TECHNICAL NOTE:

RELIABILITY OF ODOR OFFENSIVENESS QUANTIFICATION USING THE REFINED CLOTH SWATCH OLFACTOMETRIC TECHNIQUE

S. W. Gay, E. F. Wheeler, K. B. Kephart

ABSTRACT. *The cloth swatch olfactometric technique uses flannel swatches to absorb odors for presentation to a trained human sensory panel. This study examined the reliability of the refined cloth swatch technique used to facilitate the quantification of odor offensiveness of swine facility wastewater treated in subsurface–flow constructed wetlands. Reliability was based on the consistency of ratings assigned by panelists to odor sample replicate pairs. Panelists assigned odor replicate pairs ratings that differed by ± 0 or ± 1 with a 79% frequency, while differences of ± 0 and ± 1 between odor replicate pairs would occur with only a 44% frequency if ratings were randomly assigned. Although individual panelist error during individual trials varied among panelists, individual panelist error over all trials was found to be not significant (Z < 115, α =0.05). Statistical methods described here could be used to screen human sensory panelists during pre–trials. Overall, the swatch olfactometric technique used by human sensory panelists yielded highly consistent results and low overall error, indicating that odor offensiveness rating data obtained using the refined cloth swatch technique were reliable to use for various odor analyses.*

Keywords. *Frequency distribution, Odor evaluation, Odors, Sensory panel, Statistical analysis, Swine manure.*

Nuisance odor complaints have become more frequent in communities surrounding areas of high–density swine production. Mitigation of these complaints depends on the development of standards that specify acceptable amounts and frequencies of odor emission from a particular odor source. Development of effective odor standards requires an accurate method of odor measurement (NCSU, 1998). Gas chromatography, mass spectrometry, and the electronic nose can quantify concentrations of individual compounds. However, the correlation between gas concentrations and odor intensities and offensiveness is highly variable. Thus, detection and measurement of malodorous compounds by chemical means does not indicate whether the odor will be offensive to humans (Schmidt and Jacobson, 1995).

Olfactometry is an odor measurement technique that uses the human nose as the odor sensor (Schiffman and Williams, 1999). Panels of trained human “sniffers” are presented with an odor to evaluate its intensity, offensiveness, and/or hedonic tone. Olfactometry is the basis for most odor measurements because it accounts for odor characteristics as perceived by humans. The two types of olfactometry are: (1) dynamic olfactometry, in which a malodorous air stream flows toward sensory panelists’ noses, and (2) static olfactometry, which requires malodorous air to be presented to the nose in an enclosed volume (Schiffman and Williams, 1999).

Dynamic olfactometry is the most widely accepted method of subjective odor quantification. Gaseous odor samples are collected in impervious bags on site and transferred to an olfactometry laboratory. An instrument known as a dynamic olfactometer pumps a stream of malodorous air diluted with various concentrations of compressed air toward human sensory panelists. Panelists sniff the airstream and evaluate it for two values: (1) detection threshold, which is the concentration at which the odor is first detected by 50% of the panelists, and (2) recognition threshold, which is the concentration at which the odor is first identified by 50% of the panelists. Dynamic olfactometry works well; however, the technique does not evaluate odor offensiveness. Furthermore, it is a costly technique: The price of dynamic olfactometers can exceed $30,000.

The cloth swatch method of static olfactometry, developed by Miner and Licht (1981), offers an effective and inexpensive procedure to measure odor. The technique required the use of dry cotton flannel swatches to absorb odors. Swatches were exposed to an odor for 30 minutes, after which they were removed and sealed in individual plastic bags for delivery to an odor panel. Miner and Licht (1981) found that the cloth swatch technique proved satisfactory and was superior to conventional direct methods such as dynamic olfactometry.

Article was submitted for review in September 2000; approved for publication by the Structures & Environment Division of ASAE in May 2001. Presented at the 2000 ASAE Annual Meeting as Paper No. 00–4022.

The authors are **Susan Wood Gay,** ***ASAE Member Engineer,*** Assistant Professor and Extension Engineer, Department of Biological Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia; **Eileen Fabian Wheeler,** ***ASAE Member Engineer***, Associate Professor, Department of Agricultural and Biological Engineering, and **Kenneth B. Kephart,** Professor, Department of Dairy and Animal Science, Pennsylvania State University, University Park, Pennsylvania. **Corresponding author:** Susan Wood Gay, 202A Seitz Hall (0303), Virginia Tech, Blacksburg, VA 24061; phone: 540–231–1232; fax: 540–231–3199; e–mail: swgay@vt.edu.

Schiffman and Williams (1999) refined the cloth swatch technique by placing swatches in screw–capped glass bottles instead of plastic bags. Using glass bottles instead of plastic bags has two advantages: (1) glass is odorless while plastic has a distinct background odor, and (2) capped bottles allow head gas to generate in the upper portion of the bottle for sensory panelists to sniff, while plastic bags are porous and do not have a readily available gas volume for panelists to evaluate. However, no comparison of the refined cotton swatch technique has been made to the original technique.

Zhang et al. (1999) studied the cloth swatch technique by using disks of fabric contained in plastic cassettes for odor retention. Odor was well retained on swatches two days after sampling, and moistened air improved odor retention when cassettes were frequently opened. Odor losses were not found to be significant with cassette storage time. However, completion of odor sampling and sniff testing during the same day was recommended (Zhang et al., 1999).

Both static and dynamic olfactometry are subject to variation. However, error can be mitigated through implementation of several procedures. For example, human odor sensitivity varies within a large range, and selection of panelists at one extreme of the sensitivity distribution results in false odor ratings (Hangartner, 1985). Choosing a large number of panelists minimizes the probability that panelists will be from a sensitivity distribution extreme. The extent to which a human sensory panel constitutes a representative sample of the population depends directly on the number of panel members. According to Hangartner (1985), it is not feasible to select a large number of panelists, and a compromise must be sought between cost and representation of the population. An eight–member panel with an acceptable minimum of six is adequate to meet the cost and population criteria (Hangartner, 1985).

Reddell and Sweeten (1975) compared human sensory panels of five and 17 members for odor intensity index. Results indicated that that there was no significant difference between panel odor intensity geometric means or standard deviations (Reddell and Sweeten, 1975). Wijnen (1986) stated that panel size was the most important variable in determining odor strength. In the Netherlands, human sensory panel sizes have ranged from three members, according to the VDI Richtlinie guideline, to 20 members, according to AFNOR (Wijnen, 1986). Dutch guidelines prescribe a minimum panel size of six members and a minimum number of two odor sample replicates presented to each panelist (Wijnen, 1986).

In addition to panel size and selection, improving repeatability of observations between replicate samples can reduce olfactometry error. McGinley et al. (1995) reported that researchers using a standard odor intensity scale and standardized odor descriptor terminology could provide measured, dependable, and repeatable observation of odors.

McGrath (1977) studied odors from the land application of pig slurry on grasslands in Ireland, during which observers used a scale from 1 to 10 to quantify odors from the slurry (table 1). Williams (1984) investigated odor offensiveness of raw, aerobically treated, and stored piggery slurry and employed odor panelists to rate sample offensiveness on a 0–to–5 scale (table 2). Labance et al. (1999) had odor panelists rate odors from covered and uncovered mushroom substrate on a 0–to–5 scale (table 3) similar to that used by Williams (1984).

Table 1. Odor value scale used for the quantification of the odor of land–applied pig slurry (McGrath, 1977).

Odor Value	Description
1	Virtually undetectable
2 – 3	Smelly
4 – 5	Unpleasant
6 – 7	Offensive
8 – 9	Very offensive
10	Highly offensive

Table 2. Odor offensiveness scale used to rate offensiveness of raw, aerobically treated, and stored piggery slurry (Williams, 1984).

Odor Rating	Description
0	Inoffensive odor
1	Very faintly offensive odor
2	Faintly offensive odor
3	Definitely offensive odor
4	Strongly offensive odor
5	Very strongly offensive odor

Table 3. Odor levels and descriptions used by human sensory panelists (Labance et al., 1999).

Odor Level	Description
0	No odor detected
1	Faint odor – not objectionable
2	Slight odor – not objectionable
3	Moderate odor – not objectionable
4	Moderate odor – somewhat objectionable but tolerable
5	Strong odor – objectionable

In this study, a human sensory panel used the refined cloth swatch technique to test the odor offensiveness of effluent from a subsurface–flow wetland treating swine facility wastewater. The goal of this study was to determine if the refined cloth swatch technique was a reliable odor measurement technique to test odor offensiveness.

Objectives

Specific objectives of this investigation were:

1. To determine the differences between odor offensiveness ratings assigned by panelists to sample replicate pairs.
2. To determine the distribution of differences between odor offensiveness ratings assigned by panelists to sample replicate pairs.

Methods and Materials

Wastewater Treatment System

Eight wetlands were constructed to treat wastewater (feces, urine, and flushwater) from an underground storage pit that collected waste from a grower–finisher and gestating sow facility at the Pennsylvania State University Swine Center. Swine facility wastewater was transported biweekly from the underground storage pit to the wetland site during the experimental periods (May to August 1999). Submersible pumps in the transport tanks delivered wastewater to an organic filter inside the greenhouse. Filter effluent flowed into an open tank and was pumped to a storage tank. Wastewater then flowed to the feeder tank via a PVC mainline before entering the wetlands via PVC leader lines.

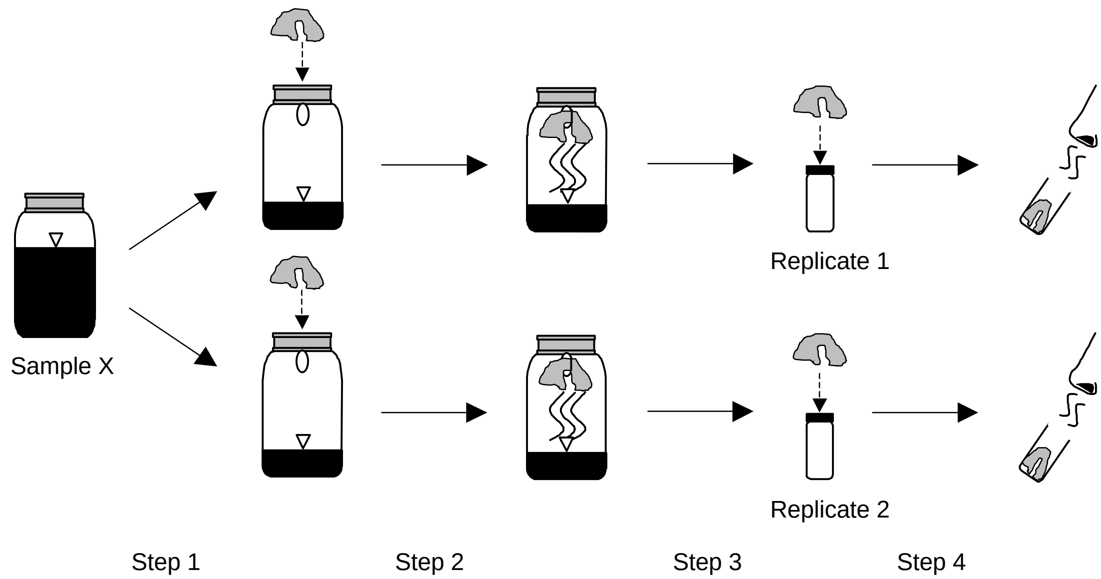

Figure 1. Sample preparation for odor offensiveness testing: Step 1. Raw swine facility wastewater is split into two 30 mL aliquots; Step 2. Cloth swatches are attached to jar lids to absorb odor for 30 min; Step 3. Cloth swatches are removed from jars and placed into 60 mL, screw–capped bottles for headspace gas generation; and Step 4. Bottles containing swatches are brought to odor panelists for testing.

SAMPLING PROCEDURES AND PREPARATIONS

Grab samples of 300 mL each were collected from the transport tanks, organic filter effluent, and each wetland effluent outlet and were then split with two 30 mL aliquots placed into glass quart jars to provide two replicates for each sample (fig. 1). Cloth flannel swatches of 10 cm × 10 cm (4 in × 4 in) dimensions were heated to 92°C (198°F) for 4 hrs prior to sampling to remove absorbed volatile gases and solids (Miner and Licht, 1981). Swatches were first attached to jar lids by a loop of Teflon–coated galvanized steel wire, then sealed in the jars for 30 min to absorb sample odor, and finally removed and placed in 60 mL glass bottles. Each bottle was randomly assigned a code consisting of a color dot sticker and one digit (e.g., Blue – 2). Bottles were capped with Teflon–lined screw caps for 30 minutes to allow head gas to generate and then immediately brought to the odor–testing laboratory (Schiffman and Williams, 1999).

SELECTION AND TRAINING FOR NON–PROFESSIONAL SENSORY PANELISTS

Eight volunteers were selected from the Department of Agricultural and Biological Engineering at Penn State. Smokers, drug dependents, pregnant women, and people who have serious allergies or frequent colds were excluded from participation (Nicolai et al., 1997). The panel consisted of four men and four women between the ages of 23 and 53. Human sensory panel training consisted of the following steps:

1. Overview – Panelists were presented with a summary of olfactometric processes, including the physiology of “sniffing” and a definition of odor terminology.
2. Introduction to Laboratory Rules – Panelists were presented with and asked to abide by sensory panel rules (table 4).
3. “Sniffing” Demonstration – Panelists were shown the recommended sniffing procedure (table 5).
4. Odor Dimensions – Panelists were presented with swatches that had been exposed for 30 minutes to distilled water (reference 1 = 0 odor rating), wetland effluent, or untreated swine facility wastewater (reference 2 = 5 odor rating) and shown how odors correspond with the odor offensiveness scale (table 6).

SAMPLE EVALUATION BY THE HUMAN SENSORY PANEL

The odor testing laboratory was a 4.4 m × 9.2 m (10 ft × 21 ft) room on the second floor of the Agricultural Engineering Building. This location was two floors removed from the sample preparation room and other areas of odor

Table 4. Odor–testing laboratory rules for human sensory panel (Nicolai et al., 1997).

1.	Must be free of colds or other physical conditions affecting the sense of smell.
2.	Must not smoke or use smokeless tobacco.
3.	Must not chew gum, eat, or consume coffee, tea, or beverages for at least one hour prior to sensory panel work.
4.	Must not eat spicy foods for at least six hours prior to odor panel work.
5.	Must not consume alcohol for at least six hours prior to panel work.
6.	Must be “fragrance–free” by not using perfume, cologne, or other scented products when evaluating odors.

Table 5. Human sensory panel odor evaluation procedure.

1.	Sniff reference 1 and consider it a “0” on the odor offensiveness rating scale.
2.	Sniff reference 5 and consider it a “5” on the odor offensiveness rating scale.
3.	Randomly choose a sample bottle.
4.	Remove cap from bottle.
5.	Place mouth of bottle so it is just touching the nose (do not remove swatch).
6.	Take a small whiff (not a deep breath) for approximately three seconds.
7.	Record rating on the rating sheet.
8.	Replace cap on bottle.
9.	Take several slow, deep breaths before sniffing the next bottle.
10.	Continue procedure with the next sample.

Table 6. Odor offensiveness scale and descriptions used by human sensory panelists.

Odor Offensiveness	Description
0	No odor detected
1	Faint odor – non–identifiable, not offensive
2	Slight odor – non–identifiable, not offensive
3	Identifiable odor – somewhat offensive, but tolerable
4	Identifiable odor – offensive, but tolerable
5	Highly identifiable odor – very offensive, intolerable

generation in the building. The room was naturally ventilated prior to each odor testing session.

Two odor testing sessions were conducted on sampling days: (1) 9:30 a.m. to noon, and (2) 1:30 p.m. to 4:00 p.m. The two sessions were held to accommodate panelists' work schedules. Each panelist was randomly assigned a half–hour time period during one of the two sessions. These time slots were staggered every half–hour to allow for the regeneration of head gas that had dissipated from the bottles during previous odor tests.

Panelists first sniffed two reference odors to calibrate their senses of smell. The references were swatches placed above distilled water (reference 1) or swine facility wastewater from an underground storage pit (reference 2). Reference 1 was to be considered a "0" and reference 2 a "5" on the odor offensiveness scale. Panelists were presented with a total of 18 swatches, two replicates for each of nine samples, in addition to the two references during each experimental trial. Sample replicate pairs were used to determine the consistency of odor ratings assigned by the human sensory panel. Bottles containing swatches were randomly arranged for presentation to the panelists. Panel members sniffed each swatch for approximately three seconds and then assigned an odor rating to that sample. Each panelist rested from 10 to 15 seconds after each sniff to attempt to minimize odor fatigue (Nicolai et al., 1997).

Testing Reliability of the Cloth Swatch Technique

Reliability of an olfactometric technique depends on the consistent representation of odorants presented to the human sensory panel. Thus, panelists would likely use the same criteria and perceptions each time an odor is rated and would assign replicate pairs the same odor rating (±0) or ratings that differ slightly (i.e., ±1).

Reliability of the refined cloth swatch technique used in this study was determined using three statistical methods. The first method examined the error of all panelists for all trials by comparing the frequency distribution of the differences between odor ratings assigned to replicate pairs by human sensory panelists to a theoretical frequency distribution of the differences between odor ratings randomly assigned to 5000 replicate pairs. Comparison of the actual frequency distribution of differences to the theoretical distribution would indicate whether panelists assigned the same rating or ratings that differed by only one with greater frequency than would occur if ratings were randomly assigned.

The second method was to determine if panelist error, calculated by summing the differences between replicate pair ratings, for each individual trial was significant. This method would indicate during which trials a particular panelist might have assigned random odor ratings. The theoretical frequency distribution of error that occurs when odor ratings are randomly assigned to 500 replicate pairs for nine events was generated, and the percentage of panelist error that comprised the distribution was calculated. Panelist error was then compared to the value corresponding to the test level (α) of 0.05. If panelist error were less than or equal to the value of the test level for a particular trial, then ratings were not randomly assigned by the panelist for that trial. If panelist error were greater than the value of the test level for a particular trial, then the panelist randomly assigned ratings and might have been considered for removal from the human sensory panel.

A panelist could appear to have assigned "random" ratings during a high percentage of trials due to the relatively small number of trials. Therefore, the third method examined panelist error over all trials to determine if panelists used consistent judging criteria overall. The theoretical frequency distribution of error that occurs when odor ratings are randomly assigned to 500 replicate pairs for 72 events was generated, and the percentage of overall panelist error that comprised the distribution was calculated. Overall panelist error was then compared to the value corresponding to the test level (α) of 0.05. If overall panelist error were less than the value of the test level, then the panelist was using consistent judging criteria to assign odor ratings. If overall panelist error were greater than the value of the test level, then the panelist was not using a consistent judging criteria overall and should have been removed from the human sensory panel.

Results and Discussion

Overall Performance of the Human Sensory Panel

Panelists assigned odor replicate pairs the same rating (±0 error) or ratings that differed by only one (±1 error) with a 79% frequency. The theoretical frequency distribution indicated that differences of ±0 and ±1 between odor replicate pairs would occur with only a 44% frequency if ratings were randomly assigned (fig. 2). Comparison of the two frequency distributions does not necessarily indicate a statistically significant difference between actual and theoretical data. However, the magnitude between the two distributions, especially at ±0 error and ±1 error, reveals that it was highly unlikely that panelists randomly assigned odor ratings.

Performance of Individual Panelists during Individual Trials

Individual panelist error for individual trials was examined to determine if each panelist was using consistent judging criterion during each trial. For a test level of 0.05, the corresponding Z value on the theoretical frequency distribution curve is 10 (fig. 3). If panelist error for each individual is less than or equal to 10, then that particular panelist used consistent judging criteria for that particular trial. If panelist error for each individual is greater than 10, then that particular panelist did not use consistent judging criteria for that particular trial and the data should be disregarded.

Individual panelist error for each trial is listed in table 7. Three out of eight had significant error for more than half the trials, indicating that their data for those trials should be disregarded. Furthermore, the overall error for each panelist having significant error during individual trials should be calculated to determine whether these panelists were fit to participate as sensory panelists.

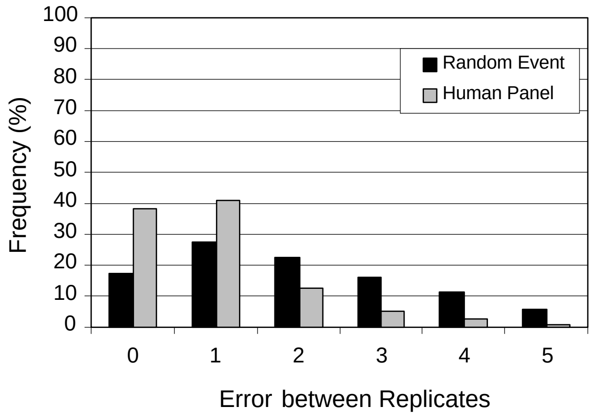

Figure 2. Frequency distributions of the error or difference between odor ratings assigned to replicate pairs by human sensory panelists and between odor ratings randomly assigned to 5000 pairs. Panelists assigned the same rating (±0 error) or ratings that differed by only one (±1 error) to replicate pairs at a much higher frequency than would occur for a random event.

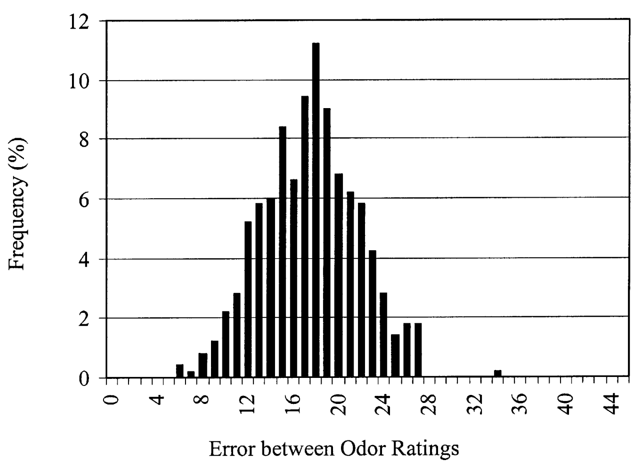

Figure 3. Theoretical frequency distribution of differences between odor ratings randomly assigned to 500 replicate pairs over 9 events. The Z value is 10 at $\alpha = 0.05$ and is used to determine if panelist error was significant during individual trials.

Table 7. Panelist error for each trial and total error for all trials. Error was calculated for each panelist by summing the differences (D) between odor ratings assigned to each replicate pair for each trial.

	Panelist Number							
Trial	1	2	3	4	5	6	7	8
1	5	7	11	5	7	14	10	6
2	12	9	10	6	6	9	6	4
3	19	12	6	9	7	7	6	4
4	15	16	8	13	5	12	3	14
5	13	12	7	9	4	8	8	2
6	15	9	14	11	7	10	10	9
7	11	6	14	10	10	8	8	2
8	10	12	11	6	6	6	8	2
Total Error	100	83	81	69	52	74	59	43

Performance of Individual Panelists over All Trials

Determining panelist error over all trials is the best method to determine if an individual panelist is randomly assigning odor ratings and is, therefore, unfit to serve as an odor panelist. For a test level of 0.05, the corresponding Z value on the theoretical frequency distribution curve is 110 (fig. 4). If panelist error for each individual over all trials is less than or equal to 110, then that particular panelist used consistent judging criteria overall. If panelist error for each individual over all trials is greater than 110, then that particular panelist did not use consistent judging criteria overall. That panelist should be removed from the human sensory panel and any data from that panelist should be discarded. Summation of panelist error over all trials indicated that all panelists had errors less than 110 and were using consistent judging criteria overall (table 7).

Conclusions

Reliability of the refined cloth swatch olfactometric technique was examined using three statistical methods. The first method examined the error of all panelists for all trials by comparing the frequency distribution of the differences between actual odor ratings assigned to replicate pairs to a theoretical frequency distribution of the differences between odor ratings randomly assigned to 5000 replicate pairs. This method demonstrated that panelists assigned odor replicate pairs the same rating (±0) or ratings that differed by only one (±1) with a 79% frequency, while differences of ±0 and ±1 between odor replicate pairs would occur with only a 44% frequency if ratings were randomly assigned. This difference in magnitude indicates that a random event did not cause panelists to assign relatively consistent odor ratings to replicate pairs.

The second method examined individual panelist error for individual trials to indicate during which trials particular panelists assigned random odor values. Data from trials where panelist error was significant should be considered erroneous and should not be used for further analyses.

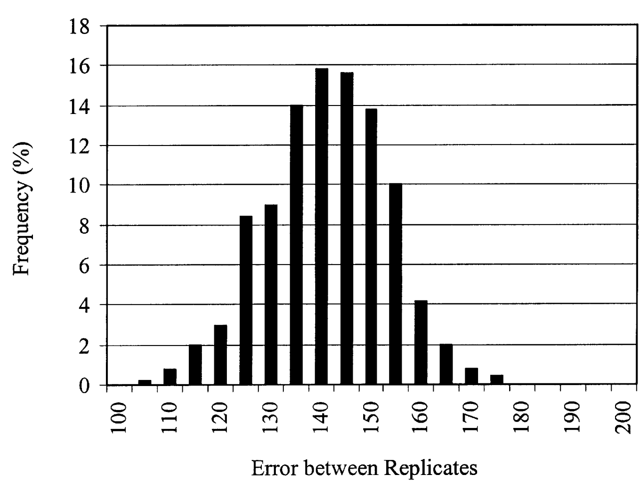

Figure 4. Theoretical frequency distribution of differences between odor ratings randomly assigned to 500 replicate pairs over 72 events. The Z value is 115 at $\alpha = 0.05$ and is used to determine if panelist error was significant during all trials.

Individual panelist error over all trials was tested by the third method. This test indicated that, although some panelists performed poorly during particular trials, their overall error was small. The human sensory panel performed well overall, thus indicating that the refined cloth swatch technique was a reliable method to test odor offensiveness. However, the method could be improved by panelist screening. Specifically, the second and third statistical methods could be used in conjunction during pre–trials as a screening procedure for potential human sensory panelists.

ACKNOWLEDGEMENTS

This research was supported in part by agricultural research funds administered by the Pennsylvania Department of Agriculture (PDA). The authors express much appreciation to the PDA for the continued interest in this project.

REFERENCES

Hangartner, M. 1985. Selection and treatment of panelists for determination of odor thresholds. In *Odour Prevention and Control of Organic Sludge and Livestock Farming*, 55–60. V. C. Nielsen, J. H. Voorburg, and P. L'Hermite, eds. New York, N.Y.: Elsevier Applied Science Publishers.

Labance, S. E., P. H. Heinemann, and D. M. Beyer. 1999. Evaluation of microporous covers for the reduction of mushroom substrate preparation odors. *Appl. Eng. in Agric.* 15(5): 559–566.

McGinley, C. M., D. L. McGinley, and K. J. McGinley. 1995. Odor school – Curriculum development for training odor investigators. In *Odors: Indoor and Environmental Air, Proc. of an International Specialty Conference*, 121–130. C. McGinley and J. Swanson, eds. Pittsburgh, Penn.: Air and Waste Management Association.

McGrath, D. 1977. Odours arising from the land–spreading of pig slurry. *Agriculture and Environment* 3(2/3): 171–177.

Miner, J. R., and M. Licht. 1981. Fabric swatches as an aid in livestock odor evaluations. In *Livestock Waste: A Renewable Source, Proc. 4th International Symposium on Livestock Wastes*, 302–305. Amarillo, Texas, 15–17 April. St. Joseph, Mich.: ASAE.

Nicolai, R. E., C. J. Clanton, P. R. Goodrich, L. D. Jacobson, K. A. Janni, V. J. Johnson, E. Lees, and D. R. Schmidt. 1997. Development of a dynamic olfactometer lab. ASAE Paper No. 97–4019. St. Joseph, Mich.: ASAE.

NCSU. 1998. Control of odor emissions from animal operations: A report from the board of governors of the University of North Carolina. Raleigh, N.C.: North Carolina State University.

Reddell, D. L., and J. M. Sweeten. 1975. Evaluation of odor intensities at livestock feeding operations in Texas. In *Proc. 3rd International Symposium on Livestock Wastes*: *Managing Livestock Wastes*, 358–361. St. Joseph, Mich.: ASAE.

Schmidt, D., and L. D. Jacobson. 1995. Odor and odor policy criteria. AEU–11. St. Paul, Minn.: University of Minnesota Extension Service.

Schiffman, S. S., and C. M Williams. 1999. Evaluation of swine odor control products using human odor panels. In *Proc. 1999 Animal Waste Management Symposium*, 110–118. Cary, N.C., 27–28 January. Raleigh, N.C.: North Carolina State University.

Wijnen, H. L. J. M. 1986. Guideline for olfactometric measurements in the Netherlands: Comparison with western European guidelines. In *Odour Prevention and Control of Organic Sludge and Livestock Farming*, 69–77. V. C. Nielsen, J. H. Voorburg, and P. L'Hermite, eds. New York: Elsevier Applied Science Publishers.

Williams, A. G. 1984. Indicators of piggery slurry odour offensiveness. *Agric. Wastes* 10(1): 15–36.

Zhang, Q., N. R. Bulley, R. K. York, A. K. Lau, and S. Baidoo. 1999. Using cloth swatches for odor measurement. ASAE Paper No. 99–4145. St. Joseph, Mich.: ASAE.