 0.30 of the threshold log values (van Harreveld, 1997; Bardsley, 2003; Oxbol, 2004; Maxeiner, 2003, 2006).

Performance Requirements of Olfactometry Standard EN13725

The internationally accepted olfactometry standard, EN13725, contains detailed quality assurance and quality control requirements to be met by the odor testing laboratory. These requirements include testing the performance of the olfactometer (calibration), qualifying and continuously monitoring assessors, and performance testing of the laboratory as a whole through proficiency testing. The later two items will be discussed further in this paper.

Assessor Selection and Monitoring

In EN13725, Section 6.7.2, *Selection of assessors on individual variability and sensitivity*, outlines the requirements for selection and monitoring of assessors. These elements involve testing with the standard odorant 1-butanol (n-butanol). The laboratory must determine the butanol threshold for potential assessors through at least ten tests collected on three separate days with at least one day between sessions. To become an assessor, the individual must meet the following criteria:

1. The antilog of the standard deviation calculated from the logarithms of the individual threshold estimates, expressed in mass concentration units, must be less than 2.3.
2. The geometric mean of the individual threshold estimates expressed in mass concentration units of the reference gas must be in the range of 20 to 80-ppb.

Individual laboratories may also determine additional specific requirement for being an assessor, e.g. non-smoker.

Once an individual is accepted as an assessor, their butanol threshold must be continuously monitored through testing at least once every twelve real odor samples for which they participate in evaluating. The assessor must continue to meet the above criteria through continual review of their most recent twenty tests.

These assessor selection and monitoring requirements have standardized the sensor used for olfactometry. This narrows the band of the general population that can be an assessor, and thus reduces the variability of results within laboratories (inter-panel variance), and more importantly between laboratories (inter-laboratory variance) (van Harreveld, 1999).

Laboratory Performance Testing

Section 5.3 in EN13725 provides requirements for testing the performance of an odor laboratory. This performance testing is based on the n-butanol standard odorant. The assumption is made that the laboratory performance to the standard odorant can be transferred to all odors tested by the laboratory.

Criteria for accuracy and precision are used to define the performance of the laboratory. Since environmental odor samples do not have a known actual or true value, accuracy can only be determined with standard reference odors or through inter-laboratory studies with a reasonable sample size to define the true value through the results of the study (CEN, 2003). For EN13725, the accuracy of a lab can be determined from testing with the n-butanol reference and the calculated value must meet the criterion defined in the standard ($A_{od} \leq 0.217$).

The precision criterion is expressed in terms of repeatability and is defined in EN13725 as being less than 0.477 ($r \leq 0.477$). Repeatability of non-reference odors is calculated using the equation:

$$r = t * \sqrt{2} * s_r$$

Where t is the student t distribution for the 95% confidence level, and s_r is the standard deviation of the test measurements. With a student t value of 1.96 ($n=\infty$) for $\alpha=0.05$ (95% confidence), the limit for repeatability refers to a laboratory's inter-panel standard deviation, s_r , equal to 0.172.

This can also be expressed through transformation as $10^r = 3.0$. This means that in the laboratory, the result from any two consecutive measurements will not be larger than three (3.0) 95% of the time.

The assumption that these performance criteria, determined from testing with butanol, are transferable to all odors must be tested by a laboratory participating in inter-laboratory studies/programs or by conducting intra-laboratory (inter-panel) testing of real environmental odor samples or reference compounds (e.g. hydrogen sulfide). This testing allows a laboratory to clearly define the precision of their non-reference odor testing results.

AN EXAMPLE QA/QC TESTING AND MONITORING PROGRAM

Within Panel Variability

St. Croix Sensory conducts periodic within panel repeatability tests by presenting a single panel of assessors with the same environmental odor sample or standard odorant multiple times during one test session. St. Croix Sensory has determined that its within panel variance can be described by using a standard deviation of 0.05 for the log dilution ratio testing result.

By definition, ninety-five percent of all values should lie between \pm two times the standard deviation (where 2 is the estimated student t -value estimate for $n=\infty$). For St. Croix Sensory this would be $2*0.05=0.10$ or mathematically analogous to the same group of people repeatedly determining sound level in a room within ± 1.0 dB.

Inter-Panel Variability

St. Croix Sensory evaluates an n -butanol test sample during every odor session conducted following olfactometry standard EN13725. This test sample is performed in accordance with Section 5.3 of this standard. The n -butanol threshold of the test panel is recorded in an active quality assurance monitoring system. The panel result is confirmed to range between 20-80 ppb as specified in EN13725. If the panel result is not in this range, then immediate action is taken to correct the performance deviation by confirming the test equipment operation or replacing a non-conforming assessor. The results of this testing provides a measure of the inter-panel variance within the laboratory.

Through 2005, St. Croix Sensory determined its inter-panel variance to be defined with a standard deviation of 0.10 ($n=50$, rolling average) for n -butanol threshold testing. This butanol testing has also shown accuracy, as defined by EN13725, to be 0.08. These values are lower than the EN13725 allowable maximum standard deviation of 0.172 and maximum accuracy value of 0.217.

As shown before, for St. Croix Sensory this measure of precision is mathematically analogous to two or more groups of people repeatedly determining sound level in a room within ± 2.0 dB.

VALIDATION OF USE OF STANDARD ODORANT FOR ESTIMATE OF PRECISION

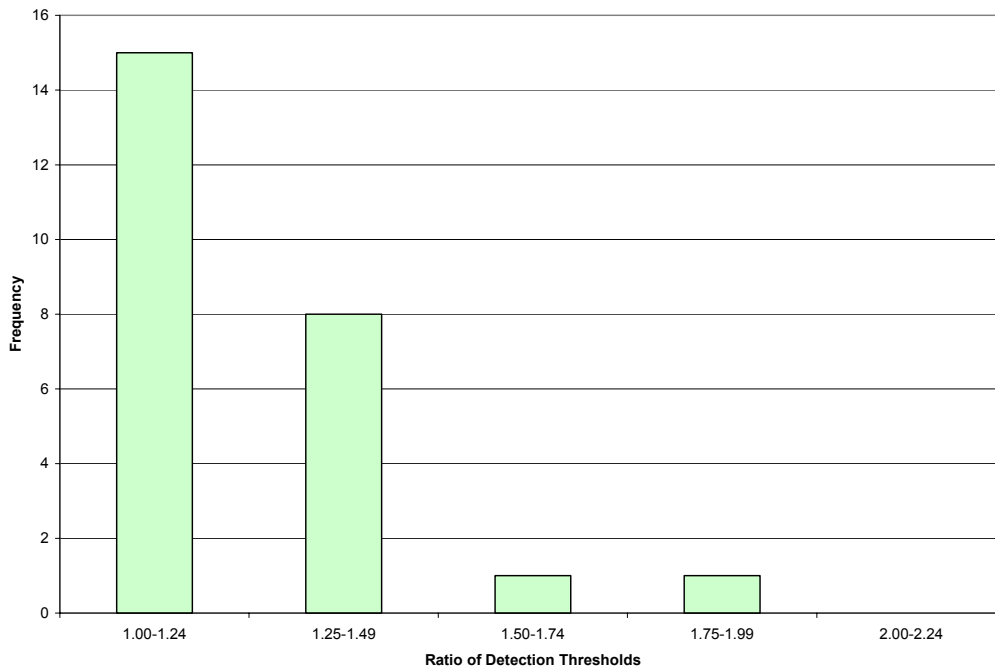
As part of a comprehensive QA/QC program, St. Croix Sensory periodically tests random environmental odor samples multiple times with the same panel of assessors and multiple times with different panels of assessors.

In 2005, St. Croix Sensory tested 27 odorous air samples 2-5 times with the same panel of assessors during the same testing session. The average detection threshold values of all samples ranged from 60 to 5,000. The standard deviations of all samples ranged from 0.00 to 0.11 with an average standard deviation of 0.05. When samples were evaluated by the panel three or more times, the standard deviation ranged from 0.00 to 0.08 with an average of 0.05.

One sample of 5-ppm hydrogen sulfide was evaluated four times as part of this testing program. The average detection threshold was 5,100 with a log standard deviation of 0.064. These results provided an average hydrogen sulfide threshold of 0.98-ppb.

In 2005, St. Croix Sensory tested 25 odorous air samples with two different panels of assessors in our laboratory on the same day. These tests were all conducted following EN13725 using one of three AC'SCENT International Olfactometers. The detection thresholds of the 25 test samples ranged from 60 to 5,000. The ratio of detection thresholds from the two panels ranged from 1.03 to 1.77. Figure 1 is a histogram summarizing all the results of the detection thresholds ratios. The log standard deviation of the two tests ranged from 0.010 to 0.176 with an average of 0.060. These results are consistent with the estimated inter-panel standard deviation of 0.10 determined through daily testing with n-butanol.

Figure 1. Histogram of ratios of odor detection threshold results (n=25) determined by multiple odor panels within the same laboratory following EN13725.



APPLICATIONS OF OLFACTOMETRY PRECISION

One must consider the ultimate use of the testing results when determining which type of variability correctly reflects the precision of the olfactometry data (Oxbol, 2004).

If the goal is to determine the threshold concentration of a pure compound or mixture, it would be most appropriate to use a variance common with inter-laboratory testing. This would take into consideration all variability between all laboratories (i.e. olfactometer design, specific laboratory procedures, etc.).

When the results of odor panel testing will be referenced over many months or years using the same odor evaluation laboratory, these threshold values should be considered with the inter-panel variability. This would include the variability between panel groups within the one laboratory.

By contrast, the within panel variability may be used when conducting testing which compares several sources during one sampling event. For example, when performing odor control efficiency testing, within panel variability correctly describes the data from the inlet and outlet (untreated and treated) samples.

95% confidence intervals

A 95% confidence interval can be determined for any odor testing result. As shown in Annex G in EN13725 (“Example of the calculation used to determine the number of odor concentration measurements required to achieve a defined precision”), the 95% confidence interval can be produced with the variability information defined by the odor laboratory. Annex G shows the confidence intervals for an example laboratory that just meets the quality criterion for precision, which is standard deviation of 0.172 determined from testing with n-butanol. This data is reproduced in the Table 1.

Table 1. Example results for one odor sample tested up to ten times (n=10) at an odor laboratory that just meets the quality criterion for precision in EN13725 (std. Dev.= 0.172). Each row of the table represents a summary of n results. The highlighted row shows odor testing results of n=3 samples (triplicate) with an average detection threshold of, m, 1000.

n	LCL	m	UCL	LCL	m	UCL
	Log(o.u./m3)	Log(o.u./m3)	Log(o.u./m3)	o.u./m3	o.u./m3	o.u./m3
1	2.656	3.000	3.344	453	1000	2209
2	2.757	3.000	3.243	571	1000	1751
3	2.801	3.000	3.199	633	1000	1580
4	2.828	3.000	3.172	673	1000	1486
5	2.846	3.000	3.154	702	1000	1425
6	2.859	3.000	3.141	724	1000	1382
7	2.870	3.000	3.130	741	1000	1349
8	2.878	3.000	3.122	756	1000	1323
9	2.885	3.000	3.115	768	1000	1302
10	2.891	3.000	3.109	778	1000	1285

Table 2 displays how the confidence intervals become narrower when the actual laboratory determined value of precision is used. Table 3 shows how these confidence intervals become even more narrow when the within panel variability of a laboratory is used.

Table 2. Example results for one odor sample tested up to ten times (n=10) at one odor laboratory. Each row of the table represents a summary of n results with the confidence interval based on the laboratory's determined inter-panel variability (std. dev. = 0.10). The highlighted row shows odor testing results of n=3 samples (triplicate) with an average detection threshold, m, of 1000.

n	LCL	m	UCL	LCL	m	UCL
	Log(o.u./m3)	Log(o.u./m3)	Log(o.u./m3)	o.u./m3	o.u./m3	o.u./m3
1	2.800	3.000	3.200	631	1000	1585
2	2.859	3.000	3.141	722	1000	1385
3	2.885	3.000	3.115	767	1000	1305
4	2.900	3.000	3.100	794	1000	1259
5	2.911	3.000	3.089	814	1000	1229
6	2.918	3.000	3.082	829	1000	1207
7	2.924	3.000	3.076	840	1000	1190
8	2.929	3.000	3.071	850	1000	1177
9	2.933	3.000	3.067	858	1000	1166
10	2.937	3.000	3.063	864	1000	1157

Table 3. Example results for one odor sample tested up to ten times (n=10) at one odor laboratory with the same odor panel. Each row of the table represents a summary of n results with the confidence interval based on the laboratory's determined within panel variability (std. dev. = 0.05). The highlighted row shows odor testing results of n=3 samples (triplicate) with an average detection threshold, m, of 1000.

n	LCL	m	UCL	LCL	m	UCL
	Log(o.u./m3)	Log(o.u./m3)	Log(o.u./m3)	o.u./m3	o.u./m3	o.u./m3
1	2.900	3.000	3.100	794	1000	1259
2	2.929	3.000	3.071	850	1000	1177
3	2.942	3.000	3.058	876	1000	1142
4	2.950	3.000	3.050	891	1000	1122
5	2.955	3.000	3.045	902	1000	1108
6	2.959	3.000	3.041	910	1000	1099
7	2.962	3.000	3.038	917	1000	1091
8	2.965	3.000	3.035	922	1000	1085
9	2.967	3.000	3.033	926	1000	1080
10	2.968	3.000	3.032	930	1000	1076

As an example, consider an engineer who goes to a facility and collects samples in triplicate (true triplicate samples collected at exactly the same time or collected from a source with confirmed stable conditions). The average detection threshold result reported by the laboratory was 1000. If the laboratory only just met the precision criterion, $s_r=0.172$, then the 95% confidence interval would range from 633 to 1580. Utilizing St. Croix Sensory's determination of inter-panel precision, $s_r=0.10$, this 95% confidence interval is reduced to a range of 767 to 1305. If the sampling event warranted the use of the within panel value of precision, $s_r=0.05$, then this would reduce the 95% confidence interval further to 876 to 1142.

The 95% confidence interval used depends on the application of the testing as described previously.

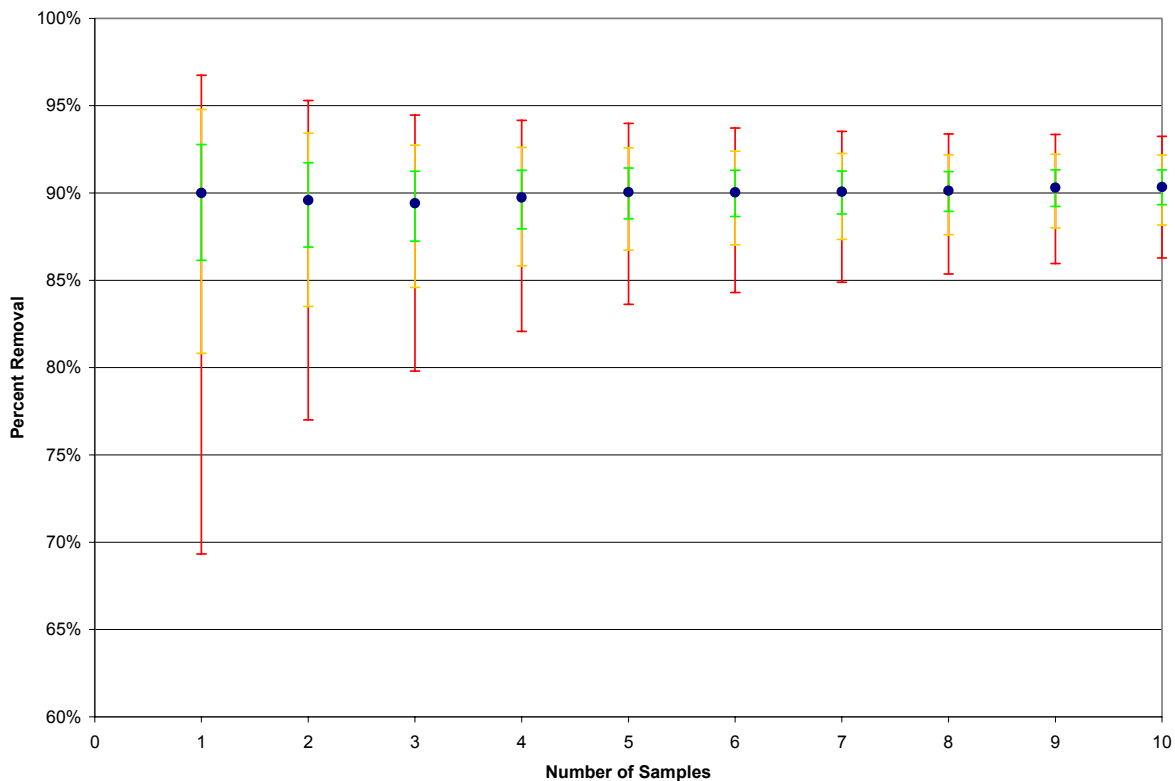
Odor Control Performance Testing

The 95% confidence interval can also be applied to odor control performance testing. Annex H in EN13725 (“Example of the calculation used to determine the number of odor concentration measurements required to detect a difference between means”) outlines the calculations for determining the confidence interval related to odor reduction efficiency. Annex H provides an example, which was again computed using the EN13725 quality criterion for precision, standard deviation of 0.172.

The confidence interval of efficiency is reduced if the specific laboratory’s precision (inter-panel variability) is defined as better than the EN13725 criteria. The confidence interval is usually even narrower if conditions allow for the use of the specific laboratory’s within panel variance.

Figure 2 is a graph which displays the results using all three forms of variability. This graph shows how multiple samples (increasing sample size, n) can provide smaller and smaller confidence intervals. The graph also shows the size of the intervals depending on the specific variance of the samples.

Figure 2. Comparison of 95% confidence intervals for 1-10 repeated odor control performance tests. The red bands (widest) represent the 95% confidence interval based on the EN13725 criteria for precision ($S_r=0.172$). The yellow bands (middle) represent the 95% confidence interval based on common inter-panel precision ($S_r=0.10$). The green bands (narrow) represent the 95% confidence interval based on common within panel precision ($S_r=0.05$).



Performance testing is often conducted as sets of samples (inlet and outlet) evaluated by the same odor panel of assessors. This fact makes it reasonable to use the within panel variability when determining the confidence interval of odor reduction efficiency. For the example, if we consider three sets of inlets and outlets provided in the results above, this would allow the use of a standard deviation on the order of 0.05, which yields a 95% confidence interval of 87.2% to 91.2%. The higher inter-panel standard deviation would yield a slightly wider interval of 84.6% to 92.7%.

CONCLUSIONS

This paper outlines and discusses the following important concepts.

1. Accuracy is the closeness of a measured value to a true or correct result.
2. Accuracy of olfactometry results cannot be determined with environmental odor samples since the true or correct value is not known.
3. Accuracy of olfactometry results can be determined for standard odorants with agreed upon reference values (e.g. 1-butanol) or through inter-laboratory programs with a sufficient number of participating labs to provide a determination of the true result.
4. Precision is a measure of the scatter (variability) of multiple results under similar testing conditions. The standard deviation of multiple results is used to define precision.
5. Precision of olfactometry results can be quantified through regular testing of reference odorants (e.g. 1-butanol) and verified through testing of environmental odor samples with the same panel of assessors (within panel precision) or with multiple panels (inter-panel precision) in the same laboratory.
6. Testing has shown that olfactometry laboratories provide precise results that are mathematically analogous to a group of assessors determining the sound level to within ± 1 to 2 dB.
7. A detailed QA/QC program is an essential component to operating an effective olfactometry laboratory.

Examples provided in the paper show how these measures of accuracy and precision can be used to provide more information about odor testing results. Depending on the specific sampling and laboratory protocols, different levels of precision may be used for computing confidence intervals and performing statistical comparisons. For example, if a facility odor study requires many samples sent to a laboratory over many days, the precision of the results would be defined by the specific laboratory's defined inter-panel standard deviation. On the other hand, if a facility collected triplicate odor control inlet and outlet samples for performance testing, these samples would be evaluated by one odor panel, thus the within panel standard deviation could correctly define the precision of the result.

EN13725 provides a framework for monitoring accuracy and precision of results in an olfactometry laboratory. Regular testing of the reference odorant 1-butanol (n-butanol) along with participation in inter-laboratory programs allows a laboratory to quantify these parameters. Periodic testing can be performed on environmental odor samples to confirm these parameters as part of a comprehensive QA/QC program. This statistical information provides users of odor testing results with a greater understanding of the meaning and application of these results.

REFERENCES

- Bardsley, T.; Demetriou, J.(2003) *Interlaboratory Odour Study Conducted with EPA Approved Method*; Publication SR2; EPA Victoria: Victoria, Australia.
- Berthouex, P.M.; Brown, L.C. (1994) *Statistics for Environmental Engineers*; CRC Press.
- CEN - Committee for European Normalization (2003) *Air Quality – Determination of odour concentration by dynamic olfactometry*; EN13725:2003; Brussels, Belgium. 2003.

- Dravnieks, A.; Schmidtsdorf, W.; M. Meilgaard (1986) Odor Thresholds by Forced-Choice Dynamic Triangle Olfactometry: Reproducibility and Methods of Calculation. *Journal of the Air Pollution Control Association*, **36**, 900-905.
- Maxeiner, B.; Mannebeck, D (2004) Round Robin Test of Olfactometry 2003; *Proceedings of the VDI Conference on Environmental Odour Management*; Köln, Germany; 17-19 November 2004; VDI-Berichte Nr 1850. 137-151.
- Maxeiner, B. (2006) Olfactometric Interlaboratory Comparison Test 2005; *Proceedings of the A&WMA/WEF 2006 Odors and Air Emissions Conference*; 9-12 April 2006; Water Environment Federation: Alexandria, VA
- Ministry of Environment – Japan (2003) *Odor Index Regulation and Triangle Odor Bag Method*; Office of Odor, Noise, and Vibration.
- Oxbol, A. (2004) Uncertainty in Odour Analysis; *Proceedings of the VDI Conference on Environmental Odour Management*; Köln, Germany; 17-19 November 2004; VDI-Berichte Nr 1850. 515-519.
- Stevens, S.S. (1960) The Psychophysics of Sensory Functions; *American Scientist*, **48**, 226-253.
- Stevens, S.S. (1962) The Surprising Simplicity of Sensory Metrics; *American Psychologist*, **17**, 29-39.
- Van Harreveld, A.P. (1997) *Interlaboratory Comparison of Olfactometry – Validation of draft CEN Standard ‘Odour concentration measurement using dynamic olfactometry,’ by working group CEN/TC264/WG2*; Document N220 submitted to CEN Technical Committee 264, Working Group 2; Project Research Environmental Consultants, Ltd.: Amsterdam, The Netherlands.
- Van Harreveld, A.P. (1999) A Review of 20 years of Standardization of Odor Concentration Measurement by Dynamic Olfactometry in Europe; *Journal of Air & Waste Management Association*, **49**, 705-715.