Odour Threshold Determination: Understanding the Assessor Response

Authored by:

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

And

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

Presented at
International Water Association
Conference on Odours and Air Emissions
Warsaw, Poland: 25-27 September 2017

St. Croix Sensory Inc.
St. Croix Sensory, Inc., 1150 Stillwater Blvd. N.
Stillwater, MN 55082, USA
800-879-9231
stcroix@fivesenses.com

Odour Threshold Determination: Understanding the Assessor Response

Michael A. McGinley, P.E., St. Croix Sensory, Inc.
Charles M. McGinley, P.E., St. Croix Sensory, Inc.
St. Croix Sensory, Inc.
1150 Stillwater Blvd. N.
Stillwater, MN 55082, USA

Abstract

Over 10 years ago the European standard test method for determination of odour concentrations made significant advancements in the field of olfactometry. Steps were taken to standardize the olfactometer and, equally important, the assessors. Testing with standard odorant n-butanol allowed laboratories to select assessors and continually monitor their performance with a universally agreed upon metric. This move to standard assessors allowed for running environmental odour threshold tests with a small sample of assessors (n=4, two rounds minimum).

Since its adoption, two key questions are continually raised within the odour testing community: 1) Is only one standard odorant sufficient for assessor selection and performance monitoring? and 2) Is a four assessors sufficient for determination of an odour threshold for environmental odour samples?

A 2009 study conducted by St. Croix Sensory showed assessor responses to a second odorant, hydrogen sulphide, were within similarly acceptable range of sensitivity. A follow-up study has been conducted to compare third and fourth environmental odorants. This study has shown an EN13725 certified panel will detect other odorants similarly. While there was no correlation between individual assessors' or a panel of assessors' n-butanol threshold concentration and the threshold concentration of the other odorants, responses were all within a narrow range of variability. While testing assessors to multiple odorants is good practice and could help in screening and training, this study suggests it may not be imperative. Additionally, a review of detection threshold calculations utilizing n=4 to n=9 assessors did not show a significant difference in testing results.

Keywords

Olfactometry; odour threshold; EN13725; ASTM E679; assessors; n-butanol

INTRODUCTION

Olfactometry Standard EN13725:2003 provides a framework for QA/QC programs in olfactometry laboratories (CEN, 2003). This standard includes requirements for important elements such as physical design of the laboratory, equipment operation, and equipment performance testing. The other principle component of this standard involves selection of the human assessors used to complete the odour observations.

Historically, ASTM E679 and other olfactometry standards recommended the use of assessors with olfactory sensitivity representative of the general population (ASTM, 2004). The expectation was that these responses obtained from eight to twelve assessors assigned to a test session would be similar to the population average.

During the 1980's and 1990's, research in Europe was being conducted as part of development of a universal olfactometry standard. The inter-laboratory studies conducted at that time showed it was impossible to represent the population while meeting agreed upon repeatability criteria with the small sample size of assessors typically convened on odour panels (Hermans, 1989; Heeres, 1990).

Van Harresveld presented a clear conclusion resulting from this study in his 1999 publication in the Journal of the Air & Waste Management Association, "A Review of 20 Years of Standardization of Odor Concentration Measurement by Dynamic Olfactometry in Europe." During the development of the new olfactometry standard, "the notion that the panel should be representative of the general population was explicitly abandoned…" (van Harresveld, 1999). It was determined that the sensor involved in the odour testing, i.e. the human noses, must be standardized.

Further work defined the use of n-butanol as a standard reference odorant for selection of odour panel assessors. Research conducted in Europe during the 1990's, including several inter-laboratory studies, arrived at an agreed upon criteria for n-butanol. It was determined that the n-butanol reference odour threshold value was 40-ppb. An assessor must have a certain sensitivity and repeatability in their responses to tests of this reference odorant. The assessor is required to have an average sensitivity to n-butanol in the range of 20 to 80-ppb, with the standard deviation of the log threshold values less than 2.3.

This selection of assessors with n-butanol creates a standardized sensor; with the intent that the assessors are essentially interchangeable. While one assessor may be slightly more sensitive and another may be slightly less sensitive, the defined range assures that on average, the panel of assessors as an instrument will provide results within a necessary range of accuracy and precision.

The EN13725 olfactometry standard states on Page 18: "The assumption made is that the sensitivity for the reference will be a predictor for sensitivity to other substances" (CEN, 2003). Some researchers have questioned the validity of this statement and have expressed concern for selecting the assessors based on one odorant when other odours are more commonly experienced. For example, a university laboratory conducting research on animal waste may question why they select assessors based on butanol while daily samples are mostly composed of hydrogen sulphide and other sulphur-based compounds.

EN13725 further states that the authors of the standard acknowledge that multiple standard odorants or one mixture of odorants would possibly provide a better standard to measure the assessor selection; however, only n-butanol was determined to be suitable at the time of publication.

St. Croix Sensory, a commercial sensory testing laboratory located in Minnesota (USA), conducts odour evaluations for various consulting firms, sanitation districts, industries, universities, and government agencies. Thousands of environmental air samples per year are evaluated from industries such as wastewater treatment, composting, municipal solid waste, agricultural, and various manufacturing.

As part of a study in 2009, St. Croix Sensory conducted threshold testing of hydrogen sulphide during multiple testing sessions. Results of individual and panel thresholds were compared. Results of this study showed the threshold results for hydrogen sulphide were within a narrow range of sensitivity with an odour detection threshold of 0.4ppb, which was consistent with published threshold values (McGinley, 2010; McGinley, 2008, McGinley, 2003; Van Gemert, 1999). While there was no direct correlation between an individual assessor's or a panel of assessors' n-butanol threshold concentration and hydrogen sulphide threshold concentration, responses were all within a narrow range of sensitivity.

The subject of this paper is to expand on this investigation of assessor sensitivity with two additional environmental odorants. The two volatile organic compounds were studied as part of proprietary contract work; therefore, the compounds cannot be specifically identified. The paper is intended to further consider assessor responses to odorous air samples and the identities of these compounds are not necessary for this discussion.

METHODOLOGIES

As part of a comprehensive QA/QC program, St. Croix Sensory presents assessors with a test sample of standard reference n-butanol during every test session. This allows documentation and tracking of the individual response of each assessor as well as the panel result. For this study, St. Croix Sensory also acquired gas cylinders of two volatile organic compounds as test environmental odorants, which will be referred to in this paper as Compound A and Compound B. All test samples were acquired from Oxygen Services, Inc. (St. Paul, MN). N-butanol used in this study was received as 38.4ppm n-butanol/balance N2, (Cylinder #UV000024). Compound A was received as 5.6ppm/balance N2, (Cylinder #UV000250). Compound B was received as 1%/balance N2, (Cylinder #1704040654).

The gaseous samples of the odorants were prepared by filling 10-L Tedlar air sample bags directly from the cylinders using a stainless steel, two-stage regulator. The dedicated sample bags were flushed before initial use and remained dedicated for tests with the specific odorants. Testing was conducted within 30-minutes of filling the Tedlar air sample bag.

Odour threshold values were determined on an AC'SCENT® International Olfactometer (St. Croix Sensory), with a presentation flow rate of 20-lpm, following dynamic dilution olfactometry standards CEN EN13725:2003, *Air Quality – Determination of Odour Concentration by Dynamic Olfactometry*, and ASTM International E679-04, *Standard Practice for Determination of Odor and Taste Threshold by a Forced-Choice Ascending Concentration Series Method of Limits* (Appendix X.3). The AC'SCENT Olfactometer has 14 dilution levels with an operating range of 60,000 down to 8 dilutions.

For testing with Compound A, in accordance with EN13725, a minimum of four assessors were utilized for each test session, with each assessor observing the sample two times (two rounds). A minimum eight assessor responses were required for a valid result. For Compound B, the protocol for the contract testing required a minimum of 6 assessors and as many as 10 assessors in any panel test session. All results for Compound B were obtained with panels of 7-9 assessors.

RESULTS

Testing of Compound A

The odour threshold of Compound A was determined with 15 testing sessions for 31 threshold tests utilizing 20 individual assessors for a total of 155 individual evaluations. The individual and panel thresholds were recorded for both n-butanol and multiple rounds of Compound A. Overall, the average detection threshold from all samples run was determined to be 4ppb, with a range of 2.3ppb to 5.6ppb.

Figure 1 shows the panel average ppb threshold result for n-butanol on the testing day matched with the corresponding Compound A ppb threshold results. All n-butanol threshold values met the EN13725 requirement of 20-80ppb. If we consider a similar criterion of 1/2x and 2x the threshold value, all threshold values for Compound A were within a 2-8ppb acceptability range.

Figure 2 shows the matched thresholds for each individual assessor. The n-butanol threshold represents the assessor's EN13725 certification result of their ppb threshold to n-butanol from their most recent 20 tests. Compound A results represent an average of at least 5 threshold tests. The individual threshold responses ranged from 1.9ppb to 7.4ppb. Only two assessor results were just below the lower limit of the 2-8ppb acceptability range. All assessors were within the EN13725 criteria of 20-80ppb. The correlation coefficient of log threshold concentration of Compound A to

Figure 1. Results of 31 matched results comparing panel detection threshold values for n-butanol and Compound A. Each point represents the panel average response to both odorants during the same test session.

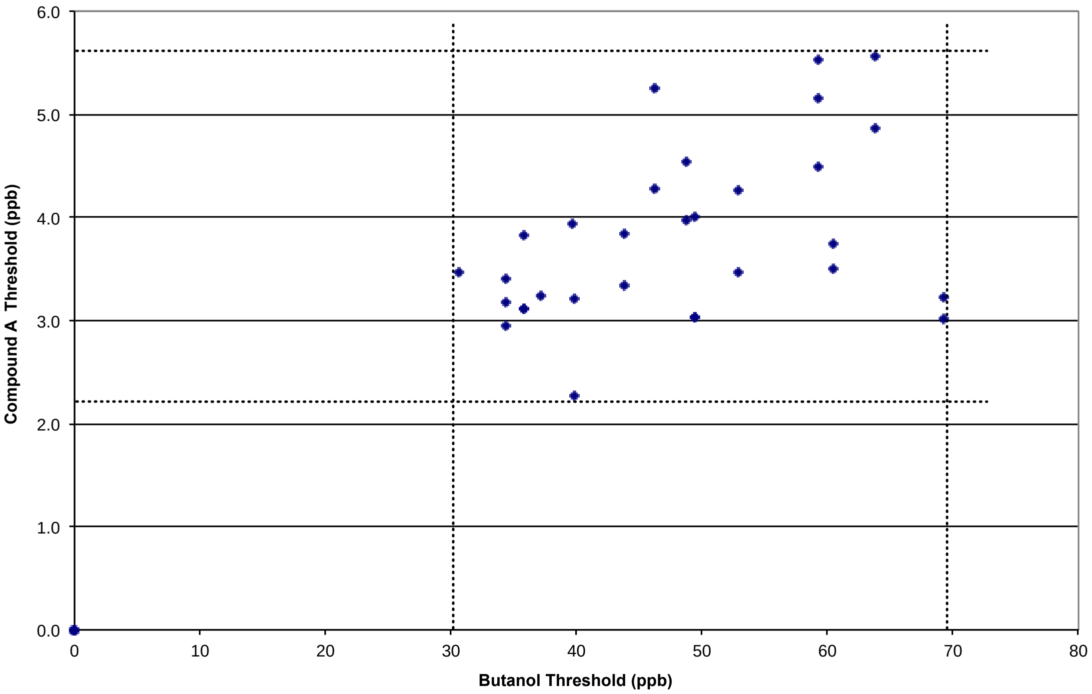
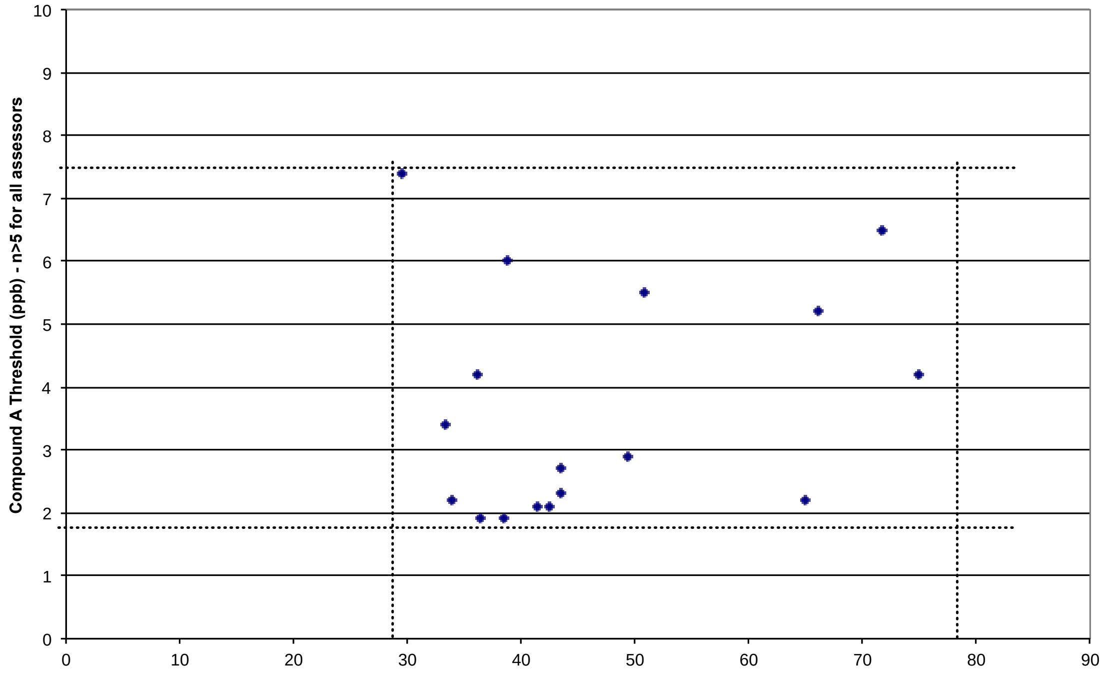
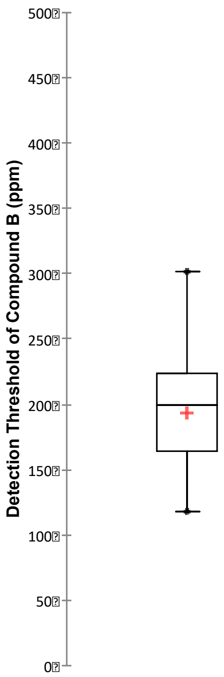


Figure 2. Results comparing assessor average threshold responses for standard odorant, n-butanol, and

Compound A for 17 assessors with more than five Compound A tests.





log threshold concentratio
n-butanol threshold canno
acceptable ranges of varial

Testing of Compound B

The odour threshold of C
utilizing 29 individual ass
The individual and panel
Compound B. Overall, th
to be 194ppm, with a rang

Figure 3 shows box plot
results are based on 7 sam
of 38ppb to 78ppb, which
Compound B shows the
values for all tests. The majority of results are in the range of 165-224ppm.

Figure 3. Box plot diagrams of the panel detection threshold values for n-butanol in ppb (left) and Compound B

in ppm (right).

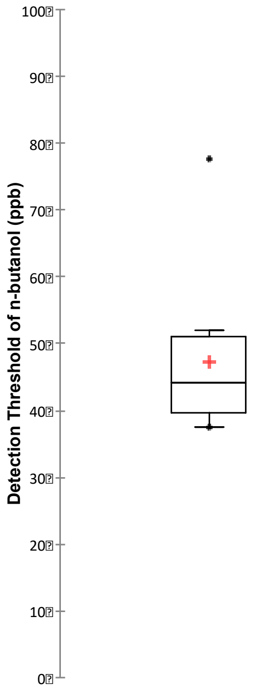
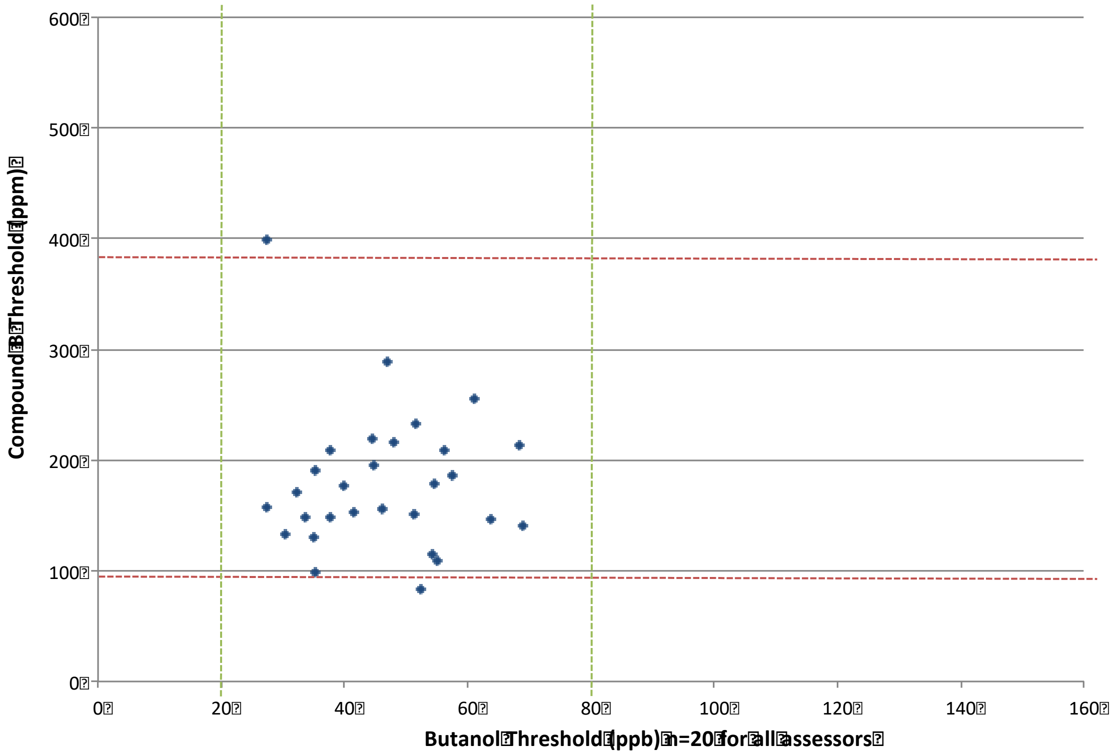


Figure 4 shows the matched threshold results for each individual assessor. The n-butanol threshold represents the assessor's EN13725 certification result of their ppb threshold to n-butanol from their most recent 20 tests. All assessors were within the EN13725 criteria of 20-80ppb. Compound B results represent an average of at least 7 threshold tests. The individual threshold responses ranged from 83ppm to 399ppm. Only two assessor results were just above and below 2x and 1/2x the average threshold. The correlation coefficient of log threshold concentration of Compound B to log threshold concentration of n-butanol shows no correlation (r^2=0.01). So, while n-butanol threshold cannot predict Compound B threshold, all but two assessor responses were within acceptable ranges of variability based on the results of this testing.

Figure 4. Results comparing assessor average threshold responses for standard odorant, n-butanol, and Compound B for 29 assessors. Vertical lines represent the EN13725 limits of 20-80ppb. Horizontal lines represent

1/2x (97ppm) and 2x (388ppm) the panel average threshold of 194ppm.



DISCUSSION

Assessor Performance

For this study, the geometric mean threshold value for Compound A and Compound B were considered to be the "reference value" for these compounds. These values were then utilized to compare the variability of individual assessor responses and panel average responses. As with n-butanol, a factor of 1/2x and 2x for both compounds were considered. Figures 1 and 3 show the panel average results were all within the ranges of 2-8ppb for Compound A and 97-388 for Compound B. Figure 2 shows that of 17 assessors tested to Compound A, 15 of the results are within this acceptability range and 2 had threshold results just below the lower limit (more sensitive). Figure 4 shows that of the 29 assessors tested to Compound B, 27 of the results are within this acceptability range; two assessor had a threshold results above and below the upper and lower bounds, respectively. A target diagram can also be utilized to visually consider the accuracy and precision of the assessor threshold responses. The polar plot shows the log value deviation of the each assessor's average response of the test compound to the reference value. The centre of the "target" would be a response equivalent to the reference value. The polar plot was created with each assessor as one spoke around the centre. Orientation around the plot does not have meaning, i.e. whether the point is on the right or left or on the top or bottom of the target is only caused by the random plotting of the assessor. Therefore, points plotted at similar distances from the centre would demonstrate precision.

The rings on the diagram represent the accuracy of the response. Typically, the inner most ring, yellow, represents one standard deviation of responses. The second ring, gold, represents 1.5 standard deviations. The third ring, orange, is two standard deviations (95% confidence interval). Finally, the outer ring represents a factor of two from the reference value, which is a log deviation value of 0.3.

Figure 5 shows accurate responses to n-butanol for the 29 assessors tested during the study of Compound B. All points had a smaller deviation than the outer factor of two from the mean (0.3) criterion. The majority of data points, 25 of 29 (86%), were less than 1.5 standard deviations from the centre reference value.

Figure 5. Polar plot "Target" diagram showing the logarithmic deviation from the reference for the average threshold response of each assessor with n-butanol. Centre of the "target" is the reference value. The rings represent 1, 1.5, and 2 standard deviations as well as the outer ring at 0.3 representing a factor of two from the reference value.

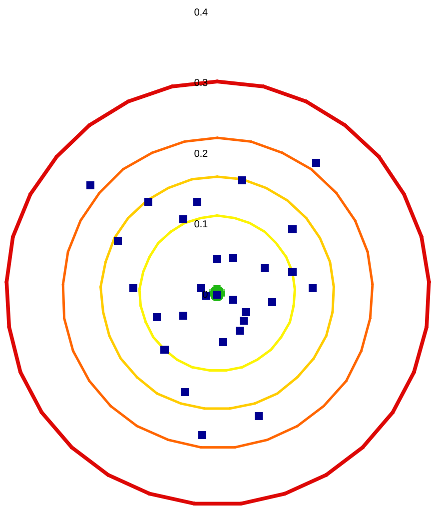


Figure 6 provides the "target" diagram for the response of 17 assessors tested with Compound A. Note that there is less precision of responses for Compound A, with the ring for two standard deviations (0.4) further out than a factor of 2 (0.3). However, even with the lower precision of the assessor responses, there was still only one data point outside the factor of two test criteria and 10 or 17 (59%) were within 1 standard deviation of the mean true value.

Figure 6. Polar plot "Target" diagram showing the logarithmic deviation from the reference for the average threshold response of each assessor with Compound A. Centre of the "target" is the reference value. The rings represent 1, 1.5, and 2 standard deviations as well as the ring at 0.3 representing a factor of two from the reference

value.

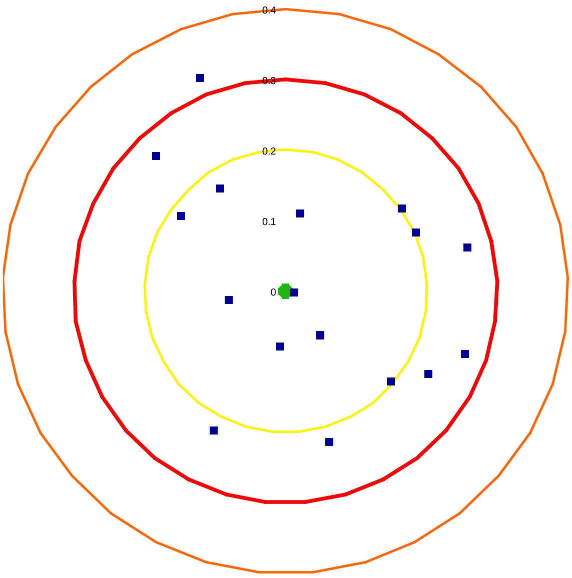
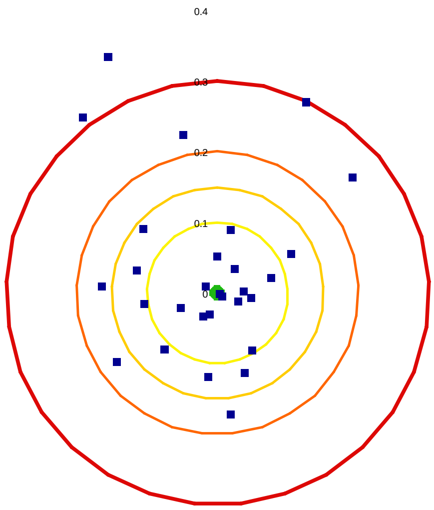


Figure 7 shows accuracy and precision of responses for the 29 assessors tested during the study of Compound B. As with n-butanol, the majority of responses, 21 of 29 (72%), are within 1.5 standard deviations from the centre reference value. Three assessor responses were outside the 1.5 standard deviations, while three average response values were outside the 2 standard deviations. As seen in previous summary graphics (Figure 4), only two assessors had average threshold values greater than a factor of two of the reference value.

Figure 7. Polar plot "Target" diagram showing the logarithmic deviation from the reference for the average threshold response of each assessor with <u>Compound B</u>. Centre of the "target" is the reference value. The rings represent 1, 1.5, and 2 standard deviations as well as the outer ring at 0.3 representing a factor of two from the reference value.



These results show the assessors are more accurate and precise for n-butanol than the other tested compounds. This may be due to the fact that the n-butanol threshold values are based on 20 tests for each assessor, while the results for Compound A and B represented fewer responses. Results for Compound A were less accurate and precise as the standard deviations of responses were much larger (rings further from the centre target). More results were closer to the factor of two times the reference value. Results for Compound B show accuracy and precision with most results near the centre and grouped within 1.5 standard deviations. Overall, the group of assessors score relatively similarly for all three tested compounds.

Panel Performance

Section 5.3 of EN13725:2003 details elements of monitoring panel performance. The result of testing the panel with n-butanol is monitored by the olfactometry laboratory to define precision (repeatability) and accuracy parameters. The results of the most recent ten tests are utilized to determine these parameters (McGinley, 2008). Table 1 provides a summary of the panel performance calculations for n-butanol, hydrogen sulphide (2009), Compound A, and Compound B.
The EN13725 acceptability criteria are also shown. All results fall well within the limits defined in EN13725.

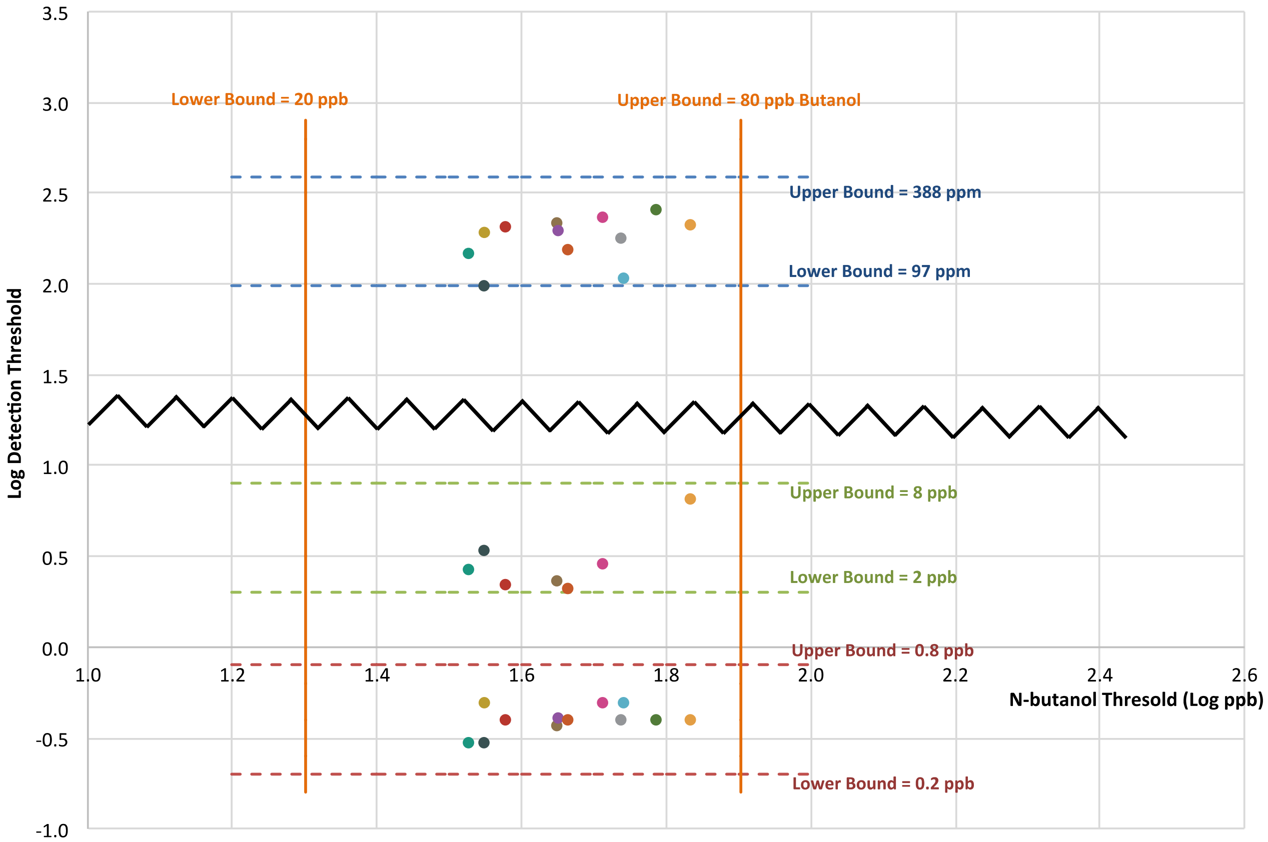
The parameter for accuracy ranged from a low value (most accurate) of 0.088 for n-butanol, to 0.131 for both hydrogen sulphide and Compound B, all below the 0.217 limit for EN13725 compliance. For repeatability, values ranged from 0.230 for n-butanol to 0.416 for Compound B. This repeatability can also be expressed as a factor, 10^r. This repeatability factor means that in 95% of tests run, the threshold value will be within a factor of 10^r of the average concentration. This value is 1.698 for n-butanol (25.5-74.0ppb), 1.817 for hydrogen sulphide (0.23-0.76ppb), 2.030 for Compound B (96-394ppm), and 2.607 for Compound A (1.5-10.4ppb). This result for Compoud A is consistent with the target diagram in Figure 6, which showed the higher standard deviation of individual assessor responses and wider spread of results in Compound A.

Table 1. Calculations of odour panel performance parameters for testing with standard odorant n-butanol as well as hydrogen sulphide (2009), Compound A, and Compound B. Calculations based on n=10.

	n-butanol	Hydrogen sulphide (2009)	Compound A	Compound B	EN13725 Criteria for n-butanol
Avg. Threshold	43.5ppb	0.42ppb	4.0ppb	194ppm	40ppb (20-80ppb)
Std. Dev. of log threshold	0.07	0.08	0.15	0.10	< 0.172
Repeatability (r)	0.230	0.259	0.416	0.307	< 0.477
Repeatability (10^r)	1.698	1.817	2.607	2.030	< 3
Accuracy (A)	0.088	0.131	0.131	0.118	< 0.217

During these studies, twelve individual assessors were tested to n-butanol, hydrogen sulphide (H_2S), and Compound B. Seven of these assessors were also tested to Compound A. Figure 8 summarizes the results from these twelve assessors for each test compound. N-butanol is shown as the x-axis with vertical lines representing the 20-80ppb limits for EN13725. The data points are grouped as threshold values for the other three compounds. Assessors are identified with data points identified by unique colour. In all cases, the assessor scores for the three compounds are all within a factor of two from the reference value as determined by the panel tests. Therefore, all were within an acceptable range of variability, even though the position of any one data point cannot predict the location of any other data point for that one assessor.

Figure 8. Average threshold responses for twelve assessors tested to n-butanol (x-axis), hydrogen sulphide (0.2 to 0.8ppb), and Compound B (97 to 388ppm), and the seven of these assessors tested to Compound A (2 to 8ppb). Note that data points of the same color in the different regions are matched as one assessor.

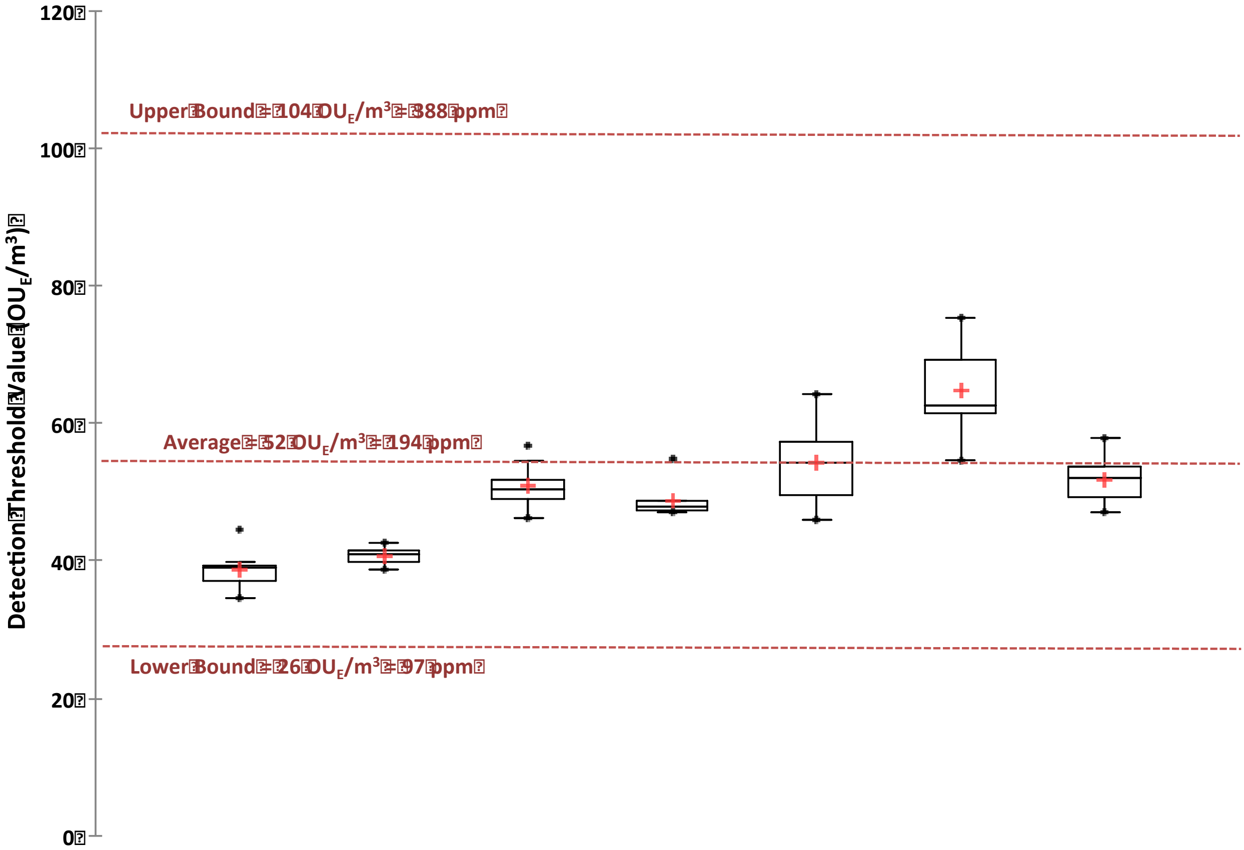


Review of Number of Assessors

For Compound B testing, all panels were run with 7-9 assessors with multiple rounds. Figure 9 shows the detection threshold results in OU_E/m^3 for multiple combinations of the 7-9 assessors from n=4 to n=9. Each box in the figure shows values for a specific sample tested with two rounds of the included assessors. In all cases the detection threshold value was between the 26-104 OU_E/m^3 acceptability range that coincides with the Compound B threshold acceptability range of 97-388ppm.

Figure 9. Box plot diagram of threshold values for Compound B as determined with multiple combinations of assessors. Each grouping is a sample tested. The box plot represents the detection threshold as determined by

combinations thresholds calculated using four to nine of the panel assessors.



CONCLUSIONS

The results of assessor performance testing to multiple compounds shows that all assessors tested, who conform to EN13275 criteria, demonstrate a similar range of acceptability for other compounds. For three other compounds, hydrogen sulphide (2009), Compound A and Compound B, almost all assessor responses were within a factor of two of the determined "accepted value", which is the average response of multiple test panels. It is likely, with more tests, all assessors would respond within the acceptability range.

The study also shows that any panel of acceptable odour assessors will provide a result for test odorants within an agreed upon range of sensitivity, 20-80ppb for n-butanol, 0.2-0.8ppb hydrogen sulphide, 2-8ppb for Compound A, and 97-388 for Compound B. Additionally, not withstanding a weak correlation between panel response to n-butanol and the other test compounds, all panel results, on average, were within a narrow range of sensitivity.

These results expand on conclusions drawn in our 2009 study that a second odorant may not be necessary for proper selection of assessors; however, a laboratory could utilize multiple odorants for documentation of performance indicators for accuracy and precision for compounds other than n-butanol. These multiple odorants can also be helpful for initial screening of assessors and for on-going, long term monitoring of assessor performance.

The panel and assessor performance testing results show that the parameters determined with testing of n-butanol were consistent for the multiple odorants. Thus, the assumption made in the EN13725 olfactometry standard that QA/QC parameters determined for n-butanol are transferable to other odorants is shown to be valid.

Additionally, in consideration of the number of assessors required for a testing session, the calculation of Compound B detection threshold results with multiple numbers of assessors (n=4 to n=9) show consistent results in all cases. This limited test did not show a significant difference in results with more than the standard four assessors.

REFERENCES

AIHA. 1989. *Odor Thresholds for Chemicals with Established Occupational Health Standards;* American Industrial Hygiene Association; Akron, OH, USA.

ASTM International. 2004. E679-04: *Standard Practice for Determination of Odor and Taste Threshold by a Forced-Choice Ascending Concentration Series Method of Limits*. Philadelphia, PA, USA.

CEN - Committee for European Normalization. 2003. *Air Quality – Determination of odour concentration by dynamic olfactometry*; EN13725:2003; Brussels, Belgium.

Heeres P.; Harssema H. 1990. Progress of the Standardization of Olfactometers in the Netherlands. *Staub Reinh. Der Luft*. Vol. **50**, 185-187.

Hermans L. 1989. *Ringonderzoek Olfactometers. Publikatiereeks Lucht No. 80*. Ministry of the Environment, Leidschendam, The Netherlands.

McGinley, M.A.; McGinley, C.M. 2010. Odor Assessor Performance to Reference and Non-Reference Odorants. *Proceedings of The Water Environment Federation / A&WMA Specialty Conference: Odors and Air Emissions*; Charlotte, NC; 21-24 March 2010.

McGinley, M.A.; McGinley, C.M. 2008. Odor Threshold Emission Factors for Common WWTP Processes. *Proceedings of The Water Environment Federation / Air & Waste Management Association Specialty Conference: Odors and Air Emissions*; Phoenix, AZ; 6-9 April 2008.

McGinley, M.A.; McGinley, C.M. 2006. Precision of Olfactometry and Odor Testing Results. *Proceedings of The Water Environment Federation / Air & Waste Management Association Specialty Conference: Odors and Air Emissions*; Hartford, CT; 9-12 April 2006.

McGinley, M.A.; McGinley, C.M. 2003. Comparison of Field Olfactometers in a Controlled Chamber using Hydrogen Sulfide as the Test Odorant. *Water Sc. Technol.* (G.B.), **50** (4), 75-82.

Van Gemert, L.J. 1999. *Compilation of Odour Threshold Values in Air and Water*; TNO Nutrition and Food Research Institute; Boelens Aroma Chemical Information Service (BACIS); 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.