This is not a peer-reviewed article.
International Symposium on Air Quality and Manure Management for Agriculture
CD-Rom Proceedings of the 13-16 September 2010 Conference
(DoubleTree Hotel, Dallas Texas) Publication date, 13 September 2010
ASABE Publication Number 711P0510cd

COMPARISON OF AMBIENT ODOR ASSESSMENT TECHNIQUES IN A CONTROLLED ENVIRONMENT

C. G. Henry[1], D. D. Schulte[1], S. J. Hoff[2], L. D. Jacobson[3], and A. M. Parkhurst[4]

ABSTRACT.

This paper compares results of using - dynamic triangular forced-choice olfactometry (DTFCO), Mask Scentometers, Nasal Rangers®, and an odor intensity reference scale (OIRS) –intensity ratings - to assess odors in a controlled-environment chamber in the Iowa State University Air Dispersion Laboratory. The methods were used to assess thirteen odor levels in the chamber where swine manure mixed with water was used to vary the odor levels. Dynamic triangular forced-choice olfactometry did not correlate well to the other ambient odor assessment methods. Predicting D/T using intensity ratings degraded R_o^2 with the other methods in all cases. Average Intensity-predicted D/T, the Mask Scentometer and the Nasal Ranger® correlated well with each other, had strong R_o^2 (greater than 0.85), had regression slopes nearest one, and the session means were not found to be significantly different (α=0.05). Using the geometric means of the device D/T settings, $(D/T)_G$, improved R_o^2 between the other methods and the Nasal Ranger® and Mask Scentometer. Average Intensity-predicted D/T values were three to four times higher than Nasal Ranger® assessment ($(D/T)_G$ and D/T, respectively), and a Nasal Ranger® $(D/T)_G$ was roughly five times higher than Mask Scentometer $(D/T)_G$.

KEYWORDS, Mask Scentometer, Nasal Ranger®, odor intensity, Dynamic triangular forced choice olfactometry, ambient odor assessment

INTRODUCTION

Primary difficulties with assessing ambient odors are the low concentrations of odor commonly experienced and the rapidly fluctuating conditions that occur over time. Laboratory-based dynamic triangular forced-choice olfactometry (DTFCO) has generally been the accepted standard method - the gold standard - for measuring odor concentrations. In the ambient atmosphere, though, odor concentrations are very low, and DTFCO typically is more effective at assessing odors at higher concentrations (> 50 D/T) than at the low concentrations encountered downwind from an odor source. Additionally the cost to analyze an air sample with DTFCO can be very expensive. Field olfactometers and odor intensity ratings have the advantage of being less expensive methods for obtaining a lot of field data over a longer period of time, making them attractive in calibrating and verifying models, as well as making general assessments of odor (Sheffield and Ndegwa, 2008). In some instances, field olfactometry may be used in conjunction with laboratory-based methods. For example, air samples from an odor source may be collected and analyzed in an olfactometry laboratory to quantify source emissions rates while field olfactometry is used to assess odor transport in the surrounding area.

[1] Biological System Engineering, University of Nebraska-Lincoln, 68583

[2] Department of Agricultural and Biosystems Engineering, Iowa State University, 50011

[3] Department of Bioporducts and Biosystems Engineering, University of Minnesota, 55108

[4] Department of Statistics, University of Nebraska-Lincoln, 68583

Field olfactometers available for use today include the Box Scentometer manufactured by the Barneby and Sutcliffe Corporation (purchased in 2004 by Calgon Carbon, (www.calgoncarbon.com), the Nasal Ranger® manufactured by St Croix Sensory (www.nasalranger.com), and the Mask Scentometer, also referred to as a facial field olfactometer, an instrument developed by Sheffield (2004) and improved by Henry (2004, 2009). Finally, Intensity ratings based on an Odor Intensity Reference Scale (OIRS), may be used as predictors of odor concentration.

PREVIOUS WORK

Sheffield et al. (2004) investigated differences between the Mask Scentometer, Nasal Ranger®, Box Scentometer, in-field intensity, and in-lab intensity (from Teldar bags) field assessment techniques with DTFCO at five agricultural and industrial sources using a group of eight assessors to make measurements. Their study evaluated the variability of responses of the devices and methods and found that the Nasal Ranger® and laboratory-based olfactometry exhibited the least amount of variability across the odor sources. Sheffield et al. (2007) performed odor assessments on 38 dairies and 15 feedlots in Idaho. They assessed odors using the Nasal Ranger® and intensity ratings using n-butanol as the reference odorant. They found a moderate correlation between D/T and H2S/Total Reduced Sulphur (TRS) which appeared to increase slightly with receptor distance from the source. McGinley and McGinley (2003) compared the Barneby and Sutcliffe Box Scentometer and Nasal Ranger® field olfactometers in an environmentally controlled room. A hydrogen sulfide generator was used to vary the odor levels while three Nasal Ranger® assessors and one Box Scentometer user evaluated the odor in the room. They found high correlation (r = 0.82, n not reported) between the Box Scentometer and the Nasal Ranger® method and no significant difference was found between assessors (p=0.309). The field olfactometers yielded hydrogen sulfide thresholds of 0.5-2.0 ppb. Laboratory olfactometry (DTFCO) yielded comparable thresholds of 0.45-0.9 ppb and the McGinley's deemed their results consistent with other published values.

Newby and McGinley (2003) compared the Nasal Ranger®, a Barnebey Sutcliffe Box Scentometer, and laboratory-based olfactometry for assessing odor in the field. They found no significant difference between a Box Scentometer and a pre-production Nasal Ranger® at a 95% confidence interval (p=0.06) and a Pearson's Correlation Coefficient of 0.82. They found that the Missouri regulatory limit of 110 D/T (their actual mean was 106.5 D/T) using laboratory olfactometry equated to 7 D/T observed with a Scentometer. According to the state statute, a 7:1 D/T observed with a scentometer is a trigger for an olfactometry sample (DTFCO) to be taken. The purpose of their work was to show that Box Scentometer readings and D/T from olfactometry analysis of samples were not comparable (i.e. that a different standard was needed for the olfactometry analysis).

PURPOSE OF WORK

In spite of the efforts reported above, the measurement of ambient odors is a crude science. One of the challenges with ambient odor assessment is that there is no standard method to relate one odor assessment technique to another. Currently, there is no agreed upon way of equating one ambient odor assessment technique or method to another; that is, the reported dilution to threshold from one instrument or method is not currently comparable to another. Much odor work has been done with a plethora of these methods, yet it is currently not possible to determine if or how the results from these various methods can be related. The objectives of this experiment were to compare the following ambient odor assessment techniques under controlled conditions: DTFCO, Nasal Ranger®, Mask Scentometer, and an Odor Intensity Reference Scale (OIRS), and to identify relationships between data produced using these methods.

MATERIALS AND METHODS

A series of thirteen odor assessment sessions were conducted in a controlled laboratory environment at the Iowa State University Air Dispersion laboratory in May and June of 2004. The number of assessments performed for each method were based on the amount of time needed to perform as many odor assessments as could be reasonably performed in the ten minute time period. In each assessment session, the following assessment methods were used:

- Dynamic triangular forced-choice olfactometry (DTFCO) DTFCO was used to analyze air samples collected in the chamber in new, un-flushed, unbaked Tedlar bags (10 L) during the first four minutes of each ten-minute assessment session. Sampling and analysis followed ASTM Standard E679-99, Standard Practice for Determination of Odor and Taste Thresholds by a Forced-Choice Ascending Concentration Series Method of Limits. Both the University of Minnesota and Iowa State University odor labs analyzed air samples using DTFCO. All samples were analyzed to determine a panel D/T within 24 hours. Both labs were in compliance with the European Standard for olfactometry (CEN, 2003).

- Nasal Ranger®. Assessors from Iowa State University were trained by St. Croix Sensory to use the Nasal Ranger® field olfactometer. Odor assessments were made twice during each 10-minute assessment session, once shortly after entering the room and again five minutes after entering the room.

- Mask Scentometer. Assessors trained by the University of Nebraska used the Mask Scentometer field olfactometer developed by Sheffield et al. (2004) and Henry (2004) to assess odors every 30 seconds during each ten-minute session. In the analysis of data, D/T settings were assigned as specified in Henry (2009).

- Intensity Rating (Odor Intensity Reference Scale). Assessors were trained by the University of Minnesota to rate odor intensity using a OIRS based on the static scale method of ASTM Standard E 544-99 Standard Practices for Referencing Suprathreshold Odor Intensity. A 0-5 scale was used in this experiment based on n-butanol in air concentrations using 25 ppm to represent I = 1; 75 ppm for I = 2; 225 ppm for I = 3; 675 ppm for I = 4; and 2,025 ppm for I = 5. Assessors could use half steps (i.e. 1.5, if they felt the odor was between a 1 and a 2), and assessments were taken every 15 seconds, which resulted in 40 assessments taken during each experiment. Field Intensity data was analyzed as raw data (Intensity), and converted to a D/T using two techniques described later and referred to as Intensity-predicted D/T and Average intensity-predicted D/T.

For the Nasal Ranger®, Mask Scentometer, and OIRS methods, three to five individuals were randomly spaced within a 20 ft by 20 ft room (6.8 m by 6.8 m) located at the Iowa State University Air Dispersion Laboratory. A swine manure odor source was placed near the inlet to the room, and air was drawn through the room using exhaust ventilation fans. A plenum was installed to create uniform airflow across the room. The odor source (raw swine manure) was diluted with water to achieve differing levels of odor in the room. Odor levels were presented in random order for each session. All panelists began their assessments at the same time (a lead assessor began and stopped all assessors).

The experiment was conducted over a period of two days with six ten-minute odor sessions conducted the first day and seven on the second day (thirteen total). On the first day (first six sessions), three assessors used Mask Scentometers, three assessors used Nasal Rangers®, and five assessors rated odor intensity. On the second day (last seven sessions), five assessors used Mask Scentometers, five assessors used Nasal Rangers®, and four assessors rated odor intensity.

A relationship first used by Sheffield et al, 2004 was used to obtain a geometric average dilutions to threshold $(D/T)_G$ for the field olfactometers (Mask Scentometer and Nasal Ranger®. The results are shown in Table 1. This was done to normalize the peaks and keep extremely high or low values from skewing the results.

Table 1. Geometric dilutions to threshold $(D/T)_G$ used for the Mask Scentometer and Nasal Ranger®

Mask Scentometer		Setting	Nasal Ranger®	
Unit D/T	Geometric D/T	n	Unit D/T	Geometric D/T
		7	60	60
18	18	6	30	42.4
4.5	9	5	15	21.2
2	3	4	7	10.2
1	1.4	3	4	5.3
0.35	0.6	2	2	2.8
0 / Non-detect	0.2	1	0 / Non-detect	1.4

Intensity data was used to predict D/T and resulting 'intensity-predicted D/T' were used to compare methods. Jacobson et al. (2000) published a relationship between intensity and D/T determined from the analysis of odor concentration using a laboratory olfactometer. For swine odors, they used the following relationship to predict dilution to threshold (D/T) as a function of odor intensity (i):

$$D/T_{swine} = 8.367\ e^{\ 1.0781i}$$

This relationship was applied to the intensity rating data in two ways. The first way used the equation to predict a D/T for each individual assessor observation (reported intensity value). Then the average D/T for each user's series of observations was then used for the session to determine an average predicted D/T and is referred to as 'intensity-predicted D/T'.

The second way, took the average of the intensity rating values, then used the same equation applied to individual's average intensity ratings (0-5) for the session to predict an 'Average-intensity-predicted D/T'. The latter (Average intensity-predicted D/T) is the same technique used by Jacobson et al. (2000), Jacobson et al. (2003), Nicolai et al. (2000), and Zhu et al. (2000).

RESULTS AND DISCUSSION

A detailed statistical analysis was completed. Raw data was checked using lack-of fit in SAS for linearity and to screen for bias, a test for interaction between days and sessions was checked, and an analysis of variance (ANOVA) was used to test for variation between methods. A few individual assessors were deemed to have bias and were removed from the dataset. Using the R statistical package (2008) the Pearson's product-moment and Spearman's Rank Correlation (ρ) were used to indicate strength and direction of the linear relationship and linear regression (forced intercept through zero) was performed to develop a relationship between methods. Developing statistical relationships between ambient odor assessment methods was complicated by the fact that a different number of odor assessors used each method. Because of the different number of observations available for each method across the sessions, only the session means for each method were used in the statistical analyses.

From the results shown in Table 2, several general trends emerge. Most notably none of the data obtained using field methods correlated well with DTFCO Lab D/T. Good correlations existed, as expected, between the intensity ratings and intensity predicted D/T and average intensity-predicted D/T. Good correlations were found between intensity ratings and Mask Scentometer $(D/T)_G$ (0.84-0.86) and between intensity ratings and Nasal Ranger® D/T and $(D/T)_G$ (0.78-0.80). Correlations were higher for $(D/T)_G$ than for D/T meaning that using the geometric mean of the unit D/T for the device provided better correlations to the other methods than did using the unit D/T directly. This difference was less pronounced for the Nasal Ranger® suggesting that using the geometric scale settings did not improve correlations between the Nasal Ranger® data and the data from the other methods. While modest correlation (0.56-0.59) was found between the Nasal Ranger® $(D/T)_G$ and the Mask Scentometer $(D/T)_G$, both of these methods correlated better to Average intensity-predicted D/T (0.74-0.79 for the Nasal Ranger® D/T and $(D/T)_G$ and 0.74-0.84 for the Mask Scentometer $(D/T)_G$).

Table 2. Pearson product-moment correlation coefficient (top) and Spearman's correlation coefficient, ρ (bottom)

	Intensity Rating (0-5)	Intensity-predicted D/T	Average-intensity-predicted D/T	Mask D/T	Mask $(D/T)_G$	DTFCO Lab D/T
Nasal Ranger® D/T	0.80* 0.76*	0.73* 0.71*	0.77* 0.74*	-0.22 0.11		-0.10 0.05
Nasal Ranger® $(D/T)_G$	0.81* 0.78*	0.77* 0.74*	0.79* 0.76*		0.59* 0.56*	-0.10 0.01
Intensity Rating (0-5)		0.93* 0.92*	0.94* 0.99*	-0.15 0.30	0.86* 0.84*	0.05 0.16
Intensity-predicted D/T			0.98* 0.92*	0.35 035	0.78* 0.87*	-0.11 0.15
Average-intensity-predicted D/T				-0.11 0.29	0.74* 0.84*	-0.09 0.15
Mask D/T						-0.31 -0.18
Mask $(D/T)_G$						0.22 0.34

** Indicates P<α=0.05, there is a significant correlation between methods.*

Table3. Slopes (top values), coefficients of determination R_o^2 (middle values), and standard errors (bottom values) from linear regression between methods (session averages, n = 13)

Independent / Predictor	Dependent /Response							
Y► X▼	DTFCO Lab D/T	Nasal Ranger® D/T	Nasal Ranger® $(D/T)_G$	Mask Scentometer D/T	Mask Scentometer $(D/T)_G$	Intensity rating (0-5)	Intensity-predicted D/T	Average intensity-predicted D/T
DTFCO Lab D/T		0.08* 0.49 0.02	0.10* 0.53 0.03	0.01* 0.28 0.005	0.02* 0.59 0.005	0.007* 0.59 0.002	0.42* 0.34 0.17	0.26* 0.43 0.09
Nasal Ranger® D/T	6.3 0.49 1.8			0.10* 0.39 0.04		0.08* 0.92 0.007	5.72 0.80 0.8	3.29 0.87 0.4
Nasal Ranger® $(D/T)_G$	5.1 0.53 1.4				0.19* 0.85 0.02	0.07* 0.94 0.004	4.5 0.81 0.6	2.6 0.88 0.3
Mask Scentometer D/T	28.4 0.28 13.0	3.79 0.39 1.35				0.37* 0.46 0.1	21.2 0.30 9.2	12.8 0.37 4.9
Mask Scentometer $(D/T)_G$	27.6 0.62 6.1	3.6 0.82 0.5	4.6 0.85 0.6			0.34* 0.94 0.02	22.8 0.83 3.0	12.8 0.86 1.5
Intensity rating (0-5)	76.6 0.56 18.3	10.7 0.92 0.92	13.7 0.94 1.0	1.26 0.46 0.39	2.8 0.94 0.2		66.6 0.88 7.3	38.2 0.94 2.8
Intensity-predicted D/T	0.82 0.34 0.3	0.14* 0.80 0.02	0.18* 0.81 0.03	0.01 0.30 0.006	0.04* 0.83 0.005			0.54 0.97 0.03
Average intensity-predicted D/T	1.65 0.43 0.5	0.26* 0.87 0.03	0.34* 0.88 0.04	0.03 0.37 0.01	0.07* 0.86 0.008		1.79* 0.97 0.09	

** Indicates stronger relationship based on lowest standard error. To scale a Nasal Ranger* $(D/T)_G$ *to Mask Scentometer* $(D/T)_G$*, take its value times 0.19 (i.e. 1 NR=0.19 MS), to scale a method below the light-grey boxes, use the inverse slope, for example to relate a Nasal Ranger* $(D/T)_G$ *to an Average-intensity-predicted D/T, the stronger relationship is 0.34 (as opposed to 2.6, because the error was lower), so multiply the D/T times 1/0.34=2.9 to obtain a relative predicted D/T for intensity, or 1 NR=2.9 Average-intensity-predicted D/T.*

Since correlation established association between methods, the next step was to establish the relationships between the methods, so that knowing one, the other could be predicted. To accomplish this, linear regression was performed. Traditionally in linear regression analysis, one variable is the independent variable or predictor (x) and a relationship can be found for the response, the dependent variable (y). One of the underlying assumptions is that the regressors (xi) are not contaminated with errors and are independent. In this experiment, this assumption is not valid. So one should base the relationship on the predictor error that is small to negligible with respect to the response variable, in order to derive the best relationship possible between methods. Thus, the standard error of the estimate was used as criterion for model selection. The standard error of estimate is a measure of error of prediction. That is the lower the standard error, the higher the precision, and the more preferred model. So each method was regressed as both an independent variable and dependent variable relative to the other methods, as shown in Table 3, and the two regression models were ranked. The model with the lowest error was the better model slope or scaling factor produced from the regression. The slope with a "*" produced the lowest error and is the more precise relationship. The resultant slopes and the goodness of fit of the relationship (coefficients of determinations, R_o^2) for the session averages from linear regression analysis are shown in Table 3. Note that the R_o^2 are the same for each of the linear models. From Table 3 one can relate one method to another and assess the scale of measurements from the different methods. For illustration, the slope between the Mask Scentometer $(D/T)_G$ and Nasal Ranger® $(D/T)_G$ is about one-fifth (0.19), so Nasal Ranger® $(D/T)_G$ readings were about 5 times higher than Mask Scentometer $(D/T)_G$.

The slope for regression of two perfectly comparable methods - methods that both produce the same result - would be 1.0 and methods that have a coefficient of determination (R_o^2) near 1.0. The coefficient of determination is the proportion of the variability that is accounted by the linear model and describes the goodness of fit of the linear estimated slope. The relationship between Intensity-predicted D/T and Average-intensity-predicted D/T is closest to a 1:1 slope at 1.79 (Table 3) and the relationship was very strong $R_o^2 = 0.97$. This good-fitting relationship is at least somewhat intuitive since both D/T are predicted from the same set of intensity data. Other methods that showed reasonably close and strong relationships, based upon this simple regression analysis, were DTFCO and intensity-predicted D/T, Mask Scentometer D/T (and $(D/T)_G$) and intensity ratings, and Nasal Ranger® $(D/T)_G$ and Average-intensity-predicted D/T. The strongest R_o^2's, beside the R_o^2's between predicted D/T as just described, all involved intensity ratings as follows: vs. Mask Scentometer $(D/T)_G$ (R_o^2=0.94), Average-intensity-predicted D/T (R_o^2=0.94), and the Nasal Ranger® $(D/T)_G$ (R_o^2=0.94) and D/T (R_o^2=0.92). The R_o^2 between the Nasal Ranger® and Mask Scentometer $(D/T)_G$ was good (0.85), as were the R_o^2's between Average-intensity-predicted D/T and Nasal Ranger® $(D/T)_G$ (0.88) and Mask Scentometer $(D/T)_G$ (0.85). In general, these methods have good fitting relationships between them.

Using geometric average D/T for the Mask Scentometer and Nasal Ranger® improved the R_o^2 data from other methods in all instances. The slopes came closer to a 1:1 slope also when $(D/T)_G$ was used. For example, R_o^2 improved from 0.34 to 0.84 between the Mask Scentometer and Nasal Ranger, and the slope increased from 0.10 to 0.19. These results are compelling for the use of $(D/T)_G$ for two reasons, first there was a dramatic increase in accountability of variation and second, because a high R_o^2 is essential, whereas a slope near one is only desirable.

In general, relationships of laboratory DTFCO had low coefficients of determination (R_o^2=0.34-0.62). The slopes between intensity-predicted D/T (0.42) and Average-intensity-predicted D/T (0.26) were nearer to one, but had low R_o^2's (not a strong relationship). Additionally the slopes of the Nasal Ranger®, Mask Scentometer, and intensity-based predictions versus laboratory-based olfactometry (DTFCO Lab D/T) were very far from a slope of one, requiring large scaling factors to relate DTFCO to these methods (top row of Table 3), a very undesirable result.

Coefficients of determination (R_o^2) for predicted D/T were degraded slightly relative to using the intensity ratings directly, meaning that using intensity ratings to predict D/T weakened the goodness of fit. R_o^2 between predicted D/T and observed intensity ratings were not as good as expected at R_o^2 = 0.88 and 0.94 for Intensity-predicted D/T and Average-intensity-predicted D/T, respectively. In fact the R_o^2 (0.94) for intensity ratings and the Nasal Ranger® $(D/T)_G$ and Mask Scentometer $(D/T)_G$ were just as good. Perhaps something is lost in the prediction or it is not robust. There are two schools of thought concerning the best application of the D/T prediction equation for intensity. Conceptually, it seems logical that when a person rates intensity, the rating corresponds directly to a predicted D/T for that assessment. Then averaging the predicted D/T, should normalize the predicted D/T. The alternative is to average the series of intensity ratings the given period of time, which has the effect of normalizing the assessment data, and then transform the intensity value to a predicted D/T. So the question becomes, should one normalize the raw data or the predictions? Average-intensity-predicted D/T was better correlated to the other methods (except for DTFCO Lab D/T) and had slopes closer to one than did intensity-predicted D/T. The prediction equation is an exponential function, so one would not expect a perfect fit to a linear model. This is the most likely reason that the exponential effect is less pronounced when the Average-intensity-predicted D/T is used. Again, the averaging of the intensity ratings is normalized first, and then transformed, rather than trying to fit the average of all the individual transformed assessments and fitting them to a linear model. It appears from this work that using predicted D/T based on averaged intensity ratings is preferable, in terms of being better correlated to other odor assessment methods, than is to averaging D/T values that were predicted from individual intensity ratings.

The Least Significant Difference multiple comparison results (Table 4) showed no significant difference between the intensity-based methods and no differences between the Average intensity-predicted D/T, Nasal Ranger® and Mask Scentometer data– with either D/T or $(D/T)_G$. Laboratory assessment (DTFCO) was significantly different from the other methods, however.

Table 4. LSD: Means for all measures of D/T for 13 sessions

Method	*Mean D/T	Standard Deviation	Maximum Session Mean	Minimum Session Mean
DTFCO lab D/T	134.36[a]	95.6	331.0	27.7
Intensity-predicted D/T	89.00[b]	78.9	290.4	7.8
Average intensity-predicted D/T	53.45[bc]	37.6	148.8	16.1
Nasal Ranger® D/T	16.20[c]	8.8	31.4	4.3
Nasal Ranger® $(D/T)_G$	21.10[c]	9.9	35.3	6.1
Mask Scentometer D/T	2.37[c]	2.0	7.1	0.5
Mask Scentometer $(D/T)_G$	4.14[c]	2.2	7.4	0.5

**Within a column, values with similar superscripts indicate means were not significantly different at alpha level of 0.05.*

While no statistically significant difference in the session means existed between the Nasal Ranger®, Mask Scentometer, and intensity-based methods, they did not produce the same results. The slope difference between the Mask Scentometer and Nasal Ranger® may be caused by the fact that their "stops" along the D/T scale are not at the same places, the range of the Mask Scentometer is limited (0.35 to 18 D/T), and the number of assessments between methods was not the same. That is, the lower D/T for the Mask Scentometer may be a result of twenty assessments compared to two assessments from the Nasal Ranger® and is likely a better representation of the room odor concentration. The researchers noted that the odor in the room decreased over the ten minute period, as the manure source equilibrated over time and less odor was generated from the source, which could explain differences between the Mask Scentometer and intensity methods to the others since these methods assessed odor during the entire session. Therefore, if we use the Nasal Ranger® $(D/T)_G$ for reference, eight of the thirteen session means were higher (19.4, 22, 22.6, 24.6, 28, 32.8, 35, and 35.3), than the maximum D/T setting (18 D/T) of the Mask Scentometer. When data from only sessions 4, 6, 7, 11, and 13 for which the Nasal Ranger® $(D/T)_G < 19$ D/T were analyzed from, R_o^2 for Mask Scentometer $(D/T)_G$ and Nasal Ranger® $(D/T)_G$ increased from 0.85 to 0.94 and the slope increased from 0.19 to 0.30 for $(D/T)_G$ and from 0.10 to 0.25 for D/T, supporting the hypothesis that the range of the Mask Scentometer is a factor in these results. This assumes that $D/T)_G$ are equivalent between a Nasal Ranger and Mask Scentometer. Additionally, it seems logical that the Mask Scentometer would "average" out a few high D/T values, where just one high or low D/T from the Nasal Ranger® could skew the results (only two assessments per session were taken). Also, there were fewer people available to take Mask Scentometer readings than for the intensity rating and Nasal Ranger®, so with more replication, the results could have improved. Therefore, the range of the Mask Scentometer is thought to have been a limitation. Nonetheless, from the regression analysis, a scaling factor appears to be necessary to compare a Mask Scentometer result to a Nasal Ranger® result, and vise versa.

Table 5. Example method comparisons

DTFCO lab D/T	Nasal Ranger® $(D/T)_G$	Mask Scentometer $(D/T)_G$	Intensity rating	Intensity-predicted D/T	Average-intensity-predicted D/T
214	15	4.5	1.5 *	100	57
286	20	6	2 *	133	76
50	5	1*	0.5	23	13
70	7 *	1.3	0.5	32	18
106 *	11	2	0.7	45	28

** Predictor used to determine other values in row.*

Newby and McGinley (2003) found that 7 D/T with a Nasal Ranger® equated to 106 D/T using DTFCO (slope of 0.07). This study found the slope to be 0.08 for a Nasal Ranger® and 0.01 for a Mask Scentometer, or 0.1 and 0.02 respectively, if the geometric means are used. A comparison of DTFCO, Nasal Ranger®, Mask Scentometer, and intensity-based methods from this work are shown in Table 5 for comparison to previous work. For 106 D/T using DTFCO, our slopes equate to 8 D/T and 11 $(D/T)_G$ for the Nasal Ranger (1 D/T and 2 $(D/T)_G$, for the Mask Scentometer). Additionally, for a Nasal Ranger® $(D/T)_G$ of 7, is equivalent to a Mask Scentometer $(D/T)_G$ of 1.3, 70 DTFCO, an intensity rating of 0.5 and an Average intensity-predicted D/T of 18.

In this study (see Table 5), an intensity of 2 equates to a Mask Scentometer $(D/T)_G$ of 6, a Nasal Ranger® $(D/T)_G$ of 20, an intensity-predicted D/T of 133, and a DTFCO D/T of 286. Newby and McGinley (2003) and Huey et al. (1960) have suggested that a D/T of 7 (the regulatory limit in Missouri at the time) is the threshold at which annoyance occurs. Clearly, we do not have a perfect picture of what D/T level is annoying, but it is clear that there are distinct differences between odor assessment methods. This work should serve as evidence that any annoyance threshold levels developed should also be referenced to the ambient odor assessment method used to determine it.

CONCLUSION

In this study, dilution-to-threshold results of dynamic triangular forced-choice olfactometry (DTFCO) are compared to D/T obtained using field olfactometers (i.e. the Mask Scentometer and Nasal Ranger®) and results based upon odor intensity ratings (using ASTM Standard E-544-99, Odor Intensity Reference Scale) under controlled conditions.

The following conclusions were made:

1. Clearly, D/T is specific to the ambient odor assessment method from which it is measured. That is, a Mask Scentometer D/T is not the same as a D/T measured with a Nasal Ranger®. When a D/T is reported, it should be referenced to the method used to measure it. This has implications to regulatory limits and odor criteria, not just in the United States, but abroad.

2. Laboratory olfactometry (DTFCO) does not correlate well with other methods when used for assessing ambient odors. DTFCO session means were significantly different from means for all of the other methods. Using intensity ratings to predict D/T (both Intensity-predicted D/T and Average intensity-predicted D/T) resulted in slopes nearest to one, (0.42 for Intensity-predicted D/T and 0.26 for Average intensity-predicted D/T) when compared to DTFCO.

3. Caution is warranted when predicting dilutions to threshold directly from odor intensity ratings since Intensity-predicted D/T were shown statistically to differ from D/T obtained using all of the other odor assessment methods. Intensity ratings and Average-intensity-predicted D/T both correlated well to D/T readings obtained using the Nasal Ranger® and Mask Scentometer methods. However, when an equation was used to predict D/T from odor intensity ratings, the results did not correlate as well to the other methods.

4. The Least Significant Difference multiple comparison results showed no significant difference between the intensity-based methods (α=0.05) and no differences between the average intensity-predicted D/T and data obtained with the Nasal Ranger® and Mask Scentometer – with either D/T or $(D/T)_G$. Laboratory assessment (DTFCO) was significantly different from the other methods, however. There was no statistically significant difference in the session means even

though D/T predicted based upon Average intensity-predicted D/T and D/T determined using the Nasal Ranger® and using the Mask Scentometer were noticeably different from each other numerically. Average-intensity-predicted D/T was roughly three times higher than D/T obtained using a Nasal Ranger® and roughly fourteen times higher than D/T obtained using a Mask Scentometer. Correspondingly, D/T obtained using a Nasal Ranger® was roughly five to ten times higher than D/T obtained using a Mask Scentometer, with geometric dilutions-to-threshold $(D/T)_G$ being more similar, 2 to 5 times that of a Nasal Ranger®. Leading candidate methods for obtaining similar ambient odor assessment results appear to be the Nasal Ranger® and the Mask Scentometer (both using the geometric dilutions to threshold $(D/T)_G$ for setting stops).

5. Results from field olfactometry methods may be more comparable to another ambient odor assessment method when the geometric average $(D/T)_G$ is used rather than the unit D/T. In this study, using $(D/T)_G$ for the Nasal Ranger® and Mask Scentometer, improved R_o^2's (compared to D/T) to other odor methods.

REFERENCES

1. ASTM. 1999. E679-99. Standard Practice for Determination of Odor and Taste Thresholds by a Forced-Choice Ascending Concentration Series Method of Limits. Annual Book of ASTM Standards. American Society of Testing and Materials. Philadelphia, PA.
2. ASTM. 1999. E544-99. Standard practices for Referencing Suprathreshold Odor Intensity. Annual Book of ASTM Standards. American Society of Testing and Materials. Philadelphia, PA.
3. Committee for European Normalization (CEN). 2003. CEN/TC264/WG2/N222/e Air Quality – Determination of Odour Concentration by Dynamic Olfactometry, European Committee for Standardization. Brussels, Belgium.
4. Henry, C. G. 2004. Instruction Manual for Mask Scentometer. University of Nebraska-Lincoln. Lincoln, NE.
5. Henry, C. G. 2009. Development of the Mask Scentometer, a comparison of Ambient Odor Assessment Methods, and their Application in Ground Truthing Atmospheric Dispersion Models. Unpublished PhD Dissertation. Lincoln, Nebraska, University of Nebraska, Biological Systems Engineering.
6. Jacobson, L. D., H. Guo, D. R. Schmidt, R. E. Nicolai, J. Zhu, and K. A. Janni. 2000. Development of an odor rating system to estimate setback distances from animal feedlots: Odor from Feedlots-Setback Estimation Tool (OFFSET). Paper No. 00-4044. Milwaukee, WI, ASAE, St. Joseph, MI.
7. Jacobson, L. D., H. Guo, D. R. Schmidt, R. E. Nicolai, and J. Zhu. 2003. Development of OFFSET (Odor from Feedlots-Setback Estimation Tool) for determination of odor annoyance free setback distance from animal production sites. In P*roc. Int. Symp. on Gaseous and Odor Emissions from Animal Production Facilities*, 320-330. Horsens, Denmark.
8. McGinley, M. A., and C. M. McGinley. 2003. Comparison of field olfactometers in a controlled chamber using hydrogen sulfide as the test odorant. Presented at the International Water Association 2nd International Conference on Odour and VOC's. Singapore. September 14-17.
9. Newby, B. D., and M. A. McGinley. 2003. Ambient odour testing of concentrated animal feeding operations using field and laboratory olfactometers. *Proceedings of the International Water Association, 2nd International Conference on Odour and VOC's: Measurement, Regulation, and Control Techniques*. 14-17, September 2003. Singapore.
10. Nicolai, R. E., C. J. Clanton, and H. Guo. 2000. Modeling the relationship between threshold and intensity of swine odors. In *Proc. Second International Conference on Air Pollution from Agricultural Operations*. St. Joseph, Mich.: ASAE.

11. R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

12. SAS Institute Inc. 2008. Statistical Analysis Software. Cary, NC, SAS Institute Inc. 2004-2008.

13. Sheffield, R. E., and P. Ndegwa. 2008. Sampling agricultural odors. PNW 595. University of Idaho.

14. Sheffield, R. E., M. Thompson, B. Dye, and D. Parker. 2004. Evaluation of field-based odor assessment methods. Presented at Water Environment Federation and the Air and Waste Management Association Conference on Odor and Air Quality. March 19-21. Tacoma, WA.

15. Sheffield, R. E., M. Chahine, B. Dye, and M. Thompson. 2007. Odor concentrations downwind of Idaho dairies and heifer facilities. *Trans. of the ASABE*, 50: 255-263.

16. Zhu, J., L. D. Jacobson, D. R. Schmidt, and R. E. Nicolai. 2000. Evaluation of Inpuff-2 model for predicting downwind odors from animal production facilities. *Appl. Engr. Agri.* 16(2): 159-164.