Odor Assessor Performance to Reference and Non-Reference Odorants

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 Pollutants 2010
Charlotte, NC: 21-24 March 2010

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

Odor Assessor Performance to Reference and Non-Reference Odorants

Michael A. McGinley, P.E., St. Croix Sensory, Inc.
Charles M. McGinley, P.E., St. Croix Sensory, Inc.
St. Croix Sensory, Inc.
3549 Lake Elmo Ave. N.
P.O. Box 313
Lake Elmo, MN 55042

ABSTRACT

The European Standard test method for the determination of odor concentrations (EN13725:2003) is based on the standardization of the assessors (test subjects) with a known sensitivity to an accepted reference odorant, n-butanol. The assumption is made that the sensitivity for the reference will be a predictor for sensitivity to other substances, i.e. non-reference odorants, odorant mixtures, and environmental odors.

Assessors from a commercial odor laboratory, certified to the EN13725 selection criteria, tested hydrogen sulfide as a non-reference odorant. Hydrogen sulfide was used as the non-reference odorant representing a common compound that is found in wastewater process odors and odors from various other industries, e.g. pulp and paper, landfill, agriculture. This paper analyzes the following:

1. Comparison of individual assessor responses to n-butanol and hydrogen sulfide during a test session;
2. Comparison of the odor panel results for n-butanol and hydrogen sulfide during a test session;
3. Comparison of assessor performance testing with n-butanol and hydrogen sulfide; and
4. Comparison of panel performance testing with n-butanol and hydrogen sulfide.

The first two objectives compare results obtained during any one test session through review of individual assessor responses and the panel average as a whole. The second two objectives will illustrate the repeatability of the individual assessors and the odor panel as a group by providing statistical evidence of parameters for accuracy and precision.

EN13725 provides a framework for monitoring accuracy and precision of results in an olfactometry laboratory. Regular testing of the reference odorant n-butanol allows a laboratory to quantify these parameters. Periodic testing can also be performed on non-reference odor samples, e.g. hydrogen sulfide, to confirm these parameters as part of a comprehensive QA/QC program. This statistical information provides a user of odor testing results greater understanding of the meaning and application of these results.

KEYWORDS

Olfactometry, EN13725, ASTM E679, n-butanol, butanol, hydrogen sulfide, H2S, odor thresholds, detection threshold.

INTRODUCTION

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

Historically, ASTM E679 and other olfactometry standards recommended the use of assessors with olfactory sensitivity representative of the general population (ASTM, 2004). Eight to twelve assessors would be assigned to a test session with the expectation that their responses were 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 and to meet agreed upon repeatability criteria with the small sample size of assessors convened on odor 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 odor 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 odor 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 odor 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.

The selection of assessors based on these criteria provides a standard sensor, and makes the assessors 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 a group 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 odors 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 sulfide and other sulfur-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.

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

St. Croix Sensory conducted threshold testing of hydrogen sulfide during test sessions in 2009 and 2010 in order to compare the individual assessor and test session panel responses with hydrogen sulfide to those from testing with n-butanol.

METHODOLOGY

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. From July 2009 through January 2010, St. Croix Sensory also ran standard samples of hydrogen sulfide as part of test sessions as time allowed. At various times, specific test sessions were also scheduled where only samples of n-butanol and hydrogen sulfide were tested.

St. Croix Sensory acquires n-butanol from Oxygen Services, Inc. (St. Paul, MN) at a nominal 40-ppm with balance nitrogen (certified at ±2%). Two cylinders were utilized during this testing. The first, 46-ppm n-butanol/balance N_2, was produced on 6/15/2009 (Cylinder #UV000247). The second, 41-ppm n-butanol/balance N_2, was produced on 9/22/2009 (Cylinder #UV000207).

Hydrogen sulfide (H_2S) was also acquired through Oxygen Services. A cylinder of hydrogen sulfide at 1.43-ppm / balance N_2 was produced on 5/29/2009 (Cylinder #UV000029). The concentration was certified to ±5%.

The gaseous samples of the odorants were prepared by filling 5-L Tedlar air sample bags directly from the cylinders using stainless steel two stage regulators. The dedicated sample bags were flushed before initial use and remained dedicated for tests with the specified odorant.

Odor threshold values were determined on an AC'SCENT® International Olfactometer 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.

According to requirements of EN13725, a minimum of four assessors were utilized for each testing session, with each assessor observing the sample two times (two rounds). A minimum of eight assessor responses were required for a valid result. In almost all instances, five assessors were utilized. Retrospective screening was conducted on the samples following specifications of EN13725, however, no assessor responses were required to be removed in any of the test sessions.

RESULTS

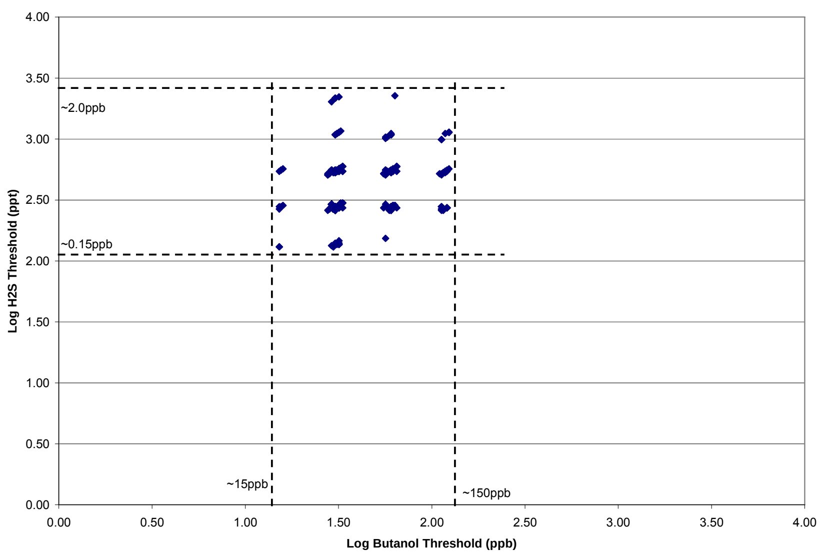
Between 14 July 2009 and 8 January 2010, St. Croix Sensory collected data from 72 matched samples of n-butanol and hydrogen sulfide. The individual responses of each assessor and the panel responses (geometric mean of participating assessors) for each odorant were recorded.

Individual Assessor Responses to Odorants

There were 355 individual responses of assessors for both odorants conducted during the same test session. These responses were made up of 42 different assessors who observed the two odorants from 1 to 20 times. Figure 1 provides a graph of the log part-per-trillion detection threshold of hydrogen sulfide versus the log part-per-billion detection threshold of n-butanol. Each point on the graph represents one of the 355 matched individual responses to both compounds.

The assessor selection criteria for n-butanol as defined by EN13725 is 20-80ppb (62-246μg/m^3), with the agreed reference value of 40ppb (123μg/m^3), based on a rolling average of the most recent 20 tests. The individual assessor responses for n-butanol ranged from a minimum of 15.2ppb to a maximum of 123.6ppb. These individual responses are all within one step factor of the acceptable range of average responses with a standard deviation of the log threshold of 0.21. When determining an assessor meets the criteria to be an assessor, the responses that fall outside the acceptable range of the average are offset by multiple responses within the range.

Figure 1. Results of 355 individual assessor responses for threshold determination testing of n-butanol and hydrogen sulfide. Each point represents an individual's response to both compounds during one test session.



The individual assessor detection threshold responses for hydrogen sulfide ranged from 0.13ppb to 2.3ppb with an average detection threshold of 0.41ppb. As with the butanol results, the lowest individual responses were within a factor of three of the overall average and only five of the 355 total responses were higher than 3 times the average.

The results show there is no correlation within this range of responses. For example, assessors who are least sensitive to butanol had hydrogen sulfide responses across the middle of the range of results. The five outlier responses that show relatively less sensitivity to hydrogen sulfide were by three assessors with responses that overall showed slightly higher sensitivity to butanol. These data show that the results of multiple odorants are within a small range of responses and those who meet the EN13725 selection criteria for n-butanol also provide a narrow range of responses to hydrogen sulfide.

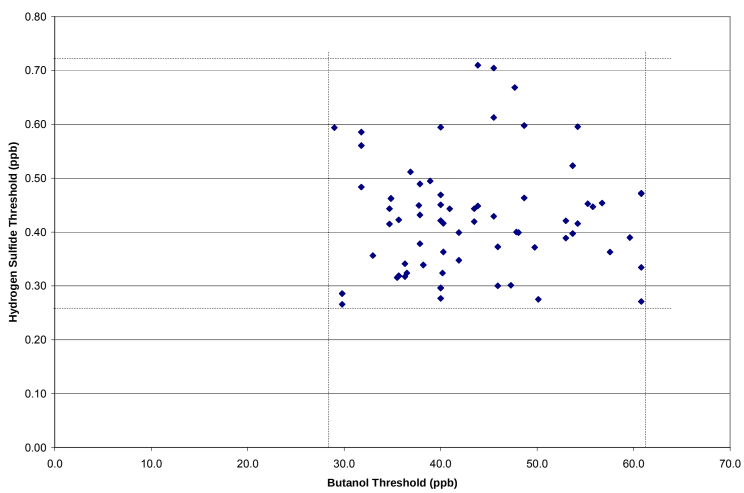
Panel Responses to Odorants

Figure 2 is a graphical summary comparing the 72 matched panel responses to both odorants during the same test session. The panel responses for n-butanol samples ranged from 29.0ppb to 60.8ppb, with an overall average of 42.3ppb. All panel results were within the 20-80ppb criteria set by EN13725. The standard deviation of the log detection threshold values was 0.074 (95% Confidence Interval = 30.1-59.4ppb).

The panel results for hydrogen sulfide during those same test sessions ranged from 0.27ppb to 0.71ppb, with an overall average response of 0.41ppb. Considering an acceptable measure of repeatability as two times and one-half of the overall average, all responses fit within this target range of 0.21-0.82ppb. The standard deviation of the log detection threshold values was 0.104 (95% Confidence Interval = 0.25-0.66ppb.

For the recognition thresholds, the butanol results ranged from 40.5ppb to 173.2ppb, with an overall average recognition threshold of 73.1ppb. The hydrogen sulfide results had a range of 0.47ppb to 1.19ppb, with an overall average recognition threshold of 0.70ppb.

Figure 2. Results of 72 matched panel responses for threshold determination of n-butanol and hydrogen sulfide. Each point represents the panel average response to both odorants during the same test session.



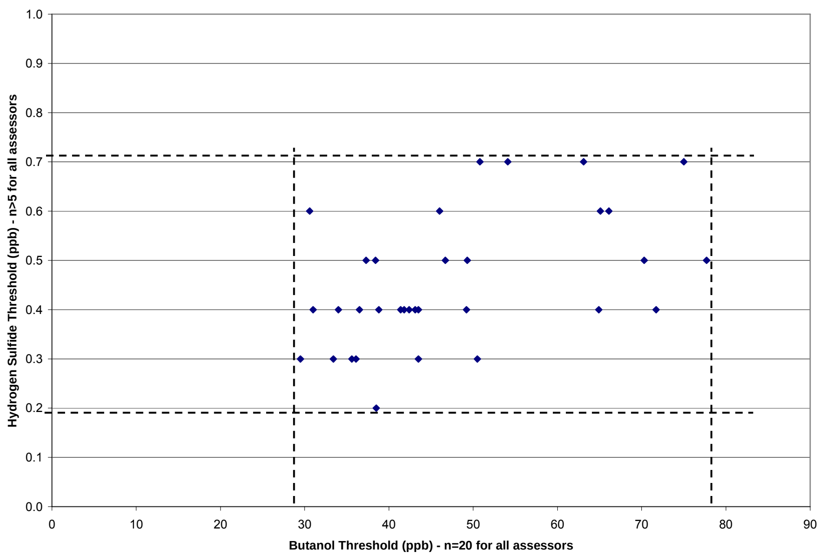
Assessor Performance Testing to Odorants

A main element of the EN13725 olfactometry standard is the standardization of the sensor through selection of assessors. Criteria are set to review a rolling average of the most recent 20 tests of each assessor with the standard odorant, n-butanol. Each assessor must have a geometric mean threshold between 20-80ppb, with a defined standard deviation of the log threshold. It is further possible to consider other odorants following this same framework.

Figure 3 compares the n-butanol testing results of 33 assessors who also had at least five tests with hydrogen sulfide (maximum of n=20 for hydrogen sulfide). Each point on the graph represents a comparison of one assessor's average responses to n-butanol and hydrogen sulfide. All n-butanol results fall within the 20-80ppb criteria, with a low of 29.5ppb and a high of 77.7ppb. As described previously, even if an assessor may have had one response outside this range, their average of the most recent 20 tests provides a result within the acceptable range.

Results to hydrogen sulfide ranged between a low of 0.2ppb and a high of 0.7ppb. As with n-butanol, this range of average assessor responses is smaller than the range of individual responses.

Figure 3. Results comparing performance testing parameters for standard odorant, n-butanol, and hydrogen sulfide for 33 assessors. Each point represents a different assessor (for each assessors, n=20 for butanol results and n>5 for hydrogen sulfide results).



Panel Performance Testing to Odorants

Section 5.3 of EN13725 details elements for monitoring panel performance. The result of the panel testing to n-butanol is monitored by the olfactometry laboratory to define precision and accuracy parameters. The results of the most recent ten tests are utilized to determine these parameters.

For EN13725, the accuracy of a laboratory can be determined from the n-butanol testing 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 repeatability can also be expressed through transformation as 10^r, or $10^{0.477} = 3.0$. This limit means that in the laboratory, the result from any two consecutive measurements will not be larger than a factor of three in 95% of the cases.

Table 1 provides a summary of the panel performance calculations for both n-butanol and hydrogen sulfide. The EN13725 acceptability criteria are also shown. For n-butanol, St. Croix Sensory calculated these parameters for the most recent 10 panel session tests. The average threshold value of n-butanol for those ten tests was 46.1ppb, with a standard deviation of the log of the threshold value computed as 0.08. The accuracy value was 0.116 and the repeatability value was 0.244 or also expressed as a factor of 1.755. Each of these parameters were well within the limits defined in EN13725. The repeatability value means that in 95% of tests run, the threshold of n-butanol will be determined to be within a factor of 1.76 of the reference value of 40ppb (95% Confidence Interval – 22.7-70.4ppb).

Table 1. Calculations of odor panel performance parameters for testing with standard odorant n-butanol as well as hydrogen sulfide. Calculations based on n=10.

	n-butanol	hydrogen sulfide	EN13725 Criteria for n-butanol
Avg. Threshold	46.1ppb	0.42ppb	40ppb (20-80ppb)
Std. Dev. of log threshold	0.08	0.08	0.172
Repeatability (r)	0.244	0.259	0.477
Repeatability (10^r)	1.755	1.817	3
Accuracy (A)	0.116	0.131	0.217

For hydrogen sulfide, the average threshold value was 0.42ppb, with a standard deviation of the log threshold value of 0.08, which was the same as for n-butanol. For calculating the other parameters, a value of 0.40 was used as the reference value of hydrogen sulfide. As with n-butanol, the repeatability and accuracy values for hydrogen sulfide were also computed to be well within the range of acceptability defined by EN13725. Notably, the repeatability value means that in 95% of the tests, the determined threshold will fall within a factor of 1.82 of the reference value (95% Confidence Interval = 0.22-0.73ppb). These values for both odorants improve even more if recalculated based on a higher number of most recent samples run (increased n).

DISCUSSION

All results show that individual assessor responses to n-butanol and hydrogen sulfide fall within a narrow range of values; however, there is no direct correlation between responses to the two odorants for one's individual responses during one panel session or the average responses of that individual. In other words, an individual assessor's sensitivity to n-butanol does not predict their exact sensitivity to hydrogen sulfide. However, the responses are all predictably within an acceptable range of sensitivity.

Figure 4 is a box plot showing the assessor n-butanol performance testing results from on-going monitoring of assessors following EN13725 olfactometry standard. Assessors' average responses ranged from 29.5 to 77.7ppb (blue dots), with a mean value of 47.8ppb (red cross), median of 43.5ppb, 1st quartile of 37.3ppb (bottom of box) and 3rd quartile of 54.1ppb (top of box). All average values for each assessor were within the 20-80ppb range of acceptability.

Figure 4. Box plot of average threshold response to n-butanol during assessor performance testing (n=20 tests per assessor).

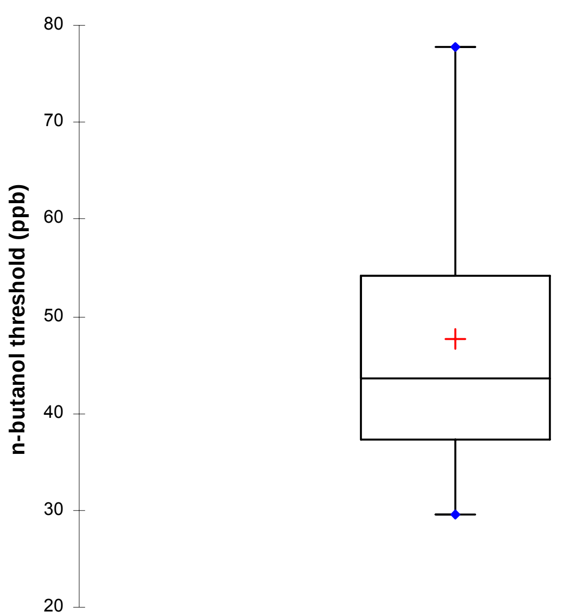
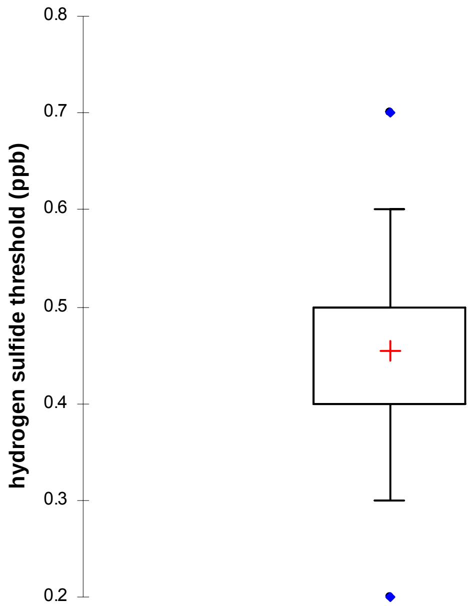


Figure 5 is a box plot showing the results of testing the sensitivity of these same assessors with hydrogen sulfide. The average responses ranged from 0.20 to 0.70ppb (blue dots), with a mean of 0.46ppb (red cross), a median of 0.40ppb, a 1st quartile also 0.40ppb (bottom of box), and 3rd quartile of 0.50ppb (top of box).

Figure 5. Box plot of average threshold response of individual assessors tested with hydrogen sulfide (n=5 - 20 tests per assessor).



There is currently no agreed upon reference value for hydrogen sulfide. Publications have most commonly reported the threshold of hydrogen sulfide as 0.5ppb to as high as 5ppb; however, these publications did not clearly differentiate a detection threshold versus recognition threshold (AIHA, 1989; van Gemert, 1999). In 2008, the authors correlated hydrogen sulfide concentrations determined through collected environmental air samples with the reported hydrogen sulfide concentration of these samples. That study showed an average hydrogen sulfide threshold of 0.52ppb (McGinley, 2008). Previous unpublished studies by these authors, with varying sources of hydrogen sulfide, have recorded hydrogen sulfide thresholds of 0.4 to 1.0ppb (McGinley, 2003).

For the assessors utilized in this study, the value determined for the 1st quartile (37.7ppb) was near the 40ppb agreed upon reference value for n-butanol. The 1st quartile response of the assessors tested with hydrogen sulfide is a value of 0.40ppb. A value of 0.40ppb is utilized throughout this paper as the reference threshold value for hydrogen sulfide. This value matches closely with threshold studies conducted to date.

Figures 6 and 7 provide a different way to view the responses of the assessors to n-butanol and hydrogen sulfide, respectively. The polar plots show the deviation of the log of the average assessor responses from the log of the reference value. The center of the “target” would be a response equivalent to the reference value. The further out from the center, the further the assessor’s average is from the reference value. The polar plot was created with each assessor as one spoke around the center. Orientation around the plot does not have meaning, i.e. whether the dot is on the right or left or top or bottom of the target is only caused by the random plotting of the assessor and does not have meaning.

Figure 6. Polar plot of the logarithmic deviation from the reference for the average threshold response of each assessor tested with n-butanol. Center of "target" is the reference value for the threshold. The rings, from inner ring moving outward, represent 1 standard deviation of responses (yellow), 1.5 standard deviations (gold), 2 standard deviations (orange), and a factor of two from the reference value (red).

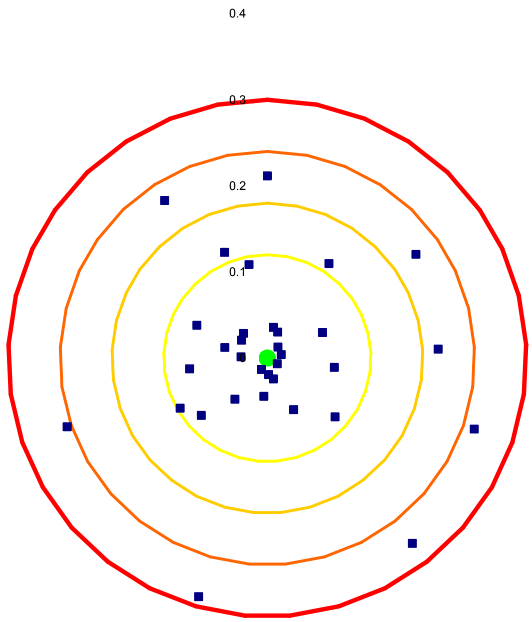
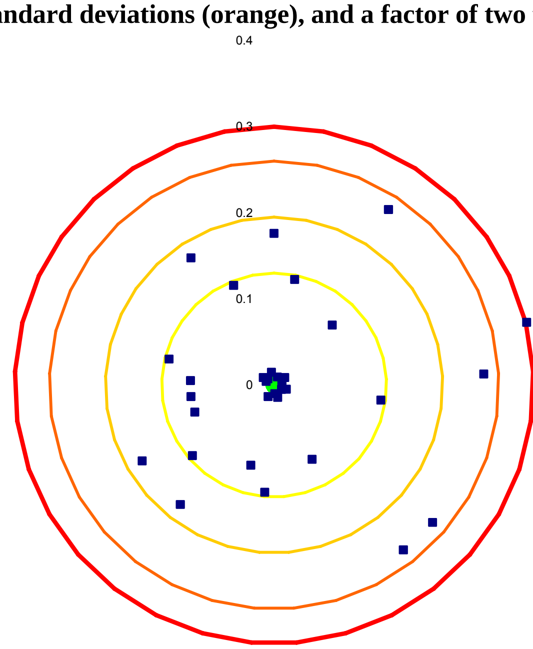


Figure 7. Polar plot of the logarithmic deviation from the reference for the average threshold response of each assessor tested with hydrogen sulfide. Center of "target" is the reference value for the threshold. The rings, from inner ring moving outward, represent 1 standard deviation of responses (yellow), 1.5 standard deviations (gold), 2 standard deviations (orange), and a factor of two from the reference value (red).



The outermost circle (red) shows a value that would be two times or one-half the reference value, which is a log value of 0.3. The inner most ring (yellow) is one standard deviation from the reference value. The next ring (gold) is 1.5 standard deviations from the reference value. The third ring (orange) is two standard deviations from the mean (95% confidence interval).

These plots show the assessors on average were more accurate, less deviation from center, for testing hydrogen sulfide than testing n-butanol. However, the results for n-butanol were slightly more precise, smaller standard deviation of all responses (inner rings of target closer to center). These results further show how all assessors who meet the EN13725 n-butanol testing criteria also provide acceptable scores for hydrogen sulfide testing that are also both accurate and precise.

The results collected from odor panels test sessions also showed a narrow range of results for both odorants. Overall panel results for n-butanol ranged from 29.0 to 60.8ppb, with a mean of 42.3ppb. For hydrogen sulfide, the results ranged from 0.27 to 0.71ppb, with a mean of 0.41ppb. Calculations of panel performance following EN13725 showed that the panels do meet accuracy and precision requirements for both the standard odorant n-butanol and for hydrogen sulfide.

The assumption that the panel performance test method criteria for n-butanol are transferable to all odors is shown to be valid for a second odorant, hydrogen sulfide. While one assessor's sensitivity to n-butanol is not a predictor of their exact sensitivity to hydrogen sulfide, the panel as a whole responds consistently to n-butanol and hydrogen sulfide within a specified accuracy and precision.

CONCLUSIONS

The data analysis of this odor assessor performance study, where all odor assessors conformed to the EN13725 criteria, demonstrates that the odor assessors also conform to a similar range of acceptability for hydrogen sulfide. That range of acceptability for hydrogen sulfide is 0.2-0.8ppb, if 0.4ppb is the assigned reference value. Not withstanding a weak correlation between one individual assessor's response to n-butanol and to hydrogen sulfide, the second odorant of hydrogen sulfide proves to be an acceptable performance indicator of an individual odor assessor's conformance to a reference value, i.e. 0.4ppb for hydrogen sulfide.

The study also shows that any panel of acceptable odor assessors will provide a result for both odorants within an agreed upon range of sensitivity, 20-80ppb for n-butanol and 0.2-0.8ppb for hydrogen sulfide. Additionally, not withstanding a weak correlation between one single panel of odor assessors to n-butanol and to hydrogen sulfide, all panel results, on average, were within a narrow range of sensitivity.

Hydrogen sulfide is an acceptable second odorant for a laboratory to use in a comprehensive QA/QC program. The results of this study suggest that this second odorant is not necessary for proper selection of assessors; however, a laboratory should utilize a second odorant, such as hydrogen sulfide, for documentation of performance indicators for accuracy and precision. This second odorant is also helpful for initial screening of assessors and for on-going, long term monitoring of assessor performance.

Panel and assessor performance testing results show that the parameters determined with testing to n-butanol are consistent with a second odorant. 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 for hydrogen sulfide.

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. (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.